https://doi.org/10.15388/vu.thesis.821
https://orcid.org/0000-0003-0928-9560

VILNIUS UNIVERSITY

Justinas Lingevičius

Fortifying Digital Europe: Agentic Security and Technocracy in the Emerging EU AI Policy

DOCTORAL DISSERTATION

Social Sciences,
Political Science (S 002)

VILNIUS 2025

The dissertation was prepared between 2020 and 2024 at Vilnius University.

Academic Supervisor – Prof. Dr. Dovilė Jakniūnaitė (Vilnius University, Social Sciences, Political Science, S 002).

This Doctoral Dissertation will be Defended in a Public Meeting of the Dissertation Defence Panel:

Chair – Prof. Dr. Ramūnas Vilpišauskas (Vilnius University, Social Sciences, Political Science, S 002).
Members:
Prof. Dr. Ingvild Bode (University of Southern Denmark, Social Sciences, Political Science, S 002),
Assoc. Prof. Dr. Marijn Hoijtink (University of Antwerp, Social Sciences, Political Science, S 002),
Prof. Dr. Tomas Janeliūnas (Vilnius University, Social Sciences, Political Science, S 002),
Dr. Florian Rabitz (Kaunas University of Technology, Social Sciences, Political Science, S 002).

The dissertation shall be defended at a public meeting of the Dissertation Defence Panel at 15:00 on the 19th September 2025 in Room 402 of the Institute of International Relations and Political Science of Vilnius University.
Address: Vokiečių St. 10, Room No. 402, Vilnius, Lithuania
Tel. +370 5 251 41 30; e-mail: tspmi@tspmi.lt

The text of this dissertation can be accessed at the library of Vilnius University, as well as on the website of Vilnius University:
www.vu.lt/lt/naujienos/ivykiu-kalendorius

https://doi.org/10.15388/vu.thesis.821
https://orcid.org/0000-0003-0928-9560

VILNIAUS UNIVERSITETAS

Justinas Lingevičius

Fortifikuojant skaitmeninę Europą: agentinis saugumas ir technokratija formuojamoje ES DI politikoje

DAKTARO DISERTACIJA

Socialiniai mokslai,
Politikos mokslai (S 002)

VILNIUS 2025

Disertacija rengta 2020–2024 metais Vilniaus universitete.

Mokslinė vadovė – prof. dr. Dovilė Jakniūnaitė (Vilniaus universitetas, socialiniai mokslai, politikos mokslai, S 002).

Gynimo taryba:
Pirmininkas – prof. dr. Ramūnas Vilpišauskas (Vilniaus universitetas, socialiniai mokslai, politikos mokslai, S 002).
Nariai:
prof. dr. Ingvild Bode (Pietų Danijos universitetas, socialiniai mokslai, politikos mokslai, S 002),
doc. dr. Marijn Hoijtink (Antverpeno universitetas, socialiniai mokslai, politikos mokslai, S 002),
prof. dr. Tomas Janeliūnas (Vilniaus universitetas, socialiniai mokslai, politikos mokslai, S 002),
dr. Florian Rabitz (Kauno technologijos universitetas, socialiniai mokslai, politikos mokslai, S 002).

Disertacija ginama viešame Gynimo tarybos posėdyje 2025 m. rugsėjo mėn. 19 d. 15 val. Vilniaus universiteto Tarptautinių santykių ir politikos mokslų instituto 402 auditorijoje. Adresas: Vokiečių g. 10, Vilnius, Lietuva, tel. +370 5 251 41 30; el. paštas tspmi@tspmi.lt.

Disertaciją galima peržiūrėti Vilniaus universiteto bibliotekoje ir VU interneto svetainėje adresu:
https://www.vu.lt/naujienos/ivykiu-kalendorius

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

ABBREVIATIONS

AI	–	Artificial Intelligence
AI Act	–	Artificial Intelligence Act
AIDA	–	Special Committee Artificial Intelligence in a Digital Age
CoE	–	Council of Europe
CSDP	–	EU Common Security and Defence Policy
DMA	–	Digital Markets Act
DSA	–	Digital Services Act
EC	–	European Commission
EDA	–	European Defence Agency
EDF	–	European Defence Fund
ENISA	–	EU Agency for Cybersecurity
EP	–	European Parliament
EU	–	European Union
EU Council	–	Council of the European Union
EUISS	–	EU Institute for Security Studies
FRA	–	EU Agency for Fundamental Rights
GDPR	–	General Data Protection Regulation
HLEG	–	High Level Expert Group on Artificial Intelligence
LAWS	–	lethal autonomous weapons systems
MEP	–	Member of the European Parliament
NGO	–	Non-Governmental Organization
STS	–	Science and Technology Studies
UK	–	United Kingdom
UN	–	United Nations
USA	–	United States of America

INTRODUCTION

> The agenda of a demiurgic humanity of this intelligence-free (as in fat-free) AI – is yet to be written (Floridi 2023, 15).

> Robots are not able to experience pain (Sharkey 2024).

Artificial intelligence (AI) has become one of the European Union's (EU) major strategic priorities in its *Digital Agenda for Europe.* The European Commission (EC) President Ursula von der Leyen emphasized that "the time has come for us to formulate a vision of where we want AI to take us, as society and as humanity […] and Europe's specific place in the global race for AI" (Ec.europa.eu 2025). In this light, the emerging EU AI policy was introduced as the world's first comprehensive legal framework for the development and use of AI, integrating it within the single market and aiming to harmonize rules and promote digitalization, while specifically excluding military applications.[1] This position raises the question of whether AI in the EU is viewed solely as an economic and innovation matter, neglecting important global concerns related to technological competition, power dynamics, and security.

At first, the political relevance that "the way we approach AI will define the world we live in the future" (Digital-strategy.ec.europa.eu, n.d.-d) was narrowed down to a technical and, in the words of the President Ursula von der Leyen, "always neutral" (Ec.europa.eu 2020) definition of AI: "a collection of technologies that combine data, algorithms, and computing power" (Eur-lex.europa.eu 2020d). Luciano Floridi (2021, 219), who was a member of the High Level Expert Group on AI (HLEG), argued that this definition highlights that AI is not "some kind of Frankenstein's monster" and the elimination of "non-scientific statements" such as "artificial consciousness" helps to avoid "sci-fi speculations about AI". This perspective views technology as a tool, focusing on the practical aspects of data processing and task execution.

However, the emphasis on AI-related risks "generated by specific uses of AI" (Digital-strategy.ec.europa.eu, n.d.-d) raises important questions about the assumption that technology can be separated from its applications.

[1] The White Paper (Eur-lex.europa.eu 2020d) states that "it does not address the development and use of AI for military purposes".

References to "risks associated with certain uses of this new technology"[2] (Eur-lex.europa.eu 2020d) and calls for "human-centric, transparent and responsible development of AI" (Ec.europa.eu 2023b) suggest that concerns extend beyond specific use cases. These points reflect uncertainties about the nature of AI itself, marked by "a sense of urgency, a sense of obligation to know, and a sense of changing and creating a future" (Manners 2024, 834). Therefore, the notion of risk indicates concerns that extend beyond the EU's traditional security agenda under the Common Security and Defence Policy (CSDP). Rather, it highlights political contestation and perceived challenges related to human-machine interaction, considering "who we might yet become" (Amoore and De Goede 2012, 5).

In addition to internal issues, the EU faces external pressures to adopt a more strategic and competitive stance on AI. These pressures have placed the EU in a "global competition for AI supremacy that has never been more intense or geopolitically contested" (Csernatoni 2025). In this context, the EU grapples with a dilemma between its normative commitments as a modernist liberal power – grounded in principles of rational governance, market integration, and rights-based regulation – and the need to respond to intensifying geopolitical dynamics related to AI. The AI-related ambitions expressed by the Commission President Ursula von der Leyen – to lead "the way on a new global framework for AI" (State-of-the-union.ec.europa.eu 2023) – reflect not only the EU's complex position on technology but also its efforts to redefine its role within a contested international landscape. These dynamics, however, expose the limitations of the EU's mandate in security politics, prompting further questions regarding established priorities within the energing AI policy framework (for example, Liebetrau 2024; Csernatoni 2024).

i. Academic debates

The discussed complexity of the EU's AI-related ambitions, concerns and external pressures brings together three elements – technology, security and risk – which intersect and guide the focus of this analysis. Further engagement with academic debates involves analysing each element to demonstrate the tendencies already identified and the questions remaining to be explored.

[2] For the purposes of this analysis, references to *technology* specifically denote artificial intelligence (AI), related developments and uses, unless otherwise specified. Similarly, the term *machine* refers to AI and related technologies, especially in contexts involving human-machine interaction.

Security has become an important dimension for the EU's self-legitimation in response to acute crises and the growing expectations of citizens (Hegemann and Schneckener 2019). Various analyses have already demonstrated that the EU subscribes to a broader view of security, such as climate change and cybercrime (Sperling and Webber 2014), or cross-over issues fixed to the protection of civilians and human rights (Calderaro and Blumfelde 2022). The case of cybersecurity is particularly illustrative, because, initially framed as an economic concern attached to the advancement of the single market and driven by the EC, it became a comprehensive security policy tackled at the EU level (Brandão and Camisão 2022; Carrapico and Barrinha 2017). These points indicate that the boundaries between conventional security and non-security issues have been blurred, as the policies have crossed into different agendas.

Going into the interrelations between technology and security, analyses demonstrate that the EU's official discourses are inconsistent with practices. For example, a closer look at EU policies involving technologies such as the European Defence Fund (EDF) – a programme for collaborative, cross-border defence research and development – as well as drones and border management, reveals that portraying these technologies as commercial or civilian does not remove their military dimension. Instead, it frames them as *solutions*, by emphasizing their positive effects and legitimacy (Csernatoni 2018; 2021a; 2019a; Csernatoni and Lavallée 2020; Lavallée and Martins 2023; Martins 2023; Martins and Jumbert 2022; Martins and Mawdsley 2021; Csernatoni and Martins 2024).

The same tendency applies to AI. Although the military is excluded from the scope of the policy, Ingvild Bode and Hendrik Huells (2023) show that the EU is both a rule-maker and a rule-taker in military AI, aiming to represent European values and influence the international landscape. The discussion about AI even involves frames of military AI, and suggests the EU's positionality as a military power (Lingevicius 2024; 2023).

These insights suggest that despite the official policy focus on the single market, both discourse and practices involving technologies in civilian and military domains get increasingly entangled, and the distinction is difficult to draw (Martins and Ahmad 2020). Additionally, the research reveals that the framing of technologies as economically profitable and politically neutral products is continuously embedded in the military discourse (Hoijtink 2014). Therefore, military and security implications are not absent from the EU's thinking and practising emerging technologies, including AI. However, the mentioned analyses do not further elaborate on or problematize civil-military

or *dual-use* contestation – what kind of security is being discussed when the EU discusses AI.

This point brings us to the apparently remaining question of how security is framed within the emerging EU AI policy. A noticeable hint is the importance of risk, which challenges the conventional definition of security focused on threats and immediate responses. However, the answer of defining and involving risk in the policy framework is not clear-cut. Following the literature, two ways of discussing the role and function of risk are noticeable: as a set of specific instruments for governance, and as a framework for producing knowledge in response to uncertainty.

Firstly, risk is approached as a mode of governance and management which moves towards European integration by overcoming national differences (Paul, Bouder and Wesseling 2016; Rothstein, Borraz and Huber 2013). Risk has also become a preferred way to develop legislative mechanisms that privilege and justify selected measures depending on the probability of risk (Niklas and Dencik 2024). For example, the EU employs risk as a policy instrument to inform risk analysis, risk-based standards, or policy enforcement across different fields, including climate, food safety, flood management and terrorism (Paul, Bouder and Wesseling 2016; Rothstein, Borraz and Huber 2013).

At the same time, the implementation of a risk-based approach has so far proven to be uneven, and related to confusion about the specific nature of the risks (Rothstein, Borraz and Huber 2013; Floridi 2021). The case of Frontex, the European Border and Coast Guard Agency, has demonstrated that the employment of risk has been associated with the EU's avoidance of securitizing border management, and rather presents it as a "governmentality of unease", which is less about the tension between the norm and the exception, but the "incremental normalization" of the issue (Neal 2009, 26).

Another way of considering the role of risk engages with critical approaches which stress that the concept of risk functions as an epistemological construct shaping regulatory identity (for example, Csernatoni 2021; Bode and Huelss 2023; Amoore and Raley 2017). Dimitar Lilkov (2021) suggests that the risk-based approach focuses on future governance and its effectiveness, rather than diving into analyses of understanding risks and their role; while Regine Paul (2024; 2017b), following the critical political economy angle, argues that risk serves as an epistemic tool to show how regulators think about the phenomenon and frame future AI in preferred ways.

Those that already note and analyse the EU's decision to frame the emerging EU AI policy through risks, share similar insights that the EU uses this notion to define perceived concerns and set its priorities in the domain. For example, the use of risk is considered as "idealistic", and matching the trajectory of the EU's identity-formation, where AI policy is attached to the protection of civil liberties (Schmid, Pham and Ferl 2024, 17). However, the proposed logic is criticized for not guaranteeing that AI will not pose other non-prioritized risks (Schuett 2024).

These analyses demonstrate that the employment of risk offers an intrigue of political motivations, as well as a process of knowledge production. The use of risk enables the EU to set its political priorities by anticipating AI-related developments and uses, identifying specific issues, and legitimising its responses. At the same time, the tension between risk and security also remains part of the discussion in the case of AI: is it "two sides of the same coin" (Methmann and Rothe 2012, 337), or does it mark different ontological and epistemological perspectives?

Technology, security, and risk should not be seen as neutral or merely framed concepts. Instead, they reflect the positions, priorities, and interpretations through which actors, such as the EU, understand related challenges and shape their responses. These responses also signal the inscription of imaginaries – defined as shared, future-oriented visions that inform how AI is approached – which influence how the EU engages with AI. From a constructivist perspective, the relationship between the EU's approach to AI and the way it frames technology, security, and risk necessitates a focus on the EU itself and an examination of how the meanings of these terms are constructed and intertwined with the EU's self-position. This focus is particularly important given the identified inconsistencies between the EU's strategies and their implementation, as well as external pressures and the significance of digitalization in shaping the EU's stance.

To do so, attention is drawn to the discussions on the kind of international actor the EU aspires to be, and the role it seeks to secure (Mügge 2024; Bellanova, Carrapico and Duez 2022a). These analyses highlight a transformation in the EU's role, particularly in the context of digitalization. Over time, the EU has been seen as a civilian or normative power, promoting values-based governance and utilizing persuasion and institutions in international relations (McNamara 2024).

However, recent discussions suggest that the EU is now adopting a more strategic approach that prioritizes security and protectionist forms of engagement. For example, Raluca Csernatoni (2019b) suggests that the EU's

ambitions in the digital domain have led to the EU's transformation from a civilian international actor into a security and technological power. This tendency has even been described as "a new role as a digital (geo)strategic power", which brings more "geopolitical posturing and the unabashed defence of interest-driven behaviour" (Broeders, Cristiano and Kaminska 2023, 1265). Linda Monsees and Daniel Lambach (Monsees and Lambach 2022) also claim that the link between technology and geopolitics in the EU is drawn more explicitly than before. This proposed shift from a normative to a strategic actor raises questions about whether the EU is adapting its normative framework to align with the realities of digitalization while also integrating a more security-oriented perspective. Alternatively, this change may indicate a redefinition of the EU's self-positioning as an adaptation to the international AI landscape, which requires further examination.

In this context, these questions relate to debates on digital sovereignty – a concept influenced by diverse political, economic, and security factors as the EU addresses the challenges of digitalization. Already discussed in various analyses, digital sovereignty can be generally understood as "a form of strategic autonomy from third countries and re-orienting relations with 'Big Tech', notably through the creation of the EU's own digital infrastructures" (Bellanova, Carrapico and Duez 2022a, 338). Here, technologies are intertwined with the EU's expectations – to increase control and manage the digital ecosystem, and decide on its future governance (Roberts, Cowls, Casolari, et al. 2021; Seidl and Schmitz 2024; Adler-Nissen and Eggeling 2024; Baur 2024; Bellanova and Glouftsios 2022; Klimburg-Witjes 2024).

At the same time, a number of articles and special issues on digital sovereignty, including *European Security* (Bellanova, Carrapico and Duez 2022a), *Journal of European Public Policy* (Falkner, et al. 2024), and *Geopolitics* (Glasze, et al. 2023), have revealed that "the EU still lacks a clear, coherent vision, with different actors from different EU institutions emphasizing different domains" (Roberts, Cowls, Casolari, et al. 2021, 18). While sovereignty here departs from its traditional association with territorial control, its meaning in relation to technology and digitalization remains fluid and open-ended. For example, how do technologies and related international trends contribute to the production of the EU's subjectivity? How do references to *AI geopolitics* entail the struggle over meanings, governance and control of AI technologies, shaped by discourses, power relations and global asymmetries?

The suggested hints of a change in the EU's positionality and the introduction of digital sovereignty seem to be explanations in themselves,

without fully detailing their interrelation with the emerging AI policy. This connection is important because, as mentioned earlier, AI is a geopolitically contested issue that raises various concerns for the EU. As a result, the EU faces challenges not only from technological developments but also from external pressures that affect its AI policy formulation, necessitating interconnected responses.

The current debates surrounding the EU's approach to technology – security ambiguities, risk heuristics in policy-making, and its assertive positioning in the digital domain – indicate that three questions remain unresolved:

First, the existing literature demonstrates that the analyses do not question *what security* is being pursued in the case of AI. Instead, they share evidence that technologies, which are often presented as being for civilian use, are increasingly co-opted into militarized frameworks, such as the notion of dual-use, which refers to applications for both civilian and military purposes. While these insights do deconstruct these trends, they still operate within the civilian-military dichotomy and suggest that technology-driven change blurs its boundaries. However, the emerging EU AI policy, excluding the military, highlights various risks that indicate the presence of security issues, framed outside the conventional military and threats logic. This specificity invites further investigation into the meaning of security in this context, particularly regarding how technology, the concerns it raises, and the justifications for policy responses are embedded within the emerging EU AI policy.

Second, the concept of risk within the context of the emerging EU AI policy has largely been taken as a given element, without sufficient scrutiny of its implications for embedding security-related elements into policy frameworks (with the notable exception of Regine Paul's analyses). As Louise Amoore (2023) suggests, the discourse of AI-related risks functions as a mode of assembling and ordering knowledge, which transforms how state and society understand itself. Yet questions remain about what this understanding entails, and how risk plays a role in shaping AI-related security – particularly when it is not framed through threats or exceptional measures.

Third, the reviewed debates have shown that the EU's interest in technology goes beyond merely preventing internal fragmentation. This policy reflects the EU's international ambitions, responses to external pressures and strategic aspirations in the digital domain. Given that the international landscape is often described in terms of AI geopolitics, a more critical analysis is needed to understand how the EU's emphasis on

normativity or digital sovereignty, as forms of self-positionality, shapes the construction of AI-related security knowledge.

Overall, the **gap**, situated across three discussed strands of literature – security ambiguities, risk heuristics, and positioning in the digital domain – requires further exploration. The EU's approach towards AI not only reflects standard policy-making practices but also shapes its understanding of security as it relates to future interactions with AI. This gap is addressed in the empirical analysis (Part II), where the relevant questions are explored, and corresponding conceptualizations are further discussed in Chapter 6.

ii. Problem statement and research goal

The thesis focuses on a noticeable inconsistency in how the EU approaches security in its emerging AI policy. While recognizing that AI introduces new synergies and concerns (Rychnovská 2020), the EU frames AI in technical terms, emphasizing risks associated with its use. This perspective reflects the EU's institutional competences, which prioritize harmonizing a level playing field within the single market and establishing rules for AI developers and providers.

At the same time, the EU aspires to become a "coherent security actor" (Carrapico and Barrinha 2017, 14) and positions itself as a global standard-setter capable of managing technologies with still uncertain effects. These dynamics unfold within broader debates about which forms of digitalization best support European values, competitiveness, and democratic governance (for example, Mügge 2025; Hoijtink and Van der Kist 2025).

However, the security dimension – particularly beyond the excluded military aspects – remains underdefined within this complex constellation of AI-related issues and ambitions. Although existing studies acknowledge that security is relevant for the EU's digital agenda (both in militarized practices and under the framework of digital sovereignty), there is limited attention to how AI-related security is constructed and what meanings are produced through the language of risk in the emerging EU AI policy.

Following the assumption that security is about boundaries of understanding, not just sovereign borders (Bigo 2001), the thesis **aims** to analyse how the EU constructs and defines AI-related security knowledge in its emerging AI policy, particularly through the framing of risk. To do so, it examines discursively articulated positions and perspectives, focusing on how

they shape the understanding of AI and constitute the EU's self-positioning within the context of the international landscape.

The thesis considers the EU as a complex actor with multiple voices contributing to its AI strategic discourse and overall subjectivity in the digital domain[3]. For the analysis, I follow the consideration of the EU as "a site of discursive authority, with enough consistency across the array of different actors to provide a common institutional language and framework for action" (Baker-Beall and Mott 2022, 1092). Although one might instinctively focus on the power dynamics between institutions and their diverging priorities, I argue that the process of knowledge production – including aspects of security and self-positionality – can be analysed as specific to the EU as a whole, and an expression of the EU's own position and representation.

While policy-making is spread across institutions with distinct competences, there is broad consensus on the main pillars and direction of the EU's digital agenda. This shared vision is exemplified by the European Declaration on Digital Rights and Principles, signed by the presidents of the Commission, the European Parliament (EP), and the European Council, and committed to by the Member States in 2022. The Declaration affirms that the EU's digital transformation should reflect the EU's values, promote a human-centric approach, and ensure the protection of citizens' rights (Digital-strategy.ec.europa.eu, n.d.-e).

In this context, the EU is progressively positioning itself as an important player in the global digital landscape. Its increasing role in framing the digital agenda and taking a coordinator's role, as opposed to "divergent" or "fragmented" approaches (Af Malmborg 2023, 3), indicate the EU's increasing focus on digitalization alongside other EU policies such as trade. Anu Bradford (2023, 6) even describes the EU, together with the USA and China, as "a digital empire", emphasizing the ambition to shape the global digital order towards its interests and values.

In addition to this, the EU is increasingly viewed as capable of performing essential security governance functions and has been defining its unique security role (Sperling and Webber 2014; Baker-Beall and Mott 2022). In this context, focusing on AI-related security knowledge further highlights the

[3] The analysis does not focus on individual institutions, member states and other stakeholders, such as lobbying groups, NGOs and associations. This exclusion is intentional, because their mapping and involvement in policy-making would shift the focus toward relationships between EU institutions and different actors. Then the scope of the analysis would require a different research question.

EU's ambitions to integrate diverse policy agendas and construct its profile on the global stage. These arguments collectively support the consideration of the EU as an actor that shapes and projects its position through digital and security policy agendas, which reflect its evolving subjectivity.

In the thesis, I refer to *the EU's approach towards AI* as a mindset and a strategic direction of the EU regarding AI. *The emerging EU AI policy* represents the concrete, formalized output and expression of this approach – communications, reports, resolutions, and even legally binding rules based on institutional contributions between 2018 and 2024. The emphasis on *emerging* also signals continuity over time and heterogenous contributions that have shaped the policy outcome focused on AI as an *emerging* technology – one that has been and continues to be developed in multiple and still-unfolding directions. Here, the main interest is directed towards the construction of meanings, pointing to a non-linear relationship between meaning, action, and the general "messiness" of the political (Matejova and Shesterinina 2023a, 277).

Considering this, the main **objectives** of the research are:

1) To distinguish the major trends and dynamics within academic and political debates on AI which help to understand what concepts are mostly used in describing AI and AI-related policies.
2) To engage with critical security debates on the intersection of technology, risk and security, that grounds a conceptual framework for the analysis.
3) To identify the AI-related security dimension in the case of the emerging EU AI policy, continuing the conversation on the EU's approach to AI and politics of security.
4) To describe the EU's proposed response through its introduced policy measures, situating this response within the broader context of existing EU governance frameworks.
5) To name the EU's self-positioning in the emerging AI policy by distinguishing its leading characteristics, and considering their possible novelty in relation to existing literature, summarized as the *Europe as power debate* and established definitions.

The thesis claims that the EU's AI-related security understanding relies on the notion of agentic security, technocratization as a way of governing, and a fortress as the EU's self-positioning in the context of the international AI landscape. In line with these insights, **four thesis statements** are proposed:

1. The process of riskification demonstrates that AI-related security is constructed through risks. The introduced categories of risk, described through the imagined level of potential harm to fundamental rights and a democratic political system, reflect the EU's established boundaries of graduating AI-related concerns and priorities. Such a risk-based approach indicates that knowledge production describes security as long-term, future-oriented, and based on imaginaries of technology.
2. The EU's AI-related security emerges as agentic security, understood through the capacity to sustain control and decision-making power vis-à-vis current and future AI, which is framed as Other. Agentic security is introduced as an additional security dimension to human and societal security, as it focuses on antagonistic relations between humans and/vs machines. This logic relies on anthropocentrism and its need to maintain the hierarchy of the human position.
3. The EU's proposed governance programme, as a response to the risks and conditions of possibility for harm, is technocratized. It seeks to ensure security through everyday routines, standardized measures, expertise, and dispersed responsibilities. Technocratization mirrors the definition of AI, where reliance on scientific knowledge is expected to guide in the face of uncertainty. Technocratized governance also allows the EU to propose its framework internationally as universally applicable and expert-driven, rather than politically contested.
4. The EU self-positionality in the emerging EU AI policy can be summarized as *a fortress.* The vision of a fortress emerges as a response to AI as Other, and to the unfavourable international landscape amplifying AI otherness. The fortress metaphor refers to a bounded space which functions in three ways: a) EU rules apply to those inside, but also to those who want to enter the fortress; b) it seeks to reduce interdependence with powers; and c) it aims to persuade others to follow the EU's approach to AI.

iii. Theoretical assumptions

This thesis follows a constructivist perspective to examine how AI-related security is constructed at the intersection of technology, security and risk. It engages with critical discussions on how power, knowledge and governance are both influenced by and shape this intersection. A multilayered conceptual framework is developed to facilitate a comprehensive analysis, based on the assumption that security is not an objective state but is constructed through shared meanings, discourses and practices.

In this analysis, technology is viewed as shaped by social, institutional, economic and material possibilities and constraints, while attention is paid to interwoven relationship between the ways technology is imagined and then framed into political decisions (Liebetrau and Christensen 2021; Leese and Hoijtink 2019). A similar approach is applied to the concept of security. From a constructivist standpoint, security prompts an exploration of how visions of the future and related concerns are structured, what boundaries are drawn, and how distinctions between "inside" and "outside" are established (Huysmans 2008; Bigo 2001). Therefore, security-related concepts, such as competition, power and interests, are not fixed, but lead to examinations of how they are produced, contested and embedded within broader political, technological and normative contexts.

In terms of the security-risk dynamic, I argue that both threats and risks are socially constructed, and depend on how and what antagonism is articulated, and what responses are proposed. Considering the different connotations, threats here relate to concerns of survival, and refer to irreversible damage that requires an immediate response, often presented as exceptional measures to deal with that threat (Methmann and Rothe 2012). Meanwhile, risk is understood as lying below the threshold of *existential* in terms of a level of concern and exceptionalism, but focuses on potential harm. Risk signals a different temporal perspective, as it turns the focus to future unknowns, while policy proposals are introduced to mitigate that future. Then a response is directed towards the longer-term management of potential harm, without going into the realm of emergency or exceptionality (Backman 2023).

The focus on risks asks what makes situations perceived as risky (Matejova and Shesterinina 2023b); and how risk is used as a "means for ordering reality" (Collective 2006, 468). Yet, the concept itself does not convey a specific meaning or level of danger. To better understand the implications of risk in this context, the thesis draws on the notion of *concerns*. Concerns, understood in a broad and flexible sense, refer to the perception of *unease* about potential issues. As Didier Bigo (2002) explains, this sense of unease is tied to the uncertainty of everyday life within a risk society, where freedom is linked to the boundaries of (in)security. In the context of this thesis, this interpretation of concerns underscores how risk indicates something perceived with *unease,* requiring further exploration of its specific connotations and associated issues.

If risks are understood as distinct from matters of *high politics* – that is, if they do not necessarily pertain to survival or demand exceptional measures – then their analysis requires a corresponding conceptual perspective that

reflects this distinction. In this context, I engage with the concept of riskification, which marks the process by which issues become viewed and acted upon in risk terms and responses (Morsut and Engen 2022). For the thesis, four analytical elements are distinguished (see Section 2.3.): a referent object, conditions of possibility for harm, a governance programme, and international engagement. Conditions of possibility for harm and a governance programme signal a major shift from securitization, because they are not about threats and extraordinary measures, but deal with potential harm and legitimising extensive governance (Corry 2012). The thesis adds the element of international engagement to the initial framework to identify how the EU positions itself in the international AI landscape.

Ultimately, riskification provides a valuable perspective for examining the tension between security and risk. This framework enables an analysis of how AI-related risks are constructed, shifting the focus from immediate challenges to an unspecified future. In this way, it addresses key aspects of the research gap and helps to ground the analysis of AI-related security knowledge.

iv. Research strategy

Considering the importance of *speech acts* as initiating the process of riskification (Harijanto 2025), the thesis employs discourse analysis to examine how meanings around AI and security are constructed, contested, and institutionalised. Drawing from a Foucauldian-inspired perspective, the focus on discourse encompasses statements, practices and rules that structure the production of knowledge and meaning. This framework examines how subjects and objects are formed and governed. The analysis focuses on specific vocabularies, established definitions, and meanings that shape the leading ways of thinking about AI. It also examines how the EU's position is mutually shaped in relation to these concepts.

The EU has already been discussed as "highly discursive", full of positioning statements and a range of policy documents, which provide a common institutional language and a framework for action (Baker-Beall 2014, 3). The emerging EU AI policy similarly reflects how the diversity and breadth of documents, positions and statements both construct the policy and shape its underlying meanings, forming what this thesis terms *the EU's AI strategic discourse* (see Chapter 3). Since the implementation of discourse analysis often depends on the specific case, the research strategy is tailored to develop key steps, such as data selection and collection, the coding scheme, and interpretation. It remains reflexive, as discourse analysis requires careful

reflection on the researcher's positionality and decisions made throughout the process.

The discourse analysis uses two types of data: 1) documents of the emerging EU AI policy, and b) semi-structured expert interviews. 75 documents released by EU institutions between 2018 and 2024 have been selected according to established criteria (the list of all the documents is provided in Annex 1). It marks the period from the EC's first Communication *Artificial Intelligence for Europe*, specifically dedicated to AI, to the enforcement of the *Artificial Intelligence Act* (AI Act) in August 2024, as closing the full circle of framing, negotiating and adopting the policy. The period in between reveals intense inter-institutional debates, discursive contributions and knowledge production, which are discussed in Chapter 4. The analysis also includes 11 semi-structured expert interviews conducted from May 2023 to February 2024 with institutional representatives and experts involved in constructing the emerging EU AI policy in the period (the list of interviews conducted is provided in Annex 3). These interviews serve as a way to advance the analysis by highlighting the perspectives by those directly involved in the policy-making. Overall, the defined research strategy and its implementation (Chapter 3) address the main questions, expectations and limitations raised to the discourse analysis in security studies.

Lastly, the question of who speaks on behalf of the EU remains methodologically sensitive, given that the EU is not monolithic, but comprises multiple institutions. That is why the empirical analysis (Chapter 5) will demonstrate and refer to institutional voices to demonstrate the EU's existing complexity and multiplicity through numerous documents, institutional set-ups, and overlapping vocabularies, while reaffirming the EU's central role in the analysis.

v. Contribution

This thesis makes **four academic contributions** to the debates on the EU's approach to technology and security, risk and riskification, and its digital ambitions and international role. It introduces the concept of *agentic security*, outlines a *technocratized governance programme*, and identifies the position of a *fortress* in the EU's AI strategic discourse. Together, these elements offer a framework for understanding how the EU constructs AI-related security and integrates responses that are intertwined with policy-making and its evolving subjectivity.

Firstly, the thesis speaks to existing analyses that highlight tensions between the EU's framing of technologies in civilian, non-military terms and the military-related practices that these technologies enable. The thesis transcends the conventional civilian-military dichotomy and contributes to the academic debate on the emerging EU AI policy by proposing the concept of **agentic security**. This concept focuses on the antagonistic human-machine interaction, framing AI as *Other*. Agentic security prioritizes the protection of human agency across diverse sociotechnical settings, contexts and groups, understood as the capacity to maintain control and decision-making power vis-à-vis current and future AI-related concerns. In this way, the thesis develops the analysis of AI-related security, which is often focused on military contexts, by demonstrating that concerns about human-machine interaction and control extend beyond the military domain. As a result, agentic security emerges as an important framework for understanding the concerns and issues raised in the face of technology, thereby broadening the scope of the security studies discourse.

Secondly, the thesis engages with security debates that view security as a terrain of *high politics*, exceptionalism, and emergency measures. The thesis develops the concept of **technocratization** to describe a form of security governance that relies on procedural, long-term strategies and routines to mitigate risks. While technocratic approaches are not new to EU policy-making, this thesis highlights their significance in the field of security. It argues that security is framed not as an exceptional space but as a risk-focused domain governed by expertise, institutional procedures, and depoliticized tools. Additionally, the thesis bridges discussions in security studies and science and technology studies (STS), particularly regarding the role of technology in security practices. The analysis demonstrates that AI serves as a justification for technocratic modes of governance, where scientific knowledge and expert judgment are relied upon for clarity and guidance.

The specificity of the response is further explored through the concept of riskification, showing how risk-based logics shape the EU's approach towards AI. The thesis extends this concept by introducing the element of international engagement, arguing that the EU internationalizes perceived risks and corresponding responses. This perspective emphasizes that both technology and risk are viewed as *borderless*, highlighting the EU's efforts to secure its subjectivity through technocratized and globally attuned responses to AI-related security challenges.

Thirdly, the thesis contributes to the debates on the EU's global positionality. Rather than recycling established concepts like normative power

or framing technology as an instrument for achieving digital sovereignty, this analysis suggests that the EU's self-positioning is shaped by portraying AI as *Other*. This results in a self-definition of a **fortress**. Unlike other analyses that describe the EU's subjectivity as being based on factors other than security – such as the market and the promotion of values – this study posits that here the EU's positionality is largely a response to security concerns. Then, various strategies, including multilateral engagement, persuasion and competition, are employed to enhance security. Therefore, rather than concentrating on discussions about the digitalization-driven shift towards a more (or less) assertive EU, this thesis demonstrates that the EU takes a protectionist stance towards both AI and the competitive international landscape, thereby amplifying AI otherness.

Finally, the thesis engages with key debates shaping the emerging EU AI policy, particularly the challenge of "how to write the rules of AI according to an ethical and human-rights agenda without hampering innovation or harming uptake of AI technologies in Europe" (Brattberg, Csernatoni and Rugova 2020, 9). The argument presented in the thesis suggests that these discussions often depict the EU as either falling behind in the global technological race or compensating for its lack of competitiveness through regulatory measures. However, this analysis indicates that such claims overlook the security dimension embedded in the emerging EU AI policy, which aims to safeguard *Europeanness* based on fundamental rights and democratic principles. The thesis contends that these efforts are not only reactive; they also aim to steer the development and use of AI in ways that promote the EU's approach as both distinctly European and universally relevant. Therefore, the thesis encourages further dialogue about the strategic direction of the emerging EU AI policy, proposing a shift away from a sole focus on economic competitiveness.

vi. Outline

The thesis has two parts:

1) *Debates on AI, conceptual framework and research strategy*. Part I overviews the major debates and trends related to AI, and formulates a conceptual framework and a research strategy. Chapter 1 focuses on different definitions of AI, to demonstrate the existing spectrum of them, and the most relevant features for the thesis. The chapter concludes that the definition of AI is not a neutral and technical debate, but reveals social and political reflections that depend on a defining actor. Chapter 2 discusses three concepts used in the

analysis – technology, security and risk – and their intersection discussed within critical security debates. The conceptual framework then focuses on riskification, and proposes four analytical elements for the empirical analysis: a referent object, conditions of possibility for harm, a governance programme, and international engagement. Chapter 3 outlines the relevance of discourse analysis, and establishes a research strategy for the empirical analysis. It details the data selection and collection process, the development of the coding scheme, and interpretation.

2) *The empirical analysis*. Part II presents the results of the empirical analysis based on the four established analytical elements (a referent object, conditions of possibility for harm, a governance programme, and international engagement) and a research strategy. To put the emerging EU AI policy into its context, Chapter 4 describes the EU's digital ambitions, the chronology, and the key characteristics of the developing emerging EU AI policy. In addition, it contextualizes the EU among other actors which have also been developing their emerging AI policies in the same period. Chapter 5 analyses the process of riskification by consistently applying the conceptual framework presented in Chapter 2. Thus, Section 5.1. investigates the ways the EU establishes categories of risks and presents them as a pyramid of a risk-based approach. Section 5.2. discusses the referent objects of fundamental rights and a democratic political system, as well as safety, as the main elements that need to be protected. Section 5.3. looks into what concerns are discussed focusing on a definition of harm, intrusion and discrimination, and the autonomy of AI as leading conditions of possibility for harm. Section 5.4. discusses the EU's proposed policy measures as the governance programme, which include the principles of human-in-the-loop, human-centrism, regulation, assessments and an institutional ecosystem. Section 5.5. discusses the EU's positionality, which is presented through the directions of international engagement – multilateralism and competition.

Next, Chapter 5 proposes three conceptualizations – agentic security, technocratization and a fortress – based on the empirical analysis. They demonstrate that the EU's AI-related security focuses on safeguarding human agency from AI as Other, and how the EU develops a response shaped through technocratic reasoning and aims for a more assertive and protectionist position in the competitive international landscape.

Chapter 6 engages with the EU's thinking revealed in Chapter 5, and critically examines the prevailing tendencies of anthropocentrism, depoliticization and Eurocentrism. Considered here as controversies, persistent yet contested viewpoints, these tendencies suggest that their

inscription in the EU's approach towards AI reflects remaining biases, not only towards technology itself, but also towards those outside the EU.

Finally, the Conclusion discusses the main findings and insights, as well as revisiting the thesis statements. Based on these observations, broader points for further reflection and discussion emerge, highlighting how AI further fosters debates around power relations, ideological struggles, competing imaginaries, and challenges in policy implementation.

I. DEBATES ON AI, CONCEPTUAL FRAMEWORK AND RESEARCH STRATEGY

1. AI: CONCEPTS AND DEBATES

The discussion of AI-related concepts, policies and their measures requires a better understanding of what AI means, and how it is described in different domains by both its proponents and opponents. Such an overview is necessary in order to show that there is no single definition, and evolving trends should be carefully considered. Following Lucy Suchman (2023; 2020), we face a largely speculative field of technological development, which is singularized under AI, but works to escape definition to maximize its suggestive power. For example, in some cases, authors decide to use "Ais" rather than "AI" to demonstrate this pluralism, and avoid impressions of considering AI as an autonomous and defined *object*. On this note, in 2024, the White House provided a list of emerging technologies where AI contains ten subfields, suggesting a range from machine and deep learning to generative AI and foundational models (NSTC 2024).

In this thesis, I refer to AI, due to its widespread use in both academic and political discourses, while remaining reflexive of the introduced complexity. AI is explicitly used among different actors which develop their AI policies, and could also be considered as an important political trend which needs to be further discussed. The EU, as a case of this thesis, does the same, and develops its emerging AI policy based on its own constructed definition of AI and its variations, as presented before. Thus, Section 1.1. introduces the debates and historical references of understanding *artificial* and *intelligence.* It demonstrates that, despite more attention and increasing awareness, a definition remains a scientific, analytical and ideological struggle.

AI also fuels both utopian and dystopian visions, reflecting uncertainties about its future impact: will AI replicate human intelligence? What are the fundamental differences between humans and machines? What level of technological advancement could transform societies, and in what ways? (Tegmark 2017; Russell and Norvig 2016) These questions emerge because they refer simultaneously to different forms of AI, AI uses, and related imaginaries, from machine learning to artificial consciousness. To approach this debate, Section 1.2. delves into the need to distinguish the meanings of *automation/automated*, *autonomous* and *autonomy*, while considering AI. The clarifications provided help to better understand the often uncritical and irreflexive mixing of them in noticeable political discussions.

Lastly, the engagement with literature has already demonstrated actively evolving academic debates, which discuss AI through different lenses and spheres: ethics, security, governance and law, to name a few. These different lenses significantly expand the scope of discussion, and indicate that priorities, issues and vocabularies may differ, depending on the field and the tradition. Following the focus of this thesis, Section 1.3. overviews the debates of the role of AI in security, and what questions are raised. It also provides relevant insights for further problematising the notion of security in military and non-military dimensions.

1.1. Focus on technical capabilities

To start with, both elements, *artificial* and *intelligence*, provoke a debate: to what extent we can discuss current technological capabilities as something nearing human intelligence. The hype of AI as challenging humans has been criticized as if proponents "were attending a magic show – they want to believe in the magic abilities" (Hunger 2023). Therefore, more *technical* notions, such as machine learning and algorithms, are proposed as being more accurate to today's technological advancement. For example, different analyses ask how algorithms should be managed in an ethical way, and if they become smart enough to operate independently (Matzner 2019). The process of using algorithms is described through portraying an engineer who chooses a mathematical structure that characterizes a range of possible decision-making rules with adjustable parameters. That is why the algorithm is sometimes understood as a "black box" that applies a rule and provides results for further interpretation (Dignum 2019, 23).

Discussions on which notion is more responsive to define current operational capacities – algorithm, machine learning or something else – do not prevent the expansive use of and references to AI. Despite variations, AI is broadly understood as the capability of a computer system to perform tasks that require human intelligence and its forms (for example, visual or speech recognition) (Cummings et al. 2018). In 1972, Hubert L. Dreyfus (1992, 143) stated that creating artificial intelligence works on the assumption that a human is "a device which calculates according to rules on data which take the form of atomic facts." However, different psychological, epistemological, and ontological assumptions indicate that the distinction between humans and computers, particularly in terms of reasoning, continues to be a topic of debate. This discussion often revolves around where exactly this separation occurs, especially as AI is envisioned as the emergence of "artificial persons"

(Héder 2020, 63), triggering controversies regarding the nature of something fundamentally different yet similar to humans.

The struggle with the definition of AI also comes from the realization that AI is a technological enabler, which is applied or used in a broader context of emerging and disruptive technologies (such as 5G, quantum engineering, cyber or biotechnology). In other words, AI is more about widely applicable solutions rather than a concrete device or function that could be easily labelled and measured in quantitative terms. In this context, the capacity to process large amounts of data in a rapid manner and to quickly provide generated results could be distinguished as the main defining characteristics of AI (Rossi 2018; Russell 2019).

To address this complexity, different categories or phases defining (possible) developments of AI are introduced. For example, the struggle to reach the expected progress of AI and its analytical capacities close to human intelligence in the late 1980s was called the "winter of AI" (Aradau and Blanke 2015, 5). This *stillness* has created path dependency for future transformations and the assessment of developments in AI. For example, recent breakthroughs in machine learning have already been considered as an *AI spring*. Another proposed categorization is symbolic AI, statistical AI, and sub-symbolic AI. These categories are based on different capabilities accordingly: AI that is able to follow in the process of problem solving, the capacity to catch trends from large data sets, repetition, experimentation and feedback, and lastly machine learning and deep learning (Woszczyna and Mania 2023).

A rather similar but clearer categorization is narrow AI, general AI, and superintelligence. These categories appear developed in relation to human intelligence and the comparative level of AI capacity. Current AI applications, such as facial recognition, the processing of a large amount of data, and intelligence gathering, are considered as narrow AI, which is already widespread. Large language models such as ChatGPT or generative AI[4], despite the increased hype and interest, still fall under the definition of narrow AI. Then general AI is supposed to be human-level intelligence, which has cognitive, creative and emotional capacities. Even though this is one of the main directions of current developments, resembling brains in terms of broader cognitive architecture remains "still a long way" (sheffield.ac.uk

[4] The EC defines generative AI systems as those that generate, in response to a user prompt, synthetic audio, image, video or text content, for a wide range of possible uses, and which can be applied to many different tasks in various fields (Ec.europa.eu 2024a).

2023). Lastly, superintelligence is the one that surpasses human-level intelligence, and remains the most future-oriented (Price, Walker and Wiley 2018). Of course, this distinction is relative, as the actual possibility of reaching general AI and superintelligence remains unclear. David Wallace-Wells suggests ironically that the idea of superintelligence is defined in such an ambiguous way that it sounds like a "benevolent genie that solves all the world's problems" (Wallace-Wells 2019, 155).

It is no surprise that superintelligence brings a major intrigue not only about reaching human intelligence but also surpassing it. As Nick Bostrom (2020) suggests, superintelligence would be the last invention of a human being, because AI would take control of further AI developments. In this case, the hierarchical structure, considering humans *on top*, would change problematizing human-machine interaction even more. Being imaginary of an undefined future, superintelligence has already received both positive and negative considerations, fuelling further debates. For example, extinctionalists would argue that superintelligence is another step in evolution, while others point out concerns related to human survival. These different perspectives are noticed in the policy paper by Allan Dafoe (2018, 10), where he states that "superintelligence offers tremendous opportunities" and, at the same time, "superintelligence may also generate catastrophic vulnerabilities." Although these scenarios sound futuristic, they already fuel public and political debates, which do not fix to the arguments of *possible* and *impossible* AI developments.

The discussions also push us to take a side over how technology and its role are perceived in creating or mitigating concerns: as a fix for insecurity, as a security challenge, security as a fix for technology, or security as a barrier to technology (Haddad, Vorlíček and Klimburg-Witjes 2024). Existing analyses show that AI is viewed not just as a security challenge, but also as a solution to various issues. For example, a potential solution to securing people's needs and addressing concerns (Schmid, Pham and Ferl 2024), or as an innovative element which becomes transversal to all policy fields (Bellanova et al. 2022).

Accepting this diversity and complexity, I focus in this thesis on AI as a socio-technical security challenge, and not as a deterministic solution. This perspective calls for deeper scrutiny of underlying political, social and security issues that are often neglected in technocratic approaches (Ulnicane and Aden 2023; Suchman 2023). It includes considering AI developments through potential risks, ethical issues, power imbalances, already-evolving international dynamics, governance structures, and social relations

(Csernatoni 2024a). The analysis of AI policy documents by Inga Ulnicane (2022b) already reveals that AI is associated with major changes, emerging global competition, and ambitions to be leaders in AI, where policies should help balancing benefits, risks and responsibilities in the development and use of AI as a revolutionary, transformative and disruptive technology. Such a broad consideration of the role of AI suggests that the definition and perception depend on individual cases, and require contextualized descriptions of AI in a particular landscape and characteristics, rather than focusing on whether one definition is better, worse, abstract, or more accurate.

As a researcher in social sciences, I do not claim any expertise in the development, application, design and test models of technology, and variations between them. The focus of the research is oriented towards the politically articulated discourse and power relations, by asking how technologies are "embedded within a particular (political) narrative" (McCarthy 2018, 34). Established vocabularies and knowledge production in defining current political actions or related speculation also require a closer look at articulated meanings of AI, asking how AI is understood and entangled in social, political, legal, moral and economic contexts (Leese 2019). That is why a pre-established definition of AI is not used in the thesis, but it is a part of investigating the ways of construction by a concrete actor and shaped by outlined priorities.

1.2. Automation – autonomous – autonomy

The introduced debates on aiming to define AI have a foundational issue of not knowing to what extent AI can be developed and can function *separately* from humans in a world of *its own* (Descombes 2010). Such a discussion, resembling different definitions, also relies on three notions – automation, autonomous and autonomy – which suggest a problematic conversation on what is already happening, and how the future of AI is imagined. Despite their differences, these stages are often mixed between each other, complicating discussions about human-machine interaction, and raising an urgency reflected in emerging AI polices. These notions are particularly important, not just for categorising different capabilities, but projecting a level of human control which is a fundamental issue in outlining legal, political and ethical boundaries.

To start with automation, which is sometimes used as *automated*, it is associated with the *workload distribution* between humans and technologies – AI being capable of processing requests, absorbing and navigating through

data, by providing answers or completing tasks by itself. Matthias Leese (2019) argues that automation is not a question of either full human autonomy or full machine autonomy, but complex and distributed dynamics of agency. It is still about the existing hierarchy between humans and machines, where the interaction is between the operator and the technology. For example, automation requires data input and the human interpretation of output, as well as potential human interventions hoping for better technological performance (Kaber 2018). Automation is seen more as *mechanization*, where the human-machine interaction is about remaining as human monitoring and potential intervention, which supports technology in a greater overall performance (Kaber 2018).

Some analyses consider AI as already being *autonomous*, and outpacing human capacity in processing information and its quantities (Suchman 2023). Autonomous technology can be described as dependent on inputs but, unlike automation, it can change its internal states without the direct intervention of humans (Taddeo 2024). Here, human-machine interaction becomes even more complicated, because it is not clear if AI can function without human involvement (Holmqvist 2013), or if autonomous technologies will move in terms of their own logic, divorced from human control (Nicholson and Reynolds 2020). Although it is still argued to be a human prerogative to *decide* whether or not to use AI, and to what extent and in what context (Christen, et al. 2023), the outcome of a given task comes from a "black box" – an increasing challenge for humans to understand and explain how AI achieved the result. Therefore, *autonomous* becomes controversial, as challenging both human control and the human capacity to grasp the implications of such a (potential) nature of AI. Compared to automation, *autonomous* technology also disrupts the human hierarchy attached to supervision and intervention.

When it comes to autonomy, it is much more focused on the matter of agency – as self-governance and an ability to carry out a task without the intervention of humans, and to demonstrate self-governance (Roff and Danks 2018). Autonomy challenges the issue of control – what is the role of humans, and who/what takes responsibility for the decisions made (H.-Y. Liu 2019); will technology be able to select between options and behaviour without external command (Dignum 2019)? Also, to what extent is the autonomy of AI possible when human *autonomy* is attached to characteristics such as dignity, meaningful control and consciousness (Vesnic-Alujevic, Nascimento and Pólvora 2020; Bhuta, Beck and Geiß 1920). The matter of "cognitive" capabilities associated with humans and human agency also remains a question of AI and autonomy (Kaber 2018, 408): if technology does not

perform them, can we still discuss AI autonomy? Therefore, the difficulty comes from the human perspective to comprehend AI autonomy, which is not necessarily the same as human autonomy, and signals something of a possible collaborator, fundamentally different from automation as tool (De Visser, Pak and Shaw 2018). Thus, considering these diverse perspectives, I follow here the description of autonomy as conflated with a lack of human control and the capacity to function without human assistance.

This brief overview of three notions is closely related to the different definitions of AI mentioned before: from narrow AI to superintelligence. They all appear to be interconnected, because they share similar considerations: what artificial is and what intelligence is, how to characterize it, and particularly, how those characteristics and future developments will *comply* with visions and concerns related to human-machine interaction. As has been mentioned, the different stages and definitions are directly attached to the human role and control, whether it moves away from immediate decision making and *manual* control, overseeing operations and having a formal ability to override decisions, to being completely absent from operating technology (Bode and Huelss 2018). These identified differences and their logic are key for the following analysis, where automation, autonomous and autonomy become mixed in AI-related imaginaries and political speculations of the future of AI.

1.3. AI in security

Issues surrounding AI in security arise across both military and non-military domains. While the contexts differ, they are linked by recurring concerns over who controls AI and who is responsible for the decisions technology enables. The following discussion suggests that these questions place human-machine interaction at the centre of the debate, making the conceptual lenses introduced in this section especially relevant across different areas of application.

In terms of non-military context, Mariarosaria Taddeo (2024), for example, suggests that security issues arise in different spaces, areas and forms, such as unjust violence online, cyberattacks, and breaches of fundamental rights. AI is also attached to the potential violation of privacy, the creation and spread of disinformation, surveillance, and social control (Agüera y Arcas 2023). This also raises questions of who develops and uses AI to amplify those challenges. Algorithms, data and computing power are primarily controlled by platform-based companies, which not only govern technological infrastructures, but also exert an influence over content

moderation, democratic processes, and the public discourse (Gu 2024; Bradford 2023). This concentration of power reveals monopolistic and oligopolistic tendencies, raising significant concerns about their implications for democracy and human rights (Hoijtink and Van der Kist 2025). In short, this diversity indicates that non-military security issues are not focused on a particular dimension (for example, societal or economic security), but cover a spectrum of issues and actors involved.

In terms of the military, AI is mostly attached to applications to weaponry, the transformation of the battlefield, and ethical concerns. For example, James Johnson (2020a; 2020b; 2019), from the strategic studies point of view, talks about a potentially intractable AI arms race, the complex interplay of advanced military technology, and military superiority in deterrence. Allan Dafoe (2018) argues that AI might have a transformative effect on international security by influencing the security of nuclear retaliation, the stability of crisis escalations, and the future balance of power. Various reports, based on existing evidence in the battlefield, suggest that AI has already been employed by the Ukrainians for counterintelligence or identifying violators of sanctions, for use by drones and their targeting, or to track troop movements and communications (Economist.com 2024; Adam 2024; Sanger 2024). At the same time, these different applications demonstrate that the spectrum of AI uses is also wide, even in just the military domain.

From a more critical perspective, the role of AI in the military is mainly contested due to blurring the lines between civilian/commercial and military uses. As Basham, Belkin and Gifkins (2015) suggest, the duality between military and civilian is not taken as natural, but as a process of construction, constitution and contestation which needs to be explored. Therefore, discourses of an AI-led arms race, or power competition, represent trends and hypes in constructing the role of AI, which still need to be processed and unpacked. Repeated claims about the inevitable doom or salvation of AI becomes a fact, the fact becomes what is sacred, and what is sacred must be defended (Schwarz 2025).

The discussions of AI in the military introduce vocabularies that often extend beyond military contexts, influencing broader conversations on AI and security. Notions such as *meaningful* human control, morality, and human dignity further continue the conversation on what forms of human-machine interaction emerge, and how they affect humans. For example, humans are presented as moral agency enactors, ensuring that decisions on life on the battlefield are not made by "non-moral artificial agents" (Amoroso and Tamburrini 2020). Morality becomes a virtue "anchored in our history of

human social relations", and functions as the "capacity to take responsibility and feel the weight of morally complex decisions" (Schwarz 2021, 63). Delegating decision-making, and control, to machines is directly associated with the human role moving away and establishing a new normality (Bode and Huelss 2018; Bode 2023). For example, the *human-in-the-loop* is introduced to define human control over technology, the *human on-the-loop* means overseeing operations conducted by technology, and the *human out-of-the-loop* is human control being absent from operating systems (Bode and Huelss 2018; Christie et al. 2024). These terms, overlapping with the notions discussed in Section 1.2., reflect key characteristics and priorities, which remain concentrated on constantly asking who controls, and to what extent. Therefore, proposals on how to meet this expectation of meaningful human control vary from establishing key boundaries on AI autonomy to safeguarding a form of human decision making, continuously considering what human agency, attached to decision-making, independent actions and moral responsibilities, is, and how it differs from technological artefacts or systems (Bode and Huelss 2019).

At the same time, the previously mentioned point of mixing current AI developments with future imaginaries remains equally relevant to security debates. Suggestions that, despite future anticipation, humans still make design decisions on technologies, data and ways of using (Christen et al. 2023), and are confronted by claims that human involvement, whether on the battlefield or in decision-making, is thought to be gradually erased (MacDonald and Howell 2019; I.G. Shaw 2017; I.G.R. Shaw 2017; I. Shaw and Akhter 2014). Then, security is not only about a battlefield and the role of technology, but what happens to humans and their agency, especially if technologies embody more than human qualities (Bourne 2012).

Conclusion

This chapter has demonstrated that the term *artificial intelligence* is a generalization, and does not necessarily represent the complexity of the issues involved. Debates on AI also depend on the specific field and use, which vary from the military to ethics. Referring to AI as a *separate tool* might also overlook the analytical sensitivity to nuances of technological functioning, enabling power and problematic present/future differentiation. However, as was mentioned earlier, I choose it because of the political landscape, specifically the emerging EU AI policy, as well as the relevant academic debates proactively referring to AI not only as technology but also a socio-

political phenomenon. Considerations of the definition of AI also represent a certain paradox between technical characteristics, where AI functions as a "technology toolbox" made up of algorithms, data and platforms; and perceptions where AI is viewed as a form of "human intelligence" (Nitzberg and Zysman 2022, 1755).

The chapter also outlined key questions problematising human-machine interaction, and further blurring the line between what users want to do and what technologies are capable of doing, in sharing tasks and making decisions (Neff and Nagy 2018). Arguments that current developments and capabilities of AI remain within the scope of *narrow AI* and signal *automation* rather than autonomy do not stop speculation over what it might become in the future, and how it could challenge human agency. Briefly overviewed concepts such as automation, autonomous and autonomy, to define technological functions, and (meaningful) human control, human in/on/out of-the-loop and morality, lead the conversation about different proposals both to describe perceived issues and establish a distinction between humans and AI. Table 1 summarizes the main points in how these different stages are described and further interpreted in the analysis.

Table 1. *Summary of AI developments and the level of control*

	Automated/ Automation	**Autonomous**	**Autonomy**
Function	Workload distribution between humans/operators and machines/system; Data input, interpretation of output and intervention by humans;	Capacity to interpret information and identify a course of action; Change of internal state without the direct intervention of humans;	Self-governance without human intervention; Capacity to function without human assistance;
Level of control	Manual control by humans; *Human-in-the-loop* principle.	Humans oversee operation and have an ability to override systems decisions; *Human-in-the-loop* principle.	Human is absent from operating systems; *Human-out-of-the-loop* principle.

Source: the author, based on the discussion in Section 1.2.

Overall, the discussions overviewed provide broad guidelines on the main themes, hypes and vocabularies. They reveal how AI-related perception ranges from being an opportunity to a concern. In most cases, the discussions centre on speculative scenarios and visions of technology (such as superintelligence and autonomy) that cannot be separated from the ways political actors define AI in preferred ways. At the same time, the production of knowledge and its embodiment in political, social or economic worlds do not come from nowhere, but are shaped by humans, institutions and imperatives that determine what they do and how they do it (Crawford 2021). Therefore, varying considerations and concerns relating to AI ae not *pre-set*, but require further investigation.

2. CONCEPTUAL FRAMEWORK

The already presented discussions have demonstrated that three elements, technology, security and risk, and their intersection, emerge in tension, and require further conceptualization in the thesis. Conventionally, security focuses on protection from threats which may cause irreversible damage (Farrand 2020). However, from a more critical point of view, the question is who decides the boundaries of security and insecurity, and how, especially when this distinction gets more difficult to draw (De Goede 2020; Evans, Leese and Rychnovská 2021). Therefore, Section 2.1. discusses principles of social construction, discourse and imagination which offer a baseline for the analysis, and continue the conversation of how the interrelation between security and technology can be conceptualized.

Risks further complicate the challenge because they mark a different level of concern than threats, as they refer to considerations of potential harm and probability occurring in the future. As David Campbell (2008) suggests, risks are not equal, and not all of them are interpreted in the same way. Therefore, like the point of security, it is a question of how risks are identified, and how the boundaries between risky and non-risky are established. To follow this conversation, Section 2.2. explores the concept of a risk society introduced by Ulrich Beck, and the continuous debates by critical scholars suggesting that risk should not be taken as a given condition of being.

Lastly, *riskification*, built on the logic of securitization, serves as a valuable conceptual framework to be further developed for this thesis. When combined with critical perspectives, riskification refers to deliberately done speech acts which signify constructing an issue as risk issue and providing an interpretation of perceived reality (Rothe 2012). Section 2.3. focuses on presenting the main claims of riskification and the proposed shift from immediate and extraordinary to long-term anticipation and governance. Based on this, I develop four analytical elements: a referent object, conditions of possibility for harm, a governance programme, and international engagement, which lead to an adapted and tailor-made analytical framework of riskification that is later applied to empirical analysis.

2.1. Social construction and knowledge production

To start with broader assumptions, references to social construction situate the analysis within the discussions considering how we create meanings and why we think of knowledge in a particular way. Following David Campbell (2008,

x), the construction evolves through the inscription of boundaries between "inside" and "outside", "self" and "other", or "domestic" and "foreign". We are not talking here about something finished, but "in the process of becoming", where the "self" is portrayed through something normal, accepted and desirable, and the other, something dangerous, different and external. For example, in terms of security, this enables us to ask how certain issues are constructed and enemies identified, and then how the security card helps to forge alliances, blame others, or raise popular identification with certain issues (Rothe 2012). In the case of technologies, this perspective is also about challenging the positivist framing of innovation as an end in itself, by asking how they are related to and shape social, institutional, political and security dimensions in specific contexts (Csernatoni and Martins 2024). Therefore, adopting a social constructivist lens in this analysis draws attention to how specific meanings of technology and security are constructed, justified and articulated; what visions of politics they promote or encode; how, where and by whom they are contested; and what broader implications they entail.

Moving to technology, the varied definitions of AI presented in Chapter 1 demonstrate that technology is shaped by social contestation, and reflects social, economic and political power (Nicholson and Reynolds 2020). I follow the argument that separation between technology and a political process creates a false dichotomy, because values are delegated to the way we understand technology, while technologies also affect the ways we understand ourselves and the world (Lidskog and Sundqvist 2015; Jasanoff 1996; Feenberg 2017). Therefore, the perspective of social constructivism is useful for this thesis, because it enables the analysis to ask how technologies are perceived, and how established definitions are reproduced or challenged (Lindekilde 2014).

Although the position of considering reality as socially constructed has been criticized for "questioning facts", the increasing denial of issues like climate change (Prasad 2022, 90), or reinforcing post-truth (Aradau and Huysmans 2019), the argument of social constructivism does not call technologies, and social and political practices, completely relational. It is about investigating relations, perceptions, meanings and power dynamics beyond the position of something being given or self-evident. Claiming that the EU's approach towards AI is different from, for example, China's, and entails EU-specific interpretations, does not change the fact of existing AI-driven systems, infrastructure and labour related to their functioning. The questions are about the embodiment of AI in a social, political and economic world and becoming an expression of power (Crawford 2021).

This argument relates closely to some critical security scholars engaging with science and technology studies (STS), who analyse how technologies actively shape security practices, and seek conceptual tools to account for their co-constitutive role in security dynamics[5]. Even though I do not directly apply any of these conceptual proposals in the analysis, the overall logic of inquiring how scientific and technical knowledge is produced, how the beliefs of science and concepts such as agency, system, power and structure to be rediscussed (McCarthy 2018; Jasanoff 1996; Dear and Jasanoff 2010) correspond to and support the overall argument of analysing technology and technology-related policies as social constructions. In other words, critical discussions on technology and security suggest that it is not a matter of a *laboratory*, technical and technocratic matters but, from the Latourian point of view, a "fierce fight to construct reality" (Latour and Woolgar 1986, 244). This leads to claims that technology, in this case AI, is shaped by cultural and societal particularities, and political structures, and participates in the production of security knowledge.

AI-related knowledge and policy-making also face an important question: is it about current forms of AI capabilities and leading understanding, or the anticipation of the future? Chapter 1 has already demonstrated that discussions and perceptions of AI vary, from narrow AI to superintelligence, from automation to autonomy. Therefore, the construction of that knowledge involves imagination and imaginaries which co-produce our understanding of which AI capabilities *exist* and which remain a matter for the future. Imagination then comes as an important (de-)legitimization tactic, since these imaginations and ideas of the future influence what actors perceive as potential pathways for technological developments and viable options for dealing with them in the present (Ferl 2024).

[5] For example, the increasing popularity of sociotechnical imaginaries in state or corporate approaches to technologies and their framing (Jasanoff and Kim 2009; Hockenhull and Cohn 2021; Klimburg-Witjes 2024). The increasing role of non-state actors and the involvement of big tech companies in shaping security (Planqué-van Hardeveld 2023; Bellanova and De Goede 2022). Attention to security infrastructures, their performativity and the role of security devices, their employment in public spaces or even surveillance (O'Grady 2021; Aradau 2010; Amicelle et al. 2015; Suchman et al. 2017). The use of actor-network theory demonstrating complex interactions between actors and technologies in different domains (Verbeek 2013; Balzacq and Cavelty 2016). Security assemblages as complex modes of knowledge, devices and institutions (Aradau and Blanke 2015). The reconsideration of warfare and the role of military due to technological developments (Gilbert 2019; Hoijtink and Planqué-van Hardeveld 2022).

Instead of employing the notion of sociotechnical imaginaries often used as conceptual exploration at the intersection of STS and critical security debates, I focus here on imagination and imaginaries in a rather loose way: the creation of new forms of knowledge, which, in the context of AI, bring a political vision of the ideas, risks and issues oriented to the future that raise the need to be addressed (Jasanoff and Kim 2015). For example, imagination could entail the conception of a desirable future or societal progress achieved through technology, or the concern of technologies transforming future wars or functioning without human supervision. Therefore, knowledge construction is not only about what we already *see* or *have*, but also which *unknown unknowns* we try to tame and already involve in policies, although it is still not possible to fully grasp and predict future developments and uses for AI (Smuha 2025).

The relevance of imagination and the future could raise further questions if we can use the same approach to make sense of the future as the approach of constructing reality. To respond to this, I argue that the future is anticipated, modelled and shaped through the same discursive tools we use to interpret the present. The way we envision the future also reflects current power structures, values and historical narratives, while the presentation of future scenarios is nothing other than imaginative constructs "no matter how 'expert' that imagination is" (Stevens 2017, 104). For example, the shift from the examination of the present state towards the imagination of future horizons and developments still characterizes contemporary AI myths (Natale and Ballatore 2020). At the same time, current debates about AI are criticized as a "failure of imagination", as they continue to rely on age-old concepts of competition, geopolitics or race to interpret the AI-related international landscape (Csernatoni 2024a, 23).

Thus, arguments of social construction demonstrate that the intersection of technology, security and risk is intersubjective, relying on shared understandings, interpretations and meanings. Even in terms of scientific *facts* associated with technology, this perspective takes a step outside the predominant dualisms of subjectivity and objectivity or right and wrong, and question these categories (Jasanoff 1996). Investigating how such knowledge is produced, and what patterns and tensions emerge, is also an inquiry into how actors establish priorities and shape the vocabularies through which issues are understood and acted on. References to power, competition, struggle or interests are therefore approached as products of specific historical and discursive contexts, revealing how security dynamics are framed, and how actors' roles and positions are shaped through these processes.

2.2. Risk society

This section discusses the relevance of risk in analysing AI-related security. As AI merges both the present and the future, risk similarly refers to the limited knowledge of future events which might be perceived as unpredictable, novel and unknown. Claudia Aradau (2014, 75) has even called this the "epistemic regime", which produces categories, boundaries and timeframes transforming unknown, unpredictable and uncertain into something identified and known that can already be acted upon in the present (Kessler 2007; Kessler and Werner 2008; Amoore 2013).

The more intense engagement with the concept of risk among critical security scholars was noticeable after 9/11, and particularly in the 2010s, as a turn to the issue of uncertainty. Berling et al. (2022) provides a highly illustrative curve suggesting how the "oscillation of conceptual tightness in security studies" has been evolving. It suggests that debates on risk and/or security peaked from 2010 to 2020, and received variations in topics such as terrorism and its financing (De Goede 2020; 2018), climate change (Rothe 2011; Benner and Rothe 2024), or border control (Broeders 2007). Analyses and debates led to conceptual novelties based on the logic of risk such as *politics of possibility* (Amoore 2013), *speculative security* (De Goede, Simon and Hoijtink 2014; Goede 2012), *degrees of riskiness* (Amoore and De Goede 2005), or *risk profiling* (Leese 2014).

These debates, and new conceptual introductions, could raise the question to what extent risk remains a relevant conceptual lens, or do contemporary issues and their complexities require proposed modifications of conceptualising and analysing risk. For example, the notion of resilience has been developed to emphasize the need to be prepared for unknown risks, adaptation, and flexibility to respond to shocks (Juncos 2017). In this case, the resilient subject (which can be both individual and collective) is not conceived in a passive way, but as an active agent which seeks resilience as a goal rather than a final state of being (Chandler 2012). Then, despite the detrimental events and their severity, stability and safety are maintained (Dunn Cavelty, Kaufmann and Søby Kristensen 2015). Even though resilience relates to the problematization of temporalities (present and future), and refers to uncertainty, it is considered as marking different epistemic regime than risk. Claudia Aradau (2014, 79) argues that risks are about endeavours to calculate, patterns, multiples and averages, where uncertainty could be "tamed"; while resilience relates to preparedness for surprises in an interconnected world "where the novel and the unexpected are always emergent". In the case of AI,

this would mean that any stage of AI developments and uses (as discussed in Section 1.2.) would not create adverse or damaging effects in states and societies considered as sufficiently resilient.

Despite evolving academic trends and shifting conceptual frameworks, ranging from resilience and prevention to precaution and resistance, I retain a focus on the concept of risk and the risk-security nexus. This decision has several reasons. Firstly, risk remains relevant because of its extensive involvement in *political* discourses. Risk has become a central concept not only in the EU's case, but an overlapping tendency in different AI-related policy proposals, involving a multitude of private and public actors, purposes and instruments, including the EU (see Section 4.3.) (Macenaite 2017). Therefore, such a tendency requires unpacking this choice and any AI-related specificities that emerge from considering technologies and issues through risks. Leaving it as *given* or *outdated* raises the question of who then scrutinizes political imagination and a toolbox employing risk.

Secondly, even though concepts such as resilience are considered as the expression of "postmodern concerns with survival" (Stevens 2017, 116), risk remains conceptually discussed in matters such as climate, disaster management, and cyber or other technology-related hazards. This proves the importance of conceptualising and understanding risk in contemporary issues which are beyond conventional security understandings. Thus, the thesis aims to continue the conversation of how risks are constructed, and how certain issues become risky and what security understanding the concept of risk signifies, if it is no longer about immediacy and an exceptional circumstance.

To understand this logic, Ulrich Beck's (1992) notion of risk society serves as a relevant framework because it defines risks as the inevitable consequences of societal, political and economic advancements tied to advanced industrialization, which, in turn generate negative environmental impacts and new globally perceived risks. Modernity is thus portrayed as a continuous and deterministic process that inherently produces and amplifies risks.

In his later writings, Beck (2006) introduced the concept of *global* risk and globalized risk society as a human condition of the 21st century. These borderless formations suggest that various social groups may face similar challenges, often without fully realizing it (Güntay 2020). The policy goal then shifts to preventing the worst outcomes and adopting self-limiting measures. According to Beck (1992), this can create dilemmas between normality and crisis, democracy and authoritarian responses leading to a state of emergency becoming the new normal. Thus, the risk society is

characterized by an ever-present awareness of potential dangers, suggesting an important tension between conditions of not knowing and *unknown* becoming the condition of being.

In the case of AI, the risk society would be defined by existing and evolving technological changes which put societies at risk by concerns of what effects these changes will have: "we do not know which risks from AI will become most salient and we do not know if AI is the most salient risk" (Boyd and Wilson 2020, 57). Therefore, the embodiment of AI-related challenges and future technological developments would be interpreted as deterministic, *the new industrialization*, which similarly reiterate the logic of risk society as a condition of being. In his article dedicated to terrorism, Beck (2002, 41) used the concept of "uncontrollable risk" to emphasize globality and borderless unpredictability. This point also reflects the case of AI, where developments and uses are *uncontrollable* by existing policy measures and become proof of the society that lives in the world of risks.

However, this perspective raises further questions as to what extent we can consider technological developments and uses as being *out there* and then directly defined through risk as a *condition* of being. Beck's conceptualization has received similar reactions from critical scholars. For example, the presentation of risk society as *objectively existing* is considered eliminating questions of biases or a lack of knowledge in defining risks which have not yet materialized (Williams 2008; Clapton 2014). Another related point is raised with risk itself: risk society does not say about risk more than risk being a condition of modernity and transformation (Elbe 2008). For example, Merryn Ekberg (2007) argues that in Beck's conceptualization, risk is associated with disaster, and the extreme of catastrophe, without any differentiation of any other possibilities, ignoring the potentially different nature of risks (technological, natural, civilizational, etc). That is why Claudia Aradau, Luis Lobo-Guerrero and Rens Van Munster (2008) criticize Beck's risk society as failing to capture the fact that risk is a social technology by means of which the uncertain future, be it of a catastrophic nature or not, is rendered knowable and actionable.

Overall, several points describing the relevance of risk for the thesis are distinguished:

- Beck's introduction of risk society marks a transformation from known to unknown, from measurable to non-measurable, and from security to risks (Petersen 2012);

- From a more critical perspective, risk is not *out there*, but fluid, flexible and political, where different forms and rationale may be used in applying this concept (Hannah-Moffat 2019);
- Instead of a *given* condition of being, risk is employed as epistemic ordering by different actors, to make sense of a future-linked phenomenon that is aimed to be understood and addressed in the face of uncertainty.

2.3. Riskification

The two previous sections have already demonstrated that technology, security and risk are not considered as technical, neutral or given, but socially constructed depending on actors, and social and political contexts. This still leaves the question how a certain issue becomes considered as risk, and how it relates to security. For example, do explicit references to risk, and the readiness to mitigate them, increase or reduce the sense of insecurity?

The initial response leads to securitization, one of the well-known constructivist mechanisms in transforming an issue into a security issue. Naming a problem as "a security problem" provides a different status and priority to an issue, compared to those "non-security problems", because it is articulated through survival claims and the urgency to act (Hansen and Nissenbaum 2009, 1156). Securitization as a mechanism contains several steps. Firstly, it is performed through speech acts where, in the case of a successful process, an issue is transformed into being a security problem (Wæver 1995). This is how political elites explicitly start the process of securitization, and justify proposed actions to address a security issue (Markiewicz 2024). Secondly, speech acts, as discursive practices, name something s an (external) enemy or threat to something or someone that needs to be secured. Thirdly, as a response to a security issue, extraordinary measures are introduced to deal with it. They are also considered as a political move, because security justifies them as exceptional and urgent, even if extraordinary measures may limit rights or impose greater control over others (Buzan and Wæver 2009).

But what happens if, within the context of established power relations and institutional structures, a "security problem" is not defined in this way. Especially, if political elites *choose* not to establish a discourse of threat, and do not search for legitimacy to adopt extraordinary measures. This possibility of various ways to construct security is already supported by analyses indicating that technologies can be both securitized and desecuritized (Sezal 2023); and varying degrees of concerns may emerge, excluding threats from

consideration (Methmann and Rothe 2012). This complexity is particularly relevant in the case of AI, where concerns and the responses to them are difficult to put into the framework of immediacy and threat, while AI remains a matter of (political) imaginaries: how and in what ways technology will be developed, used or misused.

To address this tension of who constructs AI-related security, and how, I turn to *riskification*, which identifies how risks are situated and contextualized, and how they are put through established policy measures (Harijanto 2025). I find the concept of riskification particularly relevant for two reasons. First, rather than treating risks as objective or given, it details the constructivist process of turning an issue into a risk, and deconstructs a political reasoning in articulating antagonism which constitutes the order that it is sought to be governed (Rothe 2011; Methmann and Rothe 2012). Second, I suggest that a chosen strategy to formulate and address security concerns through risks involve not only political motivations, but also the type of concerns that are perceived. In the case of AI, it is considered as an emerging phenomenon, which developments and uses are still difficult to grasp. Therefore, instead of threats and immediacy, suggested by securitization, AI-related uncertainty, future-oriented imaginaries, and a search for a response, make the concept of riskification a suitable analytical lens for understanding how security is framed in relation to AI and related risks.

Riskification as a conceptual framework has already been applied in topics such as climate change and cybersecurity: policies that extensively employ risks. For example, the analysis of riskification in climate change adaptation has demonstrated that the issue has been approached through a technocratic governance of existing risk instruments (Barquet et al. 2024). Climate change in the military has been riskified, legitimizing the use of military means to address its harmful effects, even though these measures are presented as anticipating uncertainty rather than responding to direct threats (Estève 2021). In the case of cybersecurity, riskification has revealed that the logic is focused on traits of long-term, permanent and/or future danger to societies stemming from cyberspace. Then the response is long-term social engineering and the management of causes of harm without articulating emergency (Backman 2023). These empirical cases demonstrate that riskification problematizes the employment of risk as expressing *the truth*, and demonstrates how risk is constructed to define and shape the understanding of an issue (Morsut and Engen 2022). The application of riskification also helps us to understand established power relations and institutional structures which enable or constrain proposed policy responses (Judge and Maltby 2017).

To be specific, Olaf Corry (2012, 246-247) develops and grounds riskification within the framework of securitization and its key elements, illustrating that risk, like security, is similarly constructed through speech acts, rather than being a condition or a given fact, as in the case of risk society. According to him, riskification is "a kind of second-order security politics", and "captures this idea of a social process of constructing something politically in terms of risks". It does not directly answer the question of dangers and concerns, but provides a mechanism of how risks are constructed. What is the difference here? In a conversation with securitization, Olaf Corry (2012) suggests that the difference between risk and threat is defined through a lack of immediacy. As mentioned in the Introduction, while threat is to be defended against or eliminated immediately, risk is discussed as being located in the future, which is connected to a policy proposal offering a way to prevent that risk from materializing into real harm. Thus, compared to securitization, riskification is no longer about threats, friend-enemy contestations, urgency, or the need to take extraordinary measures. The focus is on the perception of future concerns, their mitigation, and management by policies taken which characterize risk itself.

Despite these introduced differences between security and risk, Olaf Corry (2012) recycles the main elements of securitization, by transforming them into the process of riskification. He distinguishes three elements: 1) there is a referent object of which the safety is challenged through potential vulnerability; 2) those challenges can be described by identifying conditions of possibility for harm (rather than direct causes of harm) to a referent object; 3) to control those conditions, governance programmes are developed which turn a referent object into a "governance-object". Therefore, in comparison with securitization, these elements could be noticed: a referent object, something that needs to be protected, remains. A threat or enemy from securitization is transformed into conditions of possibility for harm, which marks a potentially different level of danger, less certainty and orientation towards the future (*conditions of possibility*). Extraordinary policy measures as an urgent response to threats are changed to governance programmes as longer-term mitigation, signalling a different level of urgency, timeframe and orientation towards the future. A summary of this conceptual shift and a comparison of the key elements is suggested in Table 2.

Table 2. *A comparison of key points of securitization and riskification*

Components	Securitization	Riskification
Focus	How does an issue become a security problem?	How does an issue become a risk?
Instrument	Speech act	Speech act
Actor	Elites	Elites
Timeframe	Immediacy	Long-term future orientation
Target	A referent object	A referent object
Concern	Threat	Conditions of possibility for harm
Response	Extraordinary measures	Governance programme
The goal of the response	Defence and deterrence	Mitigation and management

Source: the author, based on Olaf Corry's conceptualization of riskification

While these elements, referent objects, conditions of possibility for harm and a governance programme, explain the proposed shift, they are not detailed, and require to be further elaborated and clarified, particularly in the context of AI-related security. For this analysis, I also adapt and expand the conceptual framework from three to four elements, by adding the importance of the international engagement. I introduce 'international engagement' as a fourth analytical element of riskification for two main reasons. First, contemporary debates and policy-making around AI increasingly interrelate in articulating AI and AI-related risks as international and geopolitically contested matters. Second, this addition reflects the EU's own tendency to internationalize perceived risks, entangled with the EU's subjectivity and projection on the global stage.

Then all four can be grouped into two guiding questions: *what*, what is a referent object, what is potential harm; and *how*, how governance is constructed and how international engagement is defined. While the *what* questions specify set priorities and vocabularies, the *how* questions aim to propose responses to identified concerns and protect what is deemed to be at risk.

Overall, the conceptual framework that I develop for the thesis focuses on four key questions. What is a referent object? What conditions of possibility define harm? How is a governance programme constructed? How is international engagement framed? The search for answers through these analytical elements is expected to lead to the identification of AI-related security and its key characteristics. The summary of the questions is presented in Table 6.

1) A referent object

The question of what a referent object is requires a broader elaboration of how I consider it in this thesis, especially moving from more conventional security definitions to the logic of risk. The basic idea of a referent object refers to who or what, depending on priorities, is to be secured, what the actual or potential challenges to a referent object are, and what actions could be taken to overcome them (Shepherd 2013). Olaf Corry (2012) suggests that in the case of risk, a referent object is the primary target of governance programmes, something that needs to be changed and governed rather than defended. A referent object is not necessarily *an object*, but includes considerations of the properties of that referent object which depend on constructions of risks and established knowledge. Then the safety of a referent object becomes a function of characteristics of that referent object and its vulnerability to danger.

Based on this conceptualization, my search for a referent object relies on several points. Firstly, I reiterate the argument that riskification is rooted in speech acts as discursive practices. Therefore, the identification of a referent object is not pre-determined, but emerges from specific discursive practices within a given context, requiring an in-depth engagement with the discourse. Secondly, riskification extends beyond the identification of a referent object to include defining characteristics and political priorities. They also reveal how a referent object is framed and understood. Thirdly, the interplay between the referent object and the identified risks is embedded in specific concerns that (potentially) challenge the safety and stability of that object. Therefore, the question is what and in which ways safety is attached to a referent object.

2) Conditions of possibility for harm

Riskification does not simply refer to harm, but constructs this elaborative notion of *conditions of possibility for harm.* Olaf Corry (2012) grounds it as a way to escape conversations on the direct causality between harm and risk, and refer to potentiality, conditionality and precautionary, where risks come *as* conditions of possibility for harm. Even though the goal is clear, to create a noticeable contrast between threats, their immediacy and urgency, the definition remains challenging, especially due to the orientation towards the future. In this case, conditions of possibility for harm even require us to break down *possibility* and *harm* to grasp their relevance while conducting the analysis.

Here *possibility* is intriguing, because it refers directly to future unknowns that cannot be easily calculated and predicted: something may or may not happen. Therefore, a possibility changes a more conventional understanding

of harm as something that makes risk known or foreseeable because of already-existing effects or the considerable certainty that they will occur in the future (Smuha 2021a). In this context, possibility increases the importance of imagination, which is less objectified and measured, but much more about the future and subjective views (Amoore 2013; Adams 2005).

Focusing more on harm, it does not receive a single interpretation as well. For example, Nick Bostrom (2017) raises the point of how to define harm, and how we can weigh up harms of a completely different nature, the harm of physical pain or the harm of social injustice. In the context of AI, harm also receives different connotations: individual, collective and societal (Smuha 2021a); material or immaterial, including physical, psychological or economic, depending on how risks are articulated (Kusche 2024). Therefore, harm is not a fixed concept, but rather dynamic and contingent to political, socio-cultural and ethical dimensions underpinning the discourse and outlined political priorities.

Putting these points together, my consideration of conditions of possibility for harm in this thesis relies on several points. Firstly, attention is directed towards speech acts and leading definitions of what concerns are attached to conditions of possibility for harm. Secondly, in the case of AI, *possibility* brings us to the political imagination, where potential harm is also about the future and the process of construction, without expecting to answer how tangible, measurable or *realistic* these possibilities are. Thirdly, although this does not seem to be extensively raised in the process of riskification, it is important to ask what security understanding is implied. Is it, as discussed earlier, associated with conventional matters (such as sovereignty), the establishment of different security dimensions, or the modification of existing ones.

3) Governance programme

In the case of riskification, and in comparison with securitization, governance programmes are presented as more permanent and long-term responses to risk than extraordinary measures to deter or defeat a threat (Corry 2012). Then governance programmes may involve guidelines, laws, regulations, modes of cooperation and collaboration, all kinds of outputs, developed and inscribed as ways to address perceived risks (Morsut and Engen 2023).

Acknowledging the broad spectrum of ways to describe governance, I follow and use several major blocks to characterize it: governance as 1) structures and processes which enable actors to coordinate their needs and

interests in making policy decisions (Krahmann 2003); 2) norms, rules and ideals that shape policies (Calcara, Csernatoni and Lavallée 2020); and 3) institutional rationalizations of what should be prioritized and achieved (Rothstein, Borraz and Huber 2013; Rothstein, Huber and Gaskell 2006). These elements often intertwine and make governance a complex set of measures and principles that apply to a certain issue.

As the EU refers to the risk-based approach, it expresses how far governance should intervene rather than eliminate adverse outcomes which also share the normative rationale that adversity can be reduced to zero (Rothstein, Borraz and Huber 2013). Then risk operates as "a meta-level governance tool" to define what are acceptable and tolerable levels of harm, and rationalize chosen frameworks where certain issues are less or more risky than others (Paul 2017b). Therefore, the main challenge with the overall employment of the risk-based approach is: who decides and defines those different levels and attaches them to certain risks as a foundation of governance, and how?

Several points for the analysis should be distinguished. Firstly, I consider governance as a response to risks, and a process of construction where norms, regulations and institutions intertwine by rationalizing and legitimizing specific measures. Then the leading question becomes *how* risks are expected to be governed by tailoring certain measures based on political priorities. Secondly, the analysis of a governance programme also requires getting involved in concrete characteristics of those measures; and to what extent they are specific to a particular issue or an actor which constructs a governance programme. Thirdly, a governance programme is not only a list of policy measures, but the contribution to knowledge production: what is included and excluded in its scope, how these political choices become legitimized and normalized as a tailored response to risks.

4) International engagement

Coming back to riskification, Olaf Corry's description appears to be ambiguous in terms of the internal-external dimensions of risks. This could be explained by the fact that it depends on the risk itself and how it is framed in the process of riskification. However, for this thesis, it is important to explicitly distinguish the *international* element, as AI-related risks and their governance have an external dimension considering technology as *borderless* (Rees 2008). Even Beck's risk society involves references to globality and cosmopolitan suggesting that risk transcends the boundaries of national states (Rothe 2011). Therefore, I expand the concept of riskification and involve the

element of international engagement by arguing that both points, the distinction between inside and outside, and the positionality of an actor, need to be equally investigated under this caption.

More specifically, debates on AI already emphasize that emerging AI policies have a clear international dimension, which is described as "geopolitical imaginations and calculations" (Haddad, Vorlíček and Klimburg-Witjes 2024, 757), or the "geopolitics of AI governance" (Csernatoni 2024a). Here *AI geopolitics* denotes the struggle over the meanings, governance and control of AI technologies, shaped by discourses, power relations and global asymmetries. Different authors refer to international dynamics as a competition for political, military and economic dominance (Ulnicane 2022); the inevitable arms race (Suchman 2016; 2023); or a zero-sum game mentality, where technological dominance by a few powerful actors leads to concentrated power and inequality (Ulnicane 2023; Haner and Garcia 2019). Therefore, a more critical consideration of developing trends suggests that the international dimension is not only about *borderless* risks, but political and boundary-making activities between different actors and their positionalities.

Three key points are considered for the analysis. Firstly, it examines how the international AI landscape is perceived and defined, whether hostile, cooperative or otherwise, and what tendencies are put as the most relevant. Secondly, what directions of international engagement emerge, how they are characterized in relation to risks. Thirdly, what forms and features of self-positioning are highlighted, how they fit, contradict, challenge, or place the actor within the defined landscape.

Conclusion

This chapter situated the research question and the entire analysis in constructivist assumptions enabling the examination of how knowledge is produced about the present and the future, how chosen terms, categories and framings emerge in considering technology, security, risk and their intersection. The demonstrated academic trends searching for corresponding concepts and synergies between different agendas also encouraged me to take a more critical stance, and reconsider boundaries of established definitions, such as risk society. The overall logic of the chapter has been to develop a tailor-made conceptual framework, which both corresponds with and contributes to current analytical endeavours to conceptualize the technology-security nexus.

While doing this, I have considered technology as inherently political and further problematising conversations about security politics and risk politics. The discussions introduced have demonstrated that the issues raised, either related to technologies or broader transnational challenges, such as climate change or cybercrime, employ the concept of risk as more related to future, long-term challenges and remaining uncertainty. Therefore, I argue that the security-risk dynamic is a matter for construction and interrelation which can be depicted through discourses, frames, and positions of how meanings and mechanisms of employing risk are thought and grounded by political actors. In other words, "risk is nothing if not relative, perspectival and transformational" (Dillon 2008, 328).

As a result, riskification is employed to analyse AI-related security, aiming to demonstrate the process of how risks and the response to them is constructed in emerging EU AI policy. It is also about developing the concept of riskification itself: to what extent it deconstructs the use of risk, and what modifications are needed for future security analysis focusing on risks. The established analytical framework, based on four elements, a referent object, conditions of possibility for harm, a governance programme, and international engagement, demonstrates the ambition of this thesis to adapt riskification in the context of AI, involving both actor-specific and international perspectives. Simultaneously, the analysis is situated in the context of existing literature, and reflects both empirical and conceptual contributions to riskification, and their relevance in applying this framework for analysing emerging EU AI policy.

Questions of how reality is constructed, what ensembles of ideas and concepts, are produced, and what meanings they acquire, point to the relevance of the discourse (Rothe 2011). The thesis employs discourse analysis as a vigorous method for understanding the textual and intertextual origins of security practices (Frowd, Mutlu and Salter 2023). It does not try to find or explain motives or thoughts, psychological or cognitive elements, hidden intensions or secret plans, beyond the texts (Waever 2003).

The analysis draws on a Foucauldian perspective, in which discourse is understood as a constitutive practice through which meanings are constructed and stabilised, and political possibilities are articulated. Following Foucault (2013), discourse analysis involves examining how knowledge is produced within discursive practices that position subjects and define objects they are authorised to speak about. To note, I use the term *articulations* not in the specific sense developed by scholars such as Jutta Weldes or Stuart Hall, where articulation refers to the contingent linking of signifiers (for example, Weldes 1996), but rather in a more Foucauldian-inspired sense – to describe repetitive discursive expressions that render certain interpretations, identities, or policy directions appear self-evident over time.

The analysis focuses on the EU's AI strategic discourse, defined as the collection of official documents, statements and policy positions through which the EU's institutions suggest the vision, priorities, and approach towards AI. To do the analysis, Section 3.1 introduces a research strategy that addresses both conceptual and empirical challenges, while defining the scope of the analysis in the rapidly changing political context.

The data selection and collection, introduced in Section 3.2., presents the decision to base the analysis on the relevant EU documents and conduct semi-structured expert interviews. It describes established criteria and steps to gather this data, also reflecting on the issue of range and representation. Section 3.3. details the coding scheme, which is based on the deductive/inductive hybrid strategy and its implementation. This includes a detailed description of the coding process, followed by an explanation of how the interpretations were reached.

Overall, these sections, outlining the research strategy and its implementation, remain in line with the conceptual framework and its major assumptions related to social constructivism. They are also based on a reflexive stance, engaging with the methodological specificities and limitations, such as the required range of sources, the replicability of

interpretations and the generalization of results, associated with an interpretivist, qualitative type of research. Nevertheless, these methodological choices provide a coherent basis for the subsequent analysis, and address the mentioned challenges.

3.1. Research strategy

A research strategy outlines how research will be conducted: it becomes like a *blueprint* for how data is collected, analysed and interpreted, and what steps guide the entire research process. Following the proposal of Lene Hansen (2005, 67), four cornerstones are applied: one, the *number of Selves* which is focused on the EU as an actor shaping its approach towards AI articulated through the emerging EU AI policy; two, the *number of events*, which here concentrate on daily practices, such as issuing relevant documents, responding to them, or making statements within the chosen framework; three, *intertextual models* refer to the EU's AI strategic discourse, arguing that selected texts are intertwined and participate in their (re)production through references, similar wording or the political context (Neumann 2008; Frowd, Mutlu and Salter 2023); finally, the *temporal perspective* marks the intense period of constructing the emerging EU AI policy between 2018 and 2024, from the EC's first Communication *Artificial Intelligence for Europe*, dedicated to AI and released in April 2018, to the AI Act, the first legally binding document regulating AI, which entered into force on 1 August 2024. This period symbolizes a *full circle*, from initial strategic documents to legally binding rules entering into force to regulate AI.

The development of the emerging EU AI policy aligns with similar processes and timeframes of other AI policies worldwide (see Section 3.3.), indicating a growing global interest in this domain. Specifically, the period starting in 2018 marks a phase of active evolution, both within the EU and on a global scale. In the EU's case, the point of temporality is also relevant because of the Russian aggression against Ukraine in 2022. It marked significant changes, including a greater openness to military initiatives and their integration into EU policies. Examples include the EU Military Assistance Mission to Ukraine through the European Peace Facility and the European Commission's *Act in Support of Ammunition Production*, which aims to produce two million ammunition shells annually by the European defence industry (Defence-industry-space.ec.europa.eu, n.d.).

However, these shifts have not involved AI or impacted the EU's position on AI policy. Instead, political attention to the emerging EU AI policy focused

largely on negotiating the proposed AI Act. At the same time, the influence of evolving global developments (such as the launch of ChatGPT by OpenAI) contributed to making AI a much more pressing topic. For example, the open letter "Pause Giant AI Experiments", released by the Future of Life, instituted and signed by more than 33,000 people online, claims that "AI systems with human-competitive intelligence can pose profound risks to society and humanity" (Futureoflife.org 2023). Therefore, for this analysis, there is no specific focus on the effects of the war, nor is there a prewar or after-war classification. The thesis remains anchored within the previously described timeframe and milestones, concentrating on the emerging EU AI policy, without delving into different agendas, even though they include security-related provisions in other policy frameworks.

3.2. Data selection and collection

Following the logic of discourse analysis, texts are considered as the main data. For example, Iver B. Neumann (2008) suggests that a *text* might be of a different form, from monuments to political strategies or visuals. Then there is enough flexibility to choose the texts that are the most relevant for the analysis. Arguing that the EU's approach towards AI is articulated through the emerging EU AI policy, I focus on *written texts*, various documents introduced and released by different EU institutions on the subject of AI and generalized as the EU's AI strategic discourse. The full list of selected documents is presented in Annex 1.

The overall dataset of selected documents corresponds to Foucauldian discursive formations (Foucault 2013), from procedural ways in issuing documents to overlapping vocabulary and references, the terms and timeframe of their release, and the decision-making process. On top of the written documents, *spoken texts*, semi-structured expert interviews, are also included. They represent those closely involved in shaping the emerging AI policy through affiliations with an actual institution or contributions to forming the policy (for example, external expertise). The list of interviews and interviewees is provided in Annex 3.

Both processes, the selection of documents and interviewing, require more details explaining the process and the decisions made. In terms of selecting texts, Lene Hansen (2005) suggests three criteria: they are characterized by the clear articulation of identities and policies; they are widely read and attended to; and they have the formal authority to define a political position.

She also stresses that it is impossible to define a number of texts as a general standard, leaving it to an individual analysis.

In response, firstly, the analysis focuses on the EU and its emerging AI policy, directly referring to an actor's position, established vocabularies and the development of a particular policy. Secondly, the selected documents represent different EU institutions and their official positions, they are accessible to the public, and have already been debated throughout the defined period. Thirdly, these documents are primary sources released by the EU authorities, which not only define directions taken but also reveal the EU's priorities and self-positioning through the main characteristics of the emerging EU AI policy.

The proposed criteria were tailored to the thesis and further elaborated:

1) the document is directly related to AI and is released by an EU institution. It immediately requires decisions in defining the scope as AI appears in different fields and related documents (for example, the Communication *Launching the European Defence Fund*), as well as the crossover with other documents and policies under the umbrella of the digital agenda (the Communication *A European Strategy for Data* or the Communication *A Chips Act for Europe*). Therefore, I prefer documents that place AI as a central issue, but not as a tool for other policies and their goals. Broad inclusion throughout the various policy fields, such as ethical guidelines for using AI in teaching and learning for teachers, does not seem to bring an added value to answer the specific questions on AI-related security and its perception.

2) There is a clear affiliation with EU institutions, including agencies and ad hoc formats such as the HLEG established by the EC, which have also released documents on AI in their name. It is important to stress that specific characteristics of format, type, scope and the comprehensiveness of the documents are not distinguished, leaving the flexibility depending on institutions and their ways of discourse construction. The national AI strategies released by Member States are not included, avoiding a potential need for evaluating domestic politics or risking moving to the division of EU institutions versus Member States, which again remains outside the scope of this thesis.

3) A principle of intertextuality is followed, by demonstrating interconnectedness and the reproduction of discursive practices via repeating or overlapping vocabulary, despite the different authorship or format of the documents. This enables us to claim the *EU AI strategic discourse* and the *emerging EU AI policy*, because interconnected and noticeable overlaps

demonstrate that, despite the sometimes different institutional positions, the key notions, principles and their presentation are shared across the board.

Overall, the dataset of the strategic documents reveals that 75 documents in total were selected. To address the risk of possible overlook, a careful search through the databases of institutions was done fulfilling the overall collection. It is important to stress that the scope did not involve various reports related to AI which were requested by different committees of the EP. While they publicized on behalf of the EP, in most cases they were authored by external experts in the field. The decision not to include them relies mainly on the growing conversation of knowledge generation and positions of expertise, ethics or scientific bodies, which require another analysis on their role in creating novel synergies, relations, power structures, policies and regulations (Rychnovská 2020).

In terms of interviews, they are both relevant primary data for interpretative research and also an additional understanding of meaning-making, in which people experience and make sense of certain phenomena. The interviews also become a contextualizing part which provides additional explanatory knowledge about the subject and the policy-making process. By "explanatory knowledge", I follow the proposed definition of knowledge based on subjective relevance, points of view, interpretations, normative positions, beliefs and opinions about policy solutions (Van Audenhove and Donders 2019, 185). As was mentioned before, the aim of discourse analysis is not directed towards finding out motivations or *real* reasons behind the discourse; the same logic applies to the interviews. Interviewees are not expected to provide facts, hidden intentions or secret plans. The interviews are about insights into knowledge production, differing views, chosen directions and policy-making.

The choice of *semi-structured expert interviews* requires a few comments. Firstly, I chose this category as the *semi-structured* form provides some flexibility to approach different interviewees in various ways, while covering the same questions focusing on *how* (Azungah 2018). This strategy (see the initial questionnaire Annex 4) enabled me to start with the main questions based on a topic, and then to use some probing questions depending on the completeness and value of the answers. This gave me enough room to reorder questions or add additional ones, depending on the answers provided (Van Audenhove and Donders 2019).

In addition, the *expert* direction targets a specific group due to the focus on knowledge production within EU institutions. Leo Van Audenhove and Karen Donders (2019, 194) argue that for expert interviews, knowledge and

occupied position are central, because of their involvement in institutional processes and decision making. It is important to stress that all the interviewees were in senior and middle-management positions, involved either directly in formulating the emerging EU AI policy, or working on AI-related matters in a broader sense. Thus, it is a relatively top-heavy group, who have required a good understanding of the topic and the EU decision-making process, qualified enough to navigate so-called "eurospeak" (Kuus 2014). Therefore, I conducted these interviews during the second phase of my studies, to ensure I had a deeper understanding of the ongoing process.

At the same time, the choice of *expert* could be discussed for excluding politicians, such as members of the EP involved in policy-making and the formulation of institutional positions. The focus on expert-level interviewees presumes that they are more directly and consistently involved in the daily policy-making process, often possessing detailed institutional knowledge that politicians rely on. This strategy also creates a more equal distribution between institutions, as not all (for example, agencies) have a political level. While not interviewing political appointees may be seen as a limitation, this was partly offset by using public statements from rapporteurs and commissioners to capture the political dimension.

The primary selection criterion for interviewees was their involvement in the policymaking process, either through formal institutional roles or participation in formats such as the HLEG. To find relevant interviewees, I applied two strategies. First, I searched for people involved in the emerging EU AI policy through the websites of EU institutions. Unfortunately, in many cases this information was not accessible; therefore, I continued my search through events, media comments and public presentations. Second, the snowball strategy was also helpful, because some of the interviewees, or those who refused to speak, directed me to officials involved or gave other potentially relevant contacts. As a result, 11 interviews were conducted in the period from May 2023 to February 2024. They represent approximately 20% of the acceptance of all requests sent, six EU institutions, and independent experts involved in policy-making. Three interviews were conducted in person, the rest via MS Teams, due to the expressed preferences by interviewees. All the interviews were recorded and transcribed to be precise in the language that was used to describe views and positions. Those in institutions specified that they spoke in a personal capacity, and did not represent the official position.

All the interviewees agreed to be introduced as representatives of an institution. The use of the term *representative* here refers primarily to the

interviewees' affiliation with a particular institution or policy format, rather than implying that they speak on behalf of the institution as a whole or express an official institutional position. Recognizing this distinction is essential, as it underscores the diversity of voices and perspectives within the EU. For example, given the complexity of bodies such as the HLEG, which included participants from business, academia, law, and civil society, it would be both analytically inappropriate and empirically inaccurate to treat the group as a homogeneous entity. This diversity was also reflected in the selection of interviewees. For instance, the two interviewees affiliated with the HLEG come from different fields (one from the tech industry and one from academia). Similarly, interviewees identified as Member State representatives were drawn from countries with differing regional and economic contexts. Such logic, again, was aimed at capturing diverse perspectives, rather than generalizing a unified institutional viewpoint.

To highlight the similarities and differences among interviewees, the empirical analysis presents original and detailed quotes without paraphrasing their main ideas. When citing interviewees from the same institution, they are numbered for clarity (for example, *Representative 1 from the HLEG*). If only one interviewee from an institution is cited, they are referred to as *a representative from [institution]*. I use all the material in a manner that preserves participants' anonymity, in accordance with the signed consent forms and prior agreements. The overall process of conducting and processing interviews remains the central methodological and ethical consideration guiding this research[6].

On a reflexive note, the overall dataset, consisting of selected documents and conducted interviews, could be discussed in terms of the *sufficient* scope and related potential limitations, especially related to the overlooking of alternative discourses or a chosen time snapshot. For example, relying on a single interviewee per institution in some cases may limit the ability to capture internal diversity or differing perspectives within those organizations. However, the application of discourse analysis does not pre-set a required number of texts for the validity or replicability of the analysis.

Examples of using discourse analysis suggest varying strategies and datasets, mainly depending on the case and the research question. For instance, the study on radicalization in Denmark was based on one document called "A

[6] At the time of the research, the institution had no formal ethics committee in place; however, all the interviewees were fully informed about the study, and their participation was based on written consent, following standard ethical procedures regarding voluntary involvement, anonymity, and the right to withdraw.

Common and Safe Future", considered as the Danish government's action plan and "a programmatic statement of the radicalization discourse in Denmark" (Lindekilde 2014, 2016). In another case, the discourse analysis on the EU's counter-terrorism policy included 50 European Council documents, which, according to the author, signify "the large sample of texts" illustrating the main themes central to constituting the terrorist "other" (Baker-Beall 2014, 6). In terms of interviews, an *optimal number* similarly remains a matter for discussion among interpretative researchers (Magnusson and Marecek 2015). For example, it is considered that to "achieve depth rather than breadth", six to nine interviews are "perhaps enough" (X. Liu 2018, 3).

Taking these examples into account, and recognizing the limitations and variations in developing the research strategy, the established dataset of 75 documents and 11 interviews for this analysis gives enough confidence that a plurality of perspectives and the involvement of various institutions, positions and individual perspectives have been achieved. Compared to a focus on a particular institution or a single strategy, this scope does not focus, for example, only on the contribution of the EC, as the main policy initiator, but involves diverse institutional perspectives through different texts. Rather, this dataset reflects the involvement of various institutions, roles, and individual viewpoints within the EU's AI strategic discourse, and enables us to comprehensively understand how particular meanings of AI and AI-related security are constituted.

This analysis does not aim to provide a detailed assessment of each institution or a direct comparison between them. Instead, it situates these positions and perspectives within the context of the EU as such. At the same time, the focus remains on the emerging AI policy without claiming that the findings are universally applicable to other cases. By considering specific texts, contexts, and timeframes, the thesis is reflective of a limitation inherent in discourse analysis and does not attempt to generalize beyond the selected case.

3.3. Reading and coding process

To analyse texts, Lasse Lindekilde (2014, 2013) suggests that there are two possible ways: either an already-established coding scheme and then deductive coding, or a more inductive strategy, which is not necessarily about predefined categories but "what the data tells us on its own". Linda Ruppert (2024) argues that the coding process mainly searches for interconnections between elements that place them in relations of equivalence, opposition,

causality or temporality. Although these broad guidelines offer flexibility, it ultimately falls on the researcher to develop an appropriate coding logic that fits the research design and selected data. The process is even described as "researcher-centric concepts, themes and dimensions" (Azungah 2018, 394), as influenced by positionality, chosen theoretical lens and contextual knowledge.

In the thesis, I apply the deductive/inductive hybrid strategy, suggesting that pure deductive and inductive strategies are not feasible, and the hybrid approach harnesses advantages of both. Deduction entails a pre-determined theoretical pattern, which leads to establishing major codes. Inductive analysis then entails a generation of codes from the data itself (Proudfoot 2023). In line with the deductive/inductive hybrid strategy, the coding process was implemented by following these phases (inspired by Nowell et al. 2017): 1) familiarization with the data by reading selected texts and extracting text segments; 2) the generation of major codes following the conceptual framework of riskification; 3) a search for overlapping themes within text segments and categorising initial ideas; 4) after another review of text segments, identified themes are transformed into sub-codes. These sub-codes are distributed in line with major codes and the conceptual framework. Lastly, the analysis and interpretations are introduced.

To be more specific, the conceptual framework and the focus on technology, security and risk has led to a search for text segments which interconnect or refer directly to AI-related risks within the text. This strategy was supported not only by the application of riskification and analytical elements, but also the noticeable structuring of the emerging EU AI policy based on a risk-based approach. Therefore, following the deductive/inductive hybrid strategy, I extracted text segments (a sentence at minimum and a paragraph at maximum) directly referring to *risk* during the first reading of the documents and interview transcripts. A total of 658 text segments were extracted and coded by using the MAXQDA software (2024 version).

In terms of the code scheme, the second reading of extracted text segments led to four major codes following the conceptual framework of riskification: a referent object, conditions of possibility for harm, a governance programme, and international engagement. It is important to note that a single segment may encompass multiple overlapping codes, and, as a result, may also include several sub-issue areas (Flonk, Jachtenfuchs and Obendiek 2024). For example, those which refer directly to different risks and their definitions were coded as *a referent object*, because of naming what is considered as being at risk (for example, fundamental rights, democracy and the rule of law). When

it comes to *conditions of possibility for harm*, various text segments which directly employ a vocabulary of harm, misuses (such as intrusive surveillance technologies or biometric identification), and negative outcomes were coded here. In short, this process involved organizing the text segments by aiming to highlight connections and key issues within the logic of the conceptual framework.

Moving to the more inductive part of coding, sub-codes were developed by grouping different text segments that convey similar themes, references, or even words to depict leading tendencies and overlaps. At the same time, the *reading* of text segments does not take place in a vacuum. Different literatures and analyses also *frame* the understanding of ongoing debates related to AI, and suggest the main priorities and dominant vocabularies. For example, analyses of the emerging EU AI policy, although coming more from regulatory and public policy approaches, already focused on overlapping concepts such as "human-centred AI" (Carmel and Paul 2022), a risk-based approach and its relevance to specifying fundamental rights, harms and self-identification in emerging EU AI policy (Niklas and Dencik 2024). Then noticing and recognising these concepts in the extracted text segments indicate their importance to the EU's approach towards AI, and raise further questions of their role in constructing AI-related security.

As a result, the code of *a referent object* was sub-coded into three sub-codes: a risk-based approach (what and in which context it is introduced), high risk (how it is defined, and to which matters attached), fundamental rights (a repeating element to define a level of risk), and safety (closely attached to fundamental rights). The code of *conditions of possibility for harm* was sub-coded into two sub-codes: forms of harm (what is discussed as concerning or directly named as harm), and military concerns (how the military domain signifies potential harm). The code of *governance programme* was sub-coded into three sub-codes: measures of governance (what new measures are proposed to be established); human-centric AI (what different notions and principles are proposed to address the expectations of human centrism), and the notion of the human in the loop (in what ways this is defined and put into the policy framework). Lastly, a separate code of *international engagement* was split into sub-codes, based on dominant repetitions: international engagement (how the EU wants to participate in that dynamic), and competition (references to the international dynamics related to AI).

After coding, the analysis moved to the interpretation to understand how AI-related security knowledge is constructed within the EU's AI strategic discourse, what interlinked articulations and notions dominate. Following the

conceptual framework, the systemic examining was focused on overlapping wording, verbs, noun articles, argument style, and their repetition, suggesting meaning making and patterns within the codes and sub-codes. For example, the phrases "AI misuse may also entail risks to fundamental rights" and "to strengthen the protection of fundamental rights" within the code of *a referent object* indicated that the noun of *fundamental rights* was central to articulating concerns and establishing a contrast between AI misuse and the need to protect rights from it. This led to the identification of fundamental rights as one of the referent objects. Table 3 provides an example of how the coding process was conducted and how initial interpretive notes were created.

Table 3. *Example of the coding process*

Text segment	Code and sub-code	Interpretive note
*"The **development of bias** in algorithms over time through such feedback loops risks reinforcing or creating **discriminatory practices** that affect groups with protected characteristics (such as ethnic origin) disproportionately"* (Fra.europa.eu 2022).	Conditions of possibility for harm → forms of harm	The text segment highlights how algorithm-driven bias and discrimination increase the vulnerability of certain social groups.
*"Any policy decision [...] should be taken **with due consultation of a European-wide research and development project** dedicated to robotics and neuroscience, with **scientists and experts able to assess** all related risks and consequences"* (Europarl.europa.eu 2017).	Governance programme → measures of governance	The text segment stresses the need to rely on external expertise and scientific knowledge for risk assessment and policy implementation.
*"Underlines **the risk of European values being globally replaced**, our companies becoming **marginalised** and our living standards being drastically **reduced**"* (Europarl.europa.eu 2022a).	International engagement → competition	The text segment expresses concern about the EU's diminishing role as a global agenda and standard setter, highlighting the potential negative consequences.

Source: the author

Interpretation was also conducted to search for "organizing metaphors" which stitch a discursive formation together, and "points for legitimation", demonstrating what claims are used and presented as self-evident, natural and indisputable (Rogers 2009, 838). Examples of such metaphors could be *human-centric AI*, shared across texts and become pillars around which other discursive expressions and policy elements are organized. Following the conceptual framework, the identified organising metaphors and other dominant patterns were synthesized to develop interpretations in response to questions such as: how do the EU's aspirations to influence global AI rules define the Self? Or how are references to authoritarian regimes used to articulate concerns and conditions of possibility for harm? This interpretative approach reinforces the deductive/inductive hybrid strategy, suggesting that the process, combined with the conceptual framework, engagement with the data, and the overall logic of this thesis, results in iterative moving back and forth between the major codes and the identification of evolving patterns in the data.

All in all, the provided interpretations do not seek to present a single *truth*, but instead allow the discourse to speak for itself, acknowledging its contested notions (Baker-Beall and Mott 2022). They are intertextual, and, despite diverse contributions from EU institutions and their representatives, demonstrate the dominant meanings of the EU's AI strategic discourse. At the same time, interpretations do not make a distinction between discourse, action and materialities, but consider them as intertwined.

Conclusion

This chapter has outlined the key components of the research strategy and the steps taken to implement it. The analysis targets the EU's AI strategic discourse, and, from an interpretivist perspective, focuses on meanings and their articulations outlining AI-related security in the emerging EU AI policy. It is crucial to recognize that the EU's AI strategic discourse has been dynamic and evolving, due to the ongoing negotiations directly related to the AI Act. Consequently, adapting the research and continuously checking with both the literature and the evolving political landscape of the EU have been integral to the methodological adaptation. The analysis adopts a Foucauldian perspective, focusing on how meaning and risk are constructed through language and institutional practice.

The data selection, including various documents and interviews, has necessitated constant reflexivity regarding the process, and engagement with

existing debates on the topic. Nevertheless, the overall research strategy is tailor-made, and corresponds to major expectations of the interpretative analysis, and extends its application to further analyses focusing on AI-related vocabulary and discursive practices. On top of this, interviews, from approaching potential interviewees to analysing conversations, were conducted in line with major expectations for research ethics and the protection of participants' interests. The thesis remains attentive to the internal differences, tensions, and pluralism within and across EU institutions, recognizing that the EU is not a monolithic actor but a contested and dynamic political space.

The deductive/inductive hybrid strategy empowered the phases of engaging with the data and developing the resulting empirical analysis. The search for themes, wording, patterns and organising metaphors has been combined with the conceptual framework, where the notion of risk has been central in collecting text segments. Therefore, the entire process, from defining the main points of the research strategy, selecting texts, conducting interviews and coding text segments, to their interpretation, has been driven by broad guidelines and examples of how to do discourse analysis and address its limitations; while adapting it to the thesis and its research question. This involved interpreting texts not for their hidden intentions, but for how they discursively construct particular subjects, problems, and institutional logics in relation to AI.

Lastly, interpretations pose a challenge to detail explicitly their development, as they inevitably reflect the researcher's way of thinking and understanding of the discourse and its key elements. As a result, subjective forms of knowledge embedded in the texts become interrelated with the researcher's position in the process of interpretation. However, rather than being a limitation, these conditions are integral to the analytical process, structured and refined through the lens of existing literature and methodological choices. In this way, the analysis embraces its interpretive nature without a pretence to achieve generalising or universal claims beyond AI-related security in the emerging EU AI policy.

II. EMPIRICAL ANALYSIS

4. THE EMERGING EU AI POLICY AND INTERNATIONAL AI LANDSCAPE

This chapter is devoted to contextualising the emerging EU AI policy and the international landscape, which demonstrates evolving trends related to approaching and framing AI. Margrethe Vestager, the then Executive Vice-President for a Europe Fit for the Digital Age, highlighted the importance of ensuring that "AI technology uptake respects EU rules in Europe" (Ec.europa.eu 2024b). This represents the EU's broader digital strategy to regulate the development and use of emerging technologies, and to influence international AI standards, by reinforcing itself as a regulatory power, shaping rules and norms governing policy. However, the EU's ambitions also face other initiatives to govern and regulate AI. This dynamic is even described as the "increasingly crowded AI governance landscape" (Csernatoni 2024b, 15), suggesting that the EU's self-introduction as the first to regulate AI evolves between other initiatives.

Section 4.1. suggests that the emerging EU AI policy comes as a part of the EU's broader digital agenda, where topics range from building infrastructure to fighting against hate speech on social media. It also indicates the shift that the EU has been embracing by developing its digital agenda, towards more strategic and assertive policy goals which are not limited to the single market but cross overs, and includes security-related domains and expectations of international influence. In this way, the EU could be seen as gradually building its profile in the digital domain.

Then Section 4.2. introduces the main stages in developing the emerging EU AI policy. It shows that the process has not been linear, but received diverse contributions, such as guidelines, documents and institutional set-ups. It also involves the question of who is shaping and framing the emerging EU AI policy, relying mainly on EU institutions and their production of the discourse. This section argues that the emerging EU AI policy, introduced as a part of the digital agenda, has also been presented as a flagship initiative, inheriting the influence of the General Data Protection Regulation (GDPR).

The chapter moves on to an overview of other actors developing their emerging AI policies. This brief analysis situates the EU and the emerging EU AI policy in the international landscape, suggesting that the chosen rhetoric of the EU's exclusivity turns out to meet others claiming similar ambitions. Section 4.3. introduces several examples of state actors and organizations

which illustrate that they similarly develop their emerging AI policies and expectations to lead the conversation on AI. Even though the overview demonstrates important overlaps in discussing risks, suggesting principles for AI governance and claiming leadership, they appear to be embedded in the actors' preferred political and ideological perspectives, further reiterating the point that AI emerges not as a *technical* but a socio-political phenomenon.

4.1. The EU's evolving digital ambitions

The section briefly introduces several milestones which demonstrate ongoing debates, initiatives, and the search for their alignment with the EU's aspirations for digital leadership. This context underscores the relevance of the emerging EU AI policy as both a product and the driving force of the EU's wider digital agenda. The milestones are detailed as follows:

1) The launch of the Digital Single Market in 2015 signified a clear strategic commitment by the EU to prioritize and advance its interests in the field of digitalization;
2) Adopted in 2016, the GDPR stands as a landmark in digital regulation, solidifying the EU's reputation as a citizen rights-oriented policy-maker, and exemplifying the "Brussels effect" through its global influence;
3) Adopted in 2020–2021, the strategies *A Europe Fit for the Digital Age* and *2030 Digital Compass: The European Way for the Digital Decade* positioned digitalization as a central priority of the first von der Leyen Commission;
4) The adoption of the Digital Services Act (DSA) and the Digital Markets Act (DMA) in 2022 signals the EU's growing assertiveness in establishing rules and requirements for digital platforms, aimed at protecting fundamental rights and shaping global standards for online safety.

Despite the variety of initiatives related to the digital agenda, these events are presented as major developments, because they set the EU's tone and ambition, reflecting the recognition of technology and overall digitalization as top priorities (Martins and Mawdsley 2021; Bellanova and Glouftsios 2022). They also articulate the overarching principles guiding the EU's multilayered digital approach: an emphasis on rights, safety and democratic values, coupled with the regulatory oversight of technological developments and uses, while aiming to demonstrate a pro-innovation stance and international relevance.

The story starts with the Digital Single Market Strategy proposed by Juncker's EC in 2015 to remove "key differences between online and offline worlds, breaking down the barriers to cross-border activity" (Ec.europa.eu,

n.d.-b). The dedicated EC Communication reproduces the same elements of the Single Market (the free movement of goods, persons, services and capital) and formulates the ambition to "ensure that Europe maintains its position as a world leader in the digital economy" (Eur-lex.europa.eu 2015). In other words, the initial idea of the Digital Single Market was to replicate creation of the single market and aim for a level playing field for online businesses, consumers and investment in ICT infrastructures across the EU.

One of the most recognized outcomes from this early stage is the GDPR. It seeks to unify how others abide by a set of data management rules if it wants to trade with the EU. This is even entitled "the world's toughest data privacy law" (Daly 2025). Any corporation anywhere, if they collect data on EU citizens, can see massive penalties. Adopted in 2016, this regulation has become an international reference for digital governance worldwide, and is now agreed as a key example of the success of the EU's influence in the digital field internationally (for example, Granados Hernandez 2022; J. R. Torreblanca José Ignacio 2022; U. F. Torreblanca Carla Hobbs, Janka Oertel, Jeremy Shapiro and José Ignacio 2020). Different experts refer to this regulation as the reason for the EU becoming "a standard-setter in the field which triggered a global debate about privacy as a digital human right" (Dekker and Okano-Heijmans 2020, 9), and proving that "the EU is capable of setting rules impacting the digital economy globally" (Cervi 2022, 18).

Such influence has even received the title of the "Brussels effect", marking the EU's ability to export its rules and regulate the global market without coercion, positioning itself as "the most powerful regulator of the digital economy" (Bradford 2023; 2020). Anu Bradford (2020) has claimed that the "Brussels effect" is about peaceful and quiet power, which is norm-setting internationally and embraces the EU's role of a regulatory hegemon. For example, the Commission President Ursula von der Leyen stated that "with the General Data Protection Regulation we set the pattern for the world. We have to do the same with artificial intelligence" (Ec.europa.eu 2019). Even though it was criticized for creating tensions between competitiveness and imposed control on the availability of large data sets (Dekker and Okano-Heijmans 2020), Representative 2 from the HLEG suggested that

"the experiences with data privacy regulation supported this notion that being a front runner in certain aspects could even give a competitive advantage. They make Europe strong, give Europe a strong competitive position by ensuring legislation that allows or guarantees competition for better products."

Thus, the case of the GDPR has emerged as a pathway and a playbook for the EU in terms of shaping its digital policies and international reputation, which apparently remains the dominant logic and expectation to receive the same external recognition through other digital policies as well.

The next chapter in developing the digital agenda could be described as much more strategic, and considers digitalization as a geopolitical matter rather than merely market integration. The first strategy adopted by von der Leyen's EC in 2020 is entitled "A Europe Fit for the Digital Age", which outlines ten distinct policy directions[7] and aims to "strengthen its digital sovereignty and set standards, rather than following those of others – with a clear focus on data, technology, and infrastructure" (Commission.europa.eu 2020). In 2021, the EC also released the Communication *2030 Digital Compass: The European way for the Digital Decade*, which reiterates similar ambitions: "to pursue digital policies", to be "an assertive player", to "assess and address any strategic weaknesses", and "to be digitally sovereign" (Eur-lex.europa.eu 2021a). Both documents highlight the tendency of the EU for considering technologies and the overall process of digitalization as a matter which includes not only regulation but also infrastructures, the control of value chains, and standardization in relation to other actors.

On top of this, strategic documents on digitalization and technologies have extended beyond the mentioned strategies, permeating various security-related topics as well. For example, the Strategic Compass, the Cybersecurity Strategy, the Standardization Strategy, the European Drone Strategy 2.0 2022, the European Union Space Strategy for Security and Defence 2023, the European Economic Security Strategy 2023, and the European Defence Industrial Strategy 2024, all include and instrumentalize technologies as part of the policy implementation enabling the EU to achieve security-related objectives. The focus on different security matters (cyberspace, the economy or the defence industry) also signals a constantly noticeable combination of both internal and external dimensions of the EU, where strategic goals involve the EU's relations with those outside the EU.

Compared to the early stages of the Digital Single Market, the active proliferation of strategies across policies suggests that *digital* has evolved from the Single Market to different agendas, making technologies and digitalization a cross-overing priority. For example, the EP described these

[7] The list includes the Digital Services Act, the Digital Markets Act, the European Chips Act, the European Digital Identity, Artificial Intelligence, the European data strategy, the European industrial strategy, Contributing to European Defence, Space, and the EU-US Trade and Technology Council.

priorities as competitiveness and a functioning digital market, the protection of citizens through a safer online environment, data and algorithmic transparency, keeping pace with technological developments, and international engagement in developing governance frameworks (Europarl.europa.eu 2024c).

In this light, the DSA and the DMA could be considered as important illustrations of the EU's endeavours to use its regulatory power "to create a safe space where the fundamental rights of users are protected" (Digital-strategy.ec.europa.eu, n.d.-g), and oblige social media platforms, predominantly outside the EU, to follow established rules. The DSA, which has even been entitled "the constitution for the Internet" (Edri.org 2023), obliges online platforms and online search engines to combat harmful and illegal content, as well as the sale of illegal goods and services (Europarl.europa.eu 2024c). The DMA is described and directed to ensure contestable and fair markets when digital platforms that act as "gatekeepers" are present, and attempts to rein in digital market abuses (Edri.org 2023). A representative from the EC reiterated this point by claiming that "Digital Services and Digital Markets Act, this is something we have done first of its kind worldwide. And everybody is looking at that with great interest." The competition cases the EC initiated on the basis of the DMA with Apple, Meta, Microsoft and other platforms (Competition-cases.ec.europa.eu, n.d.) demonstrate that the EU's evolving digital initiatives transcend the market logic and the reduction of barriers as described in 2015.

The EU proactively shapes the (future) relationship to various actors involved in digitalization, and eliminates boundaries between private and public to protect the political preferences. By demanding transparency from social media platforms on how they apply algorithms, moderate content and process complaints, the EU demonstrates that *online* is not a private matter, but a legislative and public policy, and even a security concern (Schlag 2023). The reappearing reference of *being the first* also resembles the story of the GDPR suggesting the EU's expectations to reproduce the "Brussels effect" and the position of regulatory powers to perpetuate its expected influence (Heldt 2022). These digital initiatives are not only about the economy, the market and rights, but also the EU as such, about tested and patterned ways of participating and reinstating itself in a changing world.

Thus, through its digital agenda, the EU seeks to address concerns that challenge its political interpretation of *Europeanness*, particularly its commitment to freedoms and rights. At the same time, it aims to maintain its image as a regulatory power, one that introduces policy first and sets standards

both within and beyond its borders. The noticeable shift in rhetoric between the Commission under Juncker and von der Leyen, from the *digital economy* to *digital sovereignty*, marks a broader strategic turn: digitalization is no longer simply about a level playing field within the EU, but an increasingly strategic issue. References to digital sovereignty signal ambitions to develop technological capacities and define EU-specific rules across a wide range of policy domains, from platform governance to security. Together, these tendencies highlight how digitalization is increasingly embedded and prioritized in the EU's broader vision for its future.

4.2. The emerging EU AI policy

Following the brief overview of some milestones related to the EU's digital agenda, I now take a closer look at the emerging EU AI policy: its development, key nuances, and the contributors shaping its direction. Even though this initiative came as part of the broader digital strategy, the emerging EU AI policy is exclusive in that it is the most directly attached to and builds on the GDPR as a continuation of the same goals and the "Brussels effect" now in the domain of AI. It similarly refers to regulation across sectors as "guard rails" of citizens, extra-territoriality, and the influence of other actors to follow, in von der Leyen's words, "our own, distinctive approach to AI" (Ec.europa.eu 2025). Extra-territoriality and compliance with the rules here mean that the EU's AI Act will apply to any provider and deployer placing, or otherwise putting into service, an AI system on the EU market, regardless of whether the provider is established or located within the EU, in a third country, or if the output produced by the AI system is intended to be used in the EU (Whitecase.com 2025).

Sounding technocratic, such complexity demonstrates that the EU's focus on AI is not only about rapid technological developments, but simultaneously becomes a power play: to address state and corporate competition, to shape global AI standards, and to navigate the diverse governance initiatives of other actors (Csernatoni 2024b). In short, the emerging EU AI policy reflects a multifaceted response, encompassing AI developments and uses, fostering EU-wide digital integration and regulation, and actively participating in the international AI landscape.

Ronit Justo-Hanani (2022) argues that the emerging EU AI policy has been constructed in three stages: the brainstorming stage, the agenda-setting stage, and the decision-making stage, as an incremental process rather than a radical reform towards AI. While these stages become relevant conceptual

guidelines to analyse the emerging EU AI policy, I organize this section focusing on two elements: 1) important events and documents which provide a chronology, from strategic debates to the adoption of legally binding rules; and 2) institutional contributions which illustrate the diverse involvement and establishment of dedicated institutional formats. For a better sense of the process, Table 4 summarizes the key documents released and events in the logic of the proposed stages[8].

Interest in AI has been evolving for quite some time. For example, the end-of-term report of the priorities of Juncker's Commission (2014–2019), released in 2019, argued that "further initiatives are needed especially in the areas of artificial intelligence" (Europarl.europa.eu 2019a). AI as a policy area was also mentioned in documents such as the Communication on the *European Defence Fund* or the EP Resolution on *Civil Law Rules on Robotics* in 2017. The clear prioritization of AI can be traced back to the beginning of Ursula von der Leyen's first Commission, where one of the proposed strategic goals was to "ensure AI is developed in ways that respect people's rights and earn their trust" (Commission.europa.eu, n.d.-c). The President herself referred to the initiative as "the first of its kind anywhere in the world" (Weforum.org 2024), once again proving the EU's repeating obsession of *being the first* and replicating the "Brussels effect" in the case of AI.

The start of the brainstorming stage is associated with the release of the EC Communication on *Artificial Intelligence for Europe* which "shows the way forward and highlights the need to join forces at a European level, to ensure that all Europeans are part of the digital transformation" (Eur-lex.europa.eu 2018). The EC, an official initiator of the emerging EU AI policy, has provided a lot of material, which includes graphics, timelines, documents, references to websites, reports and social media channels devoted to AI since 2018. The EC's leading role is unsurprising, given that the right of initiative is embedded in its institutional competences. As the executive branch of the EU, it has a mandate to propose legislation, implement decisions, uphold treaties, and manage the day-to-day functioning of the Union (Commission.europa.eu, n.d.-b). At the same time, the rise of technologies and digitalization as prominent policy agendas has introduced more manoeuvring between institutional competences and responsibilities,

[8] Importantly, this overview does not involve processes and documents released in the Member States and the discussions among them. This decision is grounded on the research logic to focus on the EU level and institutional contributions in constructing the emerging EU AI policy.

suggesting shifting dynamics, despite the initially defined division of roles and policy domains.

The EC's efforts in initiating the emerging EU AI policy were put on the webpage *A European approach to artificial intelligence* and the more general title of *Shaping Europe's digital future.* This webpage creates an impression of multiple initiatives and channels, as well as crossovers between different policies, where links, references and visuals are supposed to come together as the EU approach towards AI. The most impressive, both visual and discursive, addition is the vertical timeline entitled "important milestones", which starts from 2018 to today. It marks every AI-related document and initiative released by the EC, the HLEG and the European Council, consultations organized, and community building, such as the launch of the European AI Alliance, an initiative to establish an open policy dialogue on AI, for which different stakeholders can sign up (Digital-strategy.ec.europa.eu, n.d.-h). Such a presentation creates the impression that the construction of the emerging EU AI policy was a linear, EC-driven and gradual progress, culminating in the adoption of legally binding rules.

Another important development of the brainstorming stage was the establishment of the HLEG by the EC in 2019. Representative 1 from the HLEG described the role of the Group:

"I think the Commission itself realized pretty quickly that in order to do that [develop AI regulation] they needed to get the first expert opinion on what these general principles of responsible trustworthy AI should be, so that they can start building their regulatory framework on top."

This group consisted of 52 experts from different fields (business, academia and law), which, in the two years of its mandate, provided the Ethical Guidelines, the Policy and Investment Recommendations for Trustworthy AI, the Assessment List for Trustworthy AI, and the Sectoral Considerations on the Policy and Investment Recommendations (Digital-strategy.ec.europa.eu, n.d.-f). The documents released by the HLEG can be considered as cornerstones of the main elements and vocabulary (such as human centrism, ethics and trustworthy AI), which are reproduced in later documents on the matter. This format also represents the EC's decision to rely on external expertise for building the EU's approach towards AI as something more grounded, objective and *evidence based.*

The brainstorming stage also includes the Coordinate Plan in 2019, which came as endeavours to avoid national fragmentation of AI strategies and

regulations on a national level. Member States were invited to "agree on common indicators to monitor AI uptake and development in the Union and the success rate of the strategies in place" (Digital-strategy.ec.europa.eu 2018). This stage was concluded by the Communication *Building Trust in Human-Centric Artificial Intelligence* in 2019, which outlined the key notions and principles, such as human-centric AI, trustworthiness, ethics, and overall ambition of "how economic competitiveness and societal trust must start from the same fundamental values" (Digital-strategy.ec.europa.eu 2019a). Even though most of this vocabulary was already proposed by the HLEG in its contributions, EC communications as strategic documents integrated it as a core of the emerging EU AI policy. Overall, this brainstorming stage centred on defining core principles and shaping the legislation's tone (Justo-Hanani 2022).

Moving further, the agenda-setting stage could be described through two major points. One is related to the White Paper released in 2020. This document marked the turn from strategic discussions towards more concrete directions of future legislation: "the White Paper presents policy options to enable a trustworthy and secure development of AI in Europe" (Eur-lex.europa.eu 2020d). The document refers broadly to industry, data, partnerships with business and public sectors, and concerns related to rights and freedoms, and suggests initial ideas of AI governance. In short, the White Paper, based on previously mentioned documents, became the main framework before the proposed AI Act.

Although the EC appeared to be in the driving seat introducing these AI-related documents, the EP became increasingly involved as well. A representative from the EU Institute for Security Studies (EUISS) suggested: "do not forget that the Parliament is also flexing its kind of normative muscle. And, you know, members of the European Parliament (MEP) are also trying to communicate back to their citizens a bit on this issue." The EP has pushed for more political debate and the inclusion of diverse perspectives, which often appear more contested and controversial compared to the EC's more linear, expert-driven approach to policy-making. Such tendencies also reflect the matter of competences between institutions where the EP comes as a directly elected body, a co-legislator in the decision-making process and a supervisor of the EC (Europarl.europa.eu, n.d.). Such a range of powers enables the EP to push for a more contested conversation and open arguments with the initial proposal introduced by the EC.

In the case of AI, the EP's role and interest were manifested in releasing different reports, resolutions, events organized, and reactions to the EC's

documents. For example, the EP brought different topics through reports: on civil liability regime for artificial intelligence (2020), on intellectual property rights for the development of artificial intelligence technologies (2020), on a framework of ethical aspects of artificial intelligence (2020), and on the interpretation and application of international law related to civil and military uses (2021). These contributions did not necessarily emerge in line with the EC, and suggested critical points to proposed directions. The most noticeable case is the EP's non-agreement to exclude the military from the scope of the policy. In different documents, the EP called and even urged to involve the military, and advocated for a more *strategic* EU presence in the AI-related international dynamic: "to take […] an active role in promoting this global framework governing the use of AI for military and other purposes" (Europarl.europa.eu 2021b). Such a discussion shows that the process was not linear, as proposed by the EC, but involved polyphonic positions and diverging priorities in framing the emerging EU AI policy.

On an institutional level, the EP established the special committee Artificial Intelligence in a Digital Age (AIDA) "with the goal of setting out a long-term EU roadmap on AI" (Europarl.europa.eu 2020d). It organized hearings on different policy fields and the role of AI, and workshops, requested studies on AI, and released an overall report of the committee's position, stating that "the EU has fallen behind in the global race for tech leadership" (Europarl.europa.eu 2022a). Again, this demonstrates continuous endeavours by the EP to frame AI as a geopolitical matter and see the EU as a participant in that struggle. On top of this, other EP committees (such as those on the Internal Market and Consumer Protection, and Civil Liberties, Justice and Home Affairs) were also actively involved, revealing the institutional tension within the EP over which bodies are responsible for leading the AI-related agenda.

This highlights a relevant tendency: different committees commissioned expert reports focusing on various aspects of AI, contributing to the EP's evolving understanding on the matter (for example, "Should we fear artificial intelligence?", "Artificial intelligence: how does it work, why does it matter, and what can we do about it?", and "Artificial intelligence diplomacy. Artificial intelligence governance as a new European Union external policy tool"). This reliance on specialized knowledge production not only reflects a desire to base political rhetoric on ostensibly impartial, scientific and objective insights, but also underscores endeavours to integrate that into the policy framework. Both the EC and the EP relied on external independent expertise (either through the HLEG or commissioned reports) as providing legitimacy

for political decision-making. However, the emphasis on expertise may also mask underlying political struggles between institutions and even committees, reflecting competing priorities and interpretations.

The agenda-setting stage could be summarized as marking a pathway from strategic thinking towards defining more concrete guidelines for future regulation. Also, it is about the institutional dynamics, where the role of the EP shows that the initial dominance of the EC in the brainstorming stage received a more political response and a call for further discussion on the EU's priorities and ambitions in the field.

The decision-making stage can be mainly associated with the EC's the EC's proposal for the AI Act and its entry into force in 2024. It marks the completion of the legislative process "to improve the functioning of the internal market by laying down a uniform legal framework" (Eur-lex.europa.eu 2024). Even though it puts the EC under the spotlight again, the period 2021 to 2024 could be considered as the most intense debates between institutions and the increasing interest in AI as such (for example, the introduction of ChatGPT by OpenAI at the end of 2022). In its Resolution "On Artificial Intelligence in a Digital Age" in 2022, the EP claimed that the EU "is still far from fulfilling its aspiration to become competitive in AI on a global scale". Representative 1 from the EU Council also stressed the difference between institutions: "the EP is speaking a very political stance. It is human rights, values, and so on, and so on. It is nothing about technology, it is nothing about competitiveness. It is nothing about the integration of data and so on." Even with a tangible legislative task, these debates and political struggles demonstrate that the emerging EU AI policy is not deterministic or linear, but comes as a contested compromise referring to the EU's complexity in policy-making.

This complexity is not limited to leading EU institutions, but also those in charge of executing AI policies. For example, the European Defence Agency (EDA) has episodically reflected various elements of AI in its magazine *European Defence Matters* and reports stating AI is "a strategically important topic" (Eda.europa.eu 2021). The European Space Agency defines AI as "one important part of the full solution, enabling scalable exploration of big data and bringing new insight and predictive capabilities" (Philab.esa.int 2018). Even though agencies are tasked to provide technical and sectorial know-how to the EC (Trondal and Jeppesen 2008), their contribution via strategies or analyses specify the set priorities and define them in more concrete policy practices. For example, the EU Agency for Fundamental Rights (FRA) claims in its paper that assessments of AI-related systems should involve an

evaluation of data quality because it may increase discriminatory situations (Fra.europa.eu 2019). Then such competence-specific contributions diversify the conversation of the EC and the EP, and once again prove the pluralism of the policy-making process in the EU.

Overall, the emerging EU AI policy reflects the EU's increasingly assertive ambitions in approaching technologies and shaping its profile in digitalization. The regulation, applied not only within the EU but also *extra-territorially*, along with the continued focus on the "Brussels effect", suggests a pathway that the EU takes and reproduces through different digital initiatives. At the same time, the emerging AI policy is driven not only by institutional debates and aspirations for influence, but also by international pressures stemming from technological developments and concerns, reiterated by the EP, about falling behind global powers. That is why this policy differs from more internal initiatives, as it asserts itself on multiple fronts, including widespread global attention and high expectations surrounding AI-related technologies, competing political visions, and strategic positioning.

Table 4. *Summary of the stages of the emerging EU AI policy*

Stages in formulating the emerging EU AI policy	**Key documents and events**
The brainstorming stage (2018-2019)	The EC releases the first Communication on *Artificial Intelligence for Europe* (2018) The EC establishes the HLEG (2018) The EC releases the Coordinate Plan to agree the terms with Member States (2018) The European AI Alliance, a platform for around 6,000 stakeholders, is launched (2018) The HLEG releases policy and investment recommendations for trustworthy AI (2019) The EC releases the Communication *Building Trust in Human-Centric Artificial Intelligence* (2019) The first European AI Alliance assembly takes place (2019) The HLEG releases the Ethics Guidelines for Trustworthy AI (2019)
The agenda-setting stage (2020-2021)	The HLEG releases the Assessment List for Trustworthy AI (2020) The EC releases the White Paper on AI (2020) The Second European AI Alliance Assembly takes place (2020)

Stages in formulating the emerging EU AI policy	Key documents and events
	The EP establishes the Special Committee on Artificial Intelligence in a Digital Age (AIDA) (2020) The EP adopts the Report on *AI in a Digital Age* (2020) The EC releases the Communication on *Fostering a European approach to AI* (2021) The EP releases resolutions focused on different elements, including calls to involve the military (2020–2021)
The decision-making stage (2021-2024)	The EC releases the Proposal for a Regulation on AI (2021) The EP releases the Resolution on *Artificial Intelligence in a Digital Age* (2022) The Council of the European Union adopts its general approach on the Proposal for a Regulation on AI (2022) The EP adopts its negotiating position on the Proposal for a Regulation on AI (2023) The EC, the EP and the Council reach a political agreement on the Proposal for a Regulation on AI (2023) The AI Act is adopted and enters into force (2024)

Source: the author, based on the analysis

4.3. Actors shape and frame their emerging AI policies

As previous sections have demonstrated, the EU presents itself as the first to regulate AI where "to be a global power means to be a leader in AI" (Europarl.europa.eu 2022b). However, such self-positioning is not in a vacuum, but part of complex regulatory initiatives. In this context, the so-called geopolitics of AI are often reduced to the USA, China and the EU proposing different approaches towards AI based on their preferences: more corporate freedom, state control or focus on protecting rights (Jakniūnaitė and Lingevičius 2021). However, the international landscape appears to be much more diverse, as "political leaders have understood AI's disruptive potential and are rushing to secure a competitive advantage in this crucial emerging domain" (Renda 2019).

Various actors, both states and organizations, have also developed their own emerging AI policies, based on issuing numerous strategic documents and aiming for legally binding rules in the same period of 2018– 2024[9]. Following the presentation of the EU's ambitions and steps taken, this section provides an international context, which explains noticeable trends often described as the competition and race for innovation, talent, and data. This context situates the EU, and demonstrates that *expected audiences* for the EU's "Brussels effect", supposedly like-minded partners and organizations, have been developing their own alternative AI policy frameworks. This context also explains that the EU's priorities related to protecting rights or mitigating AI misuses has indirect references to, for example, China's approach towards AI, and evolving autocratic practices.

A key tendency here is that state actors in particular assert leadership and influence over the international approach to AI, framing it as a source of competitive advantage (Roberts et al. 2024). As Politico aptly summarized: "governments have hit a major bump in the road: they all want to win" (Scott 2023). These claims reflect the underlying assumption that states are engaged in competition, a theme frequently emphasized in their policy documents. However, this notion needs closer scrutiny, as it frames the international system as a zero-sum game, where one actor achieves political, military and economic superiority, while others fall behind (Ulnicane 2022). The analysis suggests that competition is defined not so much in military terms, but influence to shape the international conversation, which is defined as "continuing to cement the position as a world leader in AI safety" (Gov.uk 2023b).

The following overview of several actors, the United States of America (USA), China, the United Kingdom (UK), Japan, the Council of Europe (CoE) and the United Nations (UN)[10], demonstrates the scope and interest in

[9] Acknowledging the increasing importance of big tech companies and corporations in AI-related dynamics, the analysis does not involve them in the scope of the overviewed actors. This decision is based on the predominant focus on the politics of emerging AI policies, and how they are shaped by political actors. The involvement of companies and corporations would require a different angle, and also analytical instruments for engaging with their approaches towards AI.

[10] The following actors were selected: the United States (US) and China are named as major AI powers, of which the competition will globally shape future AI dynamics (Roberts, Cowls, Morley, et al. 2021; Bächle and Bareis 2022). Among middle powers, the United Kingdom (UK) was selected for its growing ambition to be a global leader in AI. Japan, although a part of the G7, takes a different approach compared to its Western counterparts like the USA and the UK, but shares the goal

developing emerging AI policies (see the list of documents in Annex 2). Without going into a detailed analysis of riskification in each case, these documents employ the concept of risk, propose governance programmes, and define their position internationally.

While connotations of risks vary, whether described as catastrophic or substantial, risk emerges as a central way to describe AI-related concerns. For example, the US National AI Security Commission, established in 2018 to address emerging challenges in AI security, in a report released in 2021, refers to "strategic risks of AI-enabled weapons", which suggest "negative and uncertain effects to the USA security" (Reports.nscai.gov 2021). The UK Bletchley Declaration, adopted during the AI Safety Summit in 2023, mentions *significant risks* and *unforeseen risks* because "capabilities are not fully understood and are therefore hard to predict" (Gov.uk 2025b); while the UK Defence Artificial Intelligence Strategy 2020 discusses "extreme and even existential risks" that need to be managed in ways adapting to "uncertainty" (Gov.uk 2022). Japan, in its AI Strategy 2022, refers to global risks, which involve pandemics and climate change, with "a high probability that it will lead to a critical situation", and security risks related to data, privacy and "the risk of AI itself becoming an attack target" (Cao.go.jp 2022). The CoE's Convention on Artificial Intelligence, Democracy, and Rule of Law, adopted in May 2024, becoming the second legal AI framework after the EU's AI Act, involves "the consideration of wider risks and impacts related to these technologies including, but not limited to, human health and the environment, and socio-economic aspects" (Coe.int, n.d.-b). In July 2023, the UN Secretary-General Antonio Guterres stated that "we urgently need frameworks to deal with these [AI] risks" (Un.org, n.d.-b).

These examples demonstrate that risk is becoming an overlapping concept which is used to anticipate AI-related concerns, and already articulate the urgency to act. Again, it is not only the EU which frames the AI-related security understanding through risk. Even the USA and the UK, which explicitly refer to military security, similarly involve risk than the more conventional language of threat. The phrases *uncertain effects* or *hard to*

---

of influencing AI priorities. As for organizations, the Council of Europe (CoE) and the United Nations (UN) were chosen for their efforts to establish cross-border rules (CoE) and governing principles (UN). The CoE recently adopted the Convention on Artificial Intelligence, Democracy, and Rule of Law, becoming the second legal AI framework after the EU's AI Act. Meanwhile, the UN has issued a report outlining its goals to set global standards and develop a governing model.

predict prove this shift towards risk as signifying an orientation towards an undefined future and the related uncertainty.

Moving to proposals for governing AI, they rely on normative principles, which appear to be ambiguous, as both democracies and autocracies similarly refer to the importance of ethics, human centrism, and responsible AI. In July 2023, the White House released the executive order entitled "Ensuring Safe, Secure, and Trustworthy AI" which invites companies to voluntarily commit to developing AI that "have a profound obligation to behave responsibly and ensure their products are safe" (Federalregister.gov 2023). In the case of China, various exhortations to "respect the laws of AI development", "prevent the risks of algorithm abuse" or "strengthen the standards for algorithmic governance" (Digichina.stanford.edu 2021), or the need to "promote fairness, justice, harmony, and security while avoiding such problems as bias, discrimination, and privacy and information leaks" (Cset.georgetown.edu 2021), suggest that the same normative principles are shared. However, references to "carry forward the socialist core values view, uphold the correct political direction, public opinion orientation, and value orientation in the application of algorithms" (Digichina.stanford.edu 2021) indicate that, despite similar vocabularies, their interpretations are completely different.

The principles proposed to govern AI mark an ideological competition that is "appropriate" (Qiao-Franco and Bode 2023), and how AI becomes part of a pro-market vs pro-state, or pro-democratic vs pro-autocratic, clash. Anu Bradford (2023) argues that the USA aims to develop a market-driven model of AI governance where governmental intervention in regulation is supposed to be minimal; while China aims for a state-centred model where technology should empower the state, subjugating individual rights and freedoms to state control. Evolving practices such as, for example, the intensifying use of technologies for surveillance and even repressions against minorities, and their export as a form of control mechanism, do not represent *fairness* or *justice* in the way they are introduced in democratic countries (Shen 2020; Horton Zeng 2021). These differences of interpretation or mismatches between declarations and practices demonstrate why different emerging AI policies are seen as competition: whose world-view, merged with evolving practices, will become dominant.

In this state-centric dynamic, organizations, the CoE and the UN, continue the focus on supposedly global cooperation and normative principles for governing AI. In March 2024, the UN General Assembly adopted its first-ever resolution to promote "safe, secure and trustworthy" AI internationally, which calls "to refrain from or cease the use of artificial intelligence systems that are

impossible to operate in compliance with international human rights law" (Mishra 2024). In the case of the CoE, the Convention includes various references to democracy, human rights and the rule of law, stressing that signatories must ensure that AI systems "are not used to undermine the integrity, independence and effectiveness of democratic institutions and processes" (Coe.int, n.d.-b). Democracy is viewed as an essential means for ensuring transparency, accountability and responsibility in developing and using AI: "to protect that CoE fundamental values are protected in the digital environment" (Coe.int, n.d.-a). The Convention was adopted not only by CoE Member States and the EU, but also by international partners, such as Canada, Japan, the USA, Australia, Argentina and Israel.

These cases could be read as pro-Western and pro-democratic countries and organizations teaming up to make their preferred principles international standards of AI governance. However, by inscribing the mentioned principles in a rather ambiguous way, the document has already received criticism for remaining loopholes that "have the potential to enable conduct that has a negative impact on human rights protection" (Babická and Giacomin 2024). Even in this case, the ability to set standards and ensure compliance with outlined values remains a challenge.

In terms of self-positionality, there is a clear tendency for considering AI to increase power, even though it is inherently borderless, and an enabler rather than a calculable *tool.* Despite this, references to leadership through ownership and competition dominate in the documents. For example, the USA refers to competitive power and efforts to counterbalance digital authoritarianism directly associated with China: "the USA government must embrace the AI competition and organize to win it. The American approach to innovation […] must be recalibrated to account for the centrality of the competition involving AI and associated technologies to the emerging U.S-China rivalry" (Reports.nscai.gov 2021). The New Generation AI Development Plan, released in 2017, and claiming China's leadership in AI by 2030, stresses that "the world's major developed countries are taking the development of AI as a major strategy to enhance national competitiveness and protect national security", and shares ambitions to "actively participate in global governance of AI […] deepen international cooperation on AI laws and regulations, international rules and so on, and jointly cope with global challenges" (Digichina.stanford.edu 2017).

The UK and Japan, considered as middle powers compared to the USA and China, similarly claim leadership in AI. The former Prime Minister Rishi Sunak, during the launching event of the Summit on AI Safety in 2023, stated

that "I want to make the UK not just the intellectual home but the geographical home of global AI safety regulation" (Browne 2023). References to *powerhouse* or *intellectual home* suggests that the UK considers AI as a geopolitical matter, where technology can be somehow geographically placed and defined, "already home to top AI labs" (Gov.uk 2023a). The UK Defence AI Strategy also states the importance of "geostrategic competition", where the UK "will shape the development of AI in line with UK goals", and "the UK has significant strengths and is recognized as an AI powerhouse" (Gov.uk 2022). Even though Japan avoids direct language of competition, its AI strategy mentions the word *leadership* almost 20 times: for example, "establishing leadership through 'Strong and Responsible AI'", "Japan should become the world's most capable country in the AI era", "Japan should take leadership in this field and pursue a strategy to establish an AI for Nature-Positive Economy" (Cao.go.jp 2022). Therefore, regardless of their differences in status and capacity, state actors articulate leadership as their international involvement in the AI landscape. However, as mentioned earlier, these claims appear rather vague, as they do not explain more than the interest to participate and influence international debates, a form of expected leadership.

The CoE and the UN, unlike state actors, claim their leadership by proposing AI-related standards and expecting to have an influence by reaching agreements between self-claimed leaders. For example, the CoE positions itself in the role of an *honest-broker*, where the agreement on common standards is put as an ultimate goal: "the Council of Europe […] must play a pioneering role in designing procedures and formats to ensure that AI-based technologies are used to enhance, and not to damage, democracy" (Pace.coe.int 2020). The emphasis on democracy, human rights, and the rule of law aligns directly with the CoE's mandate and priorities, enabling the organization to assert a role in advocating for these principles. However, although the adoption of the Convention suggests that the CoE managed to find such an agreement, the mentioned loopholes remain a point for discussion of its effectiveness, and the organization's role in enforcing adopted rules.

The UN similarly defines itself as being "at the heart of the rules-based international order" and "uniquely positioned to address" global challenges, which provides the legitimacy to argue for global governance (Un.org, n.d.-a). The scale of the organization becomes the key point to define its role and advocate for the elaboration of norms leading to preferred AI governance. Like the CoE, the UN does not position itself within the framework of competition or rivalry. Instead, it focuses on promoting global agreements and

multilateralism in emerging AI policies, as an effort to mitigate state-level competition and move towards international standards. On the other hand, considerations to regulate autonomous weapons systems under the UN Convention on Certain Conventional Weapons have been taking place since 2014, without a clear agreement between Member States. Therefore, this example shows that these expectations to be an honest broker or facilitator in the international landscape will not necessarily materialize in a different outcome than the UN General Assembly Resolution on the Promotion of Safe, Secure, and Trustworthy AI Systems which was backed by Member States without a vote (Mishra 2024).

Overall, despite the differences between the discussed actors, their emerging AI policies demonstrate common patterns, especially in establishing and developing specific cross-overing vocabularies, through risks, ethics, responsibility, trustworthiness and leadership. These identified tendencies are presented in Table 5. Simultaneously, the ambiguity surrounding these principles indicates that they remain open to definition or adjustment based on the preferences of individual actors. The Chinese case is particularly illustrative of how the same vocabulary of ethics and responsibility is used to challenge the global West, by making them relational and depending on a *translation* into policy practices. Therefore, despite critical questions of how AI is defined, what it means to own or exercise such desired leadership in AI, documents reveal that AI is imagined as the future's technological superiority to elevate and confirm a status portraying AI as being indispensable to preferred orders in the perceived international competition (Bächle and Bareis 2022).

Anu Bradford (2023) suggests that the power struggles evolve into the direction of the battle between digital democracy, promoting rights and liberal values, and digital authoritarianism, using surveillance and state control. This section has demonstrated that the focus is not only on ideological differences, but also the issues these different actors name, and how they formulate their role in addressing them, from weaponizing AI to a desire for influence over AI governance. This context demonstrates that the EU's claims of being the first to regulate AI and export rules face the similar ambitions of others, where norms, rules and standards become imagined forms of power, regardless of the ideological standpoint. Together with these similarities and differences, this analysis also proves that AI is a matter of international dynamics that intersects between different actors and their agendas.

Table 5. *Summary of the key elements of emerging AI policies*

Actors	Importance of risks	Proposals for AI governance	Self-positioning
The USA	Strategic, substantial, potential, societal, serious risks – mainly linked to national security	Values and norms-based, trustworthy, human-centric AI, voluntary commitment to agreed principles and rules on AI (liberal-order and pro-business approach)	Democratic power, global leadership and victory in the AI race, competition with China
China	Potential safety risks as an argument for the urgency to act	Ethical and responsible AI, importance of international standards and rules, ambition to shape AI governance globally, *alternative* to the USA approach	Global domination in AI, centralized *sovereign* power, competition with the USA
The UK	Catastrophic, extreme, existential, significant, unforeseen risks – mainly linked to national security and defence	Ethical and trustworthy AI, pro-democratic values, importance of safety, international institutional set-up for governing AI	International *powerhouse* on AI, geopolitical competition, stress on the UK's role
Japan	Pandemics, geopolitical and climate-related risks	Broad examples on tackling issues, no references to pro-democratic approach towards AI	International leadership, importance of resilience, race and competition are not reflected
Council of Europe	Risks as an argument of the urgency to act	Ethical, trustworthy and human-centric AI, pro-democratic approach, international and/or regional principles and regulation, ad hoc institutional set-ups	*Honest broker* to deliver the Convention as the first legally binding international agreement on AI
UN	Reiteration of AI-related risks as an argument for urgency	Normative and based on values, trustworthy, safe and secure, human-rights oriented AI, UN institutional framework	Broker and moderator of the UN-wide agreement on global AI governance

Source: the author, based on the analysis of documents

Conclusion

This chapter is dedicated to contextualising the emerging AI policy in the EU's broader digital agenda, outlining the main stages in formulating this policy, and situating it between the emerging AI policies proposed by different actors. The main purpose of introducing these different contexts was to demonstrate that this initiative is intertextual in terms of the EU's policy-making and the international AI landscape, where different actors refer to risk, AI governance, and leadership.

The overview of the EU's digital strategy, and key initiatives, ranging from the ambition to create the digital single market to regulations controlling platforms and their compliance with the EU's values, demonstrate the EU's increasing assertiveness in the digital domain. At the same time, individual initiatives filling in the overall digital strategy are highly complex and challenging to navigate between different priorities. Then *digital* and *digitalization* are increasingly embedded in the strategic thinking, and transcend the concentration on the market or a more harmonized level playing field for Member States for future relations with both technologies and those who develop, provide and operate them.

The discussions presented demonstrate that the EC and the EP have been leading institutions in framing the emerging EU AI policy. Their contributions seem to reflect their share of competencies: the EC as a legislative initiator, introducing, arguing and developing its initiative, and the EP as co-legislator and supervisor of the EC, brought more political contestation and questions to the EC's suggested policy scope. The stages of formulating the policy have revealed that the story is not monolithic, but rather a chorus of diverse perspectives, from political and expert to ad hoc established formats, that should be continuously acknowledged and critically examined as important contributions to policy-making.

Lastly, the overview of other actors and their emerging AI policies reveals that the EU is not unique, but shares a focus on risks, the development of a response, and claims of leadership with states and other organizations. The priorities and self-positioning of different actors show that, although they use similar language, they often mean different things, revealing inconsistencies between what is claimed and what is done, as well as difficulties in reaching agreement on AI standards and the underlying ideological struggles. Therefore, this context gives a better understanding that the political interest in AI is not only about innovation and regulation, but an imaginary of future power which receives connotations depending on an actor and its preferences

and world-views. This tendency once again supports the point that AI is not *technical*, but emerges as a strategic element of which the importance varies from *owning* technology, reinforcing status, and securing the preferred international order.

5. THE PROCESS OF RISKIFICATION

The following chapters (5.1. to 5.5.) present an empirical analysis, structured according to the analytical framework introduced earlier. The analysis begins with Chapter 5.1., exploring how the EU constructs risks related to AI, and identifying the logic which underpins these constructions. Subsequent chapters (5.2. to 5.5.) systematically address each element of riskification, demonstrating how the EU defines a referent object and conditions of possibility for harm, what governance programme is introduced as a response, and international engagement strategized. The analytical elements of riskification signify knowledge production by *naming* what comes as a major focus, and why and how it is prioritized. Table 6 provides a summary of the guiding conceptual questions for the analysis.

To highlight the key findings, centrality is given to fundamental rights and a political democratic system as the referent objects. The analysis further explores the conditions of possibility for harm referring to already noticeable practices of intrusion, discrimination and the imaginary of AI autonomy. The EU proposes a governance programme based on normative and institutional policy measures which are concentrated on preserving human agency, steering AI developments and uses in preferred ways. Finally, the analysis delves into forms of international engagement, underscoring the aspirations for global leadership in setting AI standards, multilateral participation, and partnership-building, as well as competition with other powers.

On top of these interpretations, the chapters involve agentic security, technocratization and the fortress as conceptualizations of the EU's AI-related security, a proposed response, and the EU's self-positioning, that emerge from the findings. I argue that the EU's AI-related security focuses on human agency defined through control, decision-making, and expectations to keep technology subordinate. The EU puts AI-related security as latent and long-term, based on technocratic policy measures and the protection of the preferred liberal order. Importantly, while depicting the EU's thinking, these notions require further debates, as they reinstate the EU's biases and claims of universality. Therefore, instead of a conclusion, this part finishes with a discussion raising the controversies of anthropocentrism, depoliticization and Eurocentrism, as contested viewpoints that are both inscribed and justified within the EU's AI strategic discourse.

Table 6. *Leading questions for the analytical elements of riskification*

A referent object	• What discursive articulations are formulated as a referent object? • What are the leading characteristics of a referent object? • What safety concerns are attached to a referent object?
Conditions of possibility for harm	• What constitutes harm in relation to articulated risks and future uncertainty? • Who or what is associated with the causes of harm? • What security understanding arises from the identified harm?
A governance programme	• How should risks be governed? • What policy measures are proposed, and how are they characterised? • How is a governance progamme legitimised?
International engagement	• How is the international AI landscape defined? • What are the directions of international engagement? • What forms of self-positionality emerge?

Source: the author, based on the conceptual framing in Section 2.3.

5.1. Risk and a risk-based approach

Before analysing the process of riskification, our attention turns to the ways the concept of risk is introduced and articulated in the emerging EU AI policy. In one of the reports dedicated to AI, the FRA claims that "the risks are vast" (Fra.europa.eu 2022). At the same time, the employment of risk and a risk-based approach is not unique. It comes more as an established pattern, because various EU policies have already been drawing a close correlation between risk and market regulation (Fahey 2014). For example, the risk-based approach has been developed for policies of food safety and flood risk management, counter-terrorism, migration and border control. The most recent example in the digital domain is the already-mentioned DSA, which develops four categories of systemic risks: the dissemination of illegal content, the impact on fundamental rights, the negative effects on democratic processes, and the negative effects on physical and mental well-being (Digital-strategy.ec.europa.eu, n.d.-g). These examples demonstrate that the EU uses risk and a risk-based approach as a way to rationalize its policy framework and introduce potential responses as something preventive and already looking to the future (Paul 2024).

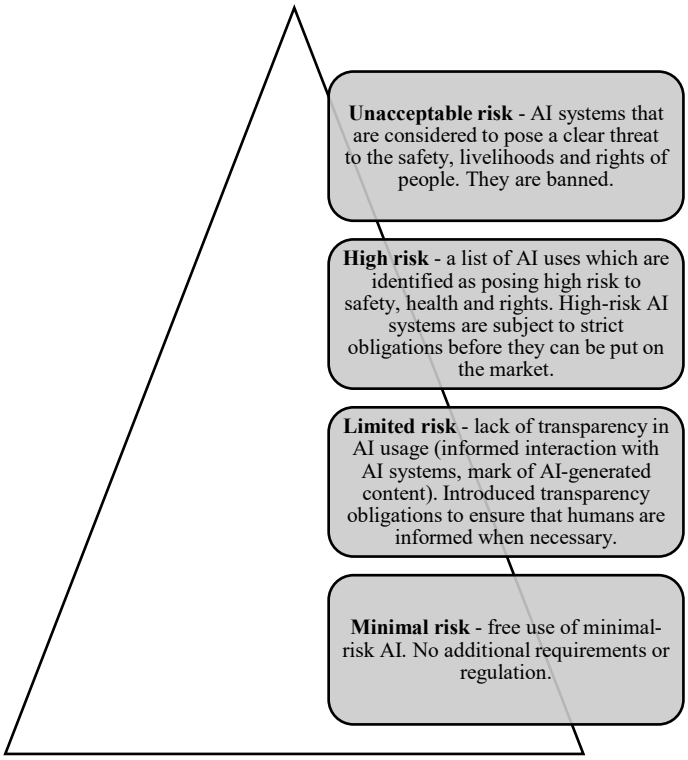
This chapter starts by introducing the pyramid of risk categories which, according to the European Union Agency for Cybersecurity (ENISA), "exhibit different degrees of risk that needs to be assessed" (Enisa.europa.eu 2020). Section 5.1.1. presents the proposed logic of risk categories which are based on different levels of potential harm and varying needs for intervention. The establishment of such a pyramid reveals that AI, as something remaining a matter of future invention, problematizes uncertainty and the urgency to act now.

The chapter then examines how risk categories are defined, and how specific AI applications are assigned to them. Section 5.1.2. reveals that the decision is based on political consensus and endeavours to balance both regulation and innovation while introducing the policy. Therefore, suggestions that definitions of risk categories are based on scientific evidence and expertise are inconsistent with the remaining ambiguity and the role of political decision-making in defining and differentiating these categories.

Overall, the analysis shows that the EU employs risk as an epistemic element, which is *flexible* in formulating the EU's own understanding of AI, setting political priorities, and narrowing down the complexity and uncertainty into a manageable process. The presented political contributions and influences involved in shaping the risk-based approach reflect not only specific imaginaries about AI, but also the EU's priorities in determining what is considered a matter for concern, and to what degree.

5.1.1. A pyramid of risks

It is crucial to understand how the notion of risk is constructed within the policy framework. For example, the HLEG suggests that risk is "broadly defined to encompass adverse impacts of all kinds, both individual and societal" (Digital-strategy.ec.europa.eu 2019b), while the AI Act defines risk as the combination of the probability of an occurrence of harm and the severity of that harm (Eur-lex.europa.eu 2024).



Source: European Commission (Ec.europa.eu)

Figure 1. *A pyramid of risk categories*

To explain the logic of the risk-based approach, the EC proposed a visualization of a pyramid of different risks, where *every step up* means a higher risk led by stricter regulation and priorities of potential concerns (Paul 2017a). A representative from the EC suggested that this logic was motivated by arguing that "the pyramid is useful in this regard, because you could say, on top is very, very dangerous stuff. And at the bottom is a totally harmless use of technology." The pyramid is structured with four categories described in rather blunt and ambiguous ways (see Figure 1). For example, the European AI Alliance Assembly claimed that it "contains applications without any risk; applications that can incorporate some risks; and at the peak of the pyramid, there are the high-level risk fields" (Futurium.ec.europa.eu 2019). The EP suggested that these categories "distinguish between a minority of 'high-risk' and the vast majority of 'low-risk' AI use cases" (Europarl.europa.eu 2022a).

Such differentiation is directly attached to the level of intervention: the higher the risk, the more regulation it requires. The EC claims that "the vast majority of AI systems currently in use are almost or even completely risk-free" and that "only a very small number of use cases can be categorized as risky and that only such cases require regulatory action and effective safeguards" (Europarl.europa.eu 2022a). Limited risks, according to the EP, "are subject to the existing legislation and transparency requirements, and additionally could choose to subscribe to voluntary, non-binding, self-regulatory schemes, such as codes of conduct" (Eur-lex.europa.eu 2021b). Then high and unacceptable risks at the top of the pyramid require the biggest intervention or even a ban. The AI Act suggests that it is "necessary to prohibit certain artificial intelligence practices, to lay down requirements for high-risk AI systems and obligations for the relevant operators" (Eur-lex.europa.eu 2024). As Regine Paul (2022b, 500) suggests, risk categories seem more like "contested semiotic constructs" which cannot be taken for granted. These different categories emerge as *degrees of riskiness,* demonstrating that the notion of risk itself does not define the level of concern, and it depends on an actor's position.

Most attention is concentrated on high-risk as requiring most of a governance response, because AI falling in the category of unacceptable risk is simply banned[11]. The EC refers to high-risk AI use cases "where the risks that the AI systems pose are particularly high. Whether an AI system is classified as high-risk depends on its intended purpose of the system and on the severity of the possible harm and the probability of its occurrence" (Eur-lex.europa.eu 2021b). The EP shares a similar definition, and suggests that "high-risk" AI systems can have "a detrimental impact on people's health, safety, or their fundamental rights" (Europarl.europa.eu 2024a). The Annex of the AI Act puts these ambiguous definitions in a more concrete way, and provides a list of high risks in eight domains[12] which target vulnerabilities of

[11] Eight practices are prohibited: harmful AI-based manipulation and deception, harmful AI-based exploitation of vulnerabilities, social scoring, individual criminal offence risk assessment or prediction, untargeted scraping of the internet or CCTV material to create or expand facial recognition databases, emotion recognition in workplaces and education institutions, biometric categorization to deduce certain protected characteristics, real-time remote biometric identification for law enforcement purposes in publicly accessible places (Digital-strategy.ec.europa.eu 2025).

[12] The list of high risks includes: biometric identification and categorization, management of critical infrastructure, education, employment and its management, access to private and public services, different forms of law enforcement, migration, asylum and border management, administration of justice and democratic processes.

different social groups or their weaknesses in power relations. These domains represent a wide spectrum of issues, but they are all related to a particular point: potential harm to fundamental rights, which might be caused by increasingly autonomous technology.

In this context, the case of the emerging EU AI policy grounds the employment of risk and the risk-based approach on both uncertainty and urgency, the tension between not knowing the future of AI but already taking action to respond. For example, a representative from the EUISS suggested "so, uncertainty is big." This could be understood in two ways: as a *condition* where the EU finds itself able to put the policy framework together; or as the source of a challenge that needs to be addressed. Either way, the EU faces limited or no knowledge about the future of AI, because, unlike national disasters and contingency plans, AI remains a question of the human ability to invent technology and transform it into different stages.

This lack of knowledge as a form of uncertainty has constantly been reflected in interviews. For example, a representative from the EP mentioned that "we are trying to reach at something which is evolving every day." A representative from the EDA considered "how can you standardize AI, since we still do not know exactly how the tools work"; while a representative from the EC claimed that "even the owners of the models and systems cannot really explain any more what is happening there." Then despite such a lack of knowledge, risk becomes something that inscribes imaginaries and defines priorities in response to uncertainty. As a representative of EU research technology initiative defined, "risk has this insurance company approach to the conversation, it also has future unknown elements to the conversation."

When it comes to urgency, it creates expectations to respond and act in the present. For example, Representative 2 from the HLEG suggested that "this was the urge we need to do something because the technology is applied in practice already and leads to several fundamental problems for democracy, for society, for individual human rights, and we need to regulate it." The point of a *need to do something* reappears in different text segments where the motivation comes from reiterating the point itself. For example, Representative 2 from the EU Council claimed that "because you have such an urgency, something needs to be done"; while an EU research technology representative stated that "it can also introduce a sense of urgency, because you see all these organizations, you know, moving towards the direction." Therefore, uncertainty and urgency are important conditions which also help us to understand the EU's thinking in employing risk and a risk-based approach for the emerging EU AI policy. Risk is used to *tame* uncertainty by

naming and structuring what is already known, what is considered as most important, and what could be left for future updates.

Thus, the definitions of risk and high risk are dependent on an individual case and perceptions in terms of its potential harm and severity. Even the provided references, dangerous stuff, certain practices, detrimental impact, cause injury or harm, do not specify what particularly characterizes risk, beyond statements that developments and uses of AI can be harmful. The analysis proves that risk is used to structure a lack of knowledge, and responds to complexity and uncertainty, where the pyramid becomes a symbol of being able to manage and regulate something that is not yet there. In this way, the EU establishes its priorities, which are not necessarily, as claimed, grounded on technical or scientific proof, but emerge in relation to imaginaries of AI and uses of AI in the future.

5.1.2. The political nature of risk

After discussing the pyramid of risk categories, the main question remains: who decides on a level of risk, and how, and which AI uses are attached to different categories? For example, Representative 1 from the EU Council reflected this by claiming that "I think, for us, there is always the question, is there a risk?" To answer, the EP raises the point that "uch a risk-based approach […] should be based on clear criteria and an appropriate definition of high risk" (Eur-lex.europa.eu 2020b).

The discourse suggests that it is not only the categories to be questioned but the very exercise of identifying and naming what risk is, particularly, what these *neutral* criteria are which enable lawmakers to establish the boundaries between these categories and attribute certain interlinkages between a level of risk and developments and uses of AI. Regine Paul (2024, 1066) claims that her respondents suggested that the decision of risk categories involved "high-level political choices about cases of unacceptable use of AI, as well as 'rule-of-thumb' categorizations of high-risk AI systems". This analysis demonstrates the same tendency.

The political nature of such decisions was explicitly suggested by a representative of the EC:

*"*This is not a scientific exercise. It is a political exercise based on some assumptions. And when I said safety and fundamental rights, this was our starting point, because from the evidence that we have collected, we said, well,

this is where the risk is most real, because we did not include hypothetical things at this stage, because that may be for future iterations and adaptations."

The reliance on political decisions was also suggested by Representative 2 from the EU Council, who claimed that "it is really a decision which technology falls in which part, and I think this is where this discussion gets very political […] also, there was a lot of lobbying, especially in this part, which technology, which cases of use in which category of risk."

I find the argument of *political nature* worth highlighting, because different claims keep referring to risk and risk-based approach as being *objective.* For example, the European AI Assembly suggests that a definition of high risk is based on "objective criteria" (Futurium.ec.europa.eu 2019). Meanwhile, Representative 2 from the HLEG argued that "I think this is the approach that is science-based, it is a sensible approach to do it in that way." The EC mentioned that the AI Act "lays down a solid risk methodology to define high-risk AI systems" (Eur-lex.europa.eu 2021c). The EP refers to "evidence-based solutions that address specific situations and sectors, where needed" (Eur-lex.europa.eu 2020b). Thus, the argument of *science*, *evidence-based* and *solid methodology* implies that the logic of risk and established categories has a *scientific* foundation, without alternative ways to frame the emerging EU AI policy. It creates the impression that claiming risks as objective and science-based gives more legitimacy to the policy framework and less contestation than explicitly communicating political preferences and decisions in establishing risk categories.

Politics is relevant, because risks are not only about technology, but also the EU aiming to address different priorities. Representatives of EU institutions emphasized the need to balance regulation with competitiveness. Representative 1 from the HLEG acknowledged that "because we also understood that if we only regulate strongly, all types of AI and all types of uses of AI, we could actually put Europe at a disadvantage against big, too big, competitive forces". A similar point was shared by Representative 2 from the EU Council, arguing that "setting some certain levels of risk helps a lot for the technology itself, for innovation." Another HLEG representative reinforced both points, by claiming that "how to make sure that Europe still has a competitive edge was to choose the risk-based approach." In addition to this, EP claims that "there is a need to establish a risk-based legal framework for AI […] while at the same time providing the private sector with enough flexibility" (Europarl.europa.eu 2022a), and that "this risk-based approach

should be developed in a way that limits the administrative burden for companies" (Europarl.europa.eu 2020a).

These points demonstrate that the EU is affected by debates about competition and regulation negatively influencing innovation. For example, the EC's proposal for the AI Act suggests that "the proposed regulation on AI combines greater safety and fundamental rights protection while supporting innovation, enabling trust without preventing innovation." Therefore, the introduction of different risk categories could also be read as a form of manoeuvring between regulatory intervention and no intervention, suggesting that the EU responds to criticism that a "heavy-handed approach could stifle innovation, weaken competitiveness, and drive some companies to rethink their operations in the region" (Borner 2024). However, such a search for a balance between different priorities has nothing to do with *scientific evidence*, but endeavours to address politically diverging priorities in AI-related hypes.

Another example of the role of politics in coming up with a pyramid comes from the EC White Paper (Eur-lex.europa.eu 2020d) referring to the German Data Ethics Commission, which "has called for a five-level risk-based system for regulation that would go from no regulation for the most innocuous AI systems to a complete ban for the most dangerous ones." Exploring this more, the German report "Opinion of the Data Ethics Commissio" released in 2019 introduces the "risk-adapted regulatory approach", which introduces a "criticality pyramid" that outlines five levels of the degree of criticality "with regard to the potential of algorithmic systems to cause harm" (Data Ethics Commission 2018). On top of this, such an approach is presented as built on "scientific and technical expertise in developing ethical guidelines for the protection of the individual, the preservation of social cohesion, and the safeguarding and promotion of prosperity in the information age" (Data Ethics Commission 2018).

The EC's references to the German case and clearly overlapping elements (the risk-based approach, a pyramid, scientific and technical expertise, the protection of individual and social cohesion) demonstrate that the EU's risk-based approach and the entire logic closely resemble the German position and the policy framework. Representative 1 from the HLEG recalled the discussions of different alternatives, suggesting that "there was a framework under the work in Germany. And I know that the Commission used this as the major baseline for developing the regulation and risk-based approach." The German example indicates that the risk-based approach is mirrored as another source for the EU in aiming to structure the uncertainty and follow the already-established framework. Therefore, the introduction of risk categories is much

more about the policy-making process, based on political inspirations, imaginaries and influences, than scientific and evidence-based relationships with technology.

Thus, the analysis suggests that the political preferences and ambiguity of risk categories turn out to be the main characteristics in understanding the construction of the pyramid and the risk-based approach. The pyramid establishes definitions, proposes boundaries between risks, and defines a role for the EU itself. However, the framing process appears to be inherently political, and based on different influences, from varying priorities to be more innovation-friendly, or more sensitive to potential harm, to the German policy framework as a source of inspiration. Therefore, the discourse suggests another inconsistency where claims of objectivity or clear criteria contradict the very logic of formulating and explaining the proposed risk-based approach.

Conclusion

This chapter has shown that risk is dynamic and multifaceted, with established risk categories defined in broad, context-dependent terms, when envisioning AI and its potential futures. The ways the EU introduces risk and the risk-based approach recalls Beck's construction of risk as already existing by the fact that AI is developed and used, *being out there*. Also, the proposed risk categories are presented as a self-evident way to classify AI developments and uses by their potential harm, even though the most serious concerns, those deemed unacceptable or high-risk, occupy only the top, and thus a small portion, of the overall pyramid.

However, the ambiguous explanations of high risk have revealed that supposed similarities with Beck's definition of the risk society are limited. The EU's risk-based approach is political, and based on a combination of different priorities: a way to tame uncertainty and make sense of the future; to balance regulation and competitiveness; to absorb political influences and trending imaginations of AI to achieve the result as being the first regulator of AI. The contrast between the EU presenting risks as given and their political nature reflects power dynamics, where the appeal to *scientific proof* can serve to downgrade the sensitivity of the directions taken and questions of who draws a line between those risk categories, and how. The power element emerges not only through the production of knowledge, what becomes risky or not, but also through interventions and constraints of others that are supposed to follow the rules.

The proposed categories of risk and their differentiation reflect how AI is understood and how the EU attaches that understanding to its own position. Compared to other actors employing risk as well (see Section 4.3.), the EU is more consistent in using a risk-based approach and established categories as an overarching policy framework. While other actors associate risk with different concerns, such as security in the USA and the UK, climate and pandemic risks in Japan, or pro-democratic initiatives from the CoE, the EU goes beyond merely identifying issues, putting risk at the core of the policy framework. Thus, for the EU, risk and the risk-based approach function as an *ordering* mechanism to establish the boundaries of the EU's priorities, and structure *everything AI* into imagined categories and their gradation of political importance.

5.2. Referent objects

Chapter 5.1. demonstrated, that despite different categories of risk, much attention is focused on two *top-tier* categories: unacceptable risks and high risks. These two categories, concentrating the highest levels of concern, serve as the main guideline for identifying what a referent object is, what the leading characteristics are, and what safety concerns are attached to a referent object.

Here, I follow the proposal to understand a referent object as not fixed, but as potentially changing, depending on social and political dynamics, without relying on pre-established categories such as state/nation or national/international levels (Hansen and Nissenbaum 2009; Mügge 2023). The analysis reveals that the EU constructs two referent objects: fundamental rights and a democratic political system, signalling that AI-related security does not rely on conventional matters such as territory or sovereignty.

Firstly, fundamental rights are one of the referent objects which demonstrate the EU's focus on human-machine interaction prioritising the human element. They serve as a litmus test to identify and describe what level of risk is attached to AI uses: the higher the risk, the more potential harm to fundamental rights. Section 5.2.1. suggests that the prioritization and indispensability of *fundamental* enable the EU to claim the universality of its approach towards AI, and establish a distinction between humans and machines.

Secondly, the chapter moves to a democratic political system as a second referent object. Intertwined with fundamental rights and safety, a democratic political system is also considered as high risk, requiring protection as representing rights-focused AI policy-making. Section 5.2.2. depicts the main

elements outlining the importance of democracy which stress that expressed concerns extend to the very foundations of the EU and its legitimacy. While fundamental rights are put in contrast with technology, a democratic political system is presented as challenged by those, primarily big tech companies, who do not follow democratic values in developing and using AI.

Lastly, safety emerges as being closely attached to both referent objects: something that needs to be ensured and worked for. Section 5.2.3. demonstrates that the EU defines safety as related to products and services, and contrasting with a more normative focus on fundamental rights and a democratic political system. However, what initially appears as a simplification ultimately reveals itself to be rooted in complex and challenging questions of human-machine interaction, such as how we will engage with robots and other autonomous technologies which move beyond the understanding of a *typical* product.

5.2.1. Fundamental rights

Both unacceptable and high risks indicate that fundamental rights are the primary focus, identifying them as the key elements to be safeguarded. The EP reiterated this point by stating that "the use of AI applications must be prohibited when incompatible with fundamental rights" (Europarl.europa.eu 2021a).

Fundamental rights as a referent object reemerge in more complex discussions which, as discussed in Chapter 5.1., overlap with the EU's endeavours to introduce different categories of risk and attached issues. For example, the EP claims that "risk of surveillance is present also in the school environment […] undermining the fundamental rights of children" (Europarl.europa.eu 2022a); and "the use and gathering of biometric data for remote identification purposes in public areas, as biometric or facial recognition, carries specific risks for fundamental rights" (Europarl.europa.eu 2020a). The European Council also specified that "the European approach to digital transformation and in particular AI should be human-centric and ensure the full respect and promotion of fundamental rights" (Consilium.europa.eu 2022). These articulations indicate that the scope of risks to fundamental rights is not limited to a pre-defined list of cases. This broadness of issues of fundamental rights was also reflected by a representative from the EC claiming that: "we again refer to the fundamental rights which is a generic term, but we also spell out sort of areas where the impact of decisions made with the help of AI would have an important impact on our lives"; and it "can

be anything from privacy violations up to fundamental rights in a more negative light […] where our lives may be really influenced by decisions that are taken by or supported by AI."

The importance of fundamental rights is explained as a longstanding EU political priority. A representative from the EUISS stressed this point by suggesting that "the EU has to always refer to fundamental rights. And the logic is you cannot refer to fundamental rights unless you are talking about humans. The treaties do not allow for anything else." This argument was also reiterated by Representative 1 from the HLEG, claiming that "you have to have human [element] which, I think, is a fundamental European value, or no technology taking over humans and the human decisions are for that." The stress on fundamental rights reflects the Charter of Fundamental Rights of the European Union, which points to necessity "to strengthen the protection of fundamental rights in the light of changes in society, social progress and scientific and technological developments" (Eur-lex.europa.eu 2012a). Also, the EU Treaty refers to fundamental rights "as guaranteed by the European Convention for the Protection of Human Rights and Fundamental Freedoms and as they result from the constitutional traditions common to the Member States" (Eur-lex.europa.eu 2012b).

These references to the EU's foundations may explain why fundamental rights appear to be presented as given, without specifying their characteristics. For example, Representative 2 from the EU Council suggested that "I think they are linked to what we as the EU see as fundamental rights, and also what we define for health and safety. So there is a sort of definition, but it is still very broad." The Charter of Fundamental Rights provides an extensive list constructed by the following chapters including more rights to each: dignity, freedoms, equality, solidarity, citizens' rights, and justice (Eur-lex.europa.eu 2012a). This list demonstrates that fundamental rights contain a broad spectrum, which becomes equally relevant if the EU does not specify particular elements, but refers to fundamental rights as a *block.*

On one hand, articulating fundamental rights as a referent object in the emerging EU AI policy suggests that the EU is consistent in its priorities, and also acts within the scope and mandate given by the EU treaties. On the other hand, centering fundamental rights as a core of AI-related concerns stresses that it is more than just compliance with EU foundational documents. The chosen phrases, that AI and AI uses are incompatible, undermining, create specific risks, and cause violations, deliberately place fundamental rights in a weaker position, highlighting their vulnerability in the face of technology. The suggested broadness of issues makes this vulnerability even more

complicated, because it strengthens the sense of uncertainty as to what extent and directions AI and AI uses might challenge fundamental rights.

The AI strategic discourse demonstrates that the EU establishes its position towards human-machine interaction as based on "mere virtue of the status as human beings" (Smuha 2021b, 594). It does not concentrate on a specific group of people (for example, only children or minority groups), or an end-list of AI uses that need to be regulated. Noticeable tendencies of contrasting surveillance, biometric data gathering, AI-driven decision making, with respect to fundamental rights and even human lives, demonstrates that the EU considers the human-machine interaction as antagonistic and as rivalry. The argument that AI may take decisions is reiterated by Representative 1 from the EU Council, suggesting that "the main element in this case is the risk against the human being. So it is all about those risks, and how much potential the AI system has to develop, for example, to a certain decision that might be detrimental or dangerous for a human being."

Thus, fundamental rights are considered to be an overarching, even universal, and most prioritized matter: the EP refers to "the intrinsically European and universal humanist values" (Europarl.europa.eu 2021b). While the respect for fundamental rights is indeed a foundational principle of the EU, in the context of AI, this principle gains renewed relevance as it is brought into contrast with a non-human technological domain that poses high risks to human integrity. The phrase *fundamental* itself implies that something foundational is challenged, and therefore requires specific attention. The emphasis on fundamental rights raises important questions about what it means to be human in the digital age. The EU's focus on these rights challenges us to consider how they are understood within the EU's definition of humanity, which is characterized as distinct and vulnerable when compared to AI, seen as a non-human entity.

5.2.2. Democratic political system

Even though fundamental rights seem to be an integral part of a democratic political system, I put it as a separate referent object because of its significance in discussing high risks. For example, the AI Act suggests that "AI systems intended for the administration of justice and democratic processes should be classified as high-risk, considering their potentially significant impact on democracy, rule of law, individual freedoms" (Eur-lex.europa.eu 2024). The EC in states the White Paper that "it is more important than ever to promote, strengthen and defend the EU's values and rules, and in particular the rights

that citizens derive from EU law" (Eur-lex.europa.eu 2020d). Meanwhile, the EP puts humans and democracy together by arguing that "any use of high-risk AI should always be ethically guided and designed to respect and allow for human agency and democratic oversight", and that AI "should seek to enhance well-being and individual freedom, as well as preserve peace" (Eur-lex.europa.eu 2021b).

Following these points, a variety of characteristics (rule of law, freedoms, values and rules, rights, and democratic oversight) define what the EU attaches to democracy. Notably, the chosen phrases create parallels with more conventional security debates: *defend* values and rules, *preserve peace*, as if the EU were in a conflict mode and technology threatens to take over. This is especially illustrative for understanding existing inconsistencies, as the EU avoids directly discussing security in the emerging EU AI policy. By such a vocabulary, the EU reiterates the antagonism of human-machine interaction, where a democratic political system needs to be defended.

In addition to this, the EU's AI strategic discourse reveals that a democratic political system is framed as a referent object, not only because of AI-related concerns, but also those who develop and use technology do not necessarily follow democratic principles. Various points suggest that the increasing role of big tech companies is seen as challenging democracy as well. For example, the HLEG claims that "digital dependency on non-European providers and the lack of a well-performing cloud infrastructure respecting European norms and values may bear risks regarding macroeconomic, economic and security policy considerations" (Digital-strategy.ec.europa.eu 2019b). The EP also argues that

"dominant tech platforms nowadays not only have significant control over access to information and its distribution, but they also use AI technologies to obtain more information on a person's identity, behaviour and knowledge of decisional history; believes that such profiling poses risks to democratic systems as well as to the safeguarding of fundamental rights and the autonomy of citizens" (Europarl.europa.eu 2022a).

The member of the EP and co-rapporteur of the proposal for the AI Act, Brando Benifei also claimed that "while big tech companies are sounding the alarm over their own creations […] we will also fight to protect our position and counter dangers to our democracies and freedoms" (Europarl.europa.eu 2023e). The distinction, and even tension, between the EU, as representing a democratic political system, and big tech companies, as representing

technology, signals that they are about different priorities: the EU advocates for the protection of rights and democracy, and big tech companies for technological developments which may go in other than democratic directions.

The contrast between European values and innovation developers, citizens' rights and significant control, democracy and private interests, reiterates the impression that the EU is in a competition whose priorities and proposed forms of human-machine interaction will become dominant. Even though the institutions do not directly name specific companies, previously mentioned EC cases with Apple, Meta and Microsoft based on the DMA suggest that the EU refers to those *outside* the EU, requiring its intervention that democratic processes remain ensured and, in a way, controlled by the EU.

The tension between a democratic political system being at stake and tech companies proliferating high risks suggests that the emerging EU AI policy is legitimized on reiterating antagonism and the distinctions between *inside* and *outside*, pro-values and pro-technologies that put a democratic political system in a vulnerable position. At the same time, these concerns enable the EU to impose rules for companies, depending on access to the EU, and request to comply with EU standards (Bradford 2020). Therefore, the suggested vulnerability could also be read as a way for the EU to reinstate its position and legitimation, arguing for safeguarding democracy as a referent object.

Thus, a focus on a democratic political system represents a lasting approach by the EU in the context of digitalization, where democratic values, economic inclusion, sustainability and market making come along in the European polity (McNamara 2024). However, in the case of the emerging EU AI policy, concerns related to a democratic political system move from a rather abstract conversation on human-machine interaction, and reveal the competition between different actors, and how the EU attaches big tech companies to the side of antagonistic technology.

5.2.3. Safety

The notion of safety is defined as something that needs to be ensured towards referent objects. Instead of something extraordinary, safety is suggested as "an inevitable policy response to the challenges for public order and domestic stability" (Huysmans 2008, 68).

The EU defines the safety of products and services meeting established standards. For example, the HLEG in the Ethics Guidelines suggests that "it is crucial for safety measures to be developed and tested proactively"

(Op.europa.eu 2019). The EP also claims that "the concept of product safety encompasses protection against all kinds of risk arising from the product, including not only mechanical, chemical, electrical risks but also cyber risks and risks related to the loss of connectivity of devices" (Eur-lex.europa.eu 2020c). The EC in the White Paper argues that in the case of AI, products and services can give rise to risks that EU legislation does not currently address explicitly (Eur-lex.europa.eu 2020d). These points introduce the logic of safety in a similar manner: something that employs AI and places it on the market needs to meet certain standards. However, such a simple phrase becomes more complicated by trying to understand what the scope and the main concerns of safety are, or, as Representative 1 from the EU Council put it, "how to ensure safety?"

Safety is noticeable in different EU policies as well. It has already been traced in domains such as security, critical infrastructure, chemicals or requirements for a range of new machinery products (Justo-Hanani 2022). The EU policy frameworks which involve risk-based approaches show that both concepts are interconnected: for example, "to enact risk-based food safety inspections" (Paul, Bouder and Wesseling 2016, 3). The close attachment of safety to the concept of risk and the risk-based approach once again demonstrates that the EU keeps reproducing already-existing notions to address different issues. This tendency creates the impression that using a similar toolbox reinstates the EU's way of policy-making and specifies its subjectivity. However, the existing pattern does not eliminate further discussion on to what extent putting these notions, risk, fundamental rights, democracy and safety, together; the EU reproduces policy routine and tailors connotations of safety to a particular case.

In the case of AI, the relevance of safety here is much more nuanced than technical and technocratic references to products and services. For example, the EC in the White Paper suggests that:

"the autonomous behaviour of certain AI systems during its life cycle may entail important product changes having an impact on safety […] citizens fear being left powerless in defending their rights and safety when facing the information asymmetries of algorithmic decision-making" (Eur-lex.europa.eu 2020d).

Representative 1 from the HLEG claimed that "there is a common understanding and agreement that when you deploy AI technology in products and services, this should not undermine the safety in use, and this is like

physical safety and, you know, non-physical safety of individuals." Representative 2 from the HLEG suggested that "safety is simply the issue that if you have an autonomous car, it must not harm anyone. And if you have robots in a care facility, you must make sure that it is safe to use the robot and the care facility. Safety is integrity of human rights."

References to AI or robots surpass the existing understanding of a *product* as an object, its use and risks as "mechanical or electrical" (Eur-lex.europa.eu 2020c). The claim of the EC on robots is not about a *typical* product, but the one that develops a different level of interaction with humans:

"proximity to humans and interaction with them requires very high safety standards to prevent accidents and injuries. Robots are also becoming more and more connected to each other and other types of devices and process more data, posing potential privacy and cybersecurity risks" (Eur-lex.europa.eu 2021b).

In this case, the major difference appears considering safety in the context of *autonomous* technology problematizing human-machine interaction even further. The use of *autonomous* resembles the discussion presented in Section 1.2., arguing that autonomous technology challenges the human hierarchy associated with exclusivity of control and decision making. For example, the EP claims that "the lack of supervision may lead to serious risks for our safety and security, as well as for the rights and values underpinning our societies" (Europarl.europa.eu 2020e). Therefore, the typical definition of safety as *products meeting required standards* in terms of toys or food, here has different connotations, because humans are considered to be in a vulnerable position. In other words, autonomous AI-driven *products* are considered as already challenging the human hierarchy because of the gradual detachment from human control, and in this way pose risks to fundamental rights.

Such a way of articulating safety relates to broader trends and hype, where AI safety is employed by various state and non-state actors, and even entitled "the AI safety epistemic community" (Ahmed et al. 2023). For example, the UK government established the AI Safety Institute as a research institution that focuses on both "opportunities and gaps in technical AI safety research" (Aisi.gov.uk, n.d.). One of its outputs, the International AI Safety Report, was based on 100 AI experts' contributions "to advance shared international understanding of the risks of advanced AI and how they can be mitigated" (Gov.uk 2025a). Meanwhile, the US AI Safety Institute is tasked to identify, measure and mitigate risks to prevent the misuse of AI "by those who seek to

undermine public safety and national security" (Nist.gov 2023). In 2024, a meeting of the International Network of AI Safety Institutes took place to "bring together technical expertise to address AI safety risks and best practices" (Digital-strategy.ec.europa.eu 2024a).

Such popularity to consider AI safety and provide recommendations could be taken as genuine concerns and agreement among diverse stakeholders to search for ways to address them. However, this diversity and a load of initiatives suggest an intensifying competition of who will participate and frame those safety standards. Even introduced as *technical expertise*, safety becomes about rules, standards, and different views of what and how they should be defined and implemented.

Lastly, safety fits into the EU's endeavours to make sense of AI through already familiar, known and applied notions, which also make the phenomenon of AI more known and *manageable.* It is also about downplaying the political debates, challenges and disagreements. The proposed logic of safety of products and services suggests that it is framed as *non-security* and presented like public good, characterized by a positive, routine and technical nature, rather than as something inherently contested, tension-filled and divisive. At the same time, references to robots and other autonomous AI-driven technologies demonstrate that the present (products and services) and the future imaginaries (robots) are merged to act upon in the present (Amoore 2013). By reproducing the same notions and their definitions from other policies, the EU normalizes future unknowns, and a dramatic change of autonomous technologies being integrated into daily life (from cars to care) as another type of (future) products and services. However, their discussion in terms of potential harm to fundamental rights and a democratic political system reveal that this normalization includes highly controversial and even existential considerations of human-machine interaction and its forms.

Conclusion

This chapter began the analysis of riskification by searching for a referent object and its leading characteristics. The analysis demonstrated that the EU introduces two referent objects, fundamental rights and a democratic political system, as both defining categories of the most concerning unacceptable and high risks. The analysis has revealed that the significance of both referent objects is grounded on constructing antagonism and even conflict-related references between the EU's political priorities, including the importance of rights and democracy, and technology, and those who develop and use

technology (such as big tech companies) that do not follow the EU's position. Therefore, through the referent objects, human-machine interaction is described by the distinction of imagined *inside* and *outside*, where the EU, concerned about rights and democracy, is inside; technology and those prioritising its development over humans or exploiting it for undemocratic purposes, are *outside.*

Fundamental rights are interpreted in a broad and fluid way, but at the same time, considered as a block. They become an overall expectation *to put human first* in the face of still-unknown and unpredictable technology and its uses. The stress on *fundamental* underscores the EU's claim to universality, even if it reflects the EU's preferences and attachment to its own interpretation of the Charter of Fundamental Rights. However, the presentation of fundamental rights as a clear and given priority that needs to be protected raises further questions of what roles or forms of adaptation the EU anticipates for humans to withstand identified risks, beyond articulating the position of vulnerability.

A democratic political system as another referent object demonstrates that concerns are not limited to still-abstract futures of human-machine interaction. Filling a democratic political system with various elements, such as the rule of law, rights, peace and freedoms, the EU suggests that all of them are challenged both by AI and those who use (or will use) AI in other than democratic ways. It is also about which actors, democratic states and pro-democratic political organizations, or big tech companies, set the tone and norms of how technologies will be developed and used. Therefore, for the EU, putting a democratic political system as a referent object is about a tension and competition whose world-views, standards and practices will take place.

In relation to both referent objects, safety appears as an element of insurance and a goal of itself, as if claiming that safety *is there* demonstrates that risks are mitigated and referent objects are protected. However, the chosen ways to portray safety as something technical and procedural (the concentration of products and services meeting regulatory standards) seem narrow and inconsistent to noticeable imaginaries of autonomous technologies. The analysis reveals that AI changes the very definition of what a product is: can we put a robot and a microwave in the same category as electronic devices? Such an unreflective presentation shows that safety is framed as a matter of fitting AI, robots and autonomous technologies into existing patterns, suggesting that their challenges can be addressed by reproducing established approaches. In this way, recycling already-existing

notions to the new phenomenon becomes an *ordering* mechanism, which defines the boundaries of EU priorities and object-subject dichotomy.

Overall, whatever the interrelation between fundamental rights, a democratic political system and safety demonstrates, they are about the *human* part in human-machine interaction. From different angles, they all represent concerns related to human agency. Notably, the analysed statements do not frame survival as a key concern, typical in the securitization process. Instead, the focus is on who has control and who will shape future constellations of human-machine interaction. The answers to these questions require not only the identification of referent objects, but also the responses of who or what challenges them, how the relations between the Self and the Other are described.

5.3. Conditions of possibility for harm

If the previous chapter introduced referent objects, this chapter focuses on *conditions of possibility for harm*. Coming back to the conceptual framework (see Chapter 2), the notion of potential harm is closely attached to risk, as referring to a different logic than the security-threat nexus. At the same time, the identified ambiguity of risks reappears in the discussion on potential harm as well. As the EP claims, "AI systems could be used to do bad things" (Europarl.europa.eu 2020c).

In this case, the following questions need to be addressed: what constitutes potential harm in relation to articulated risks; who or what is associated with the causes of harm; and what security understanding arises from identified harm. This chapter follows the conceptualization of conditions of possibility for harm (see Section 2.3.) by suggesting that the approach towards AI is heavily future-oriented, and therefore relevant articulations do not necessarily name concrete harm, but remain within the scope of *possibility.*

To start with the definition, it is discussed through considerations of how likelihood and severity are decided. Section 5.3.1. analyses these debates, and demonstrates that the answer is based on reproducing the interrelation between risk and harm as a basic definition. Then I move towards the searches for more concrete examples of conditions of possibility for harm in relation to two referent objects. Based on empirical evidence, Section 5.3.2. suggests that the EU already identifies intrusion and discrimination as forms of potential harm. By mixing different levels, individual and societal, intrusion and discrimination are presented as overarching, ultimately increasing vulnerability across various domains in relation to both referent objects. In

addition to this, AI autonomy is presented as the most radical concern, while remaining the biggest future unknown. Section 5.3.3. explains how autonomy is portrayed as a leading issue challenging human agency, and reiterating the antagonism of human-machine interaction. Even though AI autonomy is something from the imagination, it becomes an important argument reiterating the urgency to act.

I close this chapter by suggesting that the EU's AI-related security is conceptualized as agentic security (Section 5.3.4.). The proposed notion stresses that the EU's primary concerns revolve around human agency, attached to maintained control, decision making, and technology subordination. The emphasis on such a hierarchy demonstrates the established boundaries between humans as the Self and AI as the Other. This approach underscores why it is called *security*, not merely as a preventive plan to mitigate AI-related risks, but as an urgency to safeguard what it fundamentally means to be human in a European sense.

5.3.1. Definition(s) of harm

The EU's AI strategic discourse suggests that harm is not defined through one specific situation. As a representative from the EUISS argued, "you do not entirely know what will come or be harmful uses." According to the EC, technology holds the possibility to harm public and private interests, violate data privacy and information security, and introduce bias (Commission.europa.eu 2024).

Documents also refer to different spheres and levels that might experience potential harm. For example, the EC White Paper mentions harm as both material (safety and health of individuals, including loss of life, damage to property) and immaterial (loss of privacy, limitations to the right of freedom of expression, human dignity, and discrimination, for instance, in access to employment) (Eur-lex.europa.eu 2020d). Or, as the HLEG suggests, potential harm can even be emotional: "a particular risk in the case of intelligent robots with whom humans might form an intimate relationship" (Op.europa.eu 2019). The EP aims to establish several criteria of the definition: the interplay between the purpose of the use for which AI is put on the market, the manner in which it is used, the severity of the potential harm, and the degree of autonomy of decision making that can result in harm (Europarl.europa.eu 2020c). Therefore, these criteria show that the EU does not explicitly provide the definition, but it can be traced in relation to named concerns, degrees or severity. Again, references to material, immaterial and emotional harm

suggest that it is context-specific, raising further questions of how it is evaluated.

In this context, the term *significant* becomes key: "the potential to *significantly* affect the lives of individuals" (Europarl.europa.eu 2021a); "a *significant* potential to cause harm to one or more persons" (Europarl.europa.eu 2020c); and "a *significant* risk to cause injury or harm that can be expected to occur to individuals or society in breach of fundamental rights and safety" (Europarl.europa.eu 2020a). Aiming to understand what *significant* means, the judgment of a level of significance is attached to technologies becoming autonomous. For example, the EC claims that "the operation of some autonomous AI devices and services […] may cause *significant* harm to important legal interests like life, health and property, and expose the public at large to risks"; and that "the future 'behaviour' of AI applications could generate mental health risks for users deriving, for example, from their collaboration with humanoid AI robots and systems" (Eur-lex.europa.eu 2020c).

These points, although not necessarily involving the direct language of potential harm, further ground the tensions and antagonism in human-machine interaction. References to negative effects on individual physical and emotional well-being, privacy, human dignity, fundamental rights and safety imply that technology, especially its functioning as *autonomous*, is perceived as exploiting human vulnerabilities which are overarching and cannot be put on an end-list. By putting life and health together with robots, systems and devices, the EU creates a clear contrast between humans and machines like a paradox: human invention, defined as a technical object, potentially causing harm to *humanness* attached to rights and freedoms.

Thus, like risk categories, the decision on potential harm is political. Defining it through ambiguous notions such as *significant*, the EU creates a spectrum of options and room for manoeuvre in *riskifying* AI developments and uses as potentially harmful. This is also a clear difference from threats, as potential harm is put in terms of possibility, while the references to *significant* do not indicate the same level of concern as *existential.* Nevertheless, the leading and foundational axis for definition remains a matter of human agency. According to the EU, the more AI and its uses challenge, interrupt and disrupt human agency, or detach from human control, the more significant are the conditions of possibility for harm.

5.3.2. Intrusion and discrimination

The search for a definition presented above has created the impression that situations in which AI and its uses become harmful could be limitless. However, documents give relevant hints suggesting that conditions of possibility for harm are closely attached to intrusion and discrimination. They are not necessarily put in the future tense, but are argued to be already noticeable, reducing the stress on *possibility* and signalling existing precedents. For example, the EC in the White Paper (Eur-lex.europa.eu 2020d) suggests that AI "entails opaque decision-making, gender-based or other kinds of discrimination, intrusion in our private lives or being used for criminal purposes." The AI Act claims that the use of AI "for 'real-time' remote biometric identification […] is considered particularly intrusive in the rights and freedoms to the extent that it may affect the private life of a large part of the population"; and "may lead to discrimination of persons or groups and perpetuate historical patterns of discrimination, for example based on racial or ethnic origins, disabilities, age, sexual orientation, or create new forms of discriminatory impacts" (Eur-lex.europa.eu 2024).

Similar points overlap in other documents stressing different forms of intrusion and discrimination. For example, the FRA suggests that "AI systems based on incomplete or biased data can lead to inaccurate outcomes that infringe on people's fundamental rights, including non-discrimination" and "lead to a disadvantage for certain groups, such as women, ethnic minorities or people with a disability" (Fra.europa.eu 2022). Several reports by the EP reiterate that algorithms "can discriminate unfairly and perpetuate stereotypes and social biases, use toxic language (for instance inciting hate or violence), present a risk for personal and sensitive information, provide false or misleading information" (Europarl.europa.eu 2023b) or "create a risk of harm to legally protected interests, both material and immaterial ones" (Europarl.europa.eu 2020c). Therefore, as the EP puts it, AI and its uses have "a potential for bias, manipulation and spreading of disinformation, which risks weakening societies" (Europarl.europa.eu 2023a). This spectrum of forms of intrusion in privacy and discrimination directly refers to both referent objects where the importance of fundamental rights indicates potential harm to vulnerable groups, and minorities.

The focus is not only on AI and its applications, but also on those who may exploit or misuse AI, multiplying sources of potential harm. The EP refers to "interference by third parties with AI-based autonomous technology" (Europarl.europa.eu 2021b), "malevolent use of AI that may have dreadful

effects for humanity" (Europarl.europa.eu 2018a), and "malicious actors in performing known attacks such as disinformation campaigns and malware coding" (Europarl.europa.eu 2024b). Some references appear less abstract and give more details of those third parties. For example, the EP claims that, and that "many authoritarian regimes use AI systems to control, exert mass surveillance over, spy on, monitor and rank their citizens or restrict freedom of movement; is highly concerned about and condemns cases of EU companies selling biometric systems which would be illegal to use within the EU to authoritarian regimes in non-EU countries" (Europarl.europa.eu 2022a).

References to authoritarian regimes reiterate the EU's concerns of a democratic political system through the contrast with digital authoritarianism, defined as the use of technologies to repress and manipulate domestic and foreign populations (Polyakova and Meserole 2019). This, as Anu Bradford (2023) suggests, battle between techno-democracies and techno-autocracies is emphasized to distinguish the EU and its values from those who use AI for purposes misaligned with democratic principles. Such a position could also be understood not only in terms of risks and potential harm, but also the international context.

Coming back to Section 4.3., the EU can be considered as teaming up with Western pro-democratic countries, while distancing or, as this analysis suggests, riskifying an *alternative* approach developed by actors such as China. However, the EU's proposed dichotomy between democracies and autocracies seems to be simplifying the question of AI uses, while it is also about practices that can be done by democracies themselves (Yayboke and Brannen 2020). For example, tracing, hidden manipulation, targeted content, and access to personal data in unprecedented amounts are not limited to autocracies, but widely amplified by big tech companies functioning in democracies as well (Mantellassi 2023).

Identified elements (intrusion, discrimination, and the use of AI by authoritarian regimes) are considered as amplifying existing inequalities, reinforcing biases, embedding discriminatory patterns, and unfairly treating various societal groups. Examples given by EU institutions also mark different levels of harm, from individual, related to private life and the vulnerability of minorities, to societal, associated with mass surveillance or disinformation. These concerns also mix different temporalities (present and future), different imaginaries for what AI can be used, and different causes of potential harm, from algorithmic biases to authoritarian practices in seeking adversarial damage.

Thus, intrusion and discrimination direct the search for AI-related security, by demonstrating why human-machine interaction is articulated as antagonistic. AI is seen as an amplifier of already-existing challenges, in aiming to ensure human rights, equality and fair treatment. References to authoritarian regimes or other *third parties* exploiting AI, through disinformation, surveillance or control, make these concerns even more at stake, as they purposely articulate violations against the vulnerable.

5.3.3. Autonomy

The previous section demonstrated that intrusion and discrimination merge the present and the future in defining conditions of possibility for harm. Here, I go into the notion of autonomy as the most radical form of technological detachment from human control and decision making.

Autonomy shifts the paradigm, because technology becomes not a tool but a possible collaborator, which requires us to anticipate the possible adverse outcomes of these technologies (De Visser, Pak and Shaw 2018). In this case, human control is put under question, because technology is no longer dependent on human supervision. Even though autonomy remains a matter of the imagination, EU institutions already discuss it within the context of potential harm, blurring the lines between stages of *automated–autonomous–autonomy.* For example, the EP states that "AI technologies risk reducing human agency and […] should not substitute human autonomy nor assume the loss of individual freedom" (Europarl.europa.eu 2022a). Even more, the concerns are described through "the danger that conscious AI might have spurious motivations […] potentially fuelling dangerous populism" (Europarl.europa.eu 2023d), or that "AI can generate false information or spread a bias or opinions that do not represent the public sentiment" (Europarl.europa.eu 2023a).

Even though the official scope of the emerging EU AI policy excludes the military, several claims, especially focused on lethal autonomous weapons systems (LAWS), put the point of autonomy even more at stake. For example, the EP suggests that "the development of autonomous weapons is hard to control, and their proliferation is a risk. If they were actually deployed, the risk of malfunctioning, error or misuse should first be carefully addressed" (Europarl.europa.eu 2019b), and "stresses that AI-enabled systems can under no circumstances be allowed to replace human decision-making in this field" (Europarl.europa.eu 2021b). The point of human oversight was also raised by a representative from the EDA, claiming that "now we see more and more

applications with AI, especially when we are talking about unmanned vehicles and systems. How AI will at some point replace or supplement the human being at the tactical field?" These points suggest that different stages of *autonomous* and *autonomy* are blended together, stressing the detachment of technology from human control, and further reiterating the challenge.

These concerns relate to the previously discussed digital authoritarianism, as the EP refers to adversarial actors challenging not only human agency but also democracy: "military research and technological developments relating to lethal offensive weapon systems without human oversight that are pursued in countries such as Russia and China with little regard for the risk to humanity" and "non-state armed groups that can equip drones with AI software and turn them into cheap lethal offensive weapons" (Europarl.europa.eu 2022a); or "any LAWS or weapon with a high degree of autonomy can malfunction because of a badly written code or a cyber-attack perpetrated by an enemy state or a non-state actor" (Europarl.europa.eu 2021b).

The EU's perception of autonomy aligns with the concept of superintelligence described in Section 1.1., which implies that AI would not only reach but also surpass human intelligence, potentially leading to a loss of human control. Given the EU's concerns about human-machine interaction, it becomes evident why autonomy is viewed as the ultimate conditions of possibility for harm to human agency. Again, these concerns of the EU are not unique, but overlap with evolving hypes in political and public debates considering AI as a determined future challenge to humans. The discussions that humanity might permanently cede its control and depend on those who set up the computer system (Ord 2020) are actively pursued in framing related concerns by different actors, not only the EU. Even though these concerns are based on imaginaries and future uncertainty, the idea of the *possibility* that "superintelligence may generate catastrophic vulnerabilities" (Dafoe 2018, 10) already fosters a sense of danger and a need for a response. The EU's interpretations of autonomy seem to be joining the same hype.

Lastly, considerations that AI can reach autonomy contrast with the EU's proposed technical definition of AI. For example, the EP reiterates that "the capacity for self-learning and the potential autonomy of AI systems […] represent nevertheless a significant challenge to the effectiveness of the Union" (Europarl.europa.eu 2020c). A representative from the EC also put it that "because of the black box nature of AI, you simply sometimes do not know what and why and for what you know." This inconsistency between definition and perception reveal blurred boundaries between the present and

the future, existing and imagined, identifiable and obscure, technical and political.

Overall, autonomy becomes a radical condition for potential harm, *the end of humanity*, because it targets, from the EU's point of view, the very essence of human agency as a universal matter. Autonomy, whether in decision making or military applications, puts AI and human agency in a zero-sum game: AI autonomy, meaning (gradual) loss of human agency. However, in this case, the interaction becomes even more abstract: how to describe where AI autonomy starts, and human agency becomes limited or constrained.

5.3.4. Agentic security

Both referent objects of fundamental rights and a democratic political system, as well as the conditions of possibility for harm focused on intrusion and discrimination, indicate that the EU is concerned about humans and human agency in relation to AI. As discussed in Chapters 5.2. and 5.3., these elements reveal that human-machine interaction is a core aspect of framing security issues. As a representative from the EDA claimed: "one of the key questions that we expect to address […] is this state of the human in AI." Therefore, I suggest describing the EU's AI-related security as **agentic security**.

Agentic security is about protecting human agency, understood in general terms as the capacity to sustain control and decision making power vis-à-vis current and future AI and related concerns. This notion also emphasizes the maintained hierarchy of superior human position over technology, which needs to stay subordinated. Its key characteristics emerge as follows:

- First, the antagonistic distinction between the human (conscious, moral, making rational decisions and sharing pro-democratic values) and AI (product, technical, potentially enabling anti-democratic practices). In this way, the EU anchors agency in the human world as the modernist-liberal idea bound up with an ontology of rational individualism understanding the human in clear non-entangled ways that adhere to established ethical-legal categories (Leese 2019).
- Second, AI is perceived as Other, something that "will never have the essential quality of being "alive" (Stewart 2024). Conditions of possibility for harm have demonstrated that the spectrum of issues is wide: from intrusions into private lives, and discrimination based on biases and vulnerabilities, to AI autonomy, including the military. Therefore, AI's otherness amplifies concerns that position the technology not merely as

different, but as potentially harmful to the core principles of human rights and agency.

- Third, the EU's understanding further widens the concept of security, reaching to the impact of technological development. Agentic security, as another dimension, reiterates that security is no longer confined to national or territorial concerns. Instead, AI reshapes the understanding of security by introducing risks linked to technology and imaginaries of its future roles and involvements. Therefore, this dimension focuses not on variations of human or societal security, but specifically on the complex, and, from the EU's perspective, problematic symbiosis between humans and machines.

These characteristics can be further elaborated and situated within broader debates on AI and security, each revealing relevant dimensions for the importance of the notion of agentic security.

The first characteristic, antagonistic human-machine interaction, reflects the intensifying academic and political debates over forms of human-machine interaction. They mainly discuss whether agency becomes more distributed and entangled, or whether it should remain tied to the ideal of the conscious human subject defined in legal and moral terms of accountability and responsibility (Leese 2019). In either case, human-machine interaction is problematized, because it triggers a reconsideration of the human role and subordination of technology.

Those who argue that human-machine interaction implies a combination of both human and machine decision-making, define agency as distributed, meaning a blurred distinction between instances of human agency and AI (Bode and Nadibaidze 2024). This perspective claims that technology is no longer just instrumental for human actions and decision-making, but human-machine interaction becomes comparable to human-human interaction (Strasser 2022). Then incorporating AI into various processes of decision-making are not just about the delegation of tasks, but also the sharing of human functions and abilities. For example, decisions to delegate tasks to machines on the battlefield are considered as moral responsibilities linked to human agency (Taddeo 2024). Therefore, those who refuse such interpretations argue that only humans hold agency, as technology is incapable of exerting control, responsibility or dignity. Political positions that express concerns about human control seek to maintain human involvement by implementing various policies, actions, and regulatory practices (for example, Vesnic-Alujevic, Nascimento and Pólvora 2020; Leese 2019; Ferl 2024).

The EU's formulation of human-machine interaction as antagonistic, and its insistence on anchoring agency in the human, aligns with these broader debates, revealing that the EU takes a political stance in reinforcing a separation between humans and machines. At the same time, the debates introduced here are not confined to a specific sector or application of AI; rather, they demonstrate broader considerations about the nature of interaction itself, something that developments in AI compel us to reconsider. As both the referent objects of fundamental rights and the political democratic system are discussed in rather general terms, the specific type of human agency – whether individual, institutional, or societal – is not detailed. Consequently, the discussion seems to be more *universal* in how human is understood. Even in its generality, this framing suggests directions that reveal the EU's priorities and its orientation towards the European citizen and the functioning of its political system.

The second point, on AI as the Other, refers to a more specific relationship with technology, one that establishes a form of security knowledge in which both agentic and security dimensions define what is at stake. *Agentic* goes straight to the essence of the EU's perception that human decision-making is morally superior to algorithmic decision-making (Amoroso and Tamburrini 2020). Following this logic, the EU's acceptance that agency transforms into embodied, situated and dynamic perspectives between humans and AI would result in a form of *capitulation* that would necessitate a fundamental reassessment of the very pillars upon which the EU's normative and regulatory frameworks rest. For example, the EP specifies that "machines remain unable to share a goal with a human" or "machines do not share intentionality with their operator: they are tools that are unable to collaborate, regardless of how intelligent they appear" (Europarl.europa.eu 2018b). Therefore, *agentic* is about AI triggering considerations and concerns of what it means to be human in the digital age, even though the EU regards human agency as an unconditional and monolithic principle. This is how *agentic* is about the anthropocentric position that only humans have the right to retain agency, and any other forms of agentic capacities are perceived as directly and harmfully affecting that entitlement.

Security refers to what Didier Bigo (2001) proposed as a search for established boundaries. Here, human-machine interaction is defined as antagonistic, where boundaries are between humans as the Self and AI as the Other, and security is understood as protection from technology taking over. Through articulating risks of (potential) autonomy or (mis)uses to undermine the human, the EU reveals fears that AI will rob human agency by usurping

control, taking away the decision-making power and diminishing capacities (Boddington 2023). As has been mentioned, conditions of possibility for harm are not limited to a concrete domain, but focus on vulnerable groups and vulnerabilities potentially amplified by AI. The use of risks and the suggested shift towards the future and long-term mitigation, rather than immediacy and extraordinary measures, also suggest that security here is the inscription of boundaries between the Self and the Other, where, despite future uncertainty, alternatives of *distributed* agency are considered lost technology subordination.

The third characteristic, a widened understanding of security, can be situated within broader discussions on the different dimensions of security, particularly those that focus on the human and/or society. Jef Huysmans (1998) pointed out that if we add different adjectives (environmental or societal) to security, it is still to be discussed what security means in very different sectors. Barry Buzan and Lene Hansen (2018) also claimed that new security dimensions remain an area of controversy, considering whether they fragment or produce a more nuanced conversation on security. At the same time, Barry Buzan's (1991; 1984) own proposal to split security into five dimensions, political, military, economic, societal and environmental, is a good example of a range of concerns that are not limited to power and peace, and framed as security. In terms of technologies, analyses on cybersecurity already signal that security understanding changes when new issues and/or actors emerge, for example, asking of the role and impact of botnets or malware (Liebetrau and Christensen 2021). Therefore, the examples mentioned in Section 2.1. of some security scholars engaging with STS concepts indicate that emerging technologies and related security practices require corresponding ways to describe this intersection.

In this case, agentic security is crucial because the existing concepts, such as human security and societal security, do not adequately address the concerns raised by the EU. These concepts typically focus on interactions between humans or groups, regimes, and political vulnerabilities, rather than on the relationship between humans and machines. For example, a UNDP report from 1994 identified four characteristics of human security: it is universal and applies to all people; human security includes both military and non-military sources of insecurity; the importance of prevention; and it is centred around (Peoples and Vaughan-Williams 2020). These characteristics align with the idea of human centrism, which involves universal concerns for fundamental rights and the mitigation of risks that underpin the concept of agentic security.

However, the validity of human security as both a policy framework and a category for research has been questioned (Florea Hudson, Kreidenweis and Carpenter 2013). Its relevance is often seen as more closely related to developmental goals, such as climate change, and issues concerning food and water supplies (Methmann and Rothe 2012). In terms of societal security, Barry Buzan (1991, 433) defines it as "the ability of societies to reproduce their traditional patterns of language, culture, association, and religious and national identity." This concept focuses on threats to societal cohesion or the clash of different identities that do not necessarily involve varying forms of agencies and agentic capacities. In short, both concepts emphasize the significance of the human element. Yet, they address different concerns than those raised by the EU regarding AI, highlighting the need for the introduction of agentic security.

Overall, agentic security is neither about the humanitarian challenges of certain regions or violent regimes that require the protection of vulnerable groups, as in the case of human security, nor about challenges to group identities, coming, for example, from migration or globalization, as in the case of societal security. It is, from the EU's perspective, about antagonistic human-machine interaction, where AI is viewed as Other, due to the potential harm to human agency and concerns about maintaining human control and decision-making power. The key issue here relies on the distinction between humans and AI, based on anthropocentric assumptions that humans should remain in a hierarchical position as superiors, while AI needs to stay subordinated.

The relevance of agentic security, filled with various AI imaginaries, ranging from *automation* to *autonomy*, highlights contradictions in the EU's position. While the EU is hesitant to embrace the ontological shift towards distributed agency, it also acknowledges that human hierarchy and agency may be challenged by AI, especially with growing concerns about AI's autonomy and the potential loss of control. Therefore, by emphasizing the importance of human agency, this concept highlights the instability of these distinctions and reiterates evolving concerns.

Conclusion

This chapter demonstrated that conditions of possibility for harm involve a broad spectrum of concerns, from intrusion into private lives, to different forms of discrimination and biases, increased disinformation, and weakened democratic societies. Even though the list seems endless, the examples

discussed demonstrate that the focus is on AI, and uses of AI, which undermine efforts to build liberal, democratic and inclusive European societies by targeting individual freedoms, vulnerable and minority groups, or democratic societal cohesion as such.

There is a clear link with both referent objects of fundamental rights and a democratic political system, while AI and AI's uses are presented as contradicting democratic foundations. While the challenges identified may not be entirely new in themselves, their articulation in the context of AI reflects a perception of technology as a distinct phenomenon. This perception prompts a reconsideration of existing concerns – such as data exploitation, surveillance – by framing them within a technological landscape that amplifies their potential (negative) impact and complexity.

By merging the present and future through stages of automation, autonomous or autonomy, and different levels of concerns, from individual to humanity, the EU develops its security understanding by establishing and defining AI as Other. In a contrast with the initial focus on technical characteristics in defining AI, the emphasis remains on technology as an enabler of intrusion and discrimination or detachment from human control. Therefore, conditions of possibility for harm are driven by political imagination, which merges different forms of using AI, and reiterates the urgency to address these challenges.

Compared to securitization, where threats are immediate and need to be confronted, the conditions of possibility for harm rely on the remaining uncertainty of how technology may (or may not) develop and be used, and what potential harm it might create. Therefore, human-machine interaction highlights the undefined future, where the construction of security knowledge becomes a way to set boundaries, while the conditions of possibility for harm cannot be fully verified and remain a *possibility.* This means that even if the EU's AI strategic discourse refers to fundamental rights as a *block*, AI and related imaginaries also construct how "European values" are shaped and prioritized (Niklas and Dencik 2021, 22-23). For example, concerns about intrusion into private lives, discrimination, disinformation, surveillance and even human agency also affect the way human rights and democracy are understood and interpreted in the context of emerging technologies.

Lastly, the concept of agentic security introduced here demonstrates how the EU constructs this security dimension, focused on the protection of human agency. This notion involves maintaining control, ensuring decision-making capacity, and establishing a hierarchy that subordinates technology, all of which are seen as inherent to human exclusivity and as a privileged position

in antagonistic human-machine interaction. This approach to security is based on AI as the Other, where concerns about the distribution of agency – specifically, the potential for AI to assume human capacities – are viewed as the main challenges.

In this context, *agentic* becomes the core element, highlighting the importance of human agency, while *security* refers to the established boundaries between humans and AI. As a result, agentic security aims to safeguard both referent objects that are foundational to the EU itself, namely, the protection of fundamental rights, and representative democracy, as outlined in the Treaties (Eur-lex.europa.eu 2012b).

5.4. Normative and institutional governance programme

The previous two chapters focused on the question of *what*: what constitutes a referent object, and what the conditions of possibility for harm are. This chapter moves to the question of *how*: how the EU proposes to address identified AI-related security, and, following the HLEG, to "adopt adequate measures to mitigate these risks when appropriate, and proportionately to the magnitude of the risk" (Op.europa.eu 2019). It focuses on analysing an introduced governance programme, by asking how the EU aims to manage risks: what policy measures are proposed and characterized, and how a governance programme is legitimized in the EU's AI strategic discourse. As has already been mentioned, the logic of risk does not require extraordinary measures, but it is more oriented towards a long-term policy that addresses and mitigates perceived concerns. It also relates to the concept of governance (established structures, norms, rules and institutional rationalizations), suggesting a spectrum of measures that can be involved under the name of governance.

The analysis demonstrates that the EU combines multiple types of measures: normative principles of human-in-the-loop and human centrism, and regulation, as insurance that the development and uses of AI remain under human supervision and needs. Section 5.4.1. shows that both principles address agentic security, aiming to prevent AI from intruding on human agency or detaching from human control. In addition, regulation, as discussed in Section 5.4.2., becomes a part of normative measures, by putting suggested principles in legally binding rules. Thus, these measures are not only about long-term mitigation, but aim to proactively steer future AI developments to increase their adherence with the EU's approach.

The second type of measure is more focused assessments and an institutional ecosystem which are supposed to *practically* implement the normative principles and regulation. Section 5.4.3. demonstrates that debatable and complex questions, such as the decision on the level of risk, are delegated to different institutions or external expertise, to make policy claims assessable, explainable and transparent. The EU grounds these measures on its experience of other policies, and claims their relevance as being impartial and evidence-based, rather than political.

Based on the identified governance programme, Section 5.4.4. suggests the notion of technocratization, which specifies the standardization of policy measures, reliance on expertise, and dispersed responsibilities between administrators and external stakeholders, for implementing the policy. By proposing this conceptualization, I reiterate the shift from securitization and *high politics* towards riskification and normalized routine, because agentic security is to be ensured through technocratized governance rather than extraordinary measures. What is specific to the case of AI is that technocratization also reflects the EU's tendency to mirror the chosen way to define AI through technical terms, and focus on scientific knowledge as providing guidance in the face of uncertainty.

5.4.1. Human-in-the-loop and human centrism

The normative elements of the governance programme relies mainly on two principles, human-in-the-loop and human centrism, which prioritize both referent objects and set expectations for future human-machine interaction. They demonstrate a direct link with agentic security, because of the remaining focus towards humans and the human role in relation to AI. It is important to specify the difference between them. A representative from the EC provided the most concrete distinction:

"human-centric goes back to what I said about technology to serve people and not the other way around. So that would just mean that we use AI in a way that improves our society, improves our lives. The human-in-the-loop is more sort of a safeguard in the system itself, to be sure that the outcome is human-centric. Just the difference that the outcome should be human-centric technology. The human-in-the-loop is one of these mitigation steps."

Documents suggest similar interpretations. Human-in-the-loop is introduced as a principle of human oversight, meaning that humans control

AI. For example, Representative 2 from the HLEG argued that "the requirements of human oversight and control are the means to mitigate the risks of threatening human autonomy. For autonomous or semi-autonomous systems, it is very clearly assigning accountability and responsibility to the human." The HLEG suggests that "human oversight helps ensure that an AI system does not undermine human autonomy or cause other adverse effects. Oversight may be achieved through governance mechanisms such as human-in-the-loop" (Op.europa.eu 2019). This principle also refers to "the capability for human intervention in every decision cycle of the system" (Digital-strategy.ec.europa.eu 2020), and, according to the EP, "establish adequate safeguards, including design systems with human-in-the-loop control and review process" (Europarl.europa.eu 2020b). Governance, intervention, control or review: all indicate a proactively established complex process attached to human oversight as a solution that reduces AI-related risks.

The human role in decision making is articulated as a key element in ensuring the human-in-the-loop principle. For example, the EP claims:

"it is of the opinion that any decision taken by AI, robotics or related technologies within the framework of prerogatives of public power should be subject to meaningful human intervention and due process, especially following the assessment of those technologies as high-risk" (Europarl.europa.eu 2020a).

Representative 1 from the HLEG claimed that "we were concerned that you know […] basically not having human in the decision-making process will actually threaten, you know, a human"; while Representative 2 from the HLEG pointed out: "we have to make sure that no application is threatening human rights. We do not want to have that." The EP similarly suggested that "it is always up to a human to decide whether these decisions can be made automatically, without further control, or if human intervention is needed" (Europarl.europa.eu 2018b); while the AI Act claims that "high risk AI systems shall be designed and developed in such a way […] that they can be effectively overseen by natural persons" (Eur-lex.europa.eu 2024).

These varying articulations share similar points, by reiterating that human-in-the-loop is about controlling technology and maintaining a hierarchy towards AI. References to human oversight, intervention, control, review processes, appropriate involvement and meaningful action show that the EU searches for appropriate forms of AI control, while governing agentic security. The point of the EP that there is a need to have "stop buttons for

human intervention to safely and efficiently halt automated activities at any moment and ensure a human-in-the-loop approach" (Europarl.europa.eu 2022a) illustrates that the human is also closely attached to responsibility, as inherited in human agency. While AI is Othered, the reference to *stop buttons* suggests that human involvement is expected to be inscribed in human-machine interaction.

The importance of the human-in-the-loop brings us back to Section 1.3., where the same principle has been discussed in military terms. In that context, maintained human control and responsibility are put as a life-death situation on the battlefield, leading to continuous debates on how to ensure human-in-the-loop in a *meaningful* way. Even though the EU does not discuss human-in-the-loop in military terms, the same connotations and dilemmas are present. For example, Representative 2 from the EU Council stressed this, by arguing that there "also needs to be some sort of human-in-the-loop to make sure that there is not only automated decision-making, but also human that needs to do the checks and balances." A representative from the EC also stressed that "human agency comes in to supervize the system and to correct it, in case of mistake." Therefore, this conversation about the human role supports the idea that agentic security extends beyond mere single market regulation, and raises foundational concerns about human-machine interaction, which appear central regardless of the policy domain. The focus on human agency once again reiterates that the human is expected to dominate over AI, while control and supervision is exercized by the agency as such.

When it comes to human centrism, this principle is defined as "a central but multivocal concept that is mainly used to bundle together a set of ethical and human rights principles" (Sigfrids et al. 2023, 3). Unlike human-in-the-loop, which focuses on direct human involvement, human centrism emphasizes outcomes in a rather ambiguous way, promising that human-machine interaction will be more human-centred (Floridi 2021). For example, the Commission President Ursula von der Leyen suggested that "AI must serve people and, therefore, must always comply with people's rights. This is why we are promoting a responsible, human-centric approach to Artificial Intelligence" (Ec.europa.eu 2020). The similar point by the EC that "development, deployment and use should always be at the service of human beings and never the other way round" (Europarl.europa.eu 2020a) implies that AI must remain subordinate and oriented towards human needs, representing the embodiment of rights and democratic ideals.

Such expectations are detailed by suggesting that should be understandable to humans and in line with human rights. For example, the HLEG defines it as a way

"to ensure that human values are central to the way in which AI systems are developed, deployed, used and monitored, by ensuring respect for fundamental rights […] all of which are united by reference to a common foundation rooted in respect for human dignity, in which the human being enjoys a unique and inalienable moral status" (Op.europa.eu 2019).

Other institutions articulate similar elements. The EC claims that "by striving towards human-centric AI based on trust, we safeguard the respect for our core societal values and carve out a distinctive trademark for Europe and its industry as a leader in cutting-edge AI that can be trusted throughout the world" (Digital-strategy.ec.europa.eu 2019a); and that "a human-centric approach to development and use, the protection of EU values and fundamental rights […] are among key principles that guide the European approach" (Eur-lex.europa.eu 2021b). References to societal and EU values, fundamental rights, human dignity and moral status once again reiterate the contrast between human and AI as a divergence between embodied human ethics and disembodied AI.

Consequently, this concept emerges as foundational but somewhat abstract: what are the ways to measure if AI is *human-centric* enough, how does the level of human centrism relate to different levels of risks? In public policy and ethics debates, human centrism has received comments as providing limited guidance to regulate AI, or even masking diverging positions on exact meanings and challenges to adapt it to new rules (Ebers 2020; Rességuier and Rodrigues 2020; Smuha 2019; Schopmans and Cupać 2021). At the same time, the human role and decision-making in relation to technology is not questioned, but presented as an undeniable virtue to *contain* AI as Other, without assuming that what humans do is not necessarily meant only for human benefit. Such a framing of human centrism concentrating on technological wrongdoings overlooks the more critical point that issues of bias or discrimination are not just technological, but have historical, political and power aspects inscribed by humans while developing and using technology (Ulnicane and Aden 2023).

Even suggesting normative guidelines to ensure agentic security, this principle does not explain itself beyond the idea of *people first* as a response to an antagonistic Other. It is formulated as something evident: if there are

concerns related to human agency, then technology needs to be developed and used in a human-oriented way without a reflection of limitations. For example, existing evidence suggests that, even though these practices are put as high risk and remain controversial, the EU supports financially diverse AI-driven applications for border management and control which aimed to collect biometric data, detect emotion, and conduct migrant risk assessment (Desmarais 2025). Then the question is: to what extent are these evolving practices *human-centric*, or to whom are these requirements applied? In short, the Self is not scrutinized because of perceived concerns of the Other-AI, or those who develop and use it require to be closely examined, limited and subordinated. This tendency indicates an urgency, where AI is considered as already present and challenging human agency, in the worry that future developments and uses will outpace the capacity of policy-makers and the public to grasp these implications (Suchman 2023).

In this light, both human-in-the-loop and human centrism are closely related to agentic security, as they both refer to the need for human supervision, maintained control, and orientation towards human good. These principles as *response* measures to agentic security come as a promise and insurance that AI will not undermine human agency, and will not be compatible with fundamental rights and democracy. At the same time, both human centrism and previously discussed human-in-the-loop remain normative principles as guidelines and expectations defining the EU's priorities and aiming to steer future AI developments and uses in the preferred and corresponding ways. They continuously reinforce the antagonism between the Self and the Other, necessitating mechanisms, such as *stop buttons*, to maintain the existing human dominant order, while the human role in fostering AI-driven issues is not scrutinized.

5.4.2. Regulation

Regulation becomes an *embodiment* of the EU's approach towards AI, materialising in legally binding rules. The EC defines regulation as a way "to create framework that shape the context, allowing lively, dynamic and vivid ecosystems to develop" (Eur-lex.europa.eu 2020a), and "whereas European citizens could benefit from an appropriate, effective, transparent and coherent regulatory approach at Union level" (Europarl.europa.eu 2021b). The EP also suggests that regulation is expected to set "common European standards for European citizens and businesses to ensure the consistency of rights and legal

certainty" (Europarl.europa.eu 2020c). Therefore, it sets the tone, inscribing references to both fundamental rights and a democratic political system.

The need for regulation is also based on urgency, while the status quo is presented as missing and lacking a response towards perceived concerns. The EP claims that "current regulatory frameworks, both on EU and Member State level, are too fragmented, too ponderous and do not provide for legal certainty" (Europarl.europa.eu 2022a); while a representative from the EP suggested that "at the moment, there is nothing. So, my point is that it will work only if we have standards very quickly." Having standards and regulation is in line with the already-discussed tendencies of the EU adopting diverse rules on different matters and becoming one of the cornerstones of the governance programme, confirming the EU's reputation as a regulatory power.

The AI Act is a way "to ensure a consistent and high level of protection of public interests as regards health, safety and fundamental rights, common normative standards for all high-risk AI systems should be established" (Eur-lex.europa.eu 2024). References to appropriate, effective, transparent and coherent regulation, certainty and consistency, and European standards, demonstrate that regulation is presented as the antipode of uncertainty and future unknowns. Like the described normative principles, regulation then gives a promise and reassurance from the EU to mitigate that challenging complexity and address conditions of possibility for harm, even if the likelihood of potential harm is unclear. Compared to extraordinary measures as an exception, regulation comes as a structured framework, established control and guidance, which are constant and long-term.

Chapter 4 has already demonstrated that different regulations are the EU's *modus operandi*, from the famous GDPR to more recently adopted regulations, such as DMA and DSA. The same logic is also proposed here, as, for example, Representative 2 from the HLEG claimed that "we are the first European region, the first large market that introduces such a comprehensive regulatory law." In this case, at least four interviewees directly mentioned the "Brussels effect" as a self-evident characteristic and reasoning for the regulation. For example, Representative 1 from the EU Council mentioned that "the EU as such can promote this Brussels effect"; Representative 2 from the EU Council argued for EU action, "that is why we need to move fast. And we still have the Brussels effect"; a representative from the EC referred to the pathway of "this famous Brussels effect, which started with the GDPR"; and Representative 1 from the HLEG explained that "we tried to do the Brussels effect." This elevation of the EU's regulatory influence positions regulation

as a central element of a governance programme, reproduced across multiple policy initiatives. Expectations to replicate the "Brussels effect" in the case of AI also suggest that regulation and established rules represent the *European* way of addressing security concerns in the digital domain.

At the same time, expectations that regulation is a guardian of rights and democratic values could be challenged by controversy inscribed in the AI Act. For example, it states that "the use of 'real time' remote biometric identification systems" is prohibited. However, it gives exemptions that, in the name of "a substantial public interest", such systems can be used for "the search for potential victims of crime, including missing children; certain threats to the life or physical safety of natural persons or of a terrorist attack" (Eur-lex.europa.eu 2021c). This example demonstrates that the declared normative expectations and a concentration on fundamental rights are not absolute.

Such an exception received external criticism for being inconsistent with advocating for rights and leaving room for potential violations. For example, Amnesty International responded to such a decision by stating that the EU "chose to prioritize the interest of industry and law enforcement agencies over protecting people and their human rights" (Amnesty.org 2024). This case illustrates that moving from ambiguous normative principles to concrete provisions and their implementation, the regulation itself remains subject to controversy.

From the ways it is introduced, regulation should not be seen only as a legislative process, but a manifestation of the EU's chosen direction, which, according to Anu Bradford (2023, 362), could be summarized as "rights-driven." References to fundamental rights, safety, protection, and mitigation of risks, demonstrate that regulation should be protective and oriented towards the advantage of human democracy, and reinstate the EU's reputation as a creator of norms and rules in *uncharted waters*.

Regulation is not merely about protection, but also involves a desired power position: to be the first to reproduce GDPR success, to export rules where rights and democracy serve as legitimizing tools and expectations to influence others. According to the EP, the EU's own "regulatory framework in the field of AI will have the potential to become a legislative benchmark at international level" (Europarl.europa.eu 2020b). Therefore, this power element involves a form of control not only towards AI, but also control of others who are required to comply or follow the set directions for developing and using AI in the future. By expecting that "any future regulation should follow", the EU signals that the previously discussed hierarchy involves not

just relations between humans and AI in human-machine interaction, but also the EU's positionality towards others, which are expected to oblige.

5.4.3. Assessments and an institutional ecosystem

In this section, I move to assessments and an institutional ecosystem as they are both expected to *translate* normative principles, regulation and overall approach towards AI into something procedural, defined, measured, decided and managed. As a representative from the EDA claimed, "risk analysis is always part of every work that we are doing, and it makes sense that we need to assess the risks for any new technology."

An assessment is about analysis and compliance with established rules, criteria, and the process of identifying and evaluating potential risks associated with AI. For example, the AI Act talks about a conformity assessment which is based on the process of verifying that an AI-driven product or service meets specific requirements and criteria before it is placed on the market: "high-risk AI systems must be assessed for conformity with these requirements before being placed on the market or put into service" (Eur-lex.europa.eu 2024). The EP also refers to "the assessment by the Commission of whether an AI system posing a high-risk should start at the same time as the product safety assessment" (Europarl.europa.eu 2020c). Therefore, the governance programme includes different forms of assessment which create different responsibilities and purposes.

In the case of risk assessment, for example, the AI Act suggests that it "shall be established, implemented, documented, and maintained in relation to high-risk AI systems" and "would require a full, effective and properly documented ex ante compliance with all requirements of the regulation and compliance with robust quality and risk management systems" (Eur-lex.europa.eu 2024). The EP suggests that "obligatory ex ante risk self-assessments […] seem to be a sufficiently robust governance approach for AI" (Europarl.europa.eu 2022a), and that "the determination of whether artificial intelligence should be considered high-risk […] should always follow from an impartial, regulated and external ex-ante assessment based on concrete and defined criteria" (Europarl.europa.eu 2020a).

An assessment is also put as an obligation to others, as the EC claims, "to assess and mitigate the risks their models entail, comply with some design, information and environmental requirements and register such models in an EU database" (Europarl.europa.eu 2023c), and that the Member States are urged to "assess the risks related to AI-driven technologies before automating

activities connected with the exercise of state authority" (Europarl.europa.eu 2021b). In short, these articulations do not define assessments in terms of their implementation, but outline their purposes, which are oriented towards making an informed decision about a level of risk.

The answer of how assessments are supposed to provide that knowledge remains rather abstract. The EP refers to a "robust risk assessment" (Europarl.europa.eu 2017), while the HLEG, in its Ethics Guidelines, stresses that risk assessment should help to "minimize negative impact" (Op.europa.eu 2019). For example, the HLEG in the document entitled "The Assessment List for Trustworthy Artificial Intelligence for Self-Assessment", released in 2020, lists seven requirements which are explained as ways to operationalize and decide on risks, as well as their management through the assessing process. The requirements are based on these broad directions: human agency and oversight, technical robustness and safety, privacy and data governance, transparency, diversity, non-discrimination and fairness, societal and environmental well-being, and accountability (Digital-strategy.ec.europa.eu 2020). However, each requirement seems to be similarly normative, and requires further operationalization of how they translate in assessing AI-related risks and the level of *riskiness.*

The same impression comes from another document released by the HLEG entitled "Sectoral Considerations on the Policy Investment Recommendations for Trustworthy Artificial Intelligence", which provides guidelines for assessing risks depending on where AI might be used, and what issues might be caused. For example, the public sector is advised that "civil servants should be increasingly acquainted with the ethical, legal, social and economic impact of AI while remaining wary of potential adverse impacts on fundamental rights, democracy and the rule of law" (Op.europa.eu 2020). The General-Purpose AI Code of Practice, a guiding document for providers of general-purpose AI models to comply with the AI Act, was also developed together with "independent experts", consisting of 1,000 stakeholders. Its purpose is to "clearly point out the obligations" (Digital-strategy.ec.europa.eu 2024b) reiterating the argument that the governance programme is driven by endeavours to establish the assessable and competence-based process of policy implementation. Thus, continuous references to rights, democracy, transparency and other values reinforce the importance of assessment in making judgments and decisions in addressing agentic security.

This point demonstrates that knowledge about the implementation of assessments remains limited in the same way as established risk categories, and their distinctions are based on imaginaries and political considerations of

AI. For example, Representative 2 from the EU Council revealed that "we are still discussing how it will turn out to have the oversight and how it is applied." Examples from other policies applying risk assessment suggest that the exercise of assessment is more important than operationalization or clarification, because its use supports the EU's point of grounding the policy on scientific and expert evidence. For instance, in the case of health risk assessment, the EC explains it as a way to assess the magnitude of risks and determine possible options for a response, mobilising expertise to provide robust scientific advice to feed into coordinating the response (Health.ec.europa.eu 2025). The European Environment Agency suggests that the European climate risk assessment is about identifying the environmental, social and economic conditions that are most relevant for specific climate risks, including those that require consideration for adaptation policies (Eea.europa.eu 2024).

These tendencies reveal a paradox: political matters and normative principles, such as fundamental rights, democracy and human centrism, are supposed to be transformed into technocratic ways and operationalization allowing us to *assess* them in terms of risk. For example, the EC claims that "potential negative impacts of AI systems should be identified, assessed, documented and minimized. These assessments should be proportionate to the extent of the risks that the AI systems pose" (Digital-strategy.ec.europa.eu 2019a). However, the importance of different assessments reappears as a policy measure in the same words as normative principles. This tendency suggests that instead of measurable, quantified or objective criteria representing the proposed technocratic nature of assessments, they remain grounded on political imaginaries of AI.

A proposed institutional ecosystem comes together with assessments because it also contains expectations to clarify, evaluate and manage the key elements of the emerging EU AI policy. The institutional ecosystem is supposed to consist of three interrelated forms of institution: specifically dedicated national authorities in each Member State, an EU-level coordinating body, and diverse independent platforms which include different stakeholders. Beyond these official institutions, the EC also claims ambitions to set up networks of AI research excellence centres focusing on "explainability and advanced human-machine interaction", and digital innovation hubs to focus on "AI in manufacturing and on big data" (Digital-strategy.ec.europa.eu 2019a).

At the beginning of 2024, the EC already established the AI Office as an EU-level coordinating body of which the mission is to "support the

development and use of trustworthy AI, while protecting against AI risks" (Digital-strategy.ec.europa.eu, n.d.-c). However, it has already received different reactions. For example, Representative 2 from the EU Council claimed that "so there will be an AI board or an AI office and how will that work together? And what kind of formulas do we want? So, we need to see in practice, what will be the exact problem." Publicly, the AI Office has been criticized for being integrated into the EC rather than a separate independent agency, and received calls for clarification in choosing its leadership and transparency (Gkritsi 2024).

The documents outline that the institutional ecosystem is supposed to be a *guard* of established rules and obligations. For example, the HLEG discussed the need to

"institutionalize a dialogue on AI policy with affected stakeholders to define red lines and discuss AI applications that may risk generating unacceptable harms, including applications that should be prohibited and/or tightly regulated in specific situations where the risk is for people's rights and freedoms" (Digital-strategy.ec.europa.eu 2019b).

The EP refers to responsibilities which also become dedicated to institutions, such as national supervisory authorities "for ensuring, assessing, and monitoring compliance with legal obligations and ethical principles", or "an independent administrative authority to act as a supervisory authority" (Europarl.europa.eu 2020a).

References to red lines, compliance with obligations and ethical principles, independent providers and authority, demonstrate that their main role is institutionalized control. At the same time, *independent* bodies signal endeavours to eliminate the political dimension and present institutions as being predominantly expert-based, as the AI Act describes, "to exercise enhanced oversight over those AI systems posing high risks to fundamental rights" (Eur-lex.europa.eu 2024).

This reliance on institutions is not specific to the case of AI. The EU has been establishing different instruments and bodies assigned with the task of implementing newly developed policies. For example, the European Climate Law established the European Scientific Advisory Board on Climate Change to provide independent scientific advice. It also assigned the European Environment Agency responsibility for producing regular reports, such as the European Climate Risk Assessment, which is considered an "independent scientific report" (Climate.ec.europa.eu, n.d.). Additionally, Member States

were required to submit national energy and climate plans to trace progress. Similarly, the European Space Programme demonstrates how multiple institutions were established to support a policy framework. This includes the EU Agency for the Space Programme, the European Space Agency, the EU Satellite Centre, and the EU Space Support Office. Each of these entities has been tasked with different responsibilities, ranging from security to business engagement and research. Collectively, they illustrate how institutions become integral to governance, taking on key roles in policy implementation.

What is important in the case of the emerging EU AI policy is that existing institutional formats are not so different from other policies. The EU does not seem to be searching for new forms of response to articulated concerns, uncertainty and complexity. Addressing security is delegated to the routine mechanism between institutions and advisory boards, and responsibilities distributed between them, which are adopted to a specific issue.

As has been discussed, the risk-based approach and categories of risk are relational, and could be further contested on their grounds and the perception of potential harm. In this case, institutions, together with other policy measures, are tasked with solving the political debate and turning into something based on *scientific evidence*. In addition to this, through practices of assessing, regulating, clarifying, monitoring and notifying, the EU can claim *action* towards identified issues, and legitimize its own positionality of being the first to regulate and set standards on how AI can be governed.

5.4.4. Technocratization

The proposed framework for the governance programme, which includes the introduced policy measures, demonstrates that the EU frames its security response not through extraordinary measures, but as a technocratized governance plan addressing identified risks and conditions of possibility for harm. This approach suggests a deliberate strategy to enhance security related to AI. I propose that the concept of **technocratization** defines this response.

By technocratization, I refer to a chosen mode of initiating and shaping the governance programme, one that frames governance as a matter of technical management rather than political deliberation. It establishes an implementation logic grounded in expertise and rationalized authority, offering both justification and structure for how, from the EU's perspective, governance should proceed. For example, the analysis on the issue of migration across the Mediterranean demonstrated that the management of security went through specific governing techniques and routines "despite

public skirmishes at the political level" (Hegemann and Schneckener 2019, 136).

This analysis suggests that the EU reproduces the same logic for the emerging EU AI policy as well, proving that the perceived transformation by technologies and digitalization does not lead to new models of governing them, but recycles existing frameworks. Technocratization could be characterized by these points:

- First, the governance programme is grounded in the standardization of how risk-focused issues are addressed. The EU reproduces policy measures such as regulation, various forms of assessments and institutions, noticeable in other policy domains such as climate and migration. These are adapted to the emerging EU AI policy, claiming efficiency and control, formalized through already-established mechanisms. Then the policy, in turn, reinforces this strategy by becoming another example where the EU applies its established governance logic, further consolidating a technocratic approach across different policy fields.
- Second, the EU emphasizes the authority of expertise, presenting scientific knowledge as neutral and objective. This point is evident in the introduction of assessments and the institutional ecosystem, where references to *scientific proof* or *independent authority* become unquestionable characteristics of these policy measures. Therefore, arguing that the governance programme is *expert-driven* presumes its legitimization, and treats expertise as a built-in guarantee of quality and impartiality, rather than as a contestable element.
- Third, responsibility is delegated to administrators or private actors, such as in the conduct of conformity assessments, even when the decisions involve politically contested issues, like the level of risk. This delegation reframes political questions as technical matters, dispersing responsibility across established institutions at both EU and national levels, as well as among external stakeholders such as AI developers. As a result, opportunities for public scrutiny and debate are sidelined, embedding implementation in adapted procedural and institutional routines, and reinforcing a depoliticized mode of governance.

Taken together, these characteristics connect with wider discussions around the EU's way to frame the governance programme and its implementation, which help illuminate their significance and implications.

The first characteristic related to standardization suggests that technocratization is a common approach used by the EU to address various issues through established and recognized policy measures, such as regulation. However, this analysis reveals that technocratization also applies to security, a perspective that is not usually considered in the formulation of security policies.

I argue that technocratization is a deliberate choice by the EU and is one of the methods it employs to deal with concerns that have been the subject of criticism for an extended period (Juncos 2017). While it can be argued that technocratization is typical of the EU, this position can be challenged by examples illustrating that the EU has also engaged in securitizing various issues to "marshal its authority and its resources to act" (Sperling and Webber 2019, 254). For example, analyses reveal that the EU has securitized cyberterrorism, defining its role within a threat environment (Baker-Beall and Mott 2022). Similar trends have been observed in the securitization of migration (Fakhoury 2016) and organized crime, which has been framed as an increasingly dangerous enemy (Carrapico 2014). These examples, which extend beyond conventional security issues, demonstrate that the EU employs different strategies and is not hesitant to articulate threats and pursue extraordinary measures. Consequently, in the context of AI-related security, technocratization suggests that the EU seeks to address security, framed through risks, by utilizing measures traditionally reserved for non-security policies.

The second point of the importance of expertise and scientific knowledge manifests in the already noticeable extensive number of expert-driven formats. For example, the EC has established a platform called Knwoledge4Policy, which is supposed to "bridge the science-policy gap by bringing together for policy *from* scientists across Europe, *to* policymakers across Europe" (knowledge4policy.ec.europa.eu, n.d.). In 2023, the Directorate General for Research and Innovation released its report entitled "Futures of Science for Policy in Europe: Scenarios and Policy Implications", focused on "the dynamic interplay between science and policy" (Research-and-innovation.ec.europa.eu 2023).

In the field of security, at least several expert-based formats were initiated by the EC: the Group of Personalities on Security Research (2004), the Community for European Research and Innovation for Security (2014), and the Horizon 2020 Protection and Security Advisory Group (2014). These formats demonstrate that reliance on expertise, even in the field of security, is expected to yield uncontested, *rational* solutions. At the same time, this

tendency can also be understood as a response to uncertainty, where scientific knowledge and expert judgment are mobilized to provide guidance in a context where the implications of AI remain largely unclear.

The third characteristic suggests that AI-related security should be viewed as an issue that can be managed through dispersed responsibilities. This approach contrasts again with the strategy of securitization, which leads to political mobilization and immediacy. In this context, the distribution of responsibilities implies that the reassurance of uncertainty rests in the hands of experts. These experts are expected to make informed decisions and translate them into probabilistic assessments, a practice already employed across various policies and domains, such as climate, food, and product safety (Balzacq et al. 2010). This distribution, even if argued as a positive involvement of diverse representatives, emphasizes a critical shift in how security should be managed: not as a contentious issue of *high politics*, but rather as a routine process overseen by administrators and stakeholders.

Technocratization can be understood not only through its reliance on risk-based frameworks and the reproduction of existing policy measures, but also through its efforts to internationalize this mechanism. By framing the governance as technocratized rather than politically contested, the EU positions its security response as universally applicable. This approach reflects a strategic attempt to make the EU's proposed model acceptable and agreeable beyond its borders, thereby reducing friction in global negotiations. Unlike securitization, which thrives on political contestation, technocratization aims to avoid such tensions – especially internationally, where ideological divisions and state-centric rivalries could impede consensus. In this way, the technocratized governance programme serves not only to mitigate risks and guide AI developments in preferred directions, but also to export a model of AI-related security governance that bypasses confrontation in favour of rules-based cooperation.

Overall, technocratization refers to the process by which the EU's security response is developed. Unlike securitization, which involves implementing extraordinary measures in response to immediate threats, technocratization focuses on addressing risks through the governance programme. This approach suggests that security can be improved through policy measures, detached from politically expressed urgency.

This tendency aligns with current analyses of riskification, which indicate that the consideration of risk often leads to technocratic responses. These responses are based on the assumption that issues can be managed using existing risk frameworks and traditional risk instruments, such as risk-based

approaches or risk assessments (Barquet et al. 2024). Consequently, the EU's proposed measures for regulation, assessments, and various institutional formats to implement the policy and normative principles align with the overall process of riskification.

The role of technocratization demonstrates that riskification not only reveals how risks are constructed but also serves as a mechanism for responding to them. By perpetuating the idea of AI through its technical characteristics and asserting efforts to balance regulation with innovation, the EU reflects this approach to technology within its governance programme.

Conclusion

This chapter has shown that the logic of the governance programme is based on the assumption that it is possible to establish a routine process to target spaces, populations and activities considered most risky (Goede 2012). The proposed policy measures have appeared to meet typical elements of governance: principles of human-in-the-loop and human centrism come as expected norms and ideals that ground the proposed regulation; regulation is supposed to establish the process of a legally binding mechanism that frames decision making; and both the assessments and the institutional ecosystem it includes are about the implementation and rationalization of the remaining ambiguity of risk categories, normative principles and regulation. Compared to securitization, these measures refer to long-term risk mitigation and management, rather than the need for immediate extraordinary measures.

Being a response to risk and the conditions of possibility for harm, the policy measures are directly connected with agentic security and its core elements: maintaining human oversight through the human-in-the-loop principle; steering AI in preferred directions via human centrism; ensuring compliance with rights and democracy-oriented rules through regulation; and enabling enforcement and control through assessments and the broader institutional ecosystem. Therefore, in the context of riskification, the EU makes security as a routine procedural matter. This approach is facilitated by employing standard risk management tools, such as assessments, and following technocratic governance practices. This shift can be discussed for two key reasons: first, the uncertainty and unknowns associated with AI require expertise and clarification to guide decision-making; second, the EU's political motivations lead to a reluctance to securitize AI, as that might create divisions.

Compared to other actors, the EU's emphasis on human centrism, ethics and human oversight as key principles of AI governance aligns with similar proposals from others. However, as Section 4.3. has shown, these principles, while widely promoted, are open to interpretation: what is considered *ethical* by China, the USA or the CoE can lead to vastly different practices. While the EU follows global trends by employing a similar wording, it distinguishes itself by translating its principles into legally binding rules and ensuring their enforcement through discussed policy measures.

At the same time, the EU's governance programme could be seen not only as directed towards the mitigation of *internal* risk but also as an *external* dimension, because they involve and create obligations to others, especially businesses developing AI. While aiming to portray itself as distinct from others, the EU's governance programme relies on familiar measures and visions, such as regulation, that are expected to produce the same "Brussels effect" in the case of AI.

The concept of technocratization refers to a chosen mode of initiating and shaping the EU's governance programme, one that responds to AI-related risks by framing governance as a matter of technical management rather than political deliberation. It relies on standardized policy measures, the use of scientific knowledge to justify action, and the delegation of decisions to technocrats. Within this framework, policy measures are not presented as politically negotiated or ideologically contested, but as rational, objective and procedurally neutral responses to risks and conditions of possibility for harm. This reflects a broader tendency in EU policymaking to reproduce these measures, now extended to the emerging EU AI policy.

Overall, technocratization reveals the EU's ambition to craft a globally relevant, transferable governance model, which mirrors the technical definition of AI, relies on scientific knowledge to manage uncertainty, and embeds security in routine everyday practices. In doing so, it masks the political stakes involved in defining risks, deciding how they should be governed, and determining what counts as agentic security.

5.5. International engagement

This chapter examines how riskification intertwines with a reflection of the international landscape and the EU's participation there. As was mentioned before, Olaf Corry's proposed framework does not involve international engagement as a separate element of riskification. I argue that it requires attention and further develop the concept of riskification because the debates

and policy-making on AI are not only about internal policy-making, but also an internationally contested phenomenon discussed through global risks (as was already demonstrated in Section 4.3.). Additionally, the international engagement provides an important perspective on how the EU shapes the response by internationalizing risks and their governance, alongside efforts to reaffirm its subjectivity.

The EU's AI strategic discourse suggests that that the international dimension is an integral part of thinking about AI and formulating the emerging AI policy. For example, the Commission President Ursula von der Leyen claimed that "our AI Act will make a substantial contribution to the development of global rules and principles for human-centric AI", and that "we want Europe to be one of the leading AI continents. And global leadership is still up for grabs" (Ec.europa.eu 2025). Therefore, international engagement involves questions of how the EU defines the international AI landscape, what directions of international engagement are articulated, and what forms of self-positionality emerge. By proposing empirical evidence, this chapter is also a conceptual contribution to the process of riskification, because it demonstrates that different forms of the EU's international engagement are intertwined with articulating risks and the overall approach towards AI.

The following analysis demonstrates that the EU claims its ambitions to be an agenda setter in the international debate to influence AI standards. As the EP claims, "it is necessary to create a clear and fair international regime for assigning legal responsibility for adverse consequences produced by these advanced digital technologies" (Europarl.europa.eu 2021b). Section 5.5.1. argues that the EU highlights the importance of multilateralism and different forms of partnership to assert the desired influence. At the same time, partnership building becomes institutionalized, and reveals the EU's hierarchical position towards others. Section 5.5.2. suggests that the EU perceives the international landscape as competitive, challenging its own international role. However, despite presenting itself as different to major powers, the EU aims to participate in the competition, and continues the debates of how this policy serves the EU's needs for influence and reinstated international subjectivity.

Section 5.5.3. introduces the notion of a fortress to describe the EU's self-positionality in the emerging EU AI policy. This metaphor refers to a bounded space defined through established rules and obligations for others to enter it, reduced intersubjectivity from others, and expectations to persuade others to accept its approach, in order to create even more secure space. A fortress signals the EU's more assertive way of claiming subjectivity by taking a more

protectionist stance, where technology is not only strategically instrumentalized, as in the case of digital sovereignty, but is the Other that challenges the EU's self-protection. This conceptualization highlights how the EU internationalizes its AI-related risk considerations, embedding its responses within and in relation to the broader international landscape.

5.5.1. Leadership and multilateralism

The EU formulates ambitions to become an agenda setter for international AI governance and to achieve this goal through multilateral engagement with like-minded partners. Representative 2 from the EU Council claimed that "we want a global partnership on AI", while a representative from the EC stated that "we want to promote AI and our capabilities as Europe in AI". Similar positions are noticeable in the documents as well. The EC suggests that the EU "can have a leading role in developing international AI guidelines" and "contributing to relevant standardization activities" (Digital-strategy.ec.europa.eu 2019a). While the Council Presidency suggests that the EU is expected to participate "in the global debate on the use of AI with a view to shaping the international framework" (Consilium.europa.eu 2022).

Being an agenda setter interrelates with the previously discussed priorities of the EU in constructing a policy framework. For example, the Council Conclusions of 2022 and 2023, entitled "EU Digital Diplomacy", refer to the need to "actively promote universal human rights and fundamental freedoms, the rule of law and democratic principles in the digital space and advance a human-centric and human rights-based approach to digital technologies in relevant multilateral fora and other platforms" (Consilium.europa.eu 2023). The EC suggests that "Europe is well positioned to exercise global leadership in building alliances around shared values and promoting the ethical use of AI" (Eur-lex.europa.eu 2020d), and "the EU needs to shape international standards in line with values and interests" (Ec.europa.eu, n.d.-a). Even though these suggestions do not go into detail, the combination of universal human rights and a human-centric approach with global leadership, promotion and the shaping of international standards stress that the EU aims to export its approach towards AI as universally applicable and suitable to the international landscape.

Asking how that agenda is supposed to be set, and coalitions established, multilateralism is reflected as a key framework. Generally, multilateralism could be defined as a mode of cooperation and interaction which provides guiding principles on how policy can be constructed among different actors.

It also refers to the institutionalization of (security) communities by means of multilateral dialogue and community-building practices, on the basis of collective normative knowledge (Adler 1997). Then the proposed directions of the EU's international engagement suggest that the EU act as a multilateral actor searching for consensus and common interest (Christou and Simpson 2011).

In this case, the EC states that "the EU's cooperation with international bodies has also proved effective in identifying risks and malicious uses associated with AI" (Eur-lex.europa.eu 2021b), and that "the EU's approach will continue to be based on a proactive approach in various international bodies to build the strongest possible coalition of countries that share the desire for regulatory guardrails and democratic governance" (Eur-lex.europa.eu 2021b). The EP stresses that "there is a need for a consistent cooperation approach at an international level" (Europarl.europa.eu 2020a), "under the auspices of the United Nations" (Europarl.europa.eu 2017), and that the EU "is also working at the multilateral level, including in the Context of the Council of Europe, to develop common rules […] based on a high-level protection of fundamental and procedural rights" (Eur-lex.europa.eu 2020a). These expectations show that the EU advocates a rules-based and pro-liberal order, where concerns related to agentic security become an argument to build coalitions based on the EU's own preferences.

Another form of multilateralism comes with establishing partnerships and alliances with like-minded actors which are also expected to follow the EU's self-claimed lead. For example, Representative 2 from the EU Council claims that

"I think, we need to find like-minded and work with them. So that means US, Canada, Japan, Singapore, those kinds of countries that also want to have a similar perspective on how we can bring a new world I think, because and sort of with the human-centric approach as well."

Representative 2 from the EU Council suggested the same point, that:

"having in mind the economic weight that our like-minded have: the USA, Canada, Australia, New Zealand, of course, the EU. If we succeed to get India on board, there is no doubt that the rest of the world where the technologies are being developed will have to meet those requirements naturally. And at the same time, it gives us a stronger voice in international organizations that are setting the standards."

In addition, a representative from the EC argued that "the Australians and the Canadians and the South Koreans and the UK […] the Americans are pretty much on the same page […] we are also promoting this human-centric vision of technology on a global scale. So, it is not the point erecting a fortress Europe here."

These articulations are particularly rich: shaping the international framework, cooperation with international bodies, exercising global leadership and consistent cooperation, building alliances and the strongest possible coalition cooperation at an international level, or developing common rules, demonstrate that the EU does not explicitly put race or other more assertive ways of international engagement but remains within the framework of multilateral engagement. Although ambitions to lead are clearly declared, they are still put in the language of cooperation, alliances and coalition, an audience that also needs to accept the EU's proposed approach towards AI and AI governance.

The way the EU describes these partnerships and alliances (it wants to have a similar perspective, like-minded, on the same page) indicate that the EU mostly refers to the global West and pro-democratic countries, rather than truly aiming for a universal agreement. For example, a representative from the EC claimed that "we are not going to agree with China on how to do digital. It is a lost cause. But we discuss with all the others around us." Such a focus on those who *agree* suggests that despite declarations of cooperation, the EU's position towards others is hierarchical. It means that agenda-setting and partnering goals are possible only in these cases if others accept the EU's approach towards AI and follow the same priorities of rights, democratic freedoms and normative-technocratic AI governance.

In this context, the EU has already been developing and institutionalising various partnerships: the EU-USA and the EU-India trade and technology councils, and Japan-EU, Singapore-EU, Korea-EU and Canada-EU digital partnerships for "fostering a safe, secure digital space and to create a set of standards that can be used globally" (Digital-strategy.ec.europa.eu, n.d.-b). In 2021, the EU launched the Global Gateway, aimed at linking digital development investment in lower-income regions of Africa, Asia and the Pacific, and Latin America and the Caribbean. This move could also be interpreted as the EU's efforts to build multilateral reach-out and expectations of EU influence in different regions by developing global partnerships. These

formats relate to both ambitions of leadership but also pro-democratic alliances, expecting that they will be based on the EU's terms.

However, the outcomes of these partnerships, beyond the institutionalization reached, are still to be considered. For example, the results on AI from the Sixth Ministerial Meeting (April 2024) of the EU-US Trade and Technology Council suggest a "reaffirmed commitment to a risk-based approach to AI and support for safe and trustworthy AI technology […] announced a new Dialogue between the EU AI Office and the US AI Safety Institute" (Commission.europa.eu, n.d.-a). One could say that references to a "risk-based approach" and "trustworthiness" suggest that the EU's vocabulary is already embedded in the framework of this format. On the other hand, the results of this format, a generic declaration of reaffirmed commitment or a dialogue between AI offices, support the point that both the USA and the EU remain "increasingly anxious to make strides together" (Scott, Chatterjee and Volpicelli 2023).

Section 4.3. has demonstrated that most of the actors develop their own emerging AI policies, with the same ambition to be leaders in the field. Therefore, the EU is one among others sharing similar vocabularies, but acts following its own perspective. Then why is the international engagement, based on agenda setting and multilateral cooperation, so strongly articulated? One point could relate to agentic security itself. The EU aims to advocate for the protection of human agency, raising concerns related to conditions of possibility for harm. Then coalitions that agree on the same issues and the response can increase the chance that the same principles will be followed, not only by the EU, but also by other partners, and increase overall security.

On the other hand, such positionality can also be seen as a political reason, a reaction to criticism received for being a slow and reactive player, which manifests in declared ambitious efforts at global leadership (Krarup and Horst 2023). For example, Representative 1 from the HLEG mentioned that "if the global powers will follow the EU approach […] then I think Europe could see that they actually lead, you know, with this." Therefore, this analysis suggests that the EU's self-positioning as a leader is not limited to establishing and proposing pro-democratic AI governance, but also involves endeavours to receive acknowledgment from others. Again, multilateralism here works to the extent that others accept the EU's approach, rather than an interest or openness to inclusively integrate other proposals.

5.5.2. International competition

Unlike the EU's expectations of multilateralism, the international landscape is perceived as competitive and dominated by major powers. The EC suggests that "international competition is fiercer than ever" (Digital-strategy.ec.europa.eu 2018), while the EP refers to international competition as "strong" (Europarl.europa.eu 2021b). The phrases *fiercer than ever* and *strong* indicate that the EU understands the status quo as being different from its expectations of cooperation and multilateralism, where its own role is at stake. As the EP reiterates, "if the EU does not act swiftly and courageously, it will end up having to follow rules and standards set by others and risks damaging effects on political stability, social security, fundamental rights, individual liberties and economic competitiveness" (Europarl.europa.eu 2022a).

In this context, the EU's self-assessment also evolves. For example, the EP reports that "the EU currently does not meet any of the preconditions" and "has not yet met its aspirations" (Europarl.europa.eu 2022a). Additionally, the EP warns that the EU "has fallen behind in a new 'winner-takes-most' or 'superstar' economy". This situation poses a risk that "European values will be globally replaced" or that the EU might "end up becoming a digital colony of China, the USA and other states" (Europarl.europa.eu 2022a). A negative conclusion of the EU's status in the international competition was also drawn by Representative 1 from the HLEG, suggesting that "it is Europe that is lagging behind in terms of making an industrial revolution on growth from these technologies, we are still dependent on this massive use of tech developed in the USA and China."

Even compared to previous proposals for establishing like-minded partnerships with the USA to "jointly lead the coalition of techno-democracies challenging digital authoritarian norms and values" (Bradford 2023), both China and the USA are reflected as competitors, narrowing down the abstract notion of competition. A representative from the EC stated that "just imagine now, we did not become strong in AI. And we just rely on what the Americans and the Chinese are doing." Therefore, this dynamic is not the EU's preferred scenario, but considerations of how the international competition is evolving: if powers such as the USA and China refuse the European approach towards AI, ambitions for leadership and agenda-setting in international AI governance are lost.

The EU's response to the competition could be seen through two directions. One relates to reduced interdependence with powers.

Representative 1 from the EU Council claimed that "it is building, as we call it, a European fortress that we are independent of anybody anywhere in the world." Representative 1 from the HLEG suggested that "Europe has realized that they have a chance […] for leading in the digital decade, for digital sovereignty." In addition, a representative from the EUISS indicated that "sovereignty as a term has the beauty of stressing the fact that the EU needs to be in command of certain technologies, and have control over them, not necessarily own them, but at least control the way they are used."

References to independence, control and command of technologies reveal an additional layer, because they demonstrate that the EU is not only focused on principles and governance, but also on technology as a point for competition with other powerful players. In this context, agentic security is seemingly put away, because competition becomes another concern other than AI-related risks. Here, the stress is on less interdependence and detachment from others as a sign of subjectivity and also a way to enhance security. Therefore, digital sovereignty and inscribed characteristics underscore the expectations for the EU to establish a competitive stance, reiterating a point of control and its hierarchical views.

Another interrelated direction is focused on the EU's approach towards AI as a competitive advantage against other powers. For example, the EC suggests that "against a background of fierce global competition, a solid European approach is needed" (Eur-lex.europa.eu 2020d), and that "Europe's diversity will stimulate healthy competition, rather than the fragmentation of the AI community" (Eur-lex.europa.eu 2021b), while the EP argues for the need to "secure the EU's ethical principles in the global competition" (Europarl.europa.eu 2022a). In this case, as analysis has shown, the EU's approach focused on the protection of fundamental rights, and democracy remains the EU's trademark in the context of competition as well. This self-positionality reproduces the already-established role of being an advocate for a free, open, accessible and secure internet, and to promote its understanding of privacy and its application globally (Braun and Hummel 2024; Broeders, Cristiano and Kaminska 2023). In the case of AI, it turns to ethical and human-centric development and uses of technology, suggesting the importance of a normative stance in the context of international competition. As Regine Paul (2024) suggests, these aspirations for values and rights become an important part for the EU to set itself apart in a globally competitive space, something that the EU can claim as its trademark, if the competition with the USA and China creates a sense of *falling behind*.

In this context, where ongoing discussions seek to describe which orders, marked by power ambitions, competition or collaboration, are taking shape (Haddad, Vorlíček and Klimburg-Witjes 2024), the EU's international engagement reveals a tension. While advocating for its preferred liberal international order, as grounded on pro-democratic values and multilateralism, the EU simultaneously partakes in an evolving competition. These two directions suggest that interacting with and/or distancing itself from the international landscape is based not on the EU's differing ambitions but on how the EU sees itself within the context, as *lagging behind* or gaining a competitive advantage and considering different strategies: partnership building, or reducing interdependence.

The analysis has demonstrated that the main focus is on the EU's attempts to shape the international agenda, in the words of the HLEG, to "enable Europe to position itself as a global leader in cutting-edge AI" (Op.europa.eu 2019). This tendency indicates that the EU articulates a power position through influence and persuasion, even though it is justified for the importance of pro-democratic rules and principles.

5.5.3. The EU as a fortress

This chapter has shown that the EU positions itself as an agenda-setter, an advocate for multilateralism, and a participant in international competition. These tendencies indicate that the EU employs both proactive and reactive approaches in formulating its security response, including the internationalizaation of risks, their governance, and the reinforcement of its international subjectivity. Given this complexity, I suggest considering the EU's positionality through the concept of **a fortress**.

I propose the definition of a fortress as a bounded space, which is defined through a boundary between the EU's controlled, values-driven framework of AI governance, and a more unpredictable external landscape. This metaphor is important, because it demonstrates that the EU's self-positioning is entangled with agentic security as a response to perceived insecurity. It also reflects the discussions of the EU's more assertive position, suggesting that it includes ambitions, the remaining relevance of values and rights, and pressures from the competitive international AI landscape. Therefore, a fortress is also about the EU's more protectionist and competitive stance on the global stage. Its key characteristics emerge as follows:

- First, the space of the fortress establishes rules that apply equally to all actors, AI developers, users, and state or corporate leaders, whether they operate within the EU or seek to *enter* from outside. As the former Executive Vice-President of the EC Margrethe Vestager claimed, "AI uptake respects EU rules in Europe" (Ec.europa.eu 2024b). Therefore, a fortress is detached from the conventional understanding of territorial sovereignty, and refers to principles and rules which create obligations in accordance with the EU's approach towards AI.
- Second, the protection of the fortress entails reduced interdependence with powers. As the former Commissioner Thierry Breton claimed, "it is about our dependencies, preserving European interests, and avoiding technologies being used to destabilize our societies and democratic values" (Ec.europa.eu 2023a). Therefore, considering the conditions of possibility for harm, interdependence becomes a form of vulnerability, suggesting that the EU needs to detach itself from hostile actors/environments and undemocratic uses of AI that potentially undermine the main elements of agentic security (Flonk, Jachtenfuchs and Obendiek 2024).
- Third, the influence of a fortress lies in persuading others to adopt the EU's approach towards AI. This involves both the ambition to embed *Europeanness* within what is framed as universal standards and the rules-based order, in an otherwise unregulated and competitive international landscape. Therefore, when others accept the EU's proposed principles, they not only affirm the supposed universality, but also reinforce the EU's role and sense of security through broader alignment.

These characteristics also invite further reflection on how the EU's self-positionality as a fortress in the emerging AI policy resonates with broader discussions.

The first point of a fortress referring to establishing rules that create obligations both inside but also to those who want to enter the EU gains different meanings in different EU policies. In recent years, it has been associated with migration and restricted access for asylum seekers, considering that the EU's "walls are constantly raised higher" (Klemp 2024). A fortress, or fortification, has also been discussed in the context of the EU's industrial and trade policy embracing new instruments primed for intensified competition and accumulating the weaknesses of interdependence (Lavery 2024). This metaphor has been suggested by several interviewees, considering, in the words of a representative from the EC, that "the point is

not erecting barriers. That does not make sense because, again, our industries are global industries, and they export everywhere, we import, and so on and so forth."

As the analysis demonstrates, in the case of AI, the proposed dichotomy between protectionism and openness as a trading strategy does not fully grasp the complexity of the EU's self-positioning, which justifies and legitimizes expectations "to be a leading authority" (Baker-Beall and Mott 2022, 16). Therefore, created rules that apply to all actors are not only about fostering or limiting innovation and competitiveness, but also about security and a combined response to risks and conditions of possibility for harm.

The second characteristic focusing on reduced interdependence with powers signals multiplying connotations of otherness. The analysis has already demonstrated that the EU constructs security knowledge by framing AI as the Other. At the same time, articulations related to big tech, autocratic or other actors that potentially misuse or exploit AI against the referent objects suggest that they are seen as amplifying AI as the Other. Therefore, a fortress is not only a response to AI-related concerns, but also to unfavourable international competition and actors which challenge the EU through autocratic practices and strengthen the relevance of self-protection.

Following the point of intertwined relations between the Self and the Other (Campbell 2008), the EU's self-positionality reiterates the type of actor the EU aspires to be (pro-democratic, influential, competitive and safe) and through what it differentiates itself from (autocratic, unregulated, power-driven) (Baker-Beall 2014). This tendency can be seen both through more conceptual lenses of defining the EU's security knowledge, and in a more pragmatic way, as a political stance to promise the protection of the digital market and EU citizens' rights challenged by powerful actors.

The third characteristic reflects the EU's ambitions to persuade others to accept the EU's approach towards AI. The EU's international engagement has been more typically addressed by Europe as a power debate, an academic discussion on the EU's subjectivity and external participation, which proposes different definitions of power: civilian, market, normative and military.[13] Then

[13] Civilian power focuses on the necessity for cooperation with others for international goals and non-military means leaving military power as a residual instrument (Orbie 2006). However, civilian power Europe has been criticized because of a lack of a clear distinction when *civilian* ends that refuse of any military capacity does not mean only peaceful action (Stavridis 2001). As a response to civilian power Europe, Ian Manners introduced the concept of normative power Europe which is not about what the EU does or what it says, but what it is –

influencing through rules and principles or advocating for rights has been attached to Europe as a normative power. Even in the case of the emerging EU AI policy, normativity remains important, because of the nature of referent objects and policy measures, such as human centrism, human-in-the-loop and regulation, grounded in liberal values and norms.

However, I argue that the fortress extends this conversation and detaches from the power debate because it reproduces the same connotations or offers combinations of different definitions containing the same meanings. In the case of a *fortress*, normativity is far more strategic and security-focused than extending norms into the international system; or, as put by Ian Manners (2024), the leading name in developing the concept of Normative Power Europe, as empowering actions that reshape conceptions of what is normal, including openness to non-Western perspectives. Here, normativity is a way to withstand international competition, respond to practices amplifying AI otherness, and claim the EU's subjectivity through available measures in the face of the pressures of *lagging behind* and becoming *a digital colony.* In other words, the EU uses the persuasion of others to follow its approach towards AI as universally applicable, to increase its own security and address multiple concerns that are interrelated between developments and uses of AI and related international dynamics.

To summarize, the concept of the fortress represents the EU's efforts to enhance protection in response to perceived risks and conditions of possibility for harm. While the focus is on AI, this logic extends beyond just AI as the Other and includes the competitive international landscape and actors that amplify otherness. As a result, the fortress becomes a way for the EU to internationalize its security concerns and governance, aiming to present its approach to AI as universally relevant and to create a secure space that extends beyond its borders. This vision remains complicated by both conceptual and political factors regarding how the EU views AI and its position in the global context. These diverse aspects suggest a complex interplay of the EU's

transnational entity which aims to extend its norms into the international system expecting others to follow them and shape what is *normal* (Manners 2002). Some of the key characteristics of normative power Europe are presented as non-coercive, positive, and focused on commitment to international law, justice, social and political rights, and order in international relations (Scheipers and Sicurelli 2007; Merlingen 2007). In terms of market power Europe, it is also defined through the three key elements: market size, institutional features, and interest contestation. The power itself is revealed through externalization of these three elements associated with the EU's high levels of regulatory expertise, coherence, and sanctioning authority as a possible power exercise internationally (Damro 2012).

normative aspirations and strategic interests, alongside elements of protectionism and engagement.

Conclusion

This chapter has analysed international engagement as an element of riskification: how the EU engages and positions itself within the international AI landscape. This point has been added to the concept of riskification, as AI appears to be framed as an international and geopolitical matter, as well as reflecting the EU's tendency to internationalize risk as a way of solving its security issues and confirming subjectivity.

The analysis has revealed the EU's ambitions to position itself as an agenda-setter and a leader in global AI standards: through multilateral engagement with like-minded, institutionalized partners; through competition with other major powers; and through the pursuit of digital sovereignty as a response to that competition. Then self-positioning also becomes a complexity of different forms of engagement and involvement as a competitor, a leader, an alliance-builder, and/or an advocate of norm. These proposals also suggest that the EU remains committed to both referent objects. Indirect remarks about building partnerships and alliances with the global West and pro-democratic actors, while distancing itself from autocracies, demonstrate that the driving forces of international engagement are ideological, and therefore political. This claim once again questions the EU's endeavours to portray the emerging AI policy as based on science-based evidence and driven by *objective* metrics.

Coming back to the international AI landscape, claims of leadership are not unique to the EU. As was discussed in Section 4.3., all actors present themselves as leaders, particularly states, while organizations like the CoE and the UN position themselves more as honest brokers. This suggests that competition for leadership is largely state-driven, centred on power and dominance, whether it be the UK as a powerhouse, China's pursuit of global supremacy, or the USA striving to win the AI race. In this context, the EU appears to be involved in the same competitive dynamic. Even the concept of digital sovereignty mirrors China's centralized approach to sovereign power, suggesting that the striving for increased control over digitalization becomes a trend.

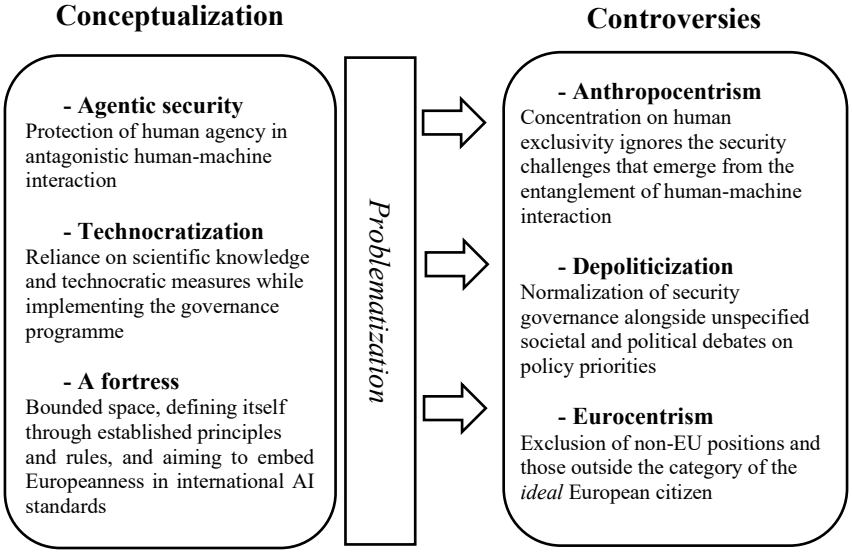
However, the EU's emphasis on multilateralism and partnerships indicates that its leadership is less about AI capability and more about setting the agenda and influencing AI standards. Considering the *international* as competitive and a power struggle, the EU grounds its vision on previously

raised points of human centrism and rules-based order that are expected to become widely applicable and leveraging the EU's role.

The notion of the fortress refers to a bounded space defined by established principles and rules that apply not only within the EU but also to those seeking entry. It entails reduced interdependence with others while simultaneously aiming to persuade the international landscape to adopt the EU's approach towards AI. This concept responds to identified risks and conditions of possibility for harm, in which international actors misusing or exploiting AI are seen as amplifying its otherness. By pursuing different strategies for protecting the Self, the EU seeks to create a secure space on its terms. Thus, the fortress is not only oriented inwards but also outwards, integrating an international dimension that both extends responses to risks and reinforces the EU's subjectivity.

6. DISCUSSION

The analysis of riskification has led to three conceptualizations: agentic security, technocratization, and a fortress. In this chapter, I problematize them, and suggest that these concepts inscribe three controversies as persistent yet contested viewpoints: 1) agentic security relies on anthropocentrism, which prioritizes humans while sidelining the security challenges evolving from that entanglement; 2) technocratization depoliticizes political contestation, and ignores alternative positions by formulating security as a broad agreement and everyday practices; 3) a fortress relies on Eurocentrism as the superior position, and promotes its approach towards AI as universal. Figure 2 summarizes the problematization of the EU's thinking and related controversies that are further elaborated.



Source: the author, based on the analysis

Figure 2. *Summary of the problematization of EU-driven concepts*

1. *Controversy of anthropocentrism.* Anthropocentrism has different interpretations. It could be defined through answers to the following questions: (1) what humans are; (2) what morality/ethics are; and (3) how we acquire knowledge about the world and ourselves (Droz 2022). The anthropocentric position relies on the belief that only human subjects can be

the subjects of security, while issues such as ecological crises, nuclear disaster or biochemical weaponry are supposed to *fit* within these existing ontological and ethical categories of human centrism (Mitchell 2014). Anthropocentrism has been extensively criticized for being "egoistic and obsessed only with the human" (Kopnina et al. 2018, 123), without acknowledging that non-human forms (such as other living species) require protection or face harm from human activities. This is why those concerned with climate change see anthropocentrism as linked to the denial of the negative influence of human activities, and the prioritization of human interests in sustainability discussions (Doudaki and Carpentier 2025).

The controversy of anthropocentrism in the case of AI goes back to the same question: what it means to be human. It also directly relates to agentic security emphasising the hierarchy of the superior human position over technology, which is expected to stay subordinated. As Vincent C. Müller (2025, 57) argues, "the standard criteria for me being the same person as that little boy in the photograph are my memory of being that boy, and the continuity of my body over time." Even pain is taken as a distinguishing element between the human and technology, "a subjective experience" that involves both body and mind not necessarily replicable to non-human (Sharkey 2024).

For the EU, anthropocentrism works as a political argument to maintain a power position, by arguing that AI has intrinsic capacities to overpower humans, and requires active intervention to avoid such a scenario. While anthropocentrism carries negative connotations in the context of climate change, signalling morally questionable human behaviour, in the case of AI, anthropocentrism is justified because the EU inscribes anthropocentric views in agentic security, and articulates humans as vulnerable, as *organisms* in need of protection from technology as the Other.

The anthropocentric position, based on the distinction between humans and technology, is problematic because it overlooks evolving security challenges and practices that come from intensifying human-machine interactions that extend human agency (Leese and Hoijtink 2019). For example, drone operation has proven to result in risks of PTSD, stigma and mental challenges, akin to those faced by ground-based soldiers (for example, Enemark 2019; Holz 2023; Chappelle et al. 2014). Reports of the Israeli army using AI targeting programs suggested that even though officials claimed to follow the human-in-the-loop principle, the human decision came after the AI-driven Lavender program made the decisions on targets, and served more "as a formal rubber stamp" (Serhan 2024). These cases suggest that the distinction

between decisions made by humans and/or machines is blurred, as both are involved in the process. Therefore, the challenge emerges not necessarily about technologies *taking* capacities from humans, as the EU suggests, but also technologies being influenced and influencing how people understand themselves and others (Nicholson and Reynolds 2020).

Even in less transformative ways, technology is seen as an extension that allows for the observation of different or unavailable areas for humans, by creating new forms of sensing and knowing (Rothe 2020). For instance, an analysis focusing on the Ukrainian President Volodymyr Zelensky's communication by selfie videos during Russia's full-scale invasion suggests that the smartphone's interaction with the human body enables new ways of seeing, representing and interpreting the world, as well as exerting agency by conditioning the actions of human actors (Markussen 2024). This interaction signifies the intersection of *culture* and *culture*, instead of anthropocentric *culture* and *nature*, where "human is integrated as an .exe file into technological ecology that is largely invisible, and which operates far beyond human capacities" (Schwarz 2021, 69).

However, the EU's proposed distinction between humans and AI eliminates the possibility of a voluntary human decision to delegate certain *properties* of agency to technology, expecting the results that outpace human capacity. This brings us to the question of control and human-machine interaction discussed in sections 5.3.3. and 5.3.4.: if humans delegate certain tasks or decision-making functions to technology in pursuit of particular (including political) objectives, can technology still be seen as merely subordinate to human, or does that delegation suggest a more complex dynamic of agency and control?

Edward A. Lee (2022, 7) claims that "we never had control and we cannot lose what we do not have." Of course, this opens another broad conversation, where the human ends and technology starts. Critical perspectives suggest that such an anthropocentric perspective is not viable, because control manifests at different levels, depending on individual situations when decisions are distributed and mediated through technological interfaces, nodes and various system components, which establish what is *appropriate* human control by certain actors (Bode 2023, 109). For example, the analysis of Russia's approach towards military AI applications demonstrated that the definition of human control is considered as a matter for a state and its own standards rather than a common agreement at the international level (Nadibaidze 2022).

Control involves not only technical means but also political ideas and interpretations, including the relations, rules, and principles that are

envisioned and displayed through the establishment of control boundaries. The EU's position that control is about subordination and intervention in the decision-making process defines its political stance that such a form of control is possible while positioning AI as the Other. The point by the President of the EP Roberta Metsola stressing the need for "constant, clear boundaries and limits to AI" (Liboreiro and Sanchez Alonso 2023) stresses that this distinction between the Self and the Other remains as defining human-machine interaction. Then the claims that humans have already been transformed into "a machine itself", redefining the human body and/or mind to meet the needs of AI (Wilcox 2015, 139), from the EU's perspective, implies a loss for control and a breakdown of the proposed hierarchy. This perspective woulve necessitate a complete reevaluation of the EU's stance on the matter.

Overall, anthropocentrism inscribed in agentic security comes as a form of resistance to human-machine dynamics, and narrows down the focus by aiming to reinstate hierarchy in a form of subordinated technology, as a passive object of *nature* and positioned as the Other. The question, then, is what effects the reliance on anthropocentric views will generate: whether the EU will manage to identify and address security issues stemming from dispersion, entanglement and the increasingly blurred distinction between humans and machines in their intensifying interactions, or whether it risks reinforcing rigid binaries that fail to account for the complexity of agency.

2. *Controversy of depoliticization*. Depoliticization emerges from the dominant position of technocrats and technocratized governance relying on scientific knowledge, alleged objectivity and technocratic neutrality (Maertens 2018). In this way, policy and related issues are portrayed not as a political debate, but as technocratic "self-evident" automatism, based on "neutral" calculations to increase legitimacy and the scope of action (Paul 2017b, 704). In the case of the emerging EU AI policy, the controversy comes not only in terms of a pretence to articulate a singular *truth* about both the issue at hand and the appropriate way to address it. Depoliticization through technocratization (see Section 5.4.4.) reflects the EU's established way of doing business: the reproduction of existing policy frameworks such as the risk-based approach, the inscription of normative principles in creating legally binding rules, the delegation of responsibilities, and policy implementation to the institutional ecosystem.

Depoliticization could be seen in two ways: 1) technocratic policy implementation while delegating responsibilities to external expertise; 2)

exclusion of the political process as diverse contributions from different actors, perspectives and arenas.

The first aspect raises questions about accountability. Even though the EU's proposed policy measures are defined as independent and expert-driven, the responsibility for protecting fundamental rights delegated to established institutional bodies also signals that the blame shifts from policy-makers to these technocratic formats which become responsible for risk assessments and decide on risk management (Macenaite 2017). Yet if responsibility and potential blame are absorbed by institutions, it becomes unclear how the public can scrutinize these processes, challenge the decisions made, or understand the basis on which such assessments are conducted.

For example, the AI Act delegates crucial functions of conformity assessment to external actors or tech industry experts: it imposes obligations on providers to provide "meaningful information about their systems and the conformity assessment carried out on those systems" (Eur-lex.europa.eu 2024). Then providers, conducting internal control to comply with high-risk requirements, will make the judgment on risk from an engineering and computer science perspective rather than expertise in fundamental rights, democracy and the rule of law (Smuha and Yeung 2024).

The same tendency to rely on the expertise of computer science while making decisions on risks to rights and democracy is fostered by EU institutions as well. For example, the EC's open call to set up "a scientific panel of independent experts to support the implementation and enforcement of the AI Act" by inviting applicants with a technical background and expertise in risk assessment methodologies, emergent systemic risks, computed measurements and thresholds, or AI technical risk mitigations and best practices (Digital-strategy.ec.europa.eu, n.d.-a). The expected composition and required expertise reinforce the observation that ethicists, human rights lawyers and other professionals capable of evaluating the broader implications for fundamental rights and democratic governance are largely absent.

The predominant emphasis on scientific knowledge and objectivity once again reflects the EU's strategy to legitimize its governance programme and its implementation through technocratic means. This approach raises further questions about the capacity to address increasingly complex challenges, especially given the continued reliance on expert knowledge, while the decision-making process itself "remains invisible to the public" (Harijanto 2025, 264).

The second aspect refers to depoliticization as the exclusion of diverse actors and their positions from discussions on decision making and governance. Here, it is defined as opening up for the political process to include a broader variety of actors, arenas, arguments and viewpoints that are debated and possibly fought over in public and show an interest towards a specific issue (Hegemann and Schneckener 2019). For example, the Proposal for the AI Act claims that it was built on an online public consultation which received contributions from businesses, individuals, academic institutions, public authorities and non-governmental organizations (NGO). Despite outlining the variety of stakeholders involved and their positions, the document states that "there were many comments underlining the importance of a technology-neutral and proportionate regulatory framework", "stakeholders mostly requested a narrow, clear and precise definition for AI", and "most of the respondents are explicitly in favour of the risk-based approach" (Eur-lex.europa.eu 2021c). These comments indicate that the EU's position and the policy framework closely intertwine with stakeholders' views, and inclusively represent their major expectations steering towards *neutral* and *precise* policy framing.

However, this supposedly broad agreement, as another point of legitimization, does not represent the noticeable contestation between human rights organizations, lobbies, industrial groups and associations publicising their positions and advocating for their prefered provisions in the AI Act. To be more specific, human rights organizations stressed different priorities pushing for stricter rules. For example, Algorithm Watch, an NGO advocating for the protection of human rights in using AI, required to include an obligation on deployers of high-risk systems to conduct a fundamental rights impact assessment before they deploy the system (Algorithmwatch.org 2023). The European Civic Forum, a pan-European network of nearly 100 associations and NGOs, argued that the AI Act "must prevent harm from AI used in migration and border control" (Goodwin 2022). European Digital Rights, a European network advocating for rights and freedoms online, focused on the need to address the structural, societal, political and economic impact of the use of AI (Edri.org 2021).

At the same time, associations representing businesses stressed the importance of innovation, asking for less regulatory intervention in the market. For example, Digital Europe, the trade association representing digitally transforming industries in Europe, claimed that "in order for Europe not to fall behind, we have to encourage innovation" (Digitaleurope.org 2023). The European Tech Alliance, representing tech companies, referred to the

need for support for "entrepreneurship and technological advancements in Europe, while fully respecting the EU values" (Eutechalliance.eu 2023). While the European AI Forum, an umbrella organization of national European AI associations, argued that the industry needs a flexible, feasible, understandable and future-oriented regulatory framework (Eaiforum.org 2023).

These examples are not exhaustive, and could be extended with even more diverse positions; nevertheless, they demonstrate clearly that the emerging EU AI policy was met with varied reactions, comments and pressures. The evident tension between human rights defenders and business representatives (stricter regulation or flexibility, protecting human rights or promoting innovation to remain competitive) underscore the sensitivity of the debates. However, the EU's introduction of stakeholder consultations, through phrases like "there is a general agreement", "most of the respondents" or "a widespread and common approach" (Eur-lex.europa.eu 2021c), reveals that the range of contributions and arguments are involved only to the extent that they support the EU's position and overall technocratic nature of the policy-making.

Politicization as an alternative to technocratization means that these positions are not necessarily *fitted* within the preferred policy framework, but resonate with broader societal groups and audiences that express their positions (Hegemann and Schneckener 2019). For the EU, the technocratic framing of AI gives an opportunity to provide solutions, while avoiding discussions about political, social and economic concerns (Ulnicane and Aden 2023). The question could be raised why depoliticization is considered controversial, especially when objectives such as balancing innovation and fundamental rights are explicitly articulated in the documents of the emerging AI policy; and also if technocratic governance is already *trademarked* as the EU's preferred mode of addressing complex issues, enabling the EU to claim results, such as global influence through GDPR and the role of a regulatory actor.

The illustrative example is Copernicus, the EU's Earth observation programme, which links technology with the EU's digital strategy. The visual products generated by this programme are presented as apolitical services for everyone's benefit, suggesting that introducing technologies as technical should eliminate the political dimension. However, such presentation does not demonstrate all the flows of money, subjective and/or algorithmic decisions, and acts of interpretation, scientific assumptions, legal norms or technical constraints behind an image or data layer which has become black-boxed (Rothe 2017).

In the case of AI, the EU similarly aims to mirror the *technical* definition of AI, and present security as an institutionalized and normalized everyday practice. However, this presentation reduces different positions and priorities, technological, economic and political pressures that arise internally and externally. Criticism from human rights organizations of the already-mentioned exemption allowing surveillance for law enforcement raises "genuine concerns that not all harmful uses of AI have been effectively guarded against" (Cabrera 2024), and suggests that the boundaries of intervention and the interpretation of risk remain a political discussion. Despite the illusion of depoliticized governance, a broad agreement and distribution of responsibilities and politics shapes the policy outcomes, as there is no actor or institution that could be defined as *pure* and *depoliticized*.

3. *Controversy of Eurocentrism.* Like anthropocentrism, Eurocentrism contains various meanings and interpretations. The term *Eurocentrism* itself comes as a criticism of Europe's hierarchical positionality towards others, and suggests that Europe defines itself as superior and distinctive, and focuses only on the European perspective. Within a Eurocentric world-view, Europe is visualized as being ahead, at the centre, and at the top all at once (Tolay 2021).

In the case of the emerging EU AI policy, the Commission President von der Leyen's references to "a distinct European brand of AI" and "our European way" (Ec.europa.eu 2025) signal that the focus is first and foremost on Europe and *European*, even though concerns of fundamental rights and democracy are presented as universal. This is the main issue of Eurocentrism: claiming universalism, while being predominantly focused on the *ideal* European citizenship. Such a position does not necessarily respond to other socially, culturally or historically different and vulnerable groups. For example, Didier Bigo (2014, 221) suggests that the EU's use of surveillance in the case of migration and border controls creates processes where some populations "end up being less human than others", where individual human beings crossing borders are less and less "real", and more and more just "numbers". Thus, despite the pursuit to protect humanness, Eurocentrism indicates that not all people are equally important and receive the EU's protection.

Eurocentrism also marks a lack of inclusivity of other perspectives, which limits the EU's ability to understand a broader picture of risks and challenges to rights and democracy that may emerge in diverse contexts. For example, Latin American scholarship suggests that AI use in governments helps automate state control and surveillance, defining vulnerable and marginalized

groups as dangerous, in contexts where democratic quality and institutional solidity remain fragile (Ricaurte, Gómez-Cruz and Siles 2024). Joana Varon and Paz Pena (2021) similarly claim that instead of seeking collective consent reinforcing multiplicity and plurality, powerful actors exploit AI for their interests, instead of protecting fundamental rights and accessibility.

In this context, the Eurocentric perspective overlooks the systemic inequalities that AI technologies can reinforce, particularly for those outside the dominant narrative: AI workers, resource supplies, lower-income communities and non-EU residents affected by AI-driven exploitation (Regilme 2025). Even though for the EU these issues could be seen as *external*, the example of Hungary's parliament banning Pride events and allowing the authorities to use facial recognition to identify attendees (Kassam 2025) suggests that concerns of technologies being used to limit rights are not related only to *the Rest*. Therefore, critically evaluating Eurocentrism in the emerging EU AI policy would mean imagining and articulating alternatives which are less anchored in race and power, but integrating already evolving challenges to which the EU is not necessarily immune.

Similarly to how anthropocentrism is justified by presenting humans as vulnerable, Eurocentrism is legitimized as a perspective to withstand global AI competition. Self-positioning of being "a role model of decision-making based on liberal values and free market principles" (Broeders et al. 2023, 1274) not only asserts its commitment to these principles, but uses them as an argument for global competition and a justification for Eurocentric views (Paul 2022a). Even the efforts to foster multilateralism and partnerships discussed in Section 5.5.1. do not necessarily align with the idea of multilateralism as a dialogue and community building among equals, but serve as a way to re-impose the EU's superiority. On top of this, Eurocentrism has a clear ideological line, which is again introduced as a support for the liberal international order, and a contrast with digital authoritarianism. This positionality is presented as being on the *right* side, self-evident, and beyond critique, while ignoring how this very order marginalizes voices that potentially challenge Western hegemony.

This EU stance underscores the fact that technology is not neutral. Rather, it is embedded in existing social and political frameworks, and often serves to reinforce longstanding power asymmetries, such as the EU's tendency to frame its approach towards AI as universally applicable, while paying limited attention to actors and perspectives outside the EU. In this way, the EU's efforts to inscribe its expectations for dominance reinforce the boundaries of who is included or excluded from shaping the global AI landscape; "what

counts as evidence, truth, and legitimacy, as well as determining who has the right to speak, act, and even be recognized as human" (Aradau, Hoijtink and Leese 2019, 199). In this context, the EU promotes its preferred *order*, rooted in modernist-liberal views that uphold existing hierarchies.

To conclude, the discussed controversies have highlighted different challenges, they have also revealed overlapping tendencies which give a clearer picture of the EU's views and expectations inscribed in the emerging EU AI policy. First, a strong objectifying tendency emerges, treating human hierarchy towards technology as fixed and the EU's position as equally beneficial to both humans and the international landscape. Second, liberal-modernist assumptions – such as a reliance on evidence-based policymaking, formal institutions, and the presumed universality of values – resurface through the forms of anthropocentrism, depoliticization, and Eurocentrism. Finally, these assumptions also reinforce boundaries and dichotomies: between humans and machines, inside and outside, democracy and autocracy, subject and object, values and interests, and ultimately between the EU and *the Rest*.

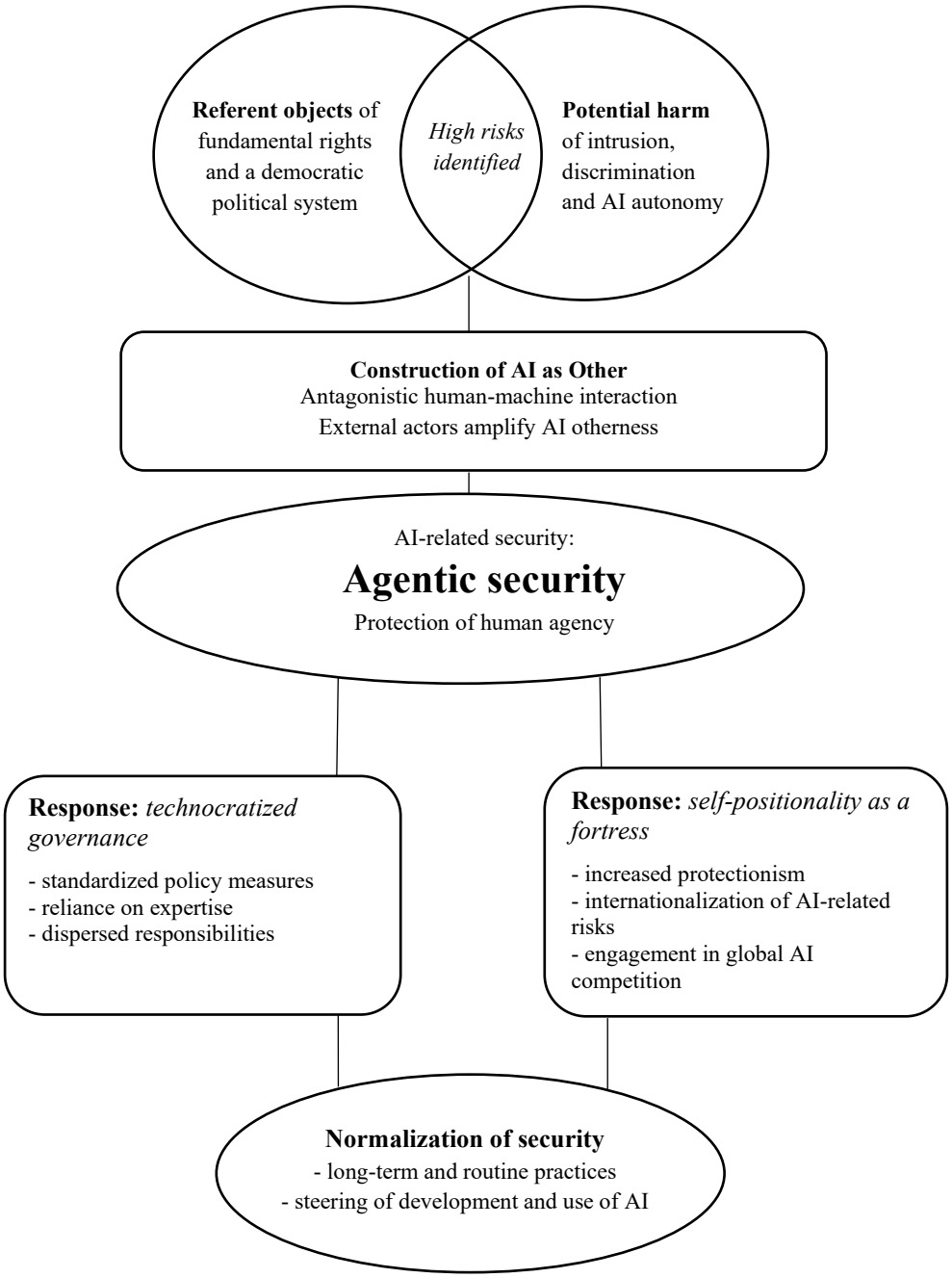
Although anthropocentrism, depoliticization, and Eurocentrism have faced long-standing criticism, the EU justifies their stance by arguing that they help reverse the logic of vulnerability: it is humans, rights, democracy and the EU itself, that must be protected and withstand the antagonism with technology. However, the presented examples suggest that the EU's self-proclaimed role does not always lead to inclusivity or openness toward those outside the EU or beyond the concept of the *ideal* European citizen. Thus, the ability to understand new issues arising from the increasing entanglement between humans and machines, to embrace diverse experiences, and to propose suitable responses becomes limited.

CONCLUSION

The thesis aimed to analyse the EU's construction of security knowledge in its emerging AI policy. This focus arose in response to existing analyses revealing inconsistencies in the EU's position on technology and security matters, the risk heuristic for governing various policies, and the EU's increasingly assertive international stance. While the policy and the EU's interest in AI are primarily viewed through the lens of the single market, the thesis argued that risk is a significant indicator in how the EU conceptualizes AI: as a long-term challenge to the conventional understanding of what it means to be human in the context of human-machine interaction.

By systematically exploring the process of riskification (chapters 5.1.-5.5.), the analysis was driven by the two guiding questions outlined in Section 2.3.: *what* priorities and vocabularies are established by articulating risks, and *how* should we address the identified concerns through the development of multilayered responses? The analysis showed that the EU's AI-related security is focused on agentic security, framing AI as Other and defining human-machine interaction as antagonistic. The proposed responses are based on the technocratized governance programme, privileging mitigation and regulation over emergency, and positioning itself as a fortress. The inclusion of international engagement as an analytical element of riskification highlights that the EU internationalizes risks and their governance within the contexts of security and its subjectivity. While considering the differences between securitization and riskification, empirical evidence revealed that AI-related security is not perceived as a need for extraordinary measures. Instead, security is framed as a domain to be addressed through normalized routine practices and policy measures.

These points are summarized in Figure 3, which shows how the key concepts are interconnected. Rather than being arranged in a linear or causal order, these concepts are mutually related and represent a co-constitutive process in the construction of security knowledge. This process involves defining security issues, identifying responsible actors, and legitimizing appropriate responses.



Source: the author, based on the analysis

Figure 3. *Elements of the process of riskification*

Building on the analysis, the thesis presents four key insights, summarized in Table 7.

Firstly, the analysis revealed that the EU actively employs risk and a risk-based approach as a form of knowledge production in defining its position on AI. The introduced categorization of risks – based on potential harm to fundamental rights and a democratic political system – an ordering mechanism: the more significant the potential harm, the higher the risk, and the greater the need for regulatory intervention. This shift from threats and immediacy towards risks and AI-related imaginaries reframes security not as a matter of *high politics* but as a process of anticipating and mitigating uncertainty. This process involves not only reactive responses but also expectations to guide AI development and applications in a desired direction.

In this context, the concept of risk emphasizes the political decisions that influence how security is defined and managed. The significance of a risk-based approach in the emerging EU AI policy illustrates that security encompasses not just conventional threats but also emerging and uncertain challenges. Consequently, analysing the process of riskification is crucial for understanding how the EU frames and addresses security issues related to AI, as well as the strategies employed to respond to perceived concerns.

Secondly, based on the EU's AI strategic discourse, the thesis proposed the concept of agentic security, which refers to the protection of human agency, understood through the capacity to sustain control and decision making power vis-à-vis current and future related concerns. While considering human-machine interaction as antagonistic and framing AI as Other, the EU emphasizes the need to maintain the superior human position over technology. Then even non-military uses of AI that potentially intrude, disrupt, or limit human exclusivity become concerns of agentic security.

This notion differs from existing concepts, such as national, human or societal security, which refer to relations between states, states and societal groups, and their vulnerabilities and identities. The focus on human agency widens the understanding of security by pointing to evolving considerations of human-machine interaction and what it means to be human. This point demonstrates that agency, which has been viewed as inherent to humans, is challenged. Therefore, agentic security emphasizes the need for clear boundaries between humans as the Self and AI as the Other, asserting that human agency must remain a prerogative, regardless of future AI developments and applications.

At the same time, this position underscores the controversy of anthropocentrism. By prioritizing human superiority, the anthropocentric perspective overlooks evolving security challenges and resists engaging with alternative forms of human-machine interaction that do not fit within established boundaries. In this context, the effort to maintain a strict hierarchy may limit the ability to fully comprehend and respond to the complex technological entanglements that increasingly characterize security today.

Thirdly, the EU's governance programme in its emerging AI policy is marked by technocratization. This concept suggests that the process of initiating and shaping governance is considered as a matter of technocratic management rather than political discussion. To shape this governance programme, the EU relies on standardized policy measures from other areas (such as climate), highlights the importance of expert authority, and implies that scientific knowledge will guide responses to uncertainty. Additionally, responsibilities for policy implementation are dispersed among administrators and stakeholders. These characteristics of technocratization indicate a shift in how AI-related security is addressed, moving from a focus on survival to a framework of technocratic governance, where security becomes normalized.

Technocratization is closely linked to the concept of riskification. When issues are framed in terms of risk, they tend to promote technocratic forms of governance. This tendency is particularly evident in the context of AI-related security, where the presentation of technology emphasizes its technical aspects while implying neutrality and objectivity. As a result, technocratization gains legitimacy not only through the perceived importance of risk but also through how technology is perceived and conceptualized. At the same time, technocratization demonstrates the controversy of depoliticization, which reduces the significance of differing viewpoints and proposals in decision-making, despite their presence. The preference for technocratized governance creates the illusion that political debate has been eliminated and that chosen security responses have not faced opposition.

Fourthly, in terms of self-positionality, the EU positions itself as a fortress, a bounded space which defines itself through established principles and rules, and aims to embed Europeanness in international AI standards. The fortress functions in several directions. It sets principles for all actors inside and those who want to *enter* the EU. While perceiving the international landscape as unfavourably competitive, the EU aims to reduce interdependence with powers, while increasing control of both technology and its governance. The EU expects to persuade others to accept its approach

towards AI, and in this way create a more secure space based on the EU's terms.

A fortress offers a different perspective on digital sovereignty: technology is not instrumentalized to enhance sovereignty, but appears as the *Other* that compels the EU to fortify itself. The competitive international landscape, including references to major powers, big tech and autocratic actors, amplifies the perception of AI otherness, making reduced interdependence and increased assertiveness shared, although differently motivated, elements of both digital sovereignty and the fortress. However, the EU's position highlights the controversy of Eurocentrism. It demonstrates the EU's inclination to portray its approach towards AI as universally applicable and beneficial for strengthening the liberal international order. These assertions of universalism fail to take into account the challenges, groups, and perspectives that lie outside the EU's predominant focus on the ideals of European citizens.

Table 7. *Summary of the key insights*

A referent object	• Fundamental rights and a democratic political system are those that face high risks and consistently remain referent objects • Safety focuses on products and services as insurance that AI does not undermine referent objects, involving more complex imaginaries of AI • The main elements represent the human dimension in human-machine interaction, where emerging concerns centre on who will have control in the future
Conditions of possibility for harm	• Potential harm signifies a spectrum of issues related to intrusion and discrimination which are caused either by AI or those misusing it • AI's autonomy is presented as the most radical concern suggesting how AI-related imaginaries foster an urgency to act • The EU's security understanding is summarised as agentic security focused on protecting human agency, defined through maintained human control and decision-making power in antagonistic human-machine interaction framed as Self-Other contestation
A governance programme	• A combination of normative and institutional measures which establish a multilayered and diverse governance framework • Policy measures of principles, regulation, assessments and the institutional ecosystem are not only to control but supposedly steer AI futures to ensure agentic security • The EU technocratizes its governance through standardized policy measures, reliance on expertise, and delegated responsibilities to instutional procedures where security becomes normalized routine practices
International engagement	• The EU aims to be an agenda-setter and advocate for the rules-based order through multilateral engagement and like-minded partnerships • The international AI-related landscape is considered as unfavourably competitive where the EU also participates to reinstate its role among powers • Self-positionality becomes like a fortress, a bounded space, which moves in directions of self-protection, reduced interdependence, and persuasion of its approach towards AI as ways to increase security and influence

Source: the author, based on empirical evidence

These insights prompt further reflection on the evolving relationship between technology, security and risk, and merit continued debate.

Firstly, AI is increasingly perceived as a source of power. As Section 4.3. suggested, various actors, particularly states, frame AI in terms of capability-building, where power is associated with winning the technological race and securing access to the most *advanced* forms of AI development. For the EU, the significance of AI as a source of power lies less in its technological capabilities and more in the ability to influence international AI standards. This influence is achieved when other entities recognize, adopt, and integrate the EU's approach towards AI through international agreements or other governance frameworks. This desired form of power raises further questions: to what extent is the case of AI different from the EU's already-established reputation of a regulatory power using the "Brussels effect"? In the context of international competition for AI as power, the EU appears concerned with diverging practices, geopolitical pressures and diverse actors involved in developing AI and amplifying AI otherness. Therefore, the analysis demonstrates that persuasion of other actors and influence of international AI standards are also about enhancing the EU's security and reaffirming its subjectivity.

Secondly, AI is about ideological tensions between democracies and autocracies, where the battle is about whose vision will shape global norms and practices. The integration of AI into the liberal international order reinforces the political embodiment where the EU, also referring to the like-minded global West, positions itself as countering digital authoritarianism and alternatives to digital democracy. However, uncritically accepting the liberal international order raises questions about the inclusion and exclusion of groups and perspectives that are not necessarily defined by the dichotomy of democracies and autocracies. The already-mentioned example of exemption for surveillance in law enforcement in the AI Act illustrates that even what is considered as a form of autocratic practices become justified in democracies as technical *solutions*.

In this context, the tension between digital democracy and digital authoritarianism is not merely a quest for ideological consistency. Rather, this tension represents an ongoing struggle to integrate AI within conflicting political, ethical, and governance frameworks that embody differing visions of power, control, and societal order. This situation sharply contrasts with the proposed technocratization, which depoliticizes AI-related security governance and removes political contestation from the process. Thus, the supposedly technical and technocratized approaches to AI and the ideological battles surrounding technology expose a significant inconsistency within the emerging EU AI policy that requires further examination.

Thirdly, AI operates as a political imaginary that, as the analysis has demonstrated, complicates conversations about capabilities (both real and imagined), technological limitations (whether AI is a functional product or a representation of non-human agentic capacities), and appropriate responses (ideological or technocratic). This imaginary is further shaped by the EU's conflation of different technological stages, *automation*, *autonomous* and *autonomy*, which co-exist within its strategic discourse on AI: from AI-driven biases to unspecified considerations of AI reaching autonomy. Therefore, despite grounding its definition in technical characteristics, the EU's approach towards AI is ultimately driven by selective imagination and argumentation. This strategy is designed to create a sense of urgency and justify the need for a response, which is reflected in the emerging AI policy. This complexity highlights how our understanding of AI is not derived solely from its material or technical features, but from the way it is imagined, framed and located within political agendas, even when those framings contain noticeable tensions and inconsistencies with practices.

Lastly, the concept of *emergence*, which refers to new or different developments from the status quo, complicates the intersection between policy and AI. The thesis discussed the EU AI policy as an *emerging* framework, being the first to regulate AI. While it is a new policy area, the analysis demonstrated that it is rooted in established areas such as the single market and the digital agenda, both of which are based on long-standing regulatory and institutional frameworks. Thus, the policy is not created in isolation but aligns with established patterns applicable to this case. The tension lies in the fact that AI is still an emerging technology, characterized by fluidity, unpredictability, and ongoing development.

These contrasting perspectives on *emergence* raise questions about how adaptable the established policy framework will be in keeping pace with technological change. As the EU constructs its security knowledge on foundational principles such as human agency, its capacity to address unforeseen AI-related challenges and adapt to a rapidly evolving landscape, still marked by uncertainty, remains a topic for further discussion. At the same time, the thesis shows that the meanings and interpretations of security and AI within the policy framework are tied to the specific period and political context being analysed. Although the key insights demonstrate how these concepts are shaped and function at this stage of emergence, these understandings may evolve as political, technological, and institutional contexts change.

BIBLIOGRAPHY

1. Adam, David. 2024. "Lethal AI Weapons Are Here: How Can We Control Them?" *Nature* 629 (8012): 521–23. https://doi.org/10.1038/d41586-024-01029-0.
2. Adams, John. 2005. "Big Ideas: Risk." *New Scientist*. https://www.newscientist.com/article/mg18725171-800-big-ideas-risk/.
3. Adler, Emanuel. 1997. "Imagined (Security) Communities: Cognitive Regions in International Relations." *Millennium: Journal of International Studies* 26 (2): 249–77. https://doi.org/10.1177/03058298970260021101.
4. Adler-Nissen, Rebecca, and Kristin Anabel Eggeling. 2024. "The Discursive Struggle for Digital Sovereignty: Security, Economy, Rights and the Cloud Project Gaia-X." *JCMS: Journal of Common Market Studies* 62 (4): 993–1011. https://doi.org/10.1111/jcms.13594.
5. Af Malmborg, Frans. 2023. "Narrative Dynamics in European Commission AI Policy – Sensemaking, Agency Construction, and Anchoring." *Review of Policy Research* 40 (5): 757–80. https://doi.org/10.1111/ropr.12529.
6. Agüera y Arcas, Blaise. 2023. "Fears about AI's Existential Risk Are Overdone, Says a Group of Experts." *The Economist*. https://www.economist.com/by-invitation/2023/07/21/fears-about-ais-existential-risk-are-overdone-says-a-group-of-experts.
7. Ahmed, Shazeda, Klaudia Jaźwińska, Archana Ahlawat, Amy Winecoff, and Mona Wang. 2023. "Building the Epistemic Community of AI Safety." *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4641526.
8. Aisi.gov.uk. n.d. "Academic Engagement." *AI Security Institute*. Accessed February 11, 2025. https://www.aisi.gov.uk/academic-engagement.
9. Algorithmwatch.org. 2023. "Civil Society Statement: We Call on Members of the EU Parliament to Ensure the AI Act Protects People and Our Rights." *AlgorithmWatch.* https://algorithmwatch.org/en/civil-society-statement-ai-act-protects-people-rights/.
10. Amicelle, Anthony, Claudia Aradau, and Julien Jeandesboz. 2015. "Questioning Security Devices: Performativity, Resistance, Politics." *Security Dialogue* 46 (4): 293–306. https://doi.org/10.1177/0967010615586964.
11. Amnesty.org. 2024. "EU: Artificial Intelligence Rulebook Fails to Stop Proliferation of Abusive Technologies." *Amnesty International*. https://www.amnesty.org/en/latest/news/2024/03/eu-artificial-

intelligence-rulebook-fails-to-stop-proliferation-of-abusive-technologies/.
12. Amoore, Louise. 2013. *The Politics of Possibility: Risk and Security Beyond Probability*. Duke University Press. https://doi.org/10.2307/j.ctv11sms8s.
13. Amoore, Louise. 2023. "Machine Learning Political Orders." *Review of International Studies* 49 (1): 20–36. https://doi.org/10.1017/S0260210522000031.
14. Amoore, Louise, and Marieke De Goede. 2005. "Governance, Risk and Dataveillance in the War on Terror." *Crime, Law and Social Change* 43 (2–3): 149–73. https://doi.org/10.1007/s10611-005-1717-8.
15. Amoore, Louise, and Marieke De Goede. 2012. "Introduction: Data and the War by Other Means." *Journal of Cultural Economy* 5 (1): 3–8. https://doi.org/10.1080/17530350.2012.640548.
16. Amoore, Louise, and Rita Raley. 2017. "Securing with Algorithms: Knowledge, Decision, Sovereignty." *Security Dialogue* 48 (1): 3–10. https://doi.org/10.1177/0967010616680753.
17. Amoroso, Daniele, and Guglielmo Tamburrini. 2020. "Autonomous Weapons Systems and Meaningful Human Control: Ethical and Legal Issues." *Current Robotics Reports* 1 (4): 187–94. https://doi.org/10.1007/s43154-020-00024-3.
18. Aradau, Claudia. 2010. "Security That Matters: Critical Infrastructure and Objects of Protection." *Security Dialogue* 41 (5): 491–514. https://doi.org/10.1177/0967010610382687.
19. Aradau, Claudia. 2014. "The Promise of Security: Resilience, Surprise and Epistemic Politics." *Resilience* 2 (2): 73–87. https://doi.org/10.1080/21693293.2014.914765.
20. Aradau, Claudia, and Tobias Blanke. 2015. "The (Big) Data-Security Assemblage: Knowledge and Critique." *Big Data & Society* 2 (2): 205395171560906. https://doi.org/10.1177/2053951715609066.
21. Aradau, Claudia, Marijn Hoijtink, and Matthias Leese. 2019. "Technology, Agency, Critique: An Interview with Claudia Aradau." In *Technology and Agency in International Relations*, edited by Marijn Hoijtink and Matthias Leese. Emerging Technologies, Ethics and International Affairs. Routledge.
22. Aradau, Claudia, and Jef Huysmans. 2019. "Assembling Credibility: Knowledge, Method and Critique in Times of 'Post-Truth.'" *Security Dialogue* 50 (1): 40–58. https://doi.org/10.1177/0967010618788996.
23. Aradau, Claudia, Luis Lobo-Guerrero, and Rens Van Munster. 2008. "Security, Technologies of Risk, and the Political: Guest Editors'

Introduction." *Security Dialogue* 39 (2–3): 147–54. https://doi.org/10.1177/0967010608089159.

24. Azungah, Theophilus. 2018. "Qualitative Research: Deductive and Inductive Approaches to Data Analysis." *Qualitative Research Journal* 18 (4): 383–400. https://doi.org/10.1108/QRJ-D-18-00035.
25. Babická, Karolina, and Cristina Giacomin. 2024. "Understanding the Scope of the Council of Europe Framework Convention on AI." Opinio Juris. https://opiniojuris.org/2024/11/05/understanding-the-scope-of-the-council-of-europe-framework-convention-on-ai/.
26. Bächle, Thomas Christian, and Jascha Bareis. 2022. "'Autonomous Weapons' as a Geopolitical Signifier in a National Power Play: Analysing AI Imaginaries in Chinese and US Military Policies." *European Journal of Futures Research* 10 (1): 20. https://doi.org/10.1186/s40309-022-00202-w.
27. Backman, Sarah. 2023. "Risk vs. Threat-Based Cybersecurity: The Case of the EU." *European Security* 32 (1): 85–103. https://doi.org/10.1080/09662839.2022.2069464.
28. Baker-Beall, Christopher. 2014. "The Evolution of the European Union's 'Fight against Terrorism' Discourse: Constructing the Terrorist 'Other.'" *Cooperation and Conflict* 49 (2): 212–38. https://doi.org/10.1177/0010836713483411.
29. Baker-Beall, Christopher, and Gareth Mott. 2022. "Understanding the European Union's Perception of the Threat of Cyberterrorism: A Discursive Analysis." *JCMS: Journal of Common Market Studies* 60 (4): 1086–105. https://doi.org/10.1111/jcms.13300.
30. Balzacq, Thierry, Tugba Basaran, Didier Bigo, Emmanuel-Pierre Guittet, and Christian Olsson. 2010. "Security Practices," edited by Thierry Balzacq, Tugba Basaran, Didier Bigo, Emmanuel-Pierre Guittet, and Christian Olsson. Oxford: Oxford University Press. https://doi.org/10.1093/acrefore/9780190846626.013.475.
31. Balzacq, Thierry, and Myriam Dunn Cavelty. 2016. "A Theory of Actor-Network for Cyber-Security." *European Journal of International Security* 1 (2): 176–98. https://doi.org/10.1017/eis.2016.8.
32. Barquet, Karina, Claudia Morsut, Mark Rhinard, et al. 2024. "Variations of Riskification: Climate Change Adaptation in Four European Cities." *Risk, Hazards & Crisis in Public Policy* 15 (4): 491–517. https://doi.org/10.1002/rhc3.12322.

33. Basham, Victoria M., Aaron Belkin, and Jess Gifkins. 2015. “What Is Critical Military Studies?” *Critical Military Studies* 1 (1): 1–2. https://doi.org/10.1080/23337486.2015.1006879.
34. Baur, Andreas. 2024. “European Dreams of the Cloud: Imagining Innovation and Political Control.” *Geopolitics* 29 (3): 796–820. https://doi.org/10.1080/14650045.2022.2151902.
35. Beck, Ulrich. 1992. *Risk Society: Towards a New Modernity*. Theory, Culture & Society. Sage Publications.
36. Beck, Ulrich. 2002. “The Terrorist Threat: World Risk Society Revisited.” *Theory, Culture & Society* 19 (4): 39–55. https://doi.org/10.1177/0263276402019004003.
37. Beck, Ulrich. 2006. “Living in the World Risk Society: A Hobhouse Memorial Public Lecture given on Wednesday 15 February 2006 at the London School of Economics.” *Economy and Society* 35 (3): 329–45. https://doi.org/10.1080/03085140600844902.
38. Bellanova, Rocco, Helena Carrapico, and Denis Duez. 2022. “Digital/Sovereignty and European Security Integration: An Introduction.” *European Security* 31 (3): 337–55. https://doi.org/10.1080/09662839.2022.2101887.
39. Bellanova, Rocco, and Marieke De Goede. 2022. “Co-Producing Security: Platform Content Moderation and European Security Integration.” *JCMS: Journal of Common Market Studies* 60 (5): 1316–34. https://doi.org/10.1111/jcms.13306.
40. Bellanova, Rocco, and Georgios Glouftsios. 2022. “Formatting European Security Integration through Database Interoperability.” *European Security* 31 (3): 454–74. https://doi.org/10.1080/09662839.2022.2101886.
41. Benner, Ann-Kathrin, and Delf Rothe. 2024. “World in the Making: On the Global Visual Politics of Climate Engineering.” *Review of International Studies* 50 (1): 79–106. https://doi.org/10.1017/S0260210523000025.
42. Berling, Trine Villumsen, Ulrik Pram Gad, Karen Lund Petersen, and Ole Wæver. 2022. *Translations of Security: A Framework for the Study of Unwanted Futures*. Routledge New Security Studies. Routledge.
43. Bhuta, Nehal, Susanne Beck, and Robin Geiß. 1920. “Present Futures: Concluding Reflections and Open Questions on Autonomous Weapons Systems.” In *Autonomous Weapons Systems*, edited by Nehal Bhuta, Susanne Beck, Robin Geiβ, Hin-Yan Liu, and Claus Kreβ. Cambridge University Press. https://doi.org/10.1017/CBO9781316597873.015.
44. Bigo, Didier. 2001. “Internal and External Security(Ies): The Möbius Ribbon.” In *Identities, Borders, Orders: Rethinking International*

Relations Theory, edited by Mathias Albert, David Jacobson, and Yosef Lapid. University of Minnesota Press.

45. Bigo, Didier. 2002. "Security and Immigration: Toward a Critique of the Governmentality of Unease." *Alternatives: Global, Local, Political*, no. 27.
46. Bigo, Didier. 2014. "The (in)Securitization Practices of the Three Universes of EU Border Control: Military/Navy – Border Guards/Police – Database Analysts." *Security Dialogue* 45 (3): 209–25. https://doi.org/10.1177/0967010614530459.
47. Bmi.bund.de. 2018. "Data Ethics Commission." *Federal Ministry of the Interior and Community*. https://www.bmi.bund.de/EN/topics/it-internet-policy/data-ethics-commission/data-ethics-commission.html?nn=11919984.
48. Boddington, Paula. 2023. *AI Ethics: A Textbook*. Artificial Intelligence: Foundations, Theory, and Algorithms. Springer Nature Singapore. https://doi.org/10.1007/978-981-19-9382-4.
49. Bode, Ingvild. 2023. "Practice-Based and Public-Deliberative Normativity: Retaining Human Control over the Use of Force." *European Journal of International Relations* 29 (4): 990–1016. https://doi.org/10.1177/13540661231163392.
50. Bode, Ingvild, and Hendrik Huelss. 2018. "Autonomous Weapons Systems and Changing Norms in International Relations." *Review of International Studies* 44 (3): 393–413. https://doi.org/10.1017/S0260210517000614.
51. Bode, Ingvild, and Hendrik Huelss. 2019. "Introduction to the Special Section: The Autonomisation of Weapons Systems: Challenges to International Relations." *Global Policy* 10 (3): 327–30. https://doi.org/10.1111/1758-5899.12704.
52. Bode, Ingvild, and Hendrik Huelss. 2023. "Constructing Expertise: The Front- and Back-Door Regulation of AI's Military Applications in the European Union." *Journal of European Public Policy* 30 (7): 1230–54. https://doi.org/10.1080/13501763.2023.2174169.
53. Bode, Ingvild, and Anna Nadibaidze. 2024. "Symposium on Military AI and the Law of Armed Conflict: Human-Machine Interaction in the Military Domain and the Responsible AI Framework." *Opinio Juris*. https://opiniojuris.org/2024/04/04/symposium-on-military-ai-and-the-law-of-armed-conflict-human-machine-interaction-in-the-military-domain-and-the-responsible-ai-framework/.

54. Borner, Peter. 2024. “EU AI Regulation - Innovation and Overregulation.” The Data Privacy Group. https://thedataprivacygroup.com/blog/eu-ai-regulation-a-balancing-act-between-innovation-and-overregulation/.
55. Bostrom, Nick. 2017. *Superintelligence: Paths, Dangers, Strategies*. Reprinted with corrections. Oxford: Oxford University Press.
56. Bostrom, Nick. 2020. “Ethical Issues in Advanced Artificial Intelligence.” In *Machine Ethics and Robot Ethics*, 1st ed., edited by Wendell Wallach, Peter Asaro, Wendell Wallach, and Peter Asaro. Routledge. https://doi.org/10.4324/9781003074991-7.
57. Bourne, Mike. 2012. “Guns Don’t Kill People, Cyborgs Do: A Latourian Provocation for Transformatory Arms Control and Disarmament.” *Global Change, Peace & Security* 24 (1): 141–63. https://doi.org/10.1080/14781158.2012.641279.
58. Boyd, Matt, and Nick Wilson. 2020. “Catastrophic Risk from Rapid Developments in Artificial Intelligence: What Is yet to Be Addressed and How Might New Zealand Policymakers Respond?” *Policy Quarterly* 16 (1). https://doi.org/10.26686/pq.v16i1.6355.
59. Bradford, Anu. 2020. *The Brussels Effect: How the European Union Rules the World*. 1st ed. New York: Oxford University Press. https://doi.org/10.1093/oso/9780190088583.001.0001.
60. Bradford, Anu. 2023. *Digital Empires: The Global Battle to Regulate Technology*. New York: Oxford University Press.
61. Brandão, Ana Paula, and Isabel Camisão. 2022. “Playing the Market Card: The Commission’s Strategy to Shape EU Cybersecurity Policy.” *JCMS: Journal of Common Market Studies* 60 (5): 1335–55. https://doi.org/10.1111/jcms.13158.
62. Brattberg, Erik, Raluca Csernatoni, and Vanesa Rugova. 2020. “Europe and AI: Leading, Lagging Behind, or Carving Its Own Way?” *Carnegie Endowment for International Peace*. https://carnegieendowment.org/research/2020/07/europe-and-ai-leading-lagging-behind-or-carving-its-own-way?lang=en.
63. Braun, Matthias, and Patrik Hummel. 2024. “Is Digital Sovereignty Normatively Desirable?” *Information, Communication & Society*, 1–14. https://doi.org/10.1080/1369118X.2024.2332624.
64. Broeders, Dennis. 2007. “The New Digital Borders of Europe: EU Databases and the Surveillance of Irregular Migrants.” *International Sociology* 22 (1): 71–92. https://doi.org/10.1177/0268580907070126.

65. Broeders, Dennis, Fabio Cristiano, and Monica Kaminska. 2023. “In Search of Digital Sovereignty and Strategic Autonomy: Normative Power Europe to the Test of Its Geopolitical Ambitions.” *JCMS: Journal of Common Market Studies* 61 (5): 1261–80. https://doi.org/10.1111/jcms.13462.
66. Browne, Ryan. 2023. “British Prime Minister Rishi Sunak Pitches UK as Home of A.I. Safety Regulation as London Bids to Be next Silicon Valley.” *CNBC*. https://www.cnbc.com/2023/06/12/pm-rishi-sunak-pitches-uk-as-geographical-home-of-ai-regulation.html.
67. Buzan, Barry. 1984. “Peace, Power, and Security: Contending Concepts in the Study of International Relations.” *Journal of Peace Research* 21 (2): 109–25. https://doi.org/10.1177/002234338402100203.
68. Buzan, Barry. 1991. “New Patterns of Global Security in the Twenty-First Century.” *International Affairs* 67 (3): 431–51. https://doi.org/10.2307/2621945.
69. Buzan, Barry, and Lene Hansen. 2018. “Defining–Redefining Security.” In *Oxford Research Encyclopedia of International Studies*, vol. 1. https://doi.org/10.1093/acrefore/9780190846626.013.382.
70. Buzan, Barry, and Ole Wæver. 2009. “Macrosecuritisation and Security Constellations: Reconsidering Scale in Securitisation Theory.” *Review of International Studies* 35 (2): 253–76. https://doi.org/10.1017/S0260210509008511.
71. Cabrera, Laura Lazaro. 2024. “EU AI Act Brief – Pt. 2, Privacy & Surveillance.” *Center for Democracy and Technology*. https://cdt.org/insights/eu-ai-act-brief-pt-2-privacy-surveillance/.
72. Calcara, Antonio, Raluca Csernatoni, and Chantal Lavallée. 2020. “Introduction. Emerging Security Technologies – an Uncharted Field for the EU.” In *Emerging Security Technologies and EU Governance: Actors, Practices and Processes*, 1st edition, edited by Antonio Calcara, Raluca Csernatoni, and Chantal Lavallée. New York: Routledge.
73. Calderaro, Andrea, and Stella Blumfelde. 2022. “Artificial Intelligence and EU Security: The False Promise of Digital Sovereignty.” *European Security* 31 (3): 415–34. https://doi.org/10.1080/09662839.2022.2101885.
74. Campbell, David. 2008. *Writing Security: United States Foreign Policy and the Politics of Identity*. Rev. ed., Minneapolis: University of Minnesota Press.
75. Cao.go.jp. 2022. “AI Strategy 2022.” *Cabinet Office*. https://www8.cao.go.jp/cstp/ai/aistratagy2022en.pdf.

76. Carmel, Emma, and Regine Paul. 2022. “Peace and Prosperity for the Digital Age? The Colonial Political Economy of European AI Governance.” *IEEE Technology and Society Magazine* 41 (2): 94–104. https://doi.org/10.1109/MTS.2022.3173340.
77. Carrapico, Helena. 2014. “Analysing the European Union’s Responses to Organized Crime through Different Securitization Lenses.” *European Security* 23 (4): 601–17. https://doi.org/10.1080/09662839.2014.949248.
78. Carrapico, Helena, and André Barrinha. 2017. “The EU as a Coherent (Cyber)Security Actor?” *JCMS: Journal of Common Market Studies* 55 (6): 1254–72. https://doi.org/10.1111/jcms.12575.
79. Cervi, Giulio Vittorio. 2022. “Why and How Does the EU Rule Global Digital Policy: An Empirical Analysis of EU Regulatory Influence in Data Protection Laws.” *Digital Society* 1 (2): 18. https://doi.org/10.1007/s44206-022-00005-3.
80. Chandler, David. 2012. “Resilience and Human Security: The Post-Interventionist Paradigm.” *Security Dialogue* 43 (3): 213–29. https://doi.org/10.1177/0967010612444151.
81. Chappelle, Wayne L., Kent D. McDonald, Lillian Prince, Tanya Goodman, Bobbie N. Ray-Sannerud, and William Thompson. 2014. “Symptoms of Psychological Distress and Post-Traumatic Stress Disorder in United States Air Force ‘Drone’ Operators.” *Military Medicine* 179 (8S): 63–70. https://doi.org/10.7205/MILMED-D-13-00501.
82. Christen, Markus, Thomas Burri, Serhiy Kandul, and Pascal Vörös. 2023. “Who Is Controlling Whom? Reframing ‘Meaningful Human Control’ of AI Systems in Security.” *Ethics and Information Technology* 25 (1): 10. https://doi.org/10.1007/s10676-023-09686-x.
83. Christie, Edward Hunter, Amy Ertan, Laurynas Adomaitis, and Matthias Klaus. 2024. “Regulating Lethal Autonomous Weapon Systems: Exploring the Challenges of Explainability and Traceability.” *AI and Ethics* 4 (2): 229–45. https://doi.org/10.1007/s43681-023-00261-0.
84. Christou, George, and Seamus Simpson. 2011. “The European Union, Multilateralism and the Global Governance of the Internet.” *Journal of European Public Policy* 18 (2): 241–57. https://doi.org/10.1080/13501763.2011.544505.
85. Clapton, William. 2014. *Risk and Hierarchy in International Society: Liberal Interventionism in the Post-Cold War Era*. Palgrave Studies in International Relations Series. Hampshire: Palgrave Macmillan.

86. Climate.ec.europa.eu. n.d. “European Climate Law.” *European Commission*. Accessed February 11, 2025. https://climate.ec.europa.eu/eu-action/european-climate-law_en.
87. Coe.int. n.d.-a. “Council of Europe and Artificial Intelligence - Artificial Intelligence.” *Council of Europe*. Accessed February 20, 2025. https://www.coe.int/en/web/artificial-intelligence.
88. Coe.int. n.d.-b. “The Framework Convention on Artificial Intelligence - Artificial Intelligence.” *Council of Europe*. Accessed February 20, 2025. https://www.coe.int/en/web/artificial-intelligence/the-framework-convention-on-artificial-intelligence.
89. Collective, C.A.S.E. 2006. “Critical Approaches to Security in Europe: A Networked Manifesto.” *Security Dialogue* 37 (4): 443–87. https://doi.org/10.1177/0967010606073085.
90. Commission.europa.eu. 2020. “A Europe Fit for the Digital Age.” *European Commission*. https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age_en.
91. Commission.europa.eu. 2024. “Artificial Intelligence in the European Commission (AI@EC) Communication.” *European Commission*. https://commission.europa.eu/publications/artificial-intelligence-european-commission-aiec-communication_en.
92. Commission.europa.eu. n.d.-a. “EU-US Trade and Technology Council.” *European Commission*. Accessed February 11, 2025. https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/stronger-europe-world/eu-us-trade-and-technology-council_en.
93. Commission.europa.eu. n.d.-b. “Role - European Commission.” *European Commission*. Accessed May 30, 2025. https://commission.europa.eu/about/role_en.
94. Commission.europa.eu. n.d.-c. “Shaping Europe’s Digital Future.” *European Commission*. Accessed February 6, 2025. https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/shaping-europes-digital-future_en.
95. Competition-cases.ec.europa.eu. n.d. “Digital Markets Act Latest Updates.” *European Commission*. Accessed February 6, 2025. https://competition-cases.ec.europa.eu/latest-updates/InstrumentDMA.
96. Consilium.europa.eu. 2022. “Artificial Intelligence: Presidency Issues Conclusions on Ensuring Respect for Fundamental Rights.” *Consilium*. https://www.consilium.europa.eu/en/press/press-releases/2020/10/21/artificial-intelligence-presidency-issues-conclusions-on-ensuring-respect-for-fundamental-rights/.

97. Consilium.europa.eu. 2023. “Digital Diplomacy: Council Sets out Priority Actions for Stronger EU Action in Global Digital Affairs.” *Consilium*. https://www.consilium.europa.eu/en/press/press-releases/2023/06/26/digital-diplomacy-council-sets-out-priority-actions-for-stronger-eu-action-in-global-digital-affairs/.
98. Corry, Olaf. 2012. “Securitisation and ‘Riskification’: Second-Order Security and the Politics of Climate Change.” *Millennium: Journal of International Studies* 40 (2): 235–58. https://doi.org/10.1177/0305829811419444.
99. Crawford, Kate. 2021. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven: Yale University Press.
100. Csernatoni, Raluca. 2018. “Constructing the EU’s High-Tech Borders: FRONTEX and Dual-Use Drones for Border Management.” *European Security* 27 (2): 175–200. https://doi.org/10.1080/09662839.2018.1481396.
101. Csernatoni, Raluca. 2019a. “Beyond the Hype: The EU and the AI Global ‘Arms Race.’” *Carnegie Endowment for International Peace*. https://carnegieendowment.org/posts/2019/08/beyond-the-hype-the-eu-and-the-ai-global-arms-race?lang=en.
102. Csernatoni, Raluca. 2019b. “The EU’s Technological Power: Harnessing Future and Emerging Technologies for European Security.” In *Peace, Security and Defence Cooperation in PostBrexit Europe. Risks and Opportunities*, edited by Cornelia-Adriana Baciu and John Doyle. Cham: Springer.
103. Csernatoni, Raluca. 2021a. “Between Rhetoric and Practice: Technological Efficiency and Defence Cooperation in the European Drone Sector.” *Critical Military Studies* 7 (2): 212–36. https://doi.org/10.1080/23337486.2019.1585652.
104. Csernatoni, Raluca. 2021b. “The Technology Challenge in the Transatlantic Relationship.” *European View* 20 (2): 157–65. https://doi.org/10.1177/17816858211059251.
105. Csernatoni, Raluca. 2022. “The EU’s Hegemonic Imaginaries: From European Strategic Autonomy in Defence to Technological Sovereignty.” *European Security* 31 (3): 395–414. https://doi.org/10.1080/09662839.2022.2103370.
106. Csernatoni, Raluca. 2024a. “Charting the Geopolitics and European Governance of Artificial Intelligence.” *Carnegie Endowment for International Peace*. https://carnegieendowment.org/research/2024/03/charting-the-geopolitics-and-european-governance-of-artificial-intelligence?lang=en.

107. Csernatoni, Raluca. 2024b. “How the EU Can Navigate the Geopolitics of AI.” Carnegie *Endowment for International Peace.* https://carnegieendowment.org/europe/strategic-europe/2024/01/how-the-eu-can-navigate-the-geopolitics-of-ai?lang=en.
108. Csernatoni, Raluca. 2025. “The EU’s AI Power Play: Between Deregulation and Innovation.” *Carnegie Endowment for International Peace.* https://carnegieendowment.org/research/2025/05/the-eus-ai-power-play-between-deregulation-and-innovation?lang=en.
109. Csernatoni, Raluca, and Chantal Lavallée. 2020. “Drones and Artificial Intelligence. The EU’s Smart Governance in Emerging Technologies.” In *Emerging Security Technologies and EU Governance. Actors, Practices and Processes*, edited by Antonio Calcara, Raluca Csernatoni, and Chantal Lavallée. New York: Routledge.
110. Csernatoni, Raluca, and Bruno Oliveira Martins. 2024. “Disruptive Technologies for Security and Defence: Temporality, Performativity and Imagination.” *Geopolitics* 29 (3): 849–72. https://doi.org/10.1080/14650045.2023.2224235.
111. Cset.georgetown.edu. 2021. “Ethical Norms for New Generation Artificial Intelligence Released.” *Center for Security and Emerging Technology*. https://cset.georgetown.edu/publication/ethical-norms-for-new-generation-artificial-intelligence-released/.
112. Cummings, M L, Heather M. Roff, Kenneth Cukier, Jacob Parakilas, and Hannah Bryce. 2018. “Artificial Intelligence and International Affairs.” *Chatham House*. https://www.chathamhouse.org/2018/06/artificial-intelligence-and-international-affairs.
113. Dafoe, Allan. 2018. “AI Governance: A Research Agenda.” *Centre for the Governance of AI Future of Humanity Institute University of Oxford*. https://www.fhi.ox.ac.uk/wp-content/uploads/GovAI-Agenda.pdf.
114. Daly, Killian. 2025. “The Geopolitics Of AI Regulation.” *The Yale Review of International Studies.* https://yris.yira.org/global-issue/the-geopolitics-of-ai-regulation/.
115. Damro, Chad. 2012. “Market Power Europe.” *Journal of European Public Policy* 19 (5): 682–99. https://doi.org/10.1080/13501763.2011.646779.
116. De Goede, Marieke. 2018. “The Chain of Security.” *Review of International Studies* 44 (1): 24–42. https://doi.org/10.1017/S0260210517000353.
117. De Goede, Marieke. 2020. “Engagement All the Way Down.” *Critical Studies on Security* 8 (2): 101–15. https://doi.org/10.1080/21624887.2020.1792158.

118. De Goede, Marieke, Stephanie Simon, and Marijn Hoijtink. 2014. "Performing Preemption." *Security Dialogue* 45 (5): 411–22. https://doi.org/10.1177/0967010614543585.
119. De Visser, Ewart J., Richard Pak, and Tyler H. Shaw. 2018. "From 'Automation' to 'Autonomy': The Importance of Trust Repair in Human–Machine Interaction." *Ergonomics* 61 (10): 1409–27. https://doi.org/10.1080/00140139.2018.1457725.
120. Dear, Peter, and Sheila Jasanoff. 2010. "Dismantling Boundaries in Science and Technology Studies." *Isis* 101 (4): 759–74. https://doi.org/10.1086/657475.
121. Defence-industry-space.ec.europa.eu. n.d. "Act in Support of Ammunition Production (ASAP)." *European Commission.* Accessed February 9, 2025. https://defence-industry-space.ec.europa.eu/eu-defence-industry/asap-boosting-defence-production_en.
122. Dekker, Brigitte, and Maaike Okano-Heijmans. 2020. "Front Matter. Europe's Digital Decade?" *Clingendael Institute*. http://www.jstor.org/stable/resrep26543.1.
123. Descombes, Vincent. 2010. *The Mind's Provisions: A Critique of Cognitivism*. Translated by Stephen Adam Schwartz. Princeton University Press.
124. Desmarais, Anna. 2025. "From Surveillance to Automation: How AI Tech Is Used at EU Borders." *Euronews*. https://www.euronews.com/next/2025/03/21/from-surveillance-to-automation-how-ai-tech-is-being-used-at-european-borders.
125. Digichina.stanford.edu. 2017. "China's 'New Generation Artificial Intelligence Development Plan.'" *University of Stanford*. https://digichina.stanford.edu/work/full-translation-chinas-new-generation-artificial-intelligence-development-plan-2017/.
126. Digichina.stanford.edu. 2021. "Guiding Opinions on Strengthening Overall Governance of Internet Information Service Algorithms." *University of Stanford*. https://digichina.stanford.edu/work/translation-guiding-opinions-on-strengthening-overall-governance-of-internet-information-service-algorithms/.
127. Digitaleurope.org. 2023. "Key Policies for Digitalization during the Spanish Presidency: The AI Act and the Data Act." *Digital Europe*. https://www.digitaleurope.org/resources/key-policies-for-digitalization-during-the-spanish-presidency-the-ai-act-and-the-data-act/.

128. Digital-strategy.ec.europa.eu. 2018. “Coordinated Plan on Artificial Intelligence.” *European Commission*. https://digital-strategy.ec.europa.eu/en/policies/plan-ai.
129. Digital-strategy.ec.europa.eu. 2019a. “Communication: Building Trust in Human Centric Artificial Intelligence.” *European Commission*. https://digital-strategy.ec.europa.eu/en/library/communication-building-trust-human-centric-artificial-intelligence.
130. Digital-strategy.ec.europa.eu. 2019b. “Policy and Investment Recommendations for Trustworthy Artificial Intelligence.” *European Commission*. https://digital-strategy.ec.europa.eu/en/library/policy-and-investment-recommendations-trustworthy-artificial-intelligence.
131. Digital-strategy.ec.europa.eu. 2020. “Assessment List for Trustworthy Artificial Intelligence (ALTAI) for Self-Assessment.” *European Commission.* https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment.
132. Digital-strategy.ec.europa.eu. 2024a. “First Meeting of the International Network of AI Safety Institutes.” *European Commission.* https://digital-strategy.ec.europa.eu/en/news/first-meeting-international-network-ai-safety-institutes.
133. Digital-strategy.ec.europa.eu. 2024b. “Second Draft of the General-Purpose AI Code of Practice Published, Written by Independent Experts.” *European Commission.* https://digital-strategy.ec.europa.eu/en/library/second-draft-general-purpose-ai-code-practice-published-written-independent-experts.
134. Digital-strategy.ec.europa.eu. 2025. “AI Act.” *European Commission*. https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai.
135. Digital-strategy.ec.europa.eu. n.d.-a. “Commission Seeks Experts for AI Scientific Panel.” *European Commission.* Accessed June 17, 2025. https://digital-strategy.ec.europa.eu/en/news/commission-seeks-experts-ai-scientific-panel.
136. Digital-strategy.ec.europa.eu. n.d.-b. “Digital Partnerships.” *European Commission.* Accessed February 11, 2025. https://digital-strategy.ec.europa.eu/en/policies/partnerships.
137. Digital-strategy.ec.europa.eu. n.d.-c. “European AI Office.” *European Commission*. Accessed February 11, 2025. https://digital-strategy.ec.europa.eu/en/policies/ai-office.
138. Digital-strategy.ec.europa.eu. n.d.-d. “European Approach to Artificial Intelligence.” *European Commission.* Accessed February 21, 2025.

https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence.
139. Digital-strategy.ec.europa.eu. n.d.-e. “European Digital Rights and Principles.” *European Commission*. Accessed May 30, 2025. https://digital-strategy.ec.europa.eu/en/policies/digital-principles.
140. Digital-strategy.ec.europa.eu. n.d.-f. “High-Level Expert Group on Artificial Intelligence.” *European Commission*. Accessed February 15, 2025. https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai.
141. Digital-strategy.ec.europa.eu. n.d.-g. “The Digital Services Act Package.” *European Commission*. Accessed February 15, 2025. https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package.
142. Digital-strategy.ec.europa.eu. n.d.-h. “The European AI Alliance.” *European Commission*. Accessed February 15, 2025. https://digital-strategy.ec.europa.eu/en/policies/european-ai-alliance.
143. Dignum, Virginia. 2019. *Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way*. Artificial Intelligence: Foundations, Theory, and Algorithms. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-30371-6.
144. Dillon, Michael. 2008. “Underwriting Security.” *Security Dialogue* 39 (2–3): 309–32. https://doi.org/10.1177/0967010608088780.
145. Doudaki, Vaia, and Nico Carpentier. 2025. “Behind the Narratives of Climate Change Denial and Rights of Nature: Sustainability and the Ideological Struggle between Anthropocentrism and Ecocentrism in Two Radical Facebook Groups in Sweden.” *Journal of Political Ideologies* 30 (1): 200–219. https://doi.org/10.1080/13569317.2023.2196506.
146. Dreyfus, Hubert Lederer. 1992. *What Computers Still Can’t Do: A Critique of Artificial Reason*. Cambridge: MIT Press.
147. Droz, L. 2022. “Anthropocentrism as the Scapegoat of the Environmental Crisis: A Review.” *Ethics in Science and Environmental Politics* 22 (May): 25–49. https://doi.org/10.3354/esep00200.
148. Dunn Cavelty, Myriam, Mareile Kaufmann, and Kristian Søby Kristensen. 2015. “Resilience and (in)Security: Practices, Subjects, Temporalities.” *Security Dialogue* 46 (1): 3–14. https://doi.org/10.1177/0967010614559637.
149. Eaiforum.org. 2023. “EAIF Support Letter to Hub France IA’s Position Paper on the AI Act.” *The European AI Forum*.

https://eaiforum.org/news-insights/eaif-support-letter-to-hub-france-ias-position-paper-on-the-ai-act.
150. Ebers, Martin. 2020. "Regulating AI and Robotics: Ethical and Legal Challenges." In *Algorithms and Law*, 1st ed., edited by Martin Ebers and Susana Navas. Cambridge University Press. https://doi.org/10.1017/9781108347846.003.
151. Ec.europa.eu. 2019. "Speech by President-elect von der Leyen in the EP." *European Commission*. https://ec.europa.eu/commission/presscorner/detail/es/speech_19_6408.
152. Ec.europa.eu. 2020. "President von der Leyen on 'Shaping Europe's Digital Future.'" *European Commission*. https://ec.europa.eu/commission/presscorner/detail/nl/speech_20_294.
153. Ec.europa.eu. 2023a. "Building European Resilience in the Digital Age." *European Commission*. https://ec.europa.eu/commission/presscorner/detail/en/speech_23_4346.
154. Ec.europa.eu. 2023b. "Statement by President von der Leyen on the AI Act." *European Commission*. https://ec.europa.eu/commission/presscorner/detail/en/statement_23_6474.
155. Ec.europa.eu. 2024a. "Competition in virtual worlds and generative AI." *European Commission*. https://ec.europa.eu/commission/presscorner/detail/en/ip_24_85.
156. Ec.europa.eu. 2024b. "European Artificial Intelligence Act comes into force." *European Commission*. https://ec.europa.eu/commission/presscorner/detail/en/ip_24_4123.
157. Ec.europa.eu. 2025. "Speech by the President: AI Action Summit." *European Commission*. https://ec.europa.eu/commission/presscorner/detail/pl/speech_25_471.
158. Ec.europa.eu. n.d.-a. "An EU approach to enhance economic security." *European Commission*. Accessed February 24, 2025. https://ec.europa.eu/commission/presscorner/detail/en/ip_23_3358.
159. Ec.europa.eu. n.d.-b. "Digital Economy and Society in the EU - What Is the Digital Single Market About?" *European Commission*. Accessed February 15, 2025. http://ec.europa.eu/eurostat/cache/infographs/ict/bloc-4.html.
160. Economist.com. 2024. "How Ukraine Is Using AI to Fight Russia." *The Economist*. https://www.economist.com/science-and-technology/2024/04/08/how-ukraine-is-using-ai-to-fight-russia.
161. Eda.europa.eu. 2021. "EDA Pursues Work on Artificial Intelligence in Defence." *European Defence Agency*. https://eda.europa.eu/news-and-

events/news/2021/06/29/eda-pursues-work-on-artificial-intelligence-in-defence.
162. Edri.org. 2021. “Civil Society Calls on the EU to Put Fundamental Rights First in the AI Act.” *European Digital Rights*. https://edri.org/our-work/civil-society-calls-on-the-eu-to-put-fundamental-rights-first-in-the-ai-act/.
163. Edri.org. 2023. “Regulating Big Tech.” *European Digital Rights*. https://edri.org/our-work/regulating-big-tech-in-europe-with-the-digital-services-act-digital-markets-act/.
164. Eea.europa.eu. 2024. “European Climate Risk Assessment.” *European Environment Agency*. https://www.eea.europa.eu/en/analysis/publications/european-climate-risk-assessment.
165. Ekberg, Merryn. 2007. “The Parameters of the Risk Society: A Review and Exploration.” *Current Sociology* 55 (3): 343–66. https://doi.org/10.1177/0011392107076080.
166. Elbe, Stefan. 2008. “Risking Lives: AIDS, Security and Three Concepts of Risk.” *Security Dialogue* 39 (2/3): 177–98. JSTOR.
167. Enemark, Christian. 2019. “Drones, Risk, and Moral Injury.” *Critical Military Studies* 5 (2): 150–67. https://doi.org/10.1080/23337486.2017.1384979.
168. Enisa.europa.eu. 2020. “Artificial Intelligence Cybersecurity Challenges.” *European Union Agency for Cybersecurity*. https://www.enisa.europa.eu/publications/artificial-intelligence-cybersecurity-challenges.
169. Estève, Adrien. 2021. “Preparing the French Military to a Warming World: Climatization through Riskification.” *International Politics* 58 (4): 600–618. https://doi.org/10.1057/s41311-020-00248-2.
170. Eur-lex.europa.eu. 2012a. “Charter of Fundamental Rights of the European Union 2012/C 326/02.” *Eur-Lex.* https://eur-lex.europa.eu/eli/treaty/char_2012/oj/eng.
171. Eur-lex.europa.eu. 2012b. “Consolidated Version of the Treaty on the Functioning of the European Union.” *Eur-Lex.* https://eur-lex.europa.eu/eli/treaty/tfeu_2012/oj/eng.
172. Eur-lex.europa.eu. 2015. “Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions A Digital Single Market Strategy for Europe.” *Eur-Lex.* https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52015DC0192.

173. Eur-lex.europa.eu. 2018. “Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions Artificial Intelligence for Europe.” *Eur-Lex*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52018DC0237.
174. Eur-lex.europa.eu. 2020a. “Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions A European Strategy for Data.” *Eur-Lex*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52020DC0066.
175. Eur-lex.europa.eu. 2020b. “European Parliament Resolution of 20 October 2020 with Recommendations to the Commission on a Civil Liability Regime for Artificial Intelligence (2020/2014(INL)).” *Eur-Lex*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=oj:JOC_2021_404_R_0006.
176. Eur-lex.europa.eu. 2020c. “Report from the Commission to the European Parliament, the Council and the European Economic and Social Committee Report on the Safety and Liability Implications of Artificial Intelligence, the Internet of Things and Robotics.” *Eur-Lex*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52020DC0064.
177. Eur-lex.europa.eu. 2020d. “White Paper.” *Eur-Lex*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=legissum:white_paper.
178. Eur-lex.europa.eu. 2021a. “Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions 2030 Digital Compass: The European Way for the Digital Decade.” *Eur-Lex*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52021DC0118.
179. Eur-lex.europa.eu. 2021b. “Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions Fostering a European Approach to Artificial Intelligence.” *Eur-Lex*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52021DC0205.
180. Eur-lex.europa.eu. 2021c. “Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts.” *Eur-Lex*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52021PC0206.
181. Eur-lex.europa.eu. 2024. “Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying down Harmonised Rules on Artificial Intelligence and Amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU)

2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)." *Eur-Lex.* http://data.europa.eu/eli/reg/2024/1689/oj/eng.

182. Europarl.europa.eu. 2017. "Civil Law Rules on Robotics." *European Parliament.* https://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.html.
183. Europarl.europa.eu. 2018a. "European Artificial Intelligence (AI) Leadership, the Path for an Integrated Vision." *European Parliament*. https://www.europarl.europa.eu/thinktank/en/document/IPOL_STU(2018)626074.
184. Europarl.europa.eu. 2018b. "Understanding Artificial Intelligence." *European Parliament.* https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2018)614654.
185. Europarl.europa.eu. 2019a. "The Juncker Commission's Ten Priorities: An End-of-Term Assessment." *European Parliament*. https://www.europarl.europa.eu/thinktank/en/document/EPRS_IDA(2019)637943.
186. Europarl.europa.eu. 2019b. "Understanding Algorithmic Decision-Making: Opportunities and Challenges." *European Parliament*. https://www.europarl.europa.eu/thinktank/en/document/EPRS_STU(2019)624261.
187. Europarl.europa.eu. 2020a. "Framework of Ethical Aspects of Artificial Intelligence, Robotics and Related Technologies." *European Parliament*. https://www.europarl.europa.eu/doceo/document/TA-9-2020-0275_EN.html.
188. Europarl.europa.eu. 2020b. "Report on Intellectual Property Rights for the Development of Artificial Intelligence Technologies A9-0176/2020." *European Parliament*. https://www.europarl.europa.eu/doceo/document/A-9-2020-0176_EN.html.
189. Europarl.europa.eu. 2020c. "Report with Recommendations to the Commission on a Civil Liability Regime for Artificial Intelligence A9-0178/2020." *European Parliament*. https://www.europarl.europa.eu/doceo/document/A-9-2020-0178_EN.html.
190. Europarl.europa.eu. 2020d. "Setting up a Special Committee on Artificial Intelligence in a Digital Age, and Defining Its Responsibilities, Numerical Strength and Term of Office." *European Parliament*. https://www.europarl.europa.eu/doceo/document/TA-9-2020-0162_EN.html.

191. Europarl.europa.eu. 2020e. "The Ethics of Artificial Intelligence: Issues and Initiatives." *European Parliament*. https://www.europarl.europa.eu/thinktank/en/document/EPRS_STU(2020)634452.
192. Europarl.europa.eu. 2021a. "Artificial Intelligence in Criminal Law and Its Use by the Police and Judicial Authorities in Criminal Matters." *European Parliament*. https://www.europarl.europa.eu/doceo/document/TA-9-2021-0405_EN.html.
193. Europarl.europa.eu. 2021b. "Report on Artificial Intelligence: Questions of Interpretation and Application of International Law in so Far as the EU Is Affected in the Areas of Civil and Military Uses and of State Authority Outside the Scope of Criminal Justice A9-0001/2021." *European Parliament*. https://www.europarl.europa.eu/doceo/document/A-9-2021-0001_EN.html.
194. Europarl.europa.eu. 2022a. "Report on Artificial Intelligence in a Digital Age A9-0088/2022." *European Parliament*. https://www.europarl.europa.eu/doceo/document/A-9-2022-0088_EN.html.
195. Europarl.europa.eu. 2022b. "The Future of AI: The Parliament's Roadmap for the EU." *European Parliament*. https://www.europarl.europa.eu/topics/en/article/20220422STO27705/the-future-of-ai-the-parliament-s-roadmap-for-the-eu.
196. Europarl.europa.eu. 2023a. "Artificial Intelligence, Democracy and Elections." *European Parliament*. https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2023)751478.
197. Europarl.europa.eu. 2023b. "General-Purpose Artificial Intelligence." *European Parliament*. https://www.europarl.europa.eu/thinktank/en/document/EPRS_ATA(2023)745708.
198. Europarl.europa.eu. 2023c. "Parliament's Negotiating Position on the Artificial Intelligence Act." *European Parliament*. https://www.europarl.europa.eu/thinktank/en/document/EPRS_ATA(2023)747926.
199. Europarl.europa.eu. 2023d. "What If Generative Artificial Intelligence Became Conscious?" *European Parliament*. https://www.europarl.europa.eu/thinktank/en/document/EPRS_ATA(2023)753162.
200. Europarl.europa.eu. 2023e. "MEPs Ready to Negotiate First-Ever Rules for Safe and Transparent AI." *European Parliament*. https://www.europarl.europa.eu/news/en/press-

room/20230609IPR96212/meps-ready-to-negotiate-first-ever-rules-for-safe-and-transparent-ai.
201. Europarl.europa.eu. 2024a. "Artificial Intelligence Act." *European Parliament.* https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2021)698792.
202. Europarl.europa.eu. 2024b. "Artificial Intelligence and Cybersecurity." *European Parliament*. https://www.europarl.europa.eu/thinktank/en/document/EPRS_ATA(2024)762292.
203. Europarl.europa.eu. 2024c. "The Global Reach of the EU's Approach to Digital Transformation." *European Parliament*. https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2024)757632.
204. Europarl.europa.eu. n.d. "Parliament's Powers." Parliament's Powers. *European Parliament.* Accessed May 30, 2025. https://www.europarl.europa.eu/about-parliament/en/parliaments-powers.
205. Eutechalliance.eu. 2023. "Position Paper on the Artificial Intelligence Act Ahead of Trilogue Negotiations - European Tech Alliance." *European Tech Alliance*. https://eutechalliance.eu/position-paper-on-the-artificial-intelligence-act-ahead-of-trilogue-negotiations/.
206. Evans, Sam Weiss, Matthias Leese, and Dagmar Rychnovská. 2021. "Science, Technology, Security: Towards Critical Collaboration." *Social Studies of Science* 51 (2): 189–213. https://doi.org/10.1177/0306312720953515.
207. Fahey, Elaine. 2014. "The EU's Cybercrime and Cyber-Security Rulemaking: Mapping the Internal and External Dimensions of EU Security." *European Journal of Risk Regulation* 5 (1): 46–60. https://doi.org/10.1017/S1867299X00002944.
208. Fakhoury, Tamirace. 2016. "Securitising Migration: The European Union in the Context of the Post-2011 Arab Upheavals." *The International Spectator* 51 (4): 67–79. https://doi.org/10.1080/03932729.2016.1245463.
209. Falkner, Gerda, Sebastian Heidebrecht, Anke Obendiek, and Timo Seidl. 2024. "Digital Sovereignty - Rhetoric and Reality." *Journal of European Public Policy* 31 (8): 2099–120. https://doi.org/10.1080/13501763.2024.2358984.
210. Farrand, Ben. 2020. "Managing Security Uncertainty with Emerging Technologies: The Example of the Governance of Neuroprosthetic

Research." In *Emerging Security Technologies and EU Governance: Actors, Practices and Processes*, edited by Antonio Calcara, Raluca Csernatoni, and Chantal Lavallée. New York: Routledge. https://www.routledge.com/Emerging-Security-Technologies-and-EU-Governance-Actors-Practices-and/Calcara-Csernatoni-Lavallee/p/book/9780367368814.

211. Federalregister.gov. 2023. "Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence." *Federal Register*. https://www.federalregister.gov/documents/2023/11/01/2023-24283/safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence.
212. Feenberg, Andrew. 2017. "Critical Theory of Technology and STS." *Thesis Eleven* 138 (1): 3–12. https://doi.org/10.1177/0725513616689388.
213. Ferl, Anna-Katharina. 2024. "Imagining Meaningful Human Control: Autonomous Weapons and the (De-) Legitimisation of Future Warfare." *Global Society* 38 (1): 139–55. https://doi.org/10.1080/13600826.2023.2233004.
214. Flonk, Daniëlle, Markus Jachtenfuchs, and Anke Obendiek. 2024. "Controlling Internet Content in the EU: Towards Digital Sovereignty." *Journal of European Public Policy* 31 (8): 2316–42. https://doi.org/10.1080/13501763.2024.2309179.
215. Florea Hudson, Natalie, Alex Kreidenweis, and Charli Carpenter. 2013. "Human Security." In *Critical Approaches to Security. An Introduction to Theories and Methods*, edited by Laura J. Shepherd. Abingdon: Routledge.
216. Floridi, Luciano. 2021. "The European Legislation on AI: A Brief Analysis of Its Philosophical Approach." *Philosophy & Technology* 34 (2): 215–22. https://doi.org/10.1007/s13347-021-00460-9.
217. Floridi, Luciano. 2023. "AI as Agency Without Intelligence: On ChatGPT, Large Language Models, and Other Generative Models." *Philosophy & Technology* 36 (1): 15, s13347-023-00621-y. https://doi.org/10.1007/s13347-023-00621-y.
218. Foucault, Michel. 2013. *Archaeology of Knowledge*. London: Routledge. https://doi.org/10.4324/9780203604168.
219. Fra.europa.eu. 2019. "Data Quality and Artificial Intelligence – Mitigating Bias and Error to Protect Fundamental Rights." *Agency for Fundamental Rights*. https://fra.europa.eu/en/publication/2019/data-quality-and-artificial-intelligence-mitigating-bias-and-error-protect.

220. Fra.europa.eu. 2022. “Bias in Algorithms - Artificial Intelligence and Discrimination.” *Agency for Fundamental Rights*. https://fra.europa.eu/en/publication/2022/bias-algorithm.
221. Frowd, Philippe M., Can E. Mutlu, and Mark B. Salter. 2023. “Discourse.” In *Research Methods in Critical Security Studies*, 2nd ed., by Mark B. Salter, Can E. Mutlu, and Philippe M. Frowd. London: Routledge. https://doi.org/10.4324/9781003108016-28.
222. Futureoflife.org. 2023. “Pause Giant AI Experiments: An Open Letter.” *Future of Life Institute*. https://futureoflife.org/open-letter/pause-giant-ai-experiments/.
223. Futurium.ec.europa.eu. 2019. “2nd European AI Alliance Assembly.” *European Commission*. https://futurium.ec.europa.eu/en/european-ai-alliance/document/2nd-european-ai-alliance-assembly-event-report.
224. Gilbert, Emily. 2019. “Military Geoeconomics: Money, Finance and War.” In *A Research Agenda for Military Geographies*, edited by Rachel Woodward. Cheltenham: Edward Elgar Publishing. https://doi.org/10.4337/9781786438874.00014.
225. Gkritsi, Eliza. 2024. “MEPs Probe Commission about AI Office Recruitment Strategy.” *Euractiv*. https://www.euractiv.com/section/digital/news/meps-probe-commission-about-ai-office-recruitment-strategy/.
226. Glasze, Georg, Amaël Cattaruzza, Frédérick Douzet, et al. 2023. “Contested Spatialities of Digital Sovereignty.” *Geopolitics* 28 (2): 919–58. https://doi.org/10.1080/14650045.2022.2050070.
227. Goede, Marieke de. 2012. *Speculative Security: The Politics of Pursuing Terrorist Monies*. Minneapolis: University of Minnesota Press.
228. Goodwin, Benjamin. 2022. *Joint Statement: The EU AI Act Must Protect People on the Move*. https://civic-forum.eu/statement/joint-statement-the-eu-ai-act-must-protect-people-on-the-move.
229. Gov.uk. 2022. “Defence Artificial Intelligence Strategy.” *Gov.uk.* https://www.gov.uk/government/publications/defence-artificial-intelligence-strategy.
230. Gov.uk. 2023a. “PM Urges Tech Leaders to Grasp Generational Opportunities and Challenges of AI.” *Gov.uk.* https://www.gov.uk/government/news/pm-urges-tech-leaders-to-grasp-generational-opportunities-and-challenges-of-ai.
231. Gov.uk. 2023b. “Prime Minister Launches New AI Safety Institute.” *Gov.uk.* https://www.gov.uk/government/news/prime-minister-launches-new-ai-safety-institute.

232. Gov.uk. 2025a. “International AI Safety Report 2025.” *Gov.uk.* https://www.gov.uk/government/publications/international-ai-safety-report-2025.
233. Gov.uk. 2025b. “The Bletchley Declaration by Countries Attending the AI Safety Summit, 1-2 November 2023.” *Gov.uk.* https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023.
234. Granados Hernandez, Marta. 2022. “Global Gateway and the EU’s Digital Ambitions.” *Center for Strategic and International Studies.* https://www.csis.org/blogs/development-dispatch/global-gateway-and-eus-digital-ambitions.
235. Gu, Hongfei. 2024. “Data, Big Tech, and the New Concept of Sovereignty.” *Journal of Chinese Political Science* 29 (4): 591–612. https://doi.org/10.1007/s11366-023-09855-1.
236. Güntay, Vahit. 2020. “Rethinking the Global Politics and Leadership: Ulrich Beck’s Risk Society Versus Postmodern Politics.” *Pamukkale University Journal of Social Sciences Institute*. https://doi.org/10.30794/pausbed.779156.
237. Haddad, Christian, Dagmar Vorlíček, and Nina Klimburg-Witjes. 2024. “The Security-Innovation Nexus in (Geo-)Political Imagination.” *Geopolitics* 29 (3): 741–64. https://doi.org/10.1080/14650045.2024.2329940.
238. Haner, Justin, and Denise Garcia. 2019. “The Artificial Intelligence Arms Race: Trends and World Leaders in Autonomous Weapons Development.” *Global Policy* 10 (3): 331–37. https://doi.org/10.1111/1758-5899.12713.
239. Hannah-Moffat, Kelly. 2019. “Algorithmic Risk Governance: Big Data Analytics, Race and Information Activism in Criminal Justice Debates.” *Theoretical Criminology* 23 (4): 453–70. https://doi.org/10.1177/1362480618763582.
240. Hansen, Lene. 2005. *Security as Practice: Discourse Analysis and the Bosnian War*. New York: Routledge.
241. Hansen, Lene, and Helen Nissenbaum. 2009. “Digital Disaster, Cyber Security, and the Copenhagen School.” *International Studies Quarterly* 53 (4): 1155–75. https://doi.org/10.1111/j.1468-2478.2009.00572.x.
242. Harijanto, Christian. 2025. “Risk, Resilience and Technocratic Exception: The Riskification of Energy Price Increases in Australia.” *Security Dialogue*. https://doi.org/10.1177/09670106251315662.

243. Health.ec.europa.eu. 2025. “Risk Assessment.” *European Commission.* https://health.ec.europa.eu/health-security-and-infectious-diseases/risk-assessment_en.
244. Héder, Mihály. 2020. “A Criticism of AI Ethics Guidelines.” *Információs Társadalom* 20 (4): 57. https://doi.org/10.22503/inftars.XX.2020.4.5.
245. Hegemann, Hendrik, and Ulrich Schneckener. 2019. “Politicising European Security: From Technocratic to Contentious Politics?” *European Security* 28 (2): 133–52. https://doi.org/10.1080/09662839.2019.1624533.
246. Heldt, Amélie P. 2022. “EU Digital Services Act: The White Hope of Intermediary Regulation.” In *Digital Platform Regulation*, edited by Terry Flew and Fiona R. Martin. Palgrave Global Media Policy and Business. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-95220-4_4.
247. Hockenhull, Michael, and Marisa Leavitt Cohn. 2021. “Hot Air and Corporate Sociotechnical Imaginaries: Performing and Translating Digital Futures in the Danish Tech Scene.” *New Media & Society* 23 (2): 302–21. https://doi.org/10.1177/1461444820929319.
248. Hoijtink, Marijn. 2014. “Capitalizing on Emergence: The ‘New’ Civil Security Market in Europe.” *Security Dialogue* 45 (5): 458–75. https://doi.org/10.1177/0967010614544312.
249. Hoijtink, Marijn, and Anneroos Planqué-van Hardeveld. 2022. “Machine Learning and the Platformization of the Military: A Study of Google’s Machine Learning Platform TensorFlow.” *International Political Sociology* 16 (2). https://doi.org/10.1093/ips/olab036.
250. Hoijtink, Marijn, and Jasper Van der Kist. 2025. “Platforms on the Frontline: The Rise of the Platform Model in Defense Tech.” *Opinio Juris*. https://opiniojuris.org/2025/02/11/platforms-on-the-frontline-the-rise-of-the-platform-model-in-defense-tech/.
251. Holmqvist, Caroline. 2013. “Undoing War: War Ontologies and the Materiality of Drone Warfare.” *Millennium: Journal of International Studies* 41 (3): 535–52. https://doi.org/10.1177/0305829813483350.
252. Holz, Jacob. 2023. “Victimhood and Trauma within Drone Warfare.” *Critical Military Studies* 9 (2): 175–90. https://doi.org/10.1080/23337486.2021.1953738.
253. Hunger, Francis. 2023. *Unhype Artificial ’Intelligence’! A Proposal to Replace the Deceiving Terminology of AI.* https://doi.org/10.5281/ZENODO.7524493.
254. Huysmans, Jef. 1998. “Security! What Do You Mean?: From Concept to Thick Signifier.” *European Journal of International Relations* 4 (2): 226–55. https://doi.org/10.1177/1354066198004002004.

255. Huysmans, Jef. 2008. *The Politics of Insecurity: Fear, Migration and Asylum in the EU*. The New International Relations Series. London: Routledge.
256. Jakniūnaitė, Dovilė, and Justinas Lingevičius. 2021. "Skaitmeninė Geopolitinė Konkurencija Dirbtinio Intelekto Amžiuje: Jungtinių Amerikos Valstijų, Kinijos Ir Europos Sąjungos Vizijos." *Deeds and Days* 76: 75–97. https://doi.org/10.7220/2335-8769.76.5.
257. Jasanoff, Sheila. 1996. "Is Science Socially Constructed - and Can It Still Inform Public Policy?" *Science and Engineering Ethics* 2 (3): 263–76. https://doi.org/10.1007/BF02583913.
258. Jasanoff, Sheila, and Sang-Hyun Kim. 2009. "Containing the Atom: Sociotechnical Imaginaries and Nuclear Power in the United States and South Korea." *Minerva* 47 (2): 119–46. https://doi.org/10.1007/s11024-009-9124-4.
259. Jasanoff, Sheila, and Sang-Hyun Kim, eds. 2015. *Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power*. Chicago: University of Chicago Press. https://doi.org/10.7208/9780226276663.
260. Johnson, James. 2019. "Artificial Intelligence & Future Warfare: Implications for International Security." *Defense & Security Analysis* 35 (2): 147–69. https://doi.org/10.1080/14751798.2019.1600800.
261. Johnson, James. 2020a. "Artificial Intelligence, Drone Swarming and Escalation Risks in Future Warfare." *The RUSI Journal* 165 (2): 26–36. https://doi.org/10.1080/03071847.2020.1752026.
262. Johnson, James. 2020b. "Artificial Intelligence in Nuclear Warfare: A Perfect Storm of Instability?" *The Washington Quarterly* 43 (2): 197–211. https://doi.org/10.1080/0163660X.2020.1770968.
263. Judge, Andrew, and Tomas Maltby. 2017. "European Energy Union? Caught between Securitisation and 'Riskification.'" *European Journal of International Security* 2 (2): 179–202. https://doi.org/10.1017/eis.2017.3.
264. Juncos, Ana E. 2017. "Resilience as the New EU Foreign Policy Paradigm: A Pragmatist Turn?" *European Security* 26 (1): 1–18. https://doi.org/10.1080/09662839.2016.1247809.
265. Justo-Hanani, Ronit. 2022. "The Politics of Artificial Intelligence Regulation and Governance Reform in the European Union." *Policy Sciences* 55 (1): 137–59. https://doi.org/10.1007/s11077-022-09452-8.

266. Kaber, David B. 2018. “A Conceptual Framework of Autonomous and Automated Agents.” *Theoretical Issues in Ergonomics Science* 19 (4): 406–30. https://doi.org/10.1080/1463922X.2017.1363314.
267. Kassam, Ashifa. 2025. “Hungary Bans Pride Events and Plans to Use Facial Recognition to Target Attenders.” *The Guardian*. https://www.theguardian.com/world/2025/mar/18/hungary-bans-pride-events-and-plans-to-use-facial-recognition-to-target-attenders.
268. Kessler, Oliver. 2007. “Risk, Power, and Authority: The Changing Politics of Global Finance.” *Review of International Political Economy* 14 (2): 357–70. https://doi.org/10.1080/09692290701203722.
269. Kessler, Oliver, and Wouter Werner. 2008. “Extrajudicial Killing as Risk Management.” *Security Dialogue* 39 (2/3): 289–308. https://doi.org/10.1177/0967010608088779.
270. Klemp, Pia. 2024. “Fortress Europe Keeps Cruelly Raising Its Walls against the Global South.” *The Guardian*. https://www.theguardian.com/commentisfree/article/2024/sep/11/europe-migration-asylum-seekers.
271. Klimburg-Witjes, Nina. 2024. “A Rocket to Protect? Sociotechnical Imaginaries of Strategic Autonomy in Controversies About the European Rocket Program.” *Geopolitics* 29 (3): 821–48. https://doi.org/10.1080/14650045.2023.2177157.
272. Knowledge4policy.ec.europa.eu. n.d. “Knowledge for Policy.” *European Commission.* Accessed February 11, 2025. https://knowledge4policy.ec.europa.eu/home_en.
273. Kopnina, Helen, Haydn Washington, Bron Taylor, and John J Piccolo. 2018. “Anthropocentrism: More than Just a Misunderstood Problem.” *Journal of Agricultural and Environmental Ethics* 31 (1): 109–27. https://doi.org/10.1007/s10806-018-9711-1.
274. Krahmann, Elke. 2003. “Conceptualizing Security Governance.” *Cooperation and Conflict* 38 (1): 5–26. https://doi.org/10.1177/0010836703038001001.
275. Krarup, Troels, and Maja Horst. 2023. “European Artificial Intelligence Policy as Digital Single Market Making.” *Big Data & Society* 10 (1): 205395172311538. https://doi.org/10.1177/20539517231153811.
276. Kusche, Isabel. 2024. “Possible Harms of Artificial Intelligence and the EU AI Act: Fundamental Rights and Risk.” *Journal of Risk Research*, May 11, 1–14. https://doi.org/10.1080/13669877.2024.2350720.
277. Kuus, Merje. 2014. *Geopolitics and Expertise: Knowledge and Authority in European Diplomacy*. Chichester: John Wiley & Sons, Inc.

278. Latour, Bruno, and Steve Woolgar. 1986. *Laboratory Life: The Construction of Scientific Facts*. Princeton: Princeton University Press.
279. Lavallée, Chantal, and Bruno Oliveira Martins. 2023. "Reframing Civil–Military Relations in the EU: Insights From the Drone Strategy 2.0." *JCMS: Journal of Common Market Studies*. https://doi.org/10.1111/jcms.13546.
280. Lavery, Scott. 2024. "Rebuilding the Fortress? Europe in a Changing World Economy." *Review of International Political Economy* 31 (1): 330–53. https://doi.org/10.1080/09692290.2023.2211281.
281. Lee, Edward A. 2022. "Are We Losing Control?" In *Perspectives on Digital Humanism*, edited by Hannes Werthner, Erich Prem, Edward A. Lee, and Carlo Ghezzi. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-86144-5_1.
282. Leese, Matthias. 2014. "The New Profiling: Algorithms, Black Boxes, and the Failure of Anti-Discriminatory Safeguards in the European Union." *Security Dialogue* 45 (5): 494–511. https://doi.org/10.1177/0967010614544204.
283. Leese, Matthias. 2019. "Configuring Warfare: Automation, Control, Agency." In *Technology and Agency in International Relations*, edited by Marijn Hoijtink and Matthias Leese. Abingdon: Routledge.
284. Leese, Matthias, and Marijn Hoijtink. 2019. "How (Not) to Talk about Technology: International Relations and the Question of Agency." In *Technology and Agency in International Relations*, edited by Marijn Hoijtink and Matthias Leese. Abingdon: Routledge.
285. Liboreiro, Jorge, and Aida Sanchez Alonso. 2023. "MEPs Endorse Blanket Ban on Live Facial Recognition in Public Spaces." Euronews. https://www.euronews.com/my-europe/2023/06/14/meps-endorse-blanket-ban-on-facial-recognition-in-public-spaces-rejecting-targeted-exempti.
286. Lidskog, Rolf, and Göran Sundqvist. 2015. "When Does Science Matter? International Relations Meets Science and Technology Studies." *Global Environmental Politics* 15 (1): 1–20. https://doi.org/10.1162/GLEP_a_00269.
287. Liebetrau, Tobias. 2024. "Problematising EU Cybersecurity: Exploring How the Single Market Functions as a Security Practice." *JCMS: Journal of Common Market Studies* 62 (3): 705–24. https://doi.org/10.1111/jcms.13523.
288. Liebetrau, Tobias, and Kristoffer Kjærgaard Christensen. 2021. "The Ontological Politics of Cyber Security: Emerging Agencies, Actors,

Sites, and Spaces." *European Journal of International Security* 6 (1): 25–43. https://doi.org/10.1017/eis.2020.10.
289. Lilkov, Dimitar. 2021. "Regulating Artificial Intelligence in the EU: A Risky Game." *European View* 20 (2): 166–74. https://doi.org/10.1177/17816858211059248.
290. Lindekilde, Lasse. 2014. "Discourse and Frame Analysis." In *Methodological Practices in Social Movement Research*, edited by Donatella Della Porta. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198719571.003.0009.
291. Lingevicius, Justinas. 2023. "Military Artificial Intelligence as Power: Consideration for European Union Actorness." *Ethics and Information Technology* 25 (1): 19. https://doi.org/10.1007/s10676-023-09684-z.
292. Lingevicius, Justinas. 2024. "Transformation, Insecurity, and Uncontrolled Automation: Frames of Military AI in the EU AI Strategic Discourse." *Critical Military Studies*: 1–22. https://doi.org/10.1080/23337486.2024.2387890.
293. Liu, Hin-Yan. 2019. "From the Autonomy Framework towards Networks and Systems Approaches for 'Autonomous' Weapons Systems." *Journal of International Humanitarian Legal Studies* 10 (1): 89–110. https://doi.org/10.1163/18781527-01001010.
294. Liu, Xu. 2018. "Interviewing Elites: Methodological Issues Confronting a Novice." *International Journal of Qualitative Methods* 17 (1): 1609406918770323. https://doi.org/10.1177/1609406918770323.
295. MacDonald, Norine, and George Howell. 2019. "Killing Me Softly: Competition in Artificial Intelligence and Unmanned Aerial Vehicles." *PRISM* 8 (3): 102–27. https://www.jstor.org/stable/26864279.
296. Macenaite, Milda. 2017. "The 'Riskification' of European Data Protection Law through a Two-Fold Shift." *European Journal of Risk Regulation* 8 (3): 506–40. https://doi.org/10.1017/err.2017.40.
297. Maertens, Lucile. 2018. "Depoliticisation as a Securitising Move: The Case of the United Nations Environment Programme." *European Journal of International Security* 3 (03): 344–63. https://doi.org/10.1017/eis.2018.5.
298. Magnusson, Eva, and Jeanne Marecek. 2015. *Doing Interview-Based Qualitative Research: A Learner's Guide*. 1st ed. Cambridge University Press. https://doi.org/10.1017/CBO9781107449893.
299. Manners, Ian. 2002. "Normative Power Europe: A Contradiction in Terms?" *JCMS: Journal of Common Market Studies* 40 (2): 235–58. https://doi.org/10.1111/1468-5965.00353.

300. Manners, Ian. 2024. “Arrival of Normative Power in Planetary Politics.” *JCMS: Journal of Common Market Studies* 62 (3): 825–44. https://doi.org/10.1111/jcms.13505.
301. Mantellassi, Federico. 2023. “Digital Authoritarianism: How Digital Technologies Can Empower Authoritarianism and Weaken Democracy.” *Geneva Centre for Security Policy.* https://www.gcsp.ch/publications/digital-authoritarianism-how-digital-technologies-can-empower-authoritarianism-and.
302. Markiewicz, Tadek. 2024. “The Vulnerability of Securitisation: The Missing Link of Critical Security Studies.” *Contemporary Politics* 30 (2): 199–220. https://doi.org/10.1080/13569775.2023.2267371.
303. Markussen, Håvard Rustad. 2024. “Inscribing Security: The Case of Zelensky’s Selfies.” *Review of International Studies* 50 (6): 1004–22. https://doi.org/10.1017/S0260210523000359.
304. Martins, Bruno Oliveira. 2023. “Security Knowledges: Circulation, Control, and Responsible Research and Innovation in EU Border Management.” *Science as Culture* 32 (3): 435–59. https://doi.org/10.1080/09505431.2023.2222739.
305. Martins, Bruno Oliveira, and Neven Ahmad. 2020. “The Security Politics of Innovation: Dual-Use Technology in the EU’s Security Research Programme.” In *Emerging Security Technologies and EU Governance. Actors, Practices and Processes*, edited by Antonio Calcara, Raluca Csernatoni, and Chantal Lavallée. New York: Routledge.
306. Martins, Bruno Oliveira, and Maria Gabrielsen Jumbert. 2022. “EU Border Technologies and the Co-Production of Security ‘Problems’ and ‘Solutions.’” *Journal of Ethnic and Migration Studies* 48 (6): 1430–47. https://doi.org/10.1080/1369183X.2020.1851470.
307. Martins, Bruno Oliveira, and Jocelyn Mawdsley. 2021. “Sociotechnical Imaginaries of EU Defence: The Past and the Future in the European Defence Fund.” *JCMS: Journal of Common Market Studies* 59 (6): 1458–74. https://doi.org/10.1111/jcms.13197.
308. Matejova, Miriam, and Anastasia Shesterinina. 2023a. “Conclusion.” In *Uncertainty in Global Politics*, 1st ed., by Anastasia Shesterinina and Miriam Matejova. Abingdon: Routledge. https://doi.org/10.4324/9781003426080-20.
309. Matejova, Miriam, and Anastasia Shesterinina. 2023b. “Introduction.” In *Uncertainty in Global Politics*, 1st ed., edited by Anastasia

Shesterinina and Miriam Matejova. Abingdon: Routledge. https://doi.org/10.4324/9781003426080-1.
310. Matzner, Tobias. 2019. "The Human Is Dead – Long Live the Algorithm! Human-Algorithmic Ensembles and Liberal Subjectivity." *Theory, Culture & Society* 36 (2): 123–44. https://doi.org/10.1177/0263276418818877.
311. McCarthy, Daniel R., ed. 2018. *Technology and World Politics: An Introduction*. Abingdon: Routledge.
312. McNamara, Kathleen R. 2024. "Transforming Europe? The EU's Industrial Policy and Geopolitical Turn." *Journal of European Public Policy* 31 (9): 2371–96. https://doi.org/10.1080/13501763.2023.2230247.
313. Merlingen, Michael. 2007. "Everything Is Dangerous: A Critique of `Normative Power Europe'." *Security Dialogue* 38 (4): 435–53. https://doi.org/10.1177/0967010607084995.
314. Methmann, Chris, and Delf Rothe. 2012. "Politics for the Day after Tomorrow: The Logic of Apocalypse in Global Climate Politics." *Security Dialogue* 43 (4): 323–44. https://doi.org/10.1177/0967010612450746.
315. Mishra, Vibhu. 2024. "General Assembly Adopts Landmark Resolution on Artificial Intelligence." https://news.un.org/en/story/2024/03/1147831.
316. Mitchell, Audra. 2014. "Only Human? A Worldly Approach to Security." *Security Dialogue* 45 (1): 5–21. https://doi.org/10.1177/0967010613515015.
317. Monsees, Linda, and Daniel Lambach. 2022. "Digital Sovereignty, Geopolitical Imaginaries, and the Reproduction of European Identity." *European Security* 31 (3): 377–94. https://doi.org/10.1080/09662839.2022.2101883.
318. Morsut, Claudia, and Ole Andreas Engen. 2022. "Climate Risk Discourses and Risk Governance in Norway." *Book of Extended Abstracts for the 32nd European Safety and Reliability Conference*, 2929–36. https://doi.org/10.3850/978-981-18-5183-4_S25-03-218-cd.
319. Morsut, Claudia, and Ole Andreas Engen. 2023. "Interfacing Risk Logic, Riskification, and Risk Governance: Some Research Implications." *Proceeding of the 33rd European Safety and Reliability Conference*, 1462–69. https://doi.org/10.3850/978-981-18-8071-1_P524-cd.
320. Mügge, Daniel. 2023. "The Securitization of the EU's Digital Tech Regulation." *Journal of European Public Policy* 30 (7): 1431–46. https://doi.org/10.1080/13501763.2023.2171090.

321. Mügge, Daniel. 2024. “EU AI Sovereignty: For Whom, to What End, and to Whose Benefit?” *Journal of European Public Policy* 31 (8): 2200–2225. https://doi.org/10.1080/13501763.2024.2318475.
322. Mügge, Daniel. 2025. “Losing the AI Race May Be a Blessing.” Linkedin, January 23. https://www.linkedin.com/pulse/losing-ai-race-may-blessing-daniel-m%25C3%25BCgge-o63hc/?trackingId=ItQeqaG9JDFhYdvJyOlI%2Fw%3D%3D.
323. Müller, Vincent C. 2025. “Philosophy of AI: A Structured Overview.” In *Cambridge Handbook on the Law, Ethics and Policy of Artificial Intelligence*, edited by Nathalie A. Smuha. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781009367783.
324. Nadibaidze, Anna. 2022. “Great Power Identity in Russia’s Position on Autonomous Weapons Systems.” *Contemporary Security Policy* 43 (3): 407–35. https://doi.org/10.1080/13523260.2022.2075665.
325. Natale, Simone, and Andrea Ballatore. 2020. “Imagining the Thinking Machine: Technological Myths and the Rise of Artificial Intelligence.” *Convergence: The International Journal of Research into New Media Technologies* 26 (1): 3–18. https://doi.org/10.1177/1354856517715164.
326. Neal, Andrew W. 2009. “Securitization and Risk at the EU Border: The Origins of FRONTEX.” *JCMS: Journal of Common Market Studies* 47 (2): 333–56. https://doi.org/10.1111/j.1468-5965.2009.00807.x.
327. Neff, Gina, and Peter Nagy. 2018. “Agency in the Digital Age: Using Symbiotic Agency to Explain Human–Technology Interaction.” In *A Networked Self and Human Augmentics, Artificial Intelligence, Sentience*, 1st ed., edited by Zizi Papacharissi. New York: Routledge. https://doi.org/10.4324/9781315202082-8.
328. Neumann, Iver B. 2008. “Discourse Analysis.” In *Qualitative Methods in International Relations*, edited by Audie Klotz and Deepa Prakash. London: Palgrave Macmillan UK. https://doi.org/10.1057/9780230584129_5.
329. Nicholson, Simon, and Jesse L. Reynolds. 2020. “Taking Technology Seriously: Introduction to the Special Issue on New Technologies and Global Environmental Politics.” *Global Environmental Politics* 20 (3): 1–8. https://doi.org/10.1162/glep_e_00576.
330. Niklas, Jędrzej, and Lina Dencik. 2021. “What Rights Matter? Examining the Place of Social Rights in the EU’s Artificial Intelligence Policy Debate.” *Internet Policy Review* 10 (3). https://doi.org/10.14763/2021.3.1579.
331. Niklas, Jędrzej, and Lina Dencik. 2024. “Data Justice in the ‘Twin Objective’ of Market and Risk: How Discrimination Is Formulated in

EU's AI Policy." *Policy & Internet* 16 (3): 509–22. https://doi.org/10.1002/poi3.392.
332. Nist.gov. 2023. "U.S. Artificial Intelligence Safety Institute." *National Institute of Standards and Technology*. https://www.nist.gov/aisi.
333. Nitzberg, Mark, and John Zysman. 2022. "Algorithms, Data, and Platforms: The Diverse Challenges of Governing AI." *Journal of European Public Policy* 29 (11): 1753–78. https://doi.org/10.1080/13501763.2022.2096668.
334. Nowell, Lorelli S., Jill M. Norris, Deborah E. White, and Nancy J. Moules. 2017. "Thematic Analysis: Striving to Meet the Trustworthiness Criteria." *International Journal of Qualitative Methods* 16 (1): 1609406917733847. https://doi.org/10.1177/1609406917733847.
335. NSTC. 2024. "Critical and Emerging Technologies List Update." *National Science and Technology Council*. https://www.govinfo.gov/content/pkg/CMR-PREX23-00185928/pdf/CMR-PREX23-00185928.pdf.
336. O'Grady, Nathaniel. 2021. "Automating Security Infrastructures: Practices, Imaginaries, Politics." *Security Dialogue* 52 (3): 231–48. https://doi.org/10.1177/0967010620933513.
337. Op.europa.eu. 2019. "Ethics Guidelines for Trustworthy AI." *Publications Office of the European Union*. https://data.europa.eu/doi/10.2759/346720.
338. Op.europa.eu. 2020. "Sectoral Considerations on the Policy and Investment Recommendations for Trustworthy Artificial Intelligence." *Publications Office of the European Union*. https://data.europa.eu/doi/10.2759/733662.
339. Orbie, Jan. 2006. "Civilian Power Europe: Review of the Original and Current Debates." *Cooperation and Conflict* 41 (1): 123–28. https://doi.org/10.1177/0010836706063503.
340. Ord, Toby. 2020. *The Precipice: Existential Risk and the Future of Humanity*. New York: Hachette Books.
341. Pace.coe.int. 2020. "Resolution." *Council of Europe*. https://pace.coe.int/en/files/28803/html.
342. Paul, Regine. 2017a. "Risk: New Issue or New Tool in Regulation and Governance Research?" In *Society, Regulation and Governance*, edited by Regine Paul, Marc Mölders, Alfons Bora, Michael Huber, and Peter Münte. Cheltenham. Edward Elgar Publishing. https://doi.org/10.4337/9781786438386.00011.

343. Paul, Regine. 2017b. “Harmonisation by Risk Analysis? Frontex and the Risk-Based Governance of European Border Control.” *Journal of European Integration* 39 (6): 689–706. https://doi.org/10.1080/07036337.2017.1320553.
344. Paul, Regine. 2022a. “The Politics of Regulating Artificial Intelligence Technologies: A Competition State Perspective.” *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4272867.
345. Paul, Regine. 2022b. “Can *Critical Policy Studies* Outsmart AI? Research Agenda on Artificial Intelligence Technologies and Public Policy.” *Critical Policy Studies* 16 (4): 497–509. https://doi.org/10.1080/19460171.2022.2123018.
346. Paul, Regine. 2024. “European Artificial Intelligence ‘Trusted throughout the World’: Risk-based Regulation and the Fashioning of a Competitive Common AI Market.” *Regulation & Governance* 18 (4): 1065–82. https://doi.org/10.1111/rego.12563.
347. Paul, Regine, Frédéric Bouder, and Mara Wesseling. 2016. “Risk-Based Governance against National Obstacles? Comparative Dynamics of Europeanization in Dutch, French, and German Flooding Policies.” *Journal of Risk Research* 19 (8): 1043–62. https://doi.org/10.1080/13669877.2015.1074936.
348. Peoples, Columba, and Nick Vaughan-Williams. 2020. *Critical Security Studies: An Introduction*. 3rd ed. Abingdon: Routledge. https://doi.org/10.4324/9780429274794.
349. Petersen, Karen Lund. 2012. “Risk Analysis – A Field within Security Studies?” *European Journal of International Relations* 18 (4): 693–717. https://doi.org/10.1177/1354066111409770.
350. Philab.esa.int. 2018. “Artificial Intelligence for Earth Observation #AI4EO White Paper.” *The European Space Agency*. https://philab.esa.int/artificial-intelligence-for-earth-observation-ai4eo-white-paper/.
351. Planqué-van Hardeveld, Anneroos. 2023. “Securing the Platform: How Google Appropriates Security.” *Critical Studies on Security* 11 (3): 161–75. https://doi.org/10.1080/21624887.2023.2239002.
352. Polyakova, Alina, and Chris Meserole. 2019. “Exporting Digital Authoritarianism: The Russian and Chinese Models.” *Brookings*. https://www.brookings.edu/wp-content/uploads/2019/08/FP_20190827_digital_authoritarianism_polyakova_meserole.pdf.

353. Prasad, Amit. 2022. “Anti-Science Misinformation and Conspiracies: COVID–19, Post-Truth, and Science & Technology Studies (STS).” *Science, Technology and Society* 27 (1): 88–112. https://doi.org/10.1177/09717218211003413.
354. Price, Matthew, Stephen Walker, and Will Wiley. 2018. “The Machine Beneath: Implications of Artificial Intelligence in Strategic Decision Making.” *PRISM* (4): 92–105.
355. Proudfoot, Kevin. 2023. “Inductive/Deductive Hybrid Thematic Analysis in Mixed Methods Research.” *Journal of Mixed Methods Research* 17 (3): 308–26. https://doi.org/10.1177/15586898221126816.
356. Qiao-Franco, Guangyu, and Ingvild Bode. 2023. “Weaponised Artificial Intelligence and Chinese Practices of Human–Machine Interaction.” *The Chinese Journal of International Politics* 16 (1): 106–28. https://doi.org/10.1093/cjip/poac024.
357. Rees, Wyn. 2008. “Inside Out: The External Face of EU Internal Security Policy.” *Journal of European Integration* 30 (1): 97–111. https://doi.org/10.1080/07036330801959515.
358. Regilme, Salvador Santino. 2025. “Tech Imperialism Reloaded: AI, Colonial Legacies, and the Global South.” *E-International Relations*. https://www.e-ir.info/2025/02/17/tech-imperialism-reloaded-ai-colonial-legacies-and-the-global-south/.
359. Renda, Andrea. 2019. “Artificial Intelligence.” *CEPS*. https://www.ceps.eu/ceps-publications/artificial-intelligence-ethics-governance-and-policy-challenges/.
360. Reports.nscai.gov. 2021. “NSCAI Final Report.” *National Security Commission on Artificial Intelligence*. https://reports.nscai.gov/final-report/.
361. Research-and-innovation.ec.europa.eu. 2023. “Futures of Science for Policy in Europe: Scenarios and Policy Implications.” *European Commission*. https://research-and-innovation.ec.europa.eu/news/all-research-and-innovation-news/futures-science-policy-europe-scenarios-and-policy-implications-2023-10-10_en.
362. Rességuier, Anaïs, and Rowena Rodrigues. 2020. “*AI Ethics Should Not Remain Toothless!* A Call to Bring Back the Teeth of Ethics.” *Big Data & Society* 7 (2): 205395172094254. https://doi.org/10.1177/2053951720942541.
363. Ricaurte, Paola, Edgar Gómez-Cruz, and Ignacio Siles. 2024. “Algorithmic Governmentality in Latin America: Sociotechnical Imaginaries, Neocolonial Soft Power, and Authoritarianism.” *Big Data & Society* 11 (1): 20539517241229697. https://doi.org/10.1177/20539517241229697.

364. Roberts, Huw, Josh Cowls, Federico Casolari, Jessica Morley, Mariarosaria Taddeo, and Luciano Floridi. 2021. "Safeguarding European Values with Digital Sovereignty: An Analysis of Statements and Policies." *Internet Policy Review* 10 (3). https://doi.org/10.14763/2021.3.1575.
365. Roberts, Huw, Josh Cowls, Jessica Morley, Mariarosaria Taddeo, Vincent Wang, and Luciano Floridi. 2021. "The Chinese Approach to Artificial Intelligence: An Analysis of Policy, Ethics, and Regulation." *AI & SOCIETY* 36 (1): 59–77. https://doi.org/10.1007/s00146-020-00992-2.
366. Roberts, Huw, Emmie Hine, Mariarosaria Taddeo, and Luciano Floridi. 2024. "Global AI Governance: Barriers and Pathways Forward." *International Affairs* 100 (3): 1275–86. https://doi.org/10.1093/ia/iiae073.
367. Roff, Heather M., and David Danks. 2018. "'Trust but Verify': The Difficulty of Trusting Autonomous Weapons Systems." *Journal of Military Ethics* 17 (1): 2–20. https://doi.org/10.1080/15027570.2018.1481907.
368. Rogers, James. 2009. "From 'Civilian Power' to 'Global Power': Explicating the European Union's 'Grand Strategy' Through the Articulation of Discourse Theory." *JCMS: Journal of Common Market Studies* 47 (4): 831–62. https://doi.org/10.1111/j.1468-5965.2009.02007.x.
369. Rossi, Francesca. 2018. "Building Trust in Artificial Intelligence." *Journal of International Affairs* 72 (1): 127–34. https://www.jstor.org/stable/26588348.
370. Rothe, Delf. 2011. "Managing Climate Risks or Risking a Managerial Climate: State, Security and Governance in the International Climate Regime." *International Relations* 25 (3): 330–45. https://doi.org/10.1177/0047117811415486.
371. Rothe, Delf. 2012. "Security as a Weapon: How Cataclysm Discourses Frame International Climate Negotiations." In *Climate Change, Human Security and Violent Conflict*, edited by Jürgen Scheffran, Michael Brzoska, Hans Günter Brauch, Peter Michael Link, and Janpeter Schilling. London: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-28626-1_12.
372. Rothe, Delf. 2017. "Seeing like a Satellite: Remote Sensing and the Ontological Politics of Environmental Security." *Security Dialogue* 48 (4): 334–53. https://doi.org/10.1177/0967010617709399.

373. Rothe, Delf. 2020. “Jellyfish Encounters: Science, Technology and Security in the Anthropocene Ocean.” *Critical Studies on Security* 8 (2): 145–59. https://doi.org/10.1080/21624887.2020.1815478.
374. Rothstein, Henry, Olivier Borraz, and Michael Huber. 2013. “Risk and the Limits of Governance: Exploring Varied Patterns of Risk-based Governance across E Urope.” *Regulation & Governance* 7 (2): 215–35. https://doi.org/10.1111/j.1748-5991.2012.01153.x.
375. Rothstein, Henry, Michael Huber, and George Gaskell. 2006. “A Theory of Risk Colonization: The Spiralling Regulatory Logics of Societal and Institutional Risk.” *Economy and Society* 35 (1): 91–112. https://doi.org/10.1080/03085140500465865.
376. Ruppert, Linda. 2024. “Geopolitics of Technological Futures: Warfare Technologies and Future Battlefields in German Security Debates.” *Geopolitics* 29 (2): 581–606. https://doi.org/10.1080/14650045.2023.2174431.
377. Russell, Stuart J. 2019. *Human Compatible: Artificial Intelligence and the Problem of Control*. Penguin.
378. Russell, Stuart J., and Peter Norvig. 2016. *Artificial Intelligence: A Modern Approach*. Boston: Pearson.
379. Rychnovská, Dagmar. 2020. “Security Meets Science Governance. The EU Politics of Dual-Use Research.” In *Emerging Security Technologies and EU Governance. Actors, Practices and Processes*, edited by Antonio Calcara, Raluca Csernatoni, and Chantal Lavallée. New York: Routledge.
380. Sanger, David E. 2024. “In Ukraine, New American Technology Won the Day. Until It Was Overwhelmed.” *The New York Times*. https://www.nytimes.com/2024/04/23/us/politics/ukraine-new-american-technology.html.
381. Scheipers, Sibylle, and Daniela Sicurelli. 2007. “Normative Power Europe: A Credible Utopia?” *JCMS: Journal of Common Market Studies* 45 (2): 435–57. https://doi.org/10.1111/j.1468-5965.2007.00717.x.
382. Schlag, Gabi. 2023. “European Union’s Regulating of Social Media: A Discourse Analysis of the Digital Services Act.” *Politics and Governance* 11 (3). https://doi.org/10.17645/pag.v11i3.6735.
383. Schmid, Stefka, Bao-Chau Pham, and Anna-Katharina Ferl. 2024. “Trust in Artificial Intelligence: Producing Ontological Security through Governmental Visions.” *Cooperation and Conflict*. https://doi.org/10.1177/00108367241288073.
384. Schopmans, Hendrik, and Jelena Cupać. 2021. “Engines of Patriarchy: Ethical Artificial Intelligence in Times of Illiberal Backlash Politics.”

Ethics & International Affairs 35 (3): 329–42. https://doi.org/10.1017/S0892679421000356.
385. Schuett, Jonas. 2024. "Risk Management in the Artificial Intelligence Act." *European Journal of Risk Regulation* 15 (2): 367–85. https://doi.org/10.1017/err.2023.1.
386. Schwarz, Elke. 2021. "Autonomous Weapons Systems, Artificial Intelligence, and the Problem of Meaningful Human Control." *Philosophical Journal of Conflict and Violence* 5 (1): 53–72. https://doi.org/10.22618/TP.PJCV.20215.1.139004.
387. Schwarz, Elke. 2025. "Conjuring the End: Techno-Eschatology and the Power of Prophecy." *Opinio Juris*. https://opiniojuris.org/2025/01/30/conjuring-the-end-techno-eschatology-and-the-power-of-prophecy/.
388. Scott, Mark. 2023. "Western Powers Argue over How to Control AI." *POLITICO*. https://www.politico.eu/article/eu-us-uk-china-artificial-intelligence-control/.
389. Scott, Mark, Mohar Chatterjee, and Gian Volpicelli. 2023. "The Struggle to Control AI." *POLITICO*. https://www.politico.eu/article/washington-eu-trade-and-tech-council-join-forces-to-stop-ai-harms/.
390. Seidl, Timo, and Luuk Schmitz. 2024. "Moving on to Not Fall behind? Technological Sovereignty and the 'Geo-Dirigiste' Turn in EU Industrial Policy." *Journal of European Public Policy* 31 (8): 2147–74. https://doi.org/10.1080/13501763.2023.2248204.
391. Serhan, Yasmeen. 2024. "What Israel's Use of AI in Gaza May Mean for the Future of War." *TIME*. https://time.com/7202584/gaza-ukraine-ai-warfare/.
392. Sezal, Mustafa Ali. 2023. "Security-Defence Nexus in Flux: (De)Securitisation of Technology in the Netherlands." *Defence Studies* 23 (4): 665–86. https://doi.org/10.1080/14702436.2023.2277456.
393. Sharkey, Amanda. 2024. "Could a Robot Feel Pain?" *AI & SOCIETY*. https://doi.org/10.1007/s00146-024-02110-y.
394. Shaw, Ian, and Majed Akhter. 2014. "The Dronification Of State Violence." *Critical Asian Studies* 46 (2): 211–34. https://doi.org/10.1080/14672715.2014.898452.
395. Shaw, Ian G.R. 2017. "The Great War of Enclosure: Securing the Skies." *Antipode* 49 (4): 883–906. https://doi.org/10.1111/anti.12309.
396. Shaw, Ian Gr. 2017. "Robot Wars: US Empire and Geopolitics in the Robotic Age." *Security Dialogue* 48 (5): 451–70. https://doi.org/10.1177/0967010617713157.

397. Sheffield.ac.uk. 2023. “AI Unlikely to Gain Human-like Cognition, Unless Connected to Real World through Robots.” *University of Sheffield.* https://www.sheffield.ac.uk/news/ai-unlikely-gain-human-cognition-unless-connected-real-world-through-robots.
398. Shepherd, Laura J., ed. 2013. *Critical Approaches to Security*. Abingdon: Routledge. https://doi.org/10.4324/9780203076873.
399. Sigfrids, Anton, Jaana Leikas, Henrikki Salo-Pöntinen, and Emmi Koskimies. 2023. “Human-Centricity in AI Governance: A Systemic Approach.” *Frontiers in Artificial Intelligence*. https://doi.org/10.3389/frai.2023.976887.
400. Smuha, Nathalie A. 2019. “From a ‘Race to AI’ to a ‘Race to AI Regulation’ - Regulatory Competition for Artificial Intelligence.” *SSRN.* https://doi.org/10.2139/ssrn.3501410.
401. Smuha, Nathalie A. 2021a. “Beyond the Individual: Governing AI’s Societal Harm.” *Internet Policy Review* 10 (3). https://doi.org/10.14763/2021.3.1574.
402. Smuha, Nathalie A. 2021b. “Beyond a Human Rights-Based Approach to AI Governance: Promise, Pitfalls, Plea.” *Philosophy & Technology* 34 (S1): 91–104. https://doi.org/10.1007/s13347-020-00403-w.
403. Smuha, Nathalie A. 2025. “Concluding Remarks.” In *The Cambridge Handbook of the Law, Ethics and Policy of Artificial Intelligence*, edited by Nathalie A. Smuha. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781009367783.
404. Smuha, Nathalie A., and Karen Yeung. 2024. “The European Union’s AI Act: Beyond Motherhood and Apple Pie?” *SSRN.* https://doi.org/10.2139/ssrn.4874852.
405. Sperling, James, and Mark Webber. 2014. “Security Governance in Europe: A Return to System.” *European Security* 23 (2): 126–44. https://doi.org/10.1080/09662839.2013.856305.
406. Sperling, James, and Mark Webber. 2019. “The European Union: Security Governance and Collective Securitisation.” *West European Politics* 42 (2): 228–60. https://doi.org/10.1080/01402382.2018.1510193.
407. State-of-the-union.ec.europa.eu. 2023. “State of the Union 2023.” *European Commission*. https://state-of-the-union.ec.europa.eu/state-union-2023_en.
408. Stavridis, Stelios. 2001. “‘Militarising’ the EU: The Concept of Civilian Power Europe Revisited.” *The International Spectator* 36 (4): 43–50. https://doi.org/10.1080/03932720108456945.

409. Stevens, Tim. 2017. *Cyber Security and the Politics of Time*. Cambridge: Cambridge University Press.
410. Stewart, William. 2024. "The Human Biological Advantage over AI." *AI & SOCIETY*. https://doi.org/10.1007/s00146-024-02112-w.
411. Strasser, Anna. 2022. "Distributed Responsibility in Human–Machine Interactions." *AI and Ethics* 2 (3): 523–32. https://doi.org/10.1007/s43681-021-00109-5.
412. Suchman, Lucy. 2016. "Configuring the Other: Sensing War through Immersive Simulation." *Catalyst: Feminism, Theory, Technoscience* 2 (1): 1–36. https://doi.org/10.28968/cftt.v2i1.28827.
413. Suchman, Lucy. 2020. "Algorithmic Warfare and the Reinvention of Accuracy." *Critical Studies on Security* 8 (2): 175–87. https://doi.org/10.1080/21624887.2020.1760587.
414. Suchman, Lucy. 2023. "The Uncontroversial 'Thingness' of AI." *Big Data & Society* 10 (2): 20539517231206794. https://doi.org/10.1177/20539517231206794.
415. Suchman, Lucy, Karolina Follis, and Jutta Weber. 2017. "Tracking and Targeting: Sociotechnologies of (In)Security." *Science, Technology, & Human Values* 42 (6): 983–1002. https://doi.org/10.1177/0162243917731524.
416. Taddeo, Mariarosaria. 2024. *The Ethics of Artificial Intelligence in Defence*. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780197745441.001.0001.
417. Tegmark, Max. 2017. *Life 3.0: Being Human in the Age of Artificial Intelligence*. Penguin UK.
418. Tolay, Juliette. 2021. "Inadvertent Reproduction of Eurocentrism in IR: The Politics of Critiquing Eurocentrism." *Review of International Studies* 47 (5): 692–713. https://doi.org/10.1017/S0260210521000176.
419. Torreblanca, Julian Ringhof, José Ignacio. 2022. "The Geopolitics of Technology: How the EU Can Become a Global Player." *ECFR*. https://ecfr.eu/publication/the-geopolitics-of-technology-how-the-eu-can-become-a-global-player/.
420. Torreblanca, Ulrike Franke, Carla Hobbs, Janka Oertel, Jeremy Shapiro, José Ignacio. 2020. "Europe's Digital Sovereignty: From Rulemaker to Superpower in the Age of US-China Rivalry." *ECFR*. https://ecfr.eu/publication/europe_digital_sovereignty_rulemaker_superpower_age_us_china_rivalry/.

421. Trondal, Jarle, and Lene Jeppesen. 2008. “Images of Agency Governance in the European Union.” *West European Politics* 31 (3): 417–41. https://doi.org/10.1080/01402380801939636.
422. Ulnicane, Inga. 2022. “Emerging Technology for Economic Competitiveness or Societal Challenges? Framing Purpose in Artificial Intelligence Policy.” *Global Public Policy and Governance* 2 (3): 326–45. https://doi.org/10.1007/s43508-022-00049-8.
423. Ulnicane, Inga. 2023. “Against the New Space Race: Global AI Competition and Cooperation for People.” *AI & SOCIETY* 38 (2): 681–83. https://doi.org/10.1007/s00146-022-01423-0.
424. Ulnicane, Inga, and Aini Aden. 2023. “Power and Politics in Framing Bias in Artificial Intelligence Policy.” *Review of Policy Research* 40 (5): 665–87. https://doi.org/10.1111/ropr.12567.
425. Ulnicane, Inga, and Tero Erkkilä. 2023. “Politics and Policy of Artificial Intelligence.” *Review of Policy Research* 40 (5): 612–25. https://doi.org/10.1111/ropr.12574.
426. Un.org. n.d.-a. “AI Advisory Body.” *United Nations*. Accessed February 25, 2025. https://www.un.org/en/ai-advisory-body.
427. Un.org. n.d.-b. “Secretary-General’s Statement at the UK AI Safety Summit.” *United Nations*. Accessed February 20, 2025. https://www.un.org/sg/en/content/sg/statement/2023-11-02/secretary-generals-statement-the-uk-ai-safety-summit.
428. Van Audenhove, Leo, and Karen Donders. 2019. “Talking to People III: Expert Interviews and Elite Interviews.” In *The Palgrave Handbook of Methods for Media Policy Research*, edited by Hilde Van Den Bulck, Manuel Puppis, Karen Donders, and Leo Van Audenhove. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-16065-4_10.
429. Varon, Joana, and Paz Peña. 2021. “Artificial Intelligence and Consent: A Feminist Anti-Colonial Critique.” *Internet Policy Review* 10 (4). https://doi.org/10.14763/2021.4.1602.
430. Verbeek, Peter-Paul. 2013. “Technology Design as Experimental Ethics.” In *Ethics on the Laboratory Floor*, edited by Simone Van Der Burg and Tsjalling Swierstra. London: Palgrave Macmillan UK. https://doi.org/10.1057/9781137002938_5.
431. Vesnic-Alujevic, Lucia, Susana Nascimento, and Alexandre Pólvora. 2020. “Societal and Ethical Impacts of Artificial Intelligence: Critical Notes on European Policy Frameworks.” *Telecommunications Policy* 44 (6): 101961. https://doi.org/10.1016/j.telpol.2020.101961.

432. Wæver, Ole. 1995. “Securitization and Desecuritization.” In *On Security*, edited by Ronnie D. Lipschutz. Columbia University Press.
433. Waever, Ole. 2003. “Identity, Communities and Foreign Policy: Discourse Analysis as Foreign Policy Theory.” In *European Integration and National Identity*, edited by Lene Hansen and Ole Waever. Abingdon: Routledge. https://doi.org/10.4324/9780203402207-10.
434. Wallace-Wells, David. 2019. *The Uninhabitable Earth: A Story of the Future*. Allen Lane.
435. Weforum.org. 2024. “Ursula von Der Leyen’s Speech to Davos.” *World Economic Forum*. https://www.weforum.org/stories/2024/01/ursula-von-der-leyen-full-speech-davos/.
436. Weldes, Jutta. 1996. “Constructing National Interests.” *European Journal of International Relations* 2 (3): 275–318. https://doi.org/10.1177/1354066196002003001.
437. Whitecase.com. 2025. “AI Watch: Global Regulatory Tracker.” *White&Case.* https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-european-union.
438. Wilcox, Lauren B. 2015. *Bodies of Violence: Theorizing Embodied Subjects in International Relations*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199384488.001.0001.
439. Williams, M.J. 2008. “(In)Security Studies, Reflexive Modernization and the Risk Society.” *Cooperation and Conflict* 43 (1): 57–79. https://doi.org/10.1177/0010836707086737.
440. Woszczyna, Karolina, and Karolina Mania. 2023. “The European Map of Artificial Intelligence Development Policies: A Comparative Analysis.” *International Journal of Contemporary Management* 59 (3): 78–87. https://doi.org/10.2478/ijcm-2023-0002.
441. Yayboke, Erol, and Samuel Brannen. 2020. “Promote and Build: A Strategic Approach to Digital Authoritarianism.” Center for Strategic and International Studies. https://www.csis.org/analysis/promote-and-build-strategic-approach-digital-authoritarianism.

ANNEXES

Annex 1. *Dataset of EU strategic documents*

Institution	Documents
European Commission (EC)	- Annexes COM(2018) 795 *Coordinated Plan on Artificial Intelligence* - Annexes COM(2021) 205 *Fostering a European approach to Artificial Intelligence* - Annexes to the Proposal for a Regulation COM (2021)206 *Laying down harmonised rules on artificial intelligence and amending certain Union legislative acts* - Communication (2019) *Building Trust in Human-Centric Artificial Intelligence* - Communication C(2024) 380 *A strategic vision to foster the development and use of lawful, safe and trustworthy Artificial Intelligence systems in the European Commission* - Communication COM (2018)237 *Artificial Intelligence for Europe* - Communication COM (2018)795 *Coordinated Plan on Artificial Intelligence* - Communication COM (2020)66 *A European Strategy for Data* - Communication COM (2021)205 *Fostering a European approach to Artificial Intelligence* - Communication COM(2024) 28 *On boosting startups and innovation in trustworthy artificial intelligence* - Joint Communication JOIN(2023)20 *European Economic Security Strategy* - Joint Research Centre (2018) *Artificial Intelligence and a European perspective* - Proposal for a Directive COM(2022) 496 *On adapting non-contractual civil liability rules to artificial intelligence (AI Liability Directive)* - Proposal for a Regulation 2024/0016 *amending Regulation (EU) 2021/1173 as regards an EuroHPC initiative for start-ups to boost European leadership in trustworthy Artificial Intelligence* - Proposal for a Regulation COM (2021)206 *Laying down harmonised rules on artificial intelligence and amending certain Union legislative acts* - Regulation 2024/1689 *Laying down harmonised rules on artificial intelligence and amending certain Union legislative acts (Artificial Intelligence Act)*

Institution	Documents
	- Report COM (2020) 64 *Report on the safety and liability implications of Artificial Intelligence, the Internet Things and robotics* - Report COM (2020)64 *on the Safety and Liability Implications of Artificial Intelligence, the Internet of Things and Robotics* - White Paper COM (2020)65 *On Artificial Intelligence – A European Approach to Excellence and Trust*
European Parliament (EP)	- Draft Report (2020/2266(INI)) *on artificial intelligence in a digital age* - European Parliamentary Research Service (2018) *Understanding artificial intelligence* - European Parliamentary Research Service (2019) *How artificial intelligence works?* - European Parliamentary Research Service (2019) *Regulating disinformation with artificial intelligence* - European Parliamentary Research Service (2019) *Why artificial intelligence matters* - European Parliamentary Research Service (2020) *an EU framework for artificial intelligence* - European Parliamentary Research Service (2020) *Artificial intelligence: how does it work, why does it matter, and what can we do about it?* - European Parliamentary Research Service (2020) *Digital sovereignty for Europe* - European Parliamentary Research Service (2021) *what if we chose new metaphors for artificial intelligence?* - European Parliamentary Research Service (2022) *investigation into the potential of artificial intelligence in the digital age* - European Parliamentary Research Service (2022) *what if machines made fairer decisions than human?* - European Parliamentary Research Service (2023) *Artificial intelligence, democracy and elections* - European Parliamentary Research Service (2023) *general-purpose artificial intelligence* - European Parliamentary Research Service (2023) *Parliament's negotiating position on the artificial intelligence act* - European Parliamentary Research Service (2023) *What if generative artificial intelligence became conscious?* - European Parliamentary Research Service (2024) *Artificial intelligence act*

Institution	Documents
	- European Parliamentary Research Service (2024) *Artificial intelligence and cybersecurity* - Report (A9-0001/2021) *on artificial intelligence: questions of interpretation and application of international law in so far as the EU is affected in the areas of civil and military uses and of state authority outside the scope of criminal justice* - Report (A9-0186/2020) with Recommendations *to the Commission on a Framework of Ethical Aspects of Artificial Intelligence, Robotics and Related Technologies* - Report (A9-0718/2020) *with Recommendations to the Commission on a Civil Liability Regime for Artificial Intelligence* - Report A9-0176/2020 *on intellectual property rights for the development of artificial intelligence technologies* - Report A9-0178/2020 *with recommendations to the Commission on the civil liability regime for artificial intelligence* - Resolution (P8_TA(2017)0051) *Civil Law Rules on Robotics* - Resolution P9_TA(2021)0009 *Artificial intelligence: questions of interpretation and application of international law* - Resolution P9_TA(2022)0140 *on artificial intelligence in a digital age*
European Council	- Council Conclusions (8098/1/20) *Shaping Europe's Digital Future* - European Council Conclusions EUCO 13/20 - Permanent Representative Committee (6177/19) *Artificial Intelligence. Conclusions on the coordinated plan on artificial intelligence* - Presidency conclusions (11481/20) *The Charter of Fundamental Rights in the Context of Artificial Intelligence and Digital Change*
Agencies	- EU Agency for Cybersecurity (2020) *AI Cybersecurity Challenges. Threat Landscape for Artificial Intelligence* - EU Agency for Cybersecurity (2023) *Artificial Intelligence and Cybersecurity Research* - EU Agency for Fundamental Rights (2019) *Data Quality and Artificial Intelligence – Mitigating Bias and Error to Protect Fundamental Rights* - EU Agency for Fundamental Rights (2020) *Getting the Future Right. Artificial Intelligence and Fundamental Rights* - EU Agency for Fundamental Rights (2022) *Bias in Algorithms – Artificial Intelligence and Discrimination*

Institution	Documents
	- EU Institute for Security Studies (2018) *Artificial Intelligence. What Implications for EU Security and Defence* - EU Institute for Security Studies (2020) *Digitalising Defence. Protecting Europe in the Age of Quantum Computing and the Cloud* - EU Institute for Security Studies (2020) *Digitalizing Defence* - EU Institute for Security Studies (2020) *Transatlantic Defence Cooperation on Artificial Intelligence* - EU Institute for Security Studies (2022) *Digital Divide? Transatlantic Defence Cooperation on Artificial Intelligence* - European Defence Agency (2018) *The EU Capability Development Priorities* - European Defence Agency (2019) *Annual Report* - European Defence Agency (2020) *European Defence Matters "Cards on the Table"* - European Defence Agency (2020) *European Defence Matters "Cards on the Table"* - European Defence Agency (2020) *European Defence Matters "Enhancing Interoperability. Train Together, Deploy Together"* - European Defence Agency (2020) *European Defence Matters "Joint Quest for Future Defence Applications"* - European Defence Agency (2021) *Annual Report* - European Defence Agency (2021) *European Defence Matters "Pushing Limits. Defence Innovation in a High-tech World"* - European Defence Agency (2022) *European Defence Matters "Investing in European Defence. Today's Promises, Tomorrow's Capabilities?"* - European Defence Agency (2023) *Annual Report*
Other	- Expert Group on Liability and New Technologies (2019) *Liability for Artificial Intelligence and Other Emerging Digital Technologies* - HLEG (2019) *A Definition of AI: Main Capabilities and Disciplines* - HLEG (2019) *Ethics Guidelines for Trustworthy AI* - HLEG (2019) *Policy and Investment Recommendations for Trustworthy AI* - HLEG (2020) *The Assessment List for Trustworthy AI* - Second European AI Alliance Assembly (2020) *Event Report*

Source: the author, based on selection criteria

Annex 2. *Dataset of strategic documents of different actors*

Actor	Documents
The United States of America (USA)	- National Artificial Intelligence Research and Development Strategic Plan (2016, 2019, 2023) - National Security Commission on Artificial Intelligence (2021) - National Security Strategy 2022 - White House document "Ensuring Safe, Secure, and Trustworthy AI" (2023) - Biden-Harris Executive Order (2023)
People's Republic of China (China)	- Document Internet+ (2015) - Strategy Made in China 2025 (2015) - New Generation AI Development Plan (2017) - Governance Principles for New Generation AI: Develop Responsible AI (2017) - Ethical Norms for New Generation AI (2021) - Opinions on Strengthening the Ethical Governance of Science and Technology (2022)
The United Kingdom (UK)	- House of Commons Science and Technology Committee Report on Robotics and AI (2016) - National AI Strategy (2021) - Defence Artificial Intelligence Strategy (2022) - Bletchley Declaration (2023)
Japan	- Social Principles of Human-Centric AI (2019) - AI Strategy (2022)
Council of Europe (CoE)	- Draft Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law (2024) - Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law (2024)
United Nations (UN)	- UNESCO Recommendation on the Ethics of Artificial Intelligence (2021) - Principles for the Ethical Use of Artificial Intelligence in the United Nations System (2022) - General Assembly Resolution Seizing the Opportunities of Safe, Secure and Trustworthy Artificial Intelligence Systems for Sustainable Development (2024) - Report Governing AI for Humanity (2024)

Source: the author

Annex 3. *List of interviews conducted*

	Interviewee	Time/format	Duration
1	Representative from the European Commission (EC)	May 2023 / live	60:57 min
2	Representative from the European Parliament (EP)	June 2023 / online	46:17 min
3	Representative 1 from the Council of the European Union (EU Council)	May 2023 / live	41:52 min
4	Representative 2 from the Council of the European Union (EU Council)	September 2023 / online	50:56 min
5	Representative from the European Defence Agency (EDA)	June 2023 / online	50:58 min
6	Representative 1 from the High Level AI Expert Group on (HLEG)	May 2023 / online	46:23 min
7	Representative 2 from the High Level AI Expert Group on (HLEG)	February 2024 / online	49:44 min
8	Institutional expert 1 on AI and security	May 2023 / live	44:59 min
9	Institutional expert 2 on AI and security	August 2023 / online	55:57 min
10	Representative from the European Union Institute for Security Studies (EUISS)	June 2023 / online	64:47 min
11	Representative of the EU research initiative on technologies	June 2023 / online	59:09 min

Names and credentials have been anonymized in accordance with the requests outlined in the consent form

Annex 4. *The questionnaire used for the interviews*

Introduction to the questionnaire

The doctoral thesis "Fortifying Digital Europe: agentic security and technocracy in the emerging EU AI policy", Vilnius University, 2020-2024

I am inviting you to participate in the research of the doctoral thesis *"Fortifying Digital Europe: agentic security and technocracy in the emerging EU AI policy",* which is conducted at the Institute of the International Relations and Political Science, Vilnius University, Lithuania. The doctoral thesis is under preparation as a monograph by myself, a PhD candidate, you can find my profile here. It is expected to finish the research by the end of 2024.

The aim of the research is to explore the AI-related security perception in the case of the EU. The research employs debates on risks as security and emerging policy considerations on AI. It is based on the discourse analysis which includes both the AI-related documents released by the EU institutions and interviews.

The information received from the interviews will provide additional understanding of the main arguments and choices to frame the AI policy, to employ the risk-based approach and taken directions for future AI governance. The research is based on the qualitative analysis of the EU documents related to the AI policy and interviews with representatives of the EU institutions.

The expected duration: up to 45 minutes. Depending on availability, the interview can take place either online or live. In case of any additional questions or comments regarding the research or the interview, please do not hesitate to contact me: justinas.lingevicius@tspmi.vu.lt.

I would like to thank you for your cooperation in advance. Please find the preliminary questions below.

AI policy and its relevance

1. To start with, what are the most important elements of AI? How would you describe the currently developed EU approach towards AI? From your perspective, what would be the main strengths and weaknesses of it?

2. The AI policy appears mainly directed towards Single Market and socio-economic matters. How would you describe various concerns raised, particularly related to EU values, fundamental rights and a matter of harm?

Risk-based approach

1. The European Commission employs different risks of AI and uses them as a major framework for the proposed AI Act. What is the reason to use the concept of risk? How different categories - unacceptable risk – high risk – transparency risk – minimal or no risk – have been introduced?

2. Risk symbolizes an issue of speculation, not exactly knowing the future advancement and applications of AI. From your perspective, what does the concept of risk talk about the EU's approach towards AI?

Preferences of AI governance

1. The overall EU approach towards AI introduces conceptual novelties such as good AI society, trustworthy and human-centric approach towards AI. From your perspective, what is the main purpose and role of inventing these concepts? What is their relationship to risks and their different levels?

2. There are various proposals of future AI governance – from the European board to national authorities, legally binding regulation and international standards or norms. How would you describe the model of AI governance in the EU? Is there a cross-institutional agreement towards its main measures?

SANTRAUKA

Tyrimo aktualumas

Dirbtinis intelektas (DI) tapo vienu iš pagrindinių Europos Sąjungos (ES) strateginių prioritetų Europos skaitmeninėje darbotvarkėje. Europos Komisijos (EK) pirmininkė Ursula von der Leyen pabrėžė: „Atėjo laikas suformuluoti viziją, atskleidžiančią, kur, mūsų noru, DI turėtų mus nuvesti kaip visuomenę ir žmoniją […] ir kokia turėtų būti konkreti Europos vieta tarptautinėse DI lenktynėse“ (Ec.europa.eu 2025). Tad formuojama ES DI politika pristatyta kaip pirmasis išsamus teisinis reguliavimas, skirtas DI plėtojimo ir naudojimo taisyklėms. Ši politika priskirta ES bendrajai rinkai ir įtraukta į skaitmeninę darbotvarkę, siekiant paspartinti skaitmeninę transformaciją ir neįtraukiant karinio dėmens.

Kartu politinis aktualumas ir tokie teiginiai kaip „mūsų požiūris į DI nulems pasaulį, kuriame gyvensime“ (Digital-strategy.ec.europa.eu, n.d.-d) kontrastuoja su pasirinktu techniniu ir, von der Leyen žodžiais tariant, „visada neutralaus“ (Ec.europa.eu 2020) DI apibrėžimu, anot kurio, tai yra „technologijų, jungiančių duomenis, algoritmus ir skaičiavimo galią, rinkinys“ (Eur-lex.europa.eu 2020d). Luciano Floridi (2021), dar buvęs Aukšto lygio ekspertų grupės DI klausimais nariu, teigė, kad dirbtinis intelektas nėra „koks nors Frankenšteino monstras“ ir kad nemokslinių formuluočių, tokių kaip „dirbtinė sąmonė“, pašalinimas padės išvengti „mokslinės fantastikos pobūdžio spekuliacijų, susijusių su DI“. Taigi pasirinkta kryptis signalizuoja, kad ES, nepaisydama išorinio spaudimo ir savo pačios politinės iniciatyvos, daugiausia dėmesio skiria su DI susijusiems techniniams procesams – funkciniams bruožams, lemiantiems, kaip apdorojami duomenys arba kokios užduotys atliekamos, – ir siekia „sutelkti Europos DI bendruomenę, kad būtų optimizuotas šios srities technologinis ir pramoninis potencialas“ (Csernatoni and Lavallée 2020).

Tačiau apibrėžimas, grįstas techninių galimybių ir bendrosios rinkos interesų argumentais, technologiją tarsi atsiejant nuo socialinių, politinių ir kultūrinių veiksnių (Ulnicane and Erkkilä 2023), neatitinka akcentuojamų rūpesčių. Skirtingos ES institucijos mini „riziką, susijusią su naujos technologijos naudojimu“ (Eur-lex.europa.eu 2020d) ir lūkesčius suformuluoti „į žmogų orientuoto, skaidraus ir atsakingo DI plėtrą“ (Ec.europa.eu 2023b). Šie teiginiai rodo, kad ES kalba apie „neatidėliotinumą, pareigą žinoti, keisti ir kurti ateitį“ (Manners 2024). Tad net ir kitapus ES bendros saugumo ir gynybos politikos (BSGP), skirtos tarptautiniam

saugumui stiprinti, ribų formuojamoje ES DI politikoje reikėtų atsižvelgti į saugumo iššūkius, formuluojamus pagal riziką, ateities vaizdinius ir norą suvaldyti dar neperprastą technologiją (pavyzdžiui, Csernatoni 2021; Bode and Huelss 2023; Amoore and Raley 2017). Šitaip DI komplikuoja tipinę perskyrą tarp saugumo ir nesaugumo ir skatina svarstyti, „kuo dar mes galime tapti“ (Amoore and De Goede 2012).

Šie, susiję, technologijų, saugumo ir rizikos klausimai taip pat verčia apmąstyti pačią ES ir jos vaidmenį tarptautinėje DI srityje. Kadangi valstybės narės vis dar rūpinasi saugumo sritimi, su DI susijusios rizikos akcentavimas skatina klausti, ar ES turi įgaliojimų formuluoti su technologijomis susijusią saugumo politiką – tiek remdamasi savo oficialiai nustatytomis kompetencijų ribomis, tiek jas peržengdama – ir, jeigu taip, kokiu mastu. Tuo pačiu metu formuojama ES DI politika sulaukė diskusijų ir net išorinio spaudimo atsakyti, kaip priimtos taisyklės paveiks ES konkurencingumą, padės išlikti atviru regionu prekybai ir investicijoms visame pasaulyje (Justo-Hanani 2022). Šie svarstymai atskleidžia sudėtingus iššūkius, neapsiribojančius technologiniais neaiškumais, apimančius ir pačios ES politinį nesaugumą, pavyzdžiui, baimę dėl vidinio susiskaldymo ir išorinio spaudimo. Šiomis aplinkybėmis ES atsiduria tarp konkuruojančių imperatyvų ir strateginių pasirinkimų: vadovauti, kaip teigė von der Leyen, „kuriant naują pasaulinę DI reguliavimo sistemą“ (State-of-the-union.ec.europa.eu 2023) ir (arba) manevruoti tarp savo įgaliojimų ribų saugumo politikos srityje (pavyzdžiui, Liebetrau 2024; Csernatoni 2024).

Remiantis mokslinėje literatūroje aptinkamomis diskusijomis, šie klausimai gali būti susieti su trimis pagrindinėmis sąvokomis – technologija, rizika ir saugumu. Jos susipina ir leidžia kalbėti apie įtampas, susijusias su formuojamos DI politikos akcentais ir ES pasirinkimais tiek sprendžiant vidinius kompetencijos ribų klausimus, tiek reaguojant į dinamišką tarptautinę aplinką. Toliau pristatomos skirtinguose tyrimuose išskirtos tendencijos liudija išliekančius klausimus ir tyrimų spragas, kurias galima nagrinėti toliau.

Saugumas tapo svarbiu ES įsiteisinimo matmeniu reaguojant į krizes ir didėjančius piliečių lūkesčius (Hegemann and Schneckener 2019). Įvairūs tyrimai jau parodė, kad ES pritarė platesniam nei vien karinis požiūriui į saugumą – pavyzdžiui, apimančiam klimato kaitą ir kibernetinius nusikaltimus (Sperling and Webber 2014) arba kompleksinius klausimus, susijusius su civilių apsauga, žmogaus teisėmis (Calderaro and Blumfelde 2022). Kibernetinio saugumo atvejis čia ypač iliustratyvus, nes tai iš pradžių buvo laikoma ekonomine problema, susijusia su bendrosios rinkos plėtojimu,

ir ją sprendė EK, o vėliau tapo saugumo politika, sprendžiama ES lygmeniu (Brandão and Camisão 2022; Carrapico and Barrinha 2017). Vadinasi, išsitrynė ribos tarp įprastinių saugumo ir su saugumu nesusijusių klausimų, o skirtingos darbotvarkės ir jų tikslai vis labiau persidengia.

Technologijų atveju situacija panaši. Įvairūs tyrimai atskleidė, kad skirtingas su technologijomis susijusias sritis ar instrumentus – tokius kaip Europos gynybos fondas ar dronai – ES linksta laikyti pramonės plėtros dalimi, o ne saugumo ar karinių pajėgumų plėtotės pavyzdžiu. Tačiau dar paaiškėjo, kad, nepaisant pasirinktos schemos, technologijos tampa militarizuotų praktikų dalimi ir net problemų sprendimu (pavyzdžiui, Csernatoni 2018; 2021a; 2019a; Csernatoni and Lavallée 2020; Lavallée and Martins 2023; Martins 2023; Martins and Jumbert 2022; Martins and Mawdsley 2021; Csernatoni and Martins 2024). Taip pat atskleista, kad, technologijas pristatant kaip ekonomiškai pelningus ir politiškai neutralius produktus, pasitelkiami ir militarizuoti diskursai (Hoijtink 2014). Todėl oficialioji retorika vis labiau susipina su veiksmais, susijusiais su civilinės ir karinės sričių perskyra (Martins and Ahmad 2020). Minėtuose tyrimuose, ypač virsmo technologijų, tokių kaip DI, atveju, plačiau neanalizuojama civilinio ir karinio, arba „dvejopo naudojimo“, problema – klausimas, apie kokį saugumą kalbama, kai ES, formuluodama savo požiūrį į DI, atsisako tiesiogiai įtraukti karinį dėmenį.

Saugumo paieškų kryptį siūlo formuojamos ES DI politikos centre esanti *rizikos* sąvoka. Ją vartojant kvestionuojama įprasta saugumo apibrėžtis, orientuota į grėsmes ir neatidėliotiną poreikį reaguoti. Tačiau rizikos apibrėžimas ir įtraukimas, ypač saugumo kontekste, nėra vienareikšmis. Remiantis teorine literatūra pastebimi du būdai kalbėti apie rizikos vaidmenį ir funkciją – arba kaip apie konkrečių valdymo priemonių rinkinį, arba kaip apie žinojimo konstravimą reaguojant į neapibrėžtumą. Čia išryškėja dar vienas ES nenuoseklumas. Nors laikoma, kad rizika nurodo tam tikrus politikos instrumentus – rizikos vertinimus, matavimo ir valdymo standartus – tokiose srityse kaip klimato kaita, maisto sauga, potvyniai ar terorizmas (pavyzdžiui: Paul, Bouder, and Wesseling 2016; Rothstein, Borraz, and Huber 2013), taip pat galima manyti, kad ji atskleidžia ateities vaizdinius, kuriais remiantis formuojama politika bei jos prioritetai. Apie tai rašo, pavyzdžiui, Regine Paul (2024; 2017b): ji teigia, kad rizika veikia kaip „epistemininis įrankis“, atskleidžiantis būdus mąstyti apie konkretų reiškinį. Kitaip tariant, rizika padeda įvardyti aktualiausias problemas, liekančias klausimu ateičiai, ir pasiūlyti formuojamos DI politikos žingsnius. Įtampa tarp rizikos ir saugumo lieka diskusijų klausimu: ar tai „dvi tos pačios monetos pusės“ (Methmann

and Rothe 2012), ar jos žymi skirtingas ontologines ir epistemologines perspektyvas?

Šiame kontekste nereikėtų manyti, kad technologijos, saugumas ir rizika žymi jau nusistovėjusias sampratas. Šie žodžiai veikiau perteikia skirtingas pozicijas, prioritetus ir interpretacijas – veikėjų, šiuo atveju ES, logiką ir siūlomą atsaką. Todėl dėmesys skiriamas ne tik pačios sąvokoms bei jų sąsajoms, bet ir jų reikšmes konstruojančiam veikėjui: klausiama, kaip technologijų, saugumo ar rizikos reikšmės susipina su ES pozicija, ypač atsižvelgiant į pastebimus neatitikimus tarp oficialiosios pozicijos ir praktikos, tarp išorės spaudimo ir teiginių, kad skaitmeninimas skatina ES saugumo politikos pokyčius.

Tyrimuose, atliekamuose žvelgiant į ES kaip tarptautinį veikėją, teigiama, kad ES skaitmeninės ambicijos keičia nusistovėjusius civilinės ir normatyvinės galios apibrėžimus, vertybėmis grindžiamo valdymo skatinimą ir įtikinėjimo bei institucijų naudojimą tarptautiniu mastu (McNamara 2024), pasitelkiant į saugumą orientuotą bei technologinę galią (pavyzdžiui, Raluca Csernatoni 2019b). Todėl vis dažniau pabrėžiama technologijų ir geopolitikos svarba ES (Monsees and Lambach 2022). Šiame kontekste dėmesys daugiausia krypsta į diskusijas apie skaitmeninį suverenitetą kaip besiplėtojančias ES ambicijas, apimančias tiek taisyklių ir demokratijos svarbą, tiek saugumo, kintančio santykio su kitomis tarptautinėmis veikėjomis, technologijų plėtros kryptis ir stipresnės kontrolės siekį (pavyzdžiui, Roberts, Cowls, Casolari, et al. 2021; Seidl and Schmitz 2024; Adler-Nissen and Eggeling 2024; Baur 2024; Bellanova and Glouftsios 2022; Klimburg-Witjes 2024).

Suverenitetas čia nesiejamas su tradicine teritorijos samprata – šios sampratos reikšmė, susijusi su technologijomis ir skaitmeninimu, išlieka kintama ir neapibrėžta. Galima klausti: kaip technologijos ir su jomis susijusios tarptautinės tendencijos prisideda prie ES subjektiškumo steigties? Ar nuorodos į „DI geopolitiką“ gali reikšti kovą dėl DI technologijų reikšmių, valdymo ir kontrolės, kurią steigia diskursai, galios santykiai ir pasaulinė asimetrija? Siūlomos užuominos, esą ES pozicijos kinta ir įvedamas skaitmeninis suverenitetas, atrodo, iki galo nedetalizuoja sąsajų su formuojama DI politika ir (ar) iš jos kylančia savęs samprata. Tokios detalės svarbios, nes, kaip jau minėta, DI, kaip tarptautinio proceso dalis, verčia ES ne tik užimti poziciją technologijos atžvilgiu, bet ir į šią tarptautinę aplinką projektuoti save.

Šie tyrimai, kuriuose kalbama apie ES vaidmens ir savęs pozicionavimo kaitą, leidžia manyti, jog skaitmeninimo kontekste ES veikiau linksta ieškoti

strateginius ir protekcionistinius prioritetus akcentuojančio vaidmens. Tačiau kyla įspūdis, kad pagrindinė tų tyrimų įžvalga ir yra pokyčio fiksavimas, o pats jis nėra plačiau detalizuojamas konkrečiame formuojamos ES DI politikos kontekste. Problema lieka aktuali, nes DI ir su juo susijęs saugumas yra ne tik vidinės, bet ir tarptautinės aplinkos dalis. Kaip jau minėta, ES siekia ne tik apibrėžti savo santykį technologijos atžvilgiu, bet ir reaguoti į išorinį spaudimą, susijusį su skirtingais interesais ir pačios ES DI politikos sprendimais.

Remiantis pristatytomis diskusijomis, galima išskirti tris aktualius klausimus.

Pirma, teorinė literatūra rodo, kad tyrimuose neklausiama, kokio saugumo siekiama kalbant apie DI: juose daugiausia dėmesio skiriama ES diskursų ir praktikos nenuoseklumams atskleisti, pabrėžiant, kaip technologijos, pristatomos prisidengiant civilių naudojimu, vis dažniau įtraukiamos į militarizuotas praktikas. Tačiau šios įžvalgos vis tiek daromos vadovaujantis karine ir civiline dichotomija: dėmesys sutelkiamas į nykstančią ribą tarp civilinės ir karinės srities. Kadangi karinis dėmuo bent jau oficialiai neįtraukiamas į DI politikos taikymo sritį, lieka neaišku, koks saugumas konstruojamas, ypač kai dėmesys sutelkiamas į rizikos sąvokos vartojimą.

Antra, rizikos sąvoka formuojamos ES DI politikos kontekste laikoma savaime suprantama, plačiau nenagrinėjamas jos vaidmuo formuojant su technologija susijusią saugumo sampratą. Kaip teigia Louise Amoore (2023), su DI susijusios rizikos diskursas veikia kaip būdas žinioms rinkti ir sisteminti, kartu keičiantis valstybės ir visuomenės savivoką. Tačiau vis dar kyla klausimų, ką tokia savivoka reiškia ir koks vaidmuo tenka rizikai DI atveju, ypač kai saugumas nėra susijęs su įprastesne grėsmių ir išskirtinių priemonių joms spręsti logika.

Trečia, apžvelgtos diskusijos parodė, kad ES domėjimasis technologijomis – tai ne tik siekis išvengti vidinio susiskaldymo. Jis rodo tarptautinius ES užmojus ir strateginius, įskaitant saugumą, siekius skaitmeninėje srityje. Kadangi tarptautinėje erdvėje dažnai kalbama apie „DI geopolitiką“, formuluojamos technologijų, saugumo ir rizikos reikšmės veikia ne tik ES vidaus kontekste, bet ir siekiant užimti norimą poziciją tarptautinėje aplinkoje. Vis dėlto ligšiolinės diskusijos, siūlančios regimą poslinkį link skaitmeninio suvereniteto kaip didesnės kontrolės siekio, labiau telkiasi į paties pokyčio įrodymą nei į jo vaidmenį konstruojant su DI susijusią saugumo sampratą.

Taigi disertacijoje atliktame tyrime identifikuojama spraga, apimanti tris dalykus – saugumo sampratos dviprasmiškumą, rizikos aktualumą ir savęs pozicionavimą skaitmeninėje erdvėje. Esami moksliniai tyrimai rodo, kad virsmo technologijos, įskaitant DI, tebėra veikiamos civilinės ir karinės sričių perskyros keliamų įtampų, tačiau tik nedaugelyje tyrimų kritiškai analizuojama, kaip DI kontekste perkuriamas ar iš naujo apibrėžiamas pats saugumas. Šis trūkumas tampa aktualus atsižvelgiant į naujausius kritinius tekstus, kuriuose teigiama, kad rizika ne vien nurodo politikos priemones, bet ir veikia kaip episteminės tvarkos elementas, atskleidžiantis, kaip apmąstoma ir įsivaizduojama DI ateitis. Be to, siūlomas poslinkis prie strategiškesnio ES vaidmens skaitmeninėje srityje taip pat ragina atidžiau išanalizuoti ES poziciją, susipinančią ne tik su tarptautine įtampa ir spaudimu, bet ir su saugumo klausimais.

Tyrimo problema

Disertacijoje dėmesys krypsta į pastebimą nenuoseklumą dėl saugumo, susijusio su DI, apibrėžimo formuojamoje ES DI politikoje. Nors pripažįstama, kad DI kuria naują vienovę, bet ir kelia daug klausimų (Rychnovská 2020), ES, pristatydama DI, linkusi pasitelkti techninius, o ne grėsmių ir saugumo terminus. Galima teigti, kad tokia logika atspindi ES kompetenciją bei apribojimus, nes ir toliau visų pirma siekiama plėtoti bendrąją vidaus rinką ir sukurti reguliavimą, kurio privalės laikytis DI kūrėjai bei paslaugų teikėjai.

Kartu formuluodama ambiciją tapti „nuoseklia saugumo srities veikėja“ (Carrapico and Barrinha 2017), ES reiškia užmojį formuoti standartus ir valdyti technologijas, kurių poveikis išlieka neapibrėžtas. Šie tikslai atskleidžia platesnes diskusijas apie tai, koks skaitmeninimas laikytinas tinkamu Europos visuomenėms vyraujančių konkurencijos ir liberaliosios demokratijos gynimo diskursų kontekste (Mügge 2025; Hoijtink and Van der Kist 2025).

Tačiau saugumo samprata, ypač atsižvelgiant į karinio dėmens neįtraukimą į politikos apimtį, šių skirtingų veiksnių ir ambicijų kontekste išlieka neapibrėžta. Nors tyrimuose pripažįstama, kad saugumas yra svarbi ES skaitmeninės darbotvarkės dalis – tiek militarizuotose praktikose, tiek diskusijose apie skaitmeninį suverenitetą – iki šiol mažai dėmesio skiriama tam, kaip ES konstruoja su DI susijusį saugumą ir kokias reikšmes suteikia formuojamoje DI politikoje, pasitelkdama rizikos sąvoką.

Tyrimo tikslas ir uždaviniai

Disertacijoje siekiama ištirti, kaip ES konstruoja ir apibrėžia su DI susijusį saugumą savo formuojamoje DI politikoje, pasitelkdama rizikos sąvoką. Šiuo tikslu analizuojamos diskurso pozicijos ir perspektyvos, daugiausia dėmesio skiriant tam, kaip jos formuoja DI suvokimą ir ES savęs pozicionavimą tarptautiniame kontekste. Čia ES laikoma veikėja, turinčia daugybę balsų, kurie prisideda prie strateginio diskurso, susijusio su DI, ir bendro subjektiškumo skaitmeninėje srityje. Nors instinktyviai norėtųsi kreipti dėmesį į institucijų galios dinamiką bei skirtingus prioritetus, disertacijoje teigiu, kad žinojimo konstravimą imant domėn saugumo ir savęs pozicionavimo aspektus galima apibendrintai laikyti būdingu ES ir atstovaujančiu ES pozicijai.

Kalbant apie skaitmeninę darbotvarkę (nors EK turi kompetenciją inicijuoti skaitmeninę politiką), atrodo, kad dėl pagrindinių šios politikos krypčių sutariama bendrai. Tai įrodo 2022 m. Europos deklaracija dėl skaitmeninių teisių ir principų – ją pasirašė EK, Europos Parlamento ir Europos Tarybos pirmininkai, o valstybės narės įsipareigojo priimti. Deklaracijoje teigiama, kad ES skaitmeninė transformacija turi atspindėti ES vertybes, skleisti į žmogų orientuotą viziją ir skatinti piliečių teisių apsaugą (Digital-strategy.ec.europa.eu, n.d.-e). Todėl stiprėjantis ES vaidmuo formuojant skaitmeninę darbotvarkę ir prisiimant koordinatorės vaidmenį, o ne „skirtingas“ ar „fragmentiškas“ nuomones (Af Malmborg 2023) rodo ES subjektiškumą skaitmeninimo srityje kartu su kitomis jos politikos sritimis, pavyzdžiui, prekyba. Anu Bradford (2023) ES netgi pavadino viena iš *skaitmeninių imperijų*, greta JAV ir Kinijos, rodančios ambicijas formuoti tarptautinę skaitmeninę tvarką. Be to, manoma, kad ES geba vis geriau atlikti saugumo valdymo funkcijas ir įsitraukti į nuolatinį unikalaus savo vaidmens saugumo srityje formulavimo procesą (Sperling and Webber 2014; Baker-Beall and Mott 2022). Todėl ir šiame tyrime daugiausia dėmesio skiriama su DI susijusio saugumo konstravimui, dar kartą teigiama, kad ES ir šioje srityje veikia kaip subjektas, integruojantis skirtingas darbotvarkes ir nuolat kintantis.

Disertacijoje laikausi nuomonės, kad ES požiūris į DI pristato ES mąstymą apie technologiją ir pasirinktą plėtojamą strateginę kryptį. Formuojama ES DI politika yra konkretus, formalizuotas šio požiūrio rezultatas ir išraiška – komunikatai, ataskaitos, rezoliucijos ar teisiškai privalomos taisyklės, grindžiamos institucijų indėliu per 2018–2024 metus. Apibūdinimas „formuojama“ taip pat nurodo tęstinumą laike ir įvairius

institucinius įnašus į politiką, skirtą *virsmo* technologijai, kurios plėtros galimybės ir kryptys išlieka ateities nežinomuoju. Todėl daugiausia dėmesio skiriama reikšmių konstravimui, matant steigiamąjį ryšį tarp reikšmės, praktikų ir bendros politinės „netvarkos“ (Matejova and Shesterinina 2023a).

Atsižvelgiant į visas šias pastabas, atliekant tyrimą buvo iškelti tokie pagrindiniai tikslai:

1. Išskirti pagrindines akademinių ir politinių diskusijų apie DI tendencijas, padedančias suprasti, kokios sąvokos dažniausiai vartojamos apibūdinant DI ir su juo susijusius vertinimus.
2. Įsitraukti į kritines saugumo diskusijas, atskleidžiančias technologijų, saugumo ir rizikos sąsajas, toliau plėtoti konceptualųjį tyrimo pagrindą.
3. Apibrėžti su DI susijusią saugumo sampratą formuojamoje ES DI politikoje ir taip pratęsti diskusijas apie ES požiūrį į saugumą bei jos pačios saugumo politiką.
4. Identifikuoti ES siūlomą atsaką įvardijamomis politikos priemonėmis, įtraukti įvardytą pasirinkimą į platesnį ES valdysenos kontekstą.
5. Nurodyti ES savęs pozicionavimą, susijusį su formuluojama saugumo samprata, išskiriant pagrindines jos ypatybes ir atsižvelgiant į esamas diskusijas apie ES kaip galią.

Disertacijoje teigiama, kad ES turimą su DI susijusio saugumo sampratą grindžia agentinio saugumo sąvoka, technokratizacija kaip valdymo būdas ir tvirtovė kaip ES pozicija tarptautinėje erdvėje. Remiantis šiomis įžvalgomis, siūlomi keturi ginamieji teiginiai:

1. „Rizikifikavimo“ procesas rodo, kad su DI susijusio saugumo samprata konstruojama pasitelkiant rizikos sąvoką. Įvestos rizikos kategorijos, apibūdinamos pagal galimas žalos sąlygas pamatinėms žmogaus teisėms ir demokratinei politinei sistemai, atspindi ES nubrėžtas ribas, pagal kurias vertinami su DI susiję klausimai ir prioritetai. Toks rizika grindžiamas požiūris rodo, kad saugumo samprata formuluojama kaip ilgalaikė, orientuota į ateitį ir grindžiama technologiniais vaizdiniais.
2. ES apibrėžia DI saugumą kaip agentinį saugumą, suprantamą kaip žmogaus gebėjimas išlaikyti kontrolę ir sprendimų priėmimo galią sąveikaujant su DI, apibrėžiamu kaip Kitas. Agentinis saugumas čia įvedamas kaip papildomas žmogaus ir visuomenės saugumo matmuo, nes vartojant šią sampratą dėmesys sutelkiamas į žmogaus agentiškumą antagonistiška laikomos žmogaus ir technologijos sąveikos kontekste.

Tokią logiką grindžia antropocentrizmas ir jo poreikis išlaikyti hierarchinę žmogaus poziciją.

3. ES siūloma valdysenos programa kaip atsakas į riziką ir galimos žalos sąlygas yra technokratizuota. Ja siekiama užtikrinti saugumą kaip kasdienę rutiną: standartizuotas priemones, ekspertinio indėlio svarbą ir paskirstytą atsakomybę. Technokratizacijos samprata atspindi ES pasirinkimą apibrėžti DI tik pagal techninius bruožus, tikintis, kad mokslinės žinios padės orientuotis neapibrėžtumo sąlygomis. Pasitelkiant technokratizuotą valdyseną taip pat siekiama įtikinti tarptautinę bendruomenę priimti ES siūlomą požiūrį į DI, apibūdinamą kaip universaliai tinkamas ir grindžiamas išoriniu ekspertiškumu, o ne politiniais ginčais.
4. ES savęs pozicionavimą formuojamoje ES DI politikoje galima apibendrintai nusakyti pasitelkus metaforą *tvirtovė*. Tvirtovės vizija atsiranda kaip atsakas į DI kaip Kito sampratą ir nepalankią tarptautinę aplinką, toliau stiprinančią DI kitoniškumą. Tvirtovė žymi apibrėžtą erdvę, veikiančią trimis būdais: a) ES taisyklės taikomos ir tiems, kurie yra viduje, ir norintiesiems pakliūti į tvirtovę; b) siekiama sumažinti priklausomybę nuo tarptautinių galių; c) siekiama įtikinti kitus perimti ES požiūrį į DI.

Teorinė prieiga

Šioje disertacijoje, remiantis konstruktyvistine teorine perspektyva, dėmesys krypsta į technologijų, saugumo ir rizikos sąveiką ir svarstoma, kaip juos formuoja galia, žinojimas ir valdysena (ir kaip jie savo ruožtu yra jų formuojami). Siekiant atlikti išsamią analizę, formuluojamas daugiasluoksnis konceptualusis modelis, grindžiamas prielaida, kad saugumas nėra objektyvi sąlyga – jis konstruojamas per bendras reikšmes, diskursus ir praktikas. Todėl susirūpinimas ir rūpesčiai plačiąja prasme interpretuojami kaip nerimas dėl galimos problemos, kuri nėra iš anksto apibrėžiama kaip pavojus, bet vis tiek reikalauja dėmesio, peržengiančio įprastas politinių praktikų ribas.

Šiame tyrime laikomasi nuomonės, kad technologijas formuoja socialinės, institucinės, ekonominės ir materialinės galimybės bei apribojimai; skiriama dėmesio tam, kaip technologijos įsivaizduojamos ir vėliau įtraukiamos į politinius sprendimus (Liebetrau and Christensen 2021; Leese and Hoijtink 2019). Panašiai žvelgiama į saugumo sąvoką. Remiantis konstruktyvistine perspektyva, klausiama, kaip struktūruojamos ateities vizijos ir su jomis susiję rūpesčiai, kokios ribos ir skirtumai nustatomi tarp

„vidaus“ ir „išorės“ (Huysmans 2008; Bigo 2001). Todėl su saugumu susijusios sąvokos, tokios kaip konkurencija, galia ir interesai, nėra pastovios; verta nagrinėti, kaip jos kuriamos, ginčijamos ir įtraukiamos į platesnį politinį, technologinį ir normatyvinį kontekstą.

Saugumo ir rizikos sąvokos taip pat laikomos socialiai konstruojamomis ir priklausančiomis nuo to, kaip ir koks antagonistiškumas artikuliuojamas. Pavyzdžiui, remiantis saugumizavimo logika, grėsmės siejamos su išlikimo klausimu ir neatitaisoma žala, reikalaujančia neatidėliotino atsako, o šis dažnai pateikiamas kaip išskirtinės priemonės formuluojamai grėsmei įveikti (Methmann and Rothe 2012). O štai rizika suprantama ne tiek akcentuojant išlikimo klausimą, kiek daugiau dėmesio skiriant galimai žalai. Rizika taip pat nurodo kitokią laiko perspektyvą, nukreiptumą į ateitį ir jos nežinomybę, siekį sušvelninti galimas pasekmes. Politiniai sprendimai savo ruožtu yra nukreipti į ilgalaikes galimos žalos valdymo priemones, nepereinančias į nepaprastosios padėties ar išskirtinių priemonių sritį (Backman 2023). Todėl kalbant apie rizikas klausiama, kokios situacijos suvokiamos kaip rizikingos (Matejova and Shesterinina 2023b), kaip rizika pasitelkiama apibrėžiant numanomus iššūkius, jeigu saugumas ima sietis ne tik su išlikimo klausimu.

Rizikos, o ne grėsmių artikuliavimas ES formuojamoje DI politikoje nepanaikina saugumo iššūkių, net jei jie nelaikomi „aukštosios politikos“, t. y. išlikimo, problemomis, dėl kurių reikėtų imtis išskirtinių priemonių. Tad tyrimui reikalingas tokią perspektyvą atitinkantis konceptualusis pagrindas, leidžiantis plėtoti minėtą grėsmių ir rizikų perskyrą. Todėl disertacijoje pasitelkiama „rizikifikavimo“ prieiga, kurios laikantis siūloma analizuoti, kaip rūpesčiai tampa rizika ir kokiomis priemonėmis siūloma ją valdyti (Morsut and Engen 2022). Disertacijoje išskiriami keturi analitiniai elementai: referentinis objektas, galimõs žalos sąlygos, valdysenos programa ir tarptautinis bendradarbiavimas. Galimos žalos sąlygos ir valdysenos programa rodo esminį poslinkį nuo saugumizavimo elementų – grėsmių ir nepaprastųjų priemonių, nes abu elementai (galimos žalos sąlygos ir valdysenos programa) susiję su ilgalaikiu valdymu, pateisinančiu politikos priemones (Corry 2012). Tarptautinio bendradarbiavimo elementas įtraukiamas kaip papildomas rizikifikavimo veiksnys, leidžiantis nustatyti, kaip ES įsitraukia į tarptautinius procesus ir sprendžia savo saugumo rūpesčius, susijusius su DI politika. Galiausiai rizikifikavimas padeda toliau analizuoti minėtąją įtampą tarp saugumo ir rizikų ir klausti, kaip konstruojamos su DI susijusios rizikos, kaip žvilgsnis krypsta nuo neatidėliotinumo į neapibrėžtą ateitį. Taigi disertacija atliepia įvardytą tyrimų spragą ir pagrindžia su DI susijusio saugumo analizę.

Lentelėje nurodomi pagrindiniai taikomų analitinių elementų motyvuojami klausimai.

1 lentelė. *Pagrindiniai rizikifikavimo analitinių elementų klausimai*

Referentinis objektas	• Kas formuojama kaip referentinis objektas? • Kokie pagrindiniai referentinio objekto bruožai? • Kokie saugumo klausimai susiję su referentiniu objektu?
Galimos žalos sąlygos	• Kas tampa žala, atsižvelgiant į artikuliuotus pavojus ir ateities neapibrėžtumą? • Kokios yra galimos žalos sąlygos? • Kokia saugumo samprata formuojasi?
Valdysenos programa	• Kaip turėtų būti valdoma rizika? • Kokios yra siūlomos politikos priemonės, kaip jos apibūdinamos? • Kaip legitimuojama valdysenos programa?
Tarptautinis bendradarbiavimas	• Kaip apibrėžiamas tarptautinis DI kontekstas? • Kokios yra tarptautinio bendradarbiavimo kryptys? • Kokios yra savęs pozicionavimo formos?

Šaltinis: autorius, remiantis teorine prieiga.

Tyrimo strategija

Kadangi tiriant rizikifikavimą svarbūs kalbos aktai (Harijanto 2025), disertacijoje atliekama diskurso analizė, siekiama suprasti pagrindines tekstuose atsiskleidžiančias reikšmes. Tokia prieiga nenukrypsta nuo pasirinktos konstruktyvistinės perspektyvos: dėmesys diskursui leidžia suprasti ir nustatyti, kaip ir kokios idėjos, sąvokos, kontekstai yra formuluojami, kokias politines funkcijas jie atlieka. Šiame kontekste ES laikoma „labai diskursyvia“ dėl pareiškimų ir įvairių politinių dokumentų gausos, atskleidžiančios institucinės kalbos ir veiksmų sistemos sąsajas (Baker-Beall 2014). ES formuojama DI politika liudija tą pačią tendenciją, nes ji apima dokumentų, pozicijų ir pareiškimų įvairovę, sudarančią ES strateginį DI diskursą. Jis ir tampa tyrimo ašimi.

Vadovaujantis interpretacine perspektyva, tyrime skiriama dėmesio ES pasirinktoms sąvokoms, nusistovėjusiems apibrėžimams ir formuluotėms, grindžiančioms mąstymą apie DI, ES tikslus bei su jais susijusias reikšmes. Turint mintyje, kad diskurso analizės strategija dažnai priklauso nuo

konkretaus atvejo, čia pirmiausia pristatomi pagrindiniai etapai, tokie kaip duomenų atranka ir rinkimas, kodavimo schema ir interpretacija. Atliekant diskurso analizę taip pat būtina kruopščiai apmąstyti tyrėjo poziciją ir per visą procesą priimtus sprendimus, todėl tyrimo strategija ir etapai pristatomi reflektuojant iškylančias problemas ir ribas. Pavyzdžiui, pabrėžiama, kad gautų rezultatų nesiekiama apibendrinti plačiau nei formuojama DI politika – linkstama koncentruotis į konkretų atvejį ir apibrėžtas tyrimo sąlygas. Vertinant surinktų šaltinių skaičių skiriama dėmesio jų įvairovei įtraukiant skirtingų institucijų pozicijas į nagrinėjamą ES strateginį DI diskursą.

Diskurso analizėje naudojami dviejų tipų duomenys: 1) ES formuojamos DI politikos dokumentai ir 2) pusiau struktūruoti interviu su ekspertais ir ekspertėmis. Vadovaujantis nustatytais kriterijais, atrinkti 75 ES institucijų dokumentai (sąrašas pateikiamas 1 disertacijos priede), paskelbti 2018–2024 metais. Pasirinktas laikotarpis nuo pirmojo komunikato „Dirbtinis intelektas Europai“, skirto konkrečiai DI, iki ES Dirbtinio intelekto akto įsigaliojimo 2024 m. rugpjūtį – savotiškos politikos formavimo, derybų ir priėmimo proceso pabaigos. Jis atskleidžia intensyvias tarpinstitucines diskusijas ir žinojimo konstravimą. Taip pat įtraukiami 11 pusiau struktūruotų interviu, atlikti nuo 2023 m. gegužės iki 2024 m. vasario, su ES institucijų ekspertėmis bei ekspertais, taip pat dalyvavusiais formuojant politiką (interviu sąrašas pateikiamas 3 priede). Šie tekstai padeda išplėsti analizę, nes juose išryškėja formuojant politiką dalyvaujančių asmenų nuomonės ir perspektyvos. Žinoma, klausimas, kas kalba, išlieka metodologiškai jautrus, nes ES nėra monolitiška. Todėl empirinėje analizėje aptariami instituciniai balsai ir jais remiamasi, kad būtų parodytas esamas ES daugialypumas, liudijamas daugybės dokumentų, sąveikų ir persidengiančių žodynų.

Apskritai tokia tyrimo strategija atitinka pagrindinius interpretacinės analizės lūkesčius, be to, ją galima taikyti tolesniuose tyrimuose, kurių akiratyje atsidurs su DI susijęs žodynas ir diskurso praktikos. Interviu, pradedant potencialių respondentų bei respondenčių paieška ir baigiant pokalbių analize, buvo atliekami laikantis pagrindinių tyrimų etikos ir dalyvių apsaugos reikalavimų. Todėl visą procesą – nuo pagrindinių strategijos punktų apibrėžimo, tekstų atrankos, interviu, tekstų segmentų ir jų kodavimo iki interpretavimo – grindė bendros gairės ir pavyzdžiai, rodantys, kaip atlikti diskurso analizę ir neperžengti jos ribų, kaip ją pritaikyti prie disertacijos tikslo. Galiausiai interpretuojant kyla iššūkis aiškiai aprašyti interpretacinius žingsnius, grindžiamus tyrėjo mąstysenos ir diskurso bei pagrindinių jo elementų sampratos. Tai ne trūkumas, o veikiau analitinio proceso dalis,

struktūruota ir patobulinta remiantis esama literatūra ir metodologiniais pasirinkimais.

Svarbiausios įžvalgos

Pristatytas disertacijos tikslas (analizuoti, kaip ES konstruoja saugumo sampratą formuojamoje DI politikoje) iškeltas reaguojant į jau atliktus tyrimus, kuriuose kalbama apie pastebėtus neatitikimus tarp technologijų įrašymo į bendrosios rinkos kontekstą ir militarizuotų praktikų, taip pat į tendencijas pasitelkti riziką kaip įvairių politikos sričių priemonę. Nors ES vengia tiesiogiai kalbėti apie saugumą, susijusį su DI, tyrime siūloma manyti, kad rizika tampa svarbiu veiksniu, parodančiu, kokiu DI vaizdiniu vadovaujasi ES, – jis suprantamas kaip ilgalaikis iššūkis vyraujančiai žmogaus ir jo reikšmės sampratai.

Pritaikius išskirtus analitinius „rizikifikavimo“ elementus, tyrimas parodė, kad ES formuluoja ir laiko DI *Kitu* – sudėtingu, kintančiu ir rūpestį keliančiu reiškiniu, kuriam reikia ilgalaikio technokratizuoto valdymo. Rūpestis pamatinėmis žmogaus teisėmis ir demokratine sistema bei galimos žalos sąlygomis parodė, kad ES suvokia DI saugumą kaip agentinį, o atsaką į jį grindžia technokratizuota valdysenos programa, pozicionuodama save kaip tvirtovę. Tarptautinio bendradarbiavimo įtraukimas kaip dar vienas analitinis rizikifikavimo elementas ne tik atskleidė ES poziciją, bet ir parodė, kad ES siekia spręsti rizikos klausimus tiek viduje, tiek perkeldama jų aktualumą į išorę. Todėl ES politinės preferencijos ir požiūris į DI susipina formuluojant su DI susijusio saugumo sampratą, apimančią reikšmes, siūlomas priemones ir praktikas, – nuo DI apibrėžimo kaip *Kito* iki savo pozicijos konkuruojančioje tarptautinėje aplinkoje nustatymo.

2 lentelė. *Pagrindinių įžvalgų santrauka*

Referentinis objektas	• Pamatinės žmogaus teisės ir demokratinė politinė sistema susiduria su didele rizika ir tampa referentiniais objektais. • Sauga yra orientuota į produktus ir paslaugas kaip garantija, kad DI nepakenks referentiniams objektams. • Dėmesys žmogaus vaidmeniui jam sąveikaujant su technologija atskleidžia, kad daugiausia rūpinamasi tuo, kas ateityje turės kontrolę.
Galimos žalos sąlygos	• Galima žala reiškia įvairias problemas, susijusias su įsibrovimu ir diskriminacija ir sukeliamas DI arba jį netinkamai naudojančių veikėjų. • DI autonomija pateikiama kaip radikali problema, rodanti, kaip turimi DI vaizdiniai skatina imtis veiksmų. • ES saugumo sampratą galima apibendrintai laikyti agentiniu saugumu, orientuotu į žmogaus veiklos apsaugą, apibrėžiamą kaip nuolatinė žmogaus kontrolė bei sprendimų priėmimo galia žmogui sąveikaujant su technologija ir tai suvokiant kaip įtampą tarp Savęs ir Kito.
Valdysenos programa	• Normatyvinių ir institucinių priemonių, kuriomis kuriama daugiasluoksnė ir įvairiapusė valdysenos sistema, derinys. • Politikos priemonės, susijusios su principais, reguliavimu, vertinimais ir institucine ekosistema, skirtos ne tik kontroliuoti, bet ir kreipti DI ateitį, siekiant užtikrinti agentinį saugumą. • ES technokratizuoja valdysenos programą taikydama standartizuotas politikos priemones, remdamasi ekspertų žiniomis ir paskirstydama atsakomybę institucinėms procedūroms, kuriose saugumas tampa įprasta rutina.
Tarptautinis bendradarbiavimas	• ES siekia formuoti tarptautinę darbotvarkę ir ginti taisyklėmis grindžiamą tvarką, dalyvaudama daugiašaliuose formatuose ir plėtodama bendraminčių partnerystę. • Tarptautinė su DI susijusi aplinka laikoma konkurencinga; čia taip pat dalyvauja ES, siekianti įtvirtinti savo vaidmenį tarp kitų galių. • Save ES pozicionuoja kaip tvirtovę – ribotą erdvę, judančią savisaugos, priklausomybės mažinimo ir kitų įtikinimo savo požiūriu į DI kaip būdo stiprinti saugumą link.

Šaltinis: autorius, remiantis atlikta analize.

Šioje lentelėje apibendrintos keturios pagrindinės disertacijos išvados.

Pirma, analizė parodė, kad ES, apibrėždama savo poziciją dėl DI, aktyviai pasitelkia rizikos ir rizika grindžiamo požiūrio sampratas. Pateikta rizikos kategorizacija, grindžiama galimu pavojumi pamatinėms žmogaus teisėms ir demokratinei politinei sistemai, iliustruoja tvarkos mechanizmą: kuo didesnis galimas pavojus, tuo didesnė rizika ir tuo didesnis reguliacinės intervencijos poreikis. Jeigu saugumizavimą palygintume su rizikifikavimu, perėjimas nuo grėsmių ir neatidėliotinos būtinybės prie rizikos ir su DI susijusių vaizdinių saugumą kaip „aukštąją politiką“ perorientuotų prie neapibrėžtumo numatymo ir mažinimo, siekiant pageidaujama linkme nukreipti DI plėtrą ir panaudojimą. Todėl vietoj ypatingų priemonių ES rengia valdysenos programą, grindžiamą saugumo *normalizavimu* ir rutina. Kartu su tokiais klausimais kaip klimato kaita ir kibernetinis saugumas DI prisideda prie saugumo sampratos plėtros – parodo, kad technologijos ir su jomis susijęs nežinojimas keičia grėsmės diskursus. Todėl „rizikifikavimo“ proceso analizė tampa konceptualiai svarbi dekonstruojant ES mąstymą ir politiką, susijusią su DI saugumo apibrėžimu ir formavimu, kur rizika reiškia politinius pasirinkimus bei prioritetus apibūdinant ir sprendžiant saugumo klausimus, nukreiptus į ateitį bei su ja susijusį neapibrėžtumą.

Antra, remiantis ES strateginiu DI diskursu, disertacijoje pasiūlyta *agentinio saugumo* sąvoka. Ji reiškia žmogaus agentiškumo apsaugą, suprantamą kaip gebėjimas išlaikyti kontrolę ir sprendimų priėmimo galią esamų ir būsimų su DI susijusių problemų atžvilgiu. Žmogaus ir technologijos sąveiką laikydama grįsta priešiškumu, o DI traktuodama kaip Kitą, ES pabrėžia būtinybę išlaikyti žmogaus viršenybę technologijų atžvilgiu. Tuomet agentinio saugumo klausimais tampa netgi nekariniai DI naudojimo atvejai, galintys pažeisti, sutrikdyti ar apriboti žmogaus išskirtinumą. Ši sąvoka skiriasi nuo jau vartojamų, tokių kaip nacionalinis, žmogaus ar visuomenės saugumas, ir žyminčių santykius tarp valstybių, valstybių ir visuomenės grupių. Dėmesys žmogaus agentiškumui išplečia saugumo sampratą, nes manoma, kad DI atkreipia dėmesį į svarstymus apie tai, ką reiškia būti žmogumi technologijų atžvilgiu, – tai, kas anksčiau laikyta savaime suprantama prerogatyva. Tad saugumas čia susijęs su siekiu nustatyti ribas ir skirtumus tarp žmonių kaip Savęs ir DI kaip Kito, atsižvelgiant į tai, kad agentiškumas turi likti žmogaus išskirtinumu nepaisant (būsimų) DI plėtros ir naudojimo etapų. Sykiu tokia pozicija liudija antropocentrizmo prieštaringumą: nepaisoma kintančių saugumo iššūkių, kylančių iš žmogaus ir technologijos sąveikos, tačiau atsisakoma svarstyti alternatyvias šios sąveikos formas, veikiančias žmogaus vietą hierarchijoje.

Trečia, technokratizacija reiškia, kad valdysenos programos formavimas laikomas techniniu, o ne politinių svarstymų klausimu. Technokratizacija, grindžiama rizikų sprendimo standartizavimu, ekspertų autoritetu ir atsakomybės paskirstymu administratoriams bei suinteresuotosioms šalims, reiškia pasirinktą strategiją, kuria vadovaudamasi ES pateisina ir struktūruoja savo atsaką į iššūkius, susijusius su agentiniu saugumu. Be to, remiantis esamais kitų politikos sričių modeliais – pavyzdžiui, reguliavimu, vertinimais ir institucijomis – ir kartojant siūlomą techninį DI apibrėžimą pabrėžiamas ekspertinių ir mokslinių žinių aktualumas sufleruoja lūkesčius, kad tos žinios pasiūlys objektyvias ir veiksmingas gaires, rodančias, kaip reaguoti į neapibrėžtumą, susijusį su DI. Todėl ES, atrodo, teikia pirmenybę tikėtinam nešališkumui, o ne atviroms politinėms diskusijoms. Toks mąstymas atitinka rizikifikavimo logiką, kai nuo mobilizacijos ir išgyvenimo klausimų pereinama prie technokratinio valdymo, kur saugumas įgyvendinamas per kasdieninę rutiną. Technokratizacija rodo depolitizacijos prieštaringumą, nes mažina skirtingų pozicijų ir pasiūlymų svarbą priimant sprendimus, o pasiūlytas politikos priemones pristato kaip savaime suprantamas.

Ketvirta, ES pozicionuoja save kaip tvirtovę – apribotą erdvę, apibrėžiančią save pagal nustatytus principus ir taisykles, siekiančią tarptautiniuose DI standartuose įtvirtinti savo požiūrį. Tvirtovė veikia keliais būdais: ji nustato principus visiems veikėjams ES viduje ir norintiesiems *patekti* į ES; tarptautinę aplinką laikydama nepalankiai konkuruojančia, siekia sumažinti priklausomybę nuo tarptautinių galių, sustiprinti technologijų bei jų valdymo kontrolę; tikisi įtikinti kitus laikytis jos požiūrio į DI ir taip sukurti saugesnę erdvę, vertinant pagal ES. Tvirtovė numano ir kitokį požiūrį į skaitmeninį suverenitetą: joje technologijos nėra naudojamos suverenitetui stiprinti, jos pasirodo kaip *Kitas*, verčiantis ES stiprinti savo apsaugą. Konkurencinga tarptautinė aplinka, įskaitant didžiąsias valstybes, didžiąsias technologijų įmones ir autokratinius veikėjus, sustiprina DI kaip kitoniško sampratą, todėl mažesnė priklausomybė ir kontrolė tampa bendromis, nors ir skirtingai motyvuotomis skaitmeninio suvereniteto ir tvirtovės dalimis.

Tokia ES pozicija atskleidžia eurocentrizmo prieštaringumą. Ji liudija ES tendenciją siūlyti savo požiūrį į DI kaip visuotinai taikomą ir priimtiną, neatsižvelgiant į iššūkius, grupes ir pozicijas, nepatenkančias į ES dėmesio centrą – Europos piliečių ir demokratinės politinės sistemos idealus.

Mokslinis reikšmingumas

Ši disertacija siūlo ketveriopą akademinį įnašą. Pirma, joje plėtojamos akademinės diskusijos apie ES formuojamą DI politiką, sutelkiamas dėmesys į saugumo, susijusio su DI, sampratą. Disertacijoje reaguojama į esamus tyrimus, kuriuose pabrėžiama įtampa tarp to, kaip ES civiliniu, nekariniu būdu apibrėžia technologijas, ir militarizuotų praktikų. Peržengiant literatūroje įprastą civilinės ir karinės sričių dichotomiją, disertacijoje siūloma *agentinio saugumo* sąvoka – saugumo matmuo, kurio centre atsiduria antagonistiška žmogaus ir technologijos sąveika, DI suvokiant kaip *Kitą*. Ji sutelkta į žmogaus veikimo apsaugą, suprantamą kaip gebėjimas išlaikyti kontrolę ir sprendimų priėmimo galią, atsižvelgiant į dabartinius ir būsimus su DI susijusius klausimus. Be to, šis tyrimas išplečia su DI susijusio saugumo analizę, dažnai sutelkiamą į karinį kontekstą, ir parodo, kad rūpestis dėl žmogaus ir technologijų sąveikos bei kontrolės peržengia karinę sritį. Taigi *agentinis saugumas* tampa atskaitos tašku, padedančiu suprasti žmogaus veikimo apsaugą įvairiose sociotechninėse aplinkose, ir plečia saugumo studijų diskursą.

Antra, disertacijoje plėtojama technokratizacijos sąvoka, susijusi su DI saugumu, kartu parodoma, kaip siekiama užtikrinti saugumą – latentiškai, ilgalaikiškai ir procedūromis mažinant riziką bei galimą žalą. Disertacijoje parodoma, *kaip* veikia technokratizacija: standartizuojant politikos priemones, pasitikint ekspertų žiniomis ir perduodant atsakomybę administratoriams bei suinteresuotosioms šalims. Taip pat svarstoma, *kokios* yra ES požiūrio į DI kaip saugumo klausimą ypatybės. Ši disertacija prisideda prie saugumo studijų ir mokslo ir technologijų studijų diskusijų, dažnai sutelkiančių dėmesį į tai, kaip į saugumo praktikas yra integruotos technologijos. Disertacija parodo, kad DI laikomas ne tik priemone, bet ir saugumo problemų šaltiniu bei technokratizuoto atsako pateisinimu. Daroma prielaida, jog mokslinės žinios ir ekspertiniai vertinimai gali pasiūlyti aiškumo ir veikimo gaires neapibrėžtomis sąlygomis.

Disertaciijoje pritaikoma ir plėtojama rizikifikavimo prieiga, jau atskleidusi, kad dėmesys valdysenos programoms tiesiogiai siejasi su technokratiniais problemų sprendimais. Disertacijoje teigiama, kad rizikos logika – ne tik tipiškas ir jau įrodytas ES politikos formavimo būdas, bet ir strategija, skirta įtvirtinti saugumo klausimams, sprendžiamiems technokratizuotomis valdysenos programomis. Rizikifikavimo sąvoka taip pat išplečiama: svarbia analitine dalimi tampa tarptautinis bendradarbiavimas. Jis parodo, kad su DI susijusių rizikų tarptautinimas tampa neatskiriama ES

saugumo logikos dalimi ir prisideda prie mąstymo apie technologijas ir rizikas kaip *neturinčias sienų*. Taip siekiama įtvirtinti ES subjektiškumą.

Trečia, disertacija prisideda prie diskusijų apie ES savęs pozicionavimą pasaulyje. Užuot pritaikius jau nusistovėjusias sampratas, pavyzdžiui, ES kaip normatyvinė galia ar technologijos kaip priemonė siekti skaitmeninio suvereniteto, šiame tyrime laikomasi prielaidos, kad ES pozicija formuojama vaizduojant DI kaip *Kitą*, o tai lemia savęs kaip *tvirtovės* apibrėžimą. Skirtingai nuo tyrimų, kuriuose ES subjektiškumas apibūdinamas kaip grindžiamas ne saugumo prioritetais (tokiais kaip rinka ir vertybių skatinimas), čia ES pozicija apibrėžiama kaip atsakas į saugumo problemas, kur saugumui stiprinti pasitelkiamos įvairios strategijos (daugiašalis bendradarbiavimas, įtaka ir konkurencija). Todėl, užuot sutelkus dėmesį į diskusijas apie skaitmeninimo skatinamą pokytį ir požiūrio į jį formulavimą, disertacijoje rodoma, kad ES užima protekcionistinę poziciją DI ir nepalankios tarptautinės aplinkos atžvilgiu, o su DI susijusią valdyseną laiko kova dėl galios.

Galiausiai tyrime nagrinėjamos platesnės diskusijos, veikiančios formuojamą ES DI politiką, grindžiamą dilema „kaip parengti DI taisykles atsižvelgiant į etikos ir žmogaus teisių darbotvarkę, nevaržant inovacijų ir nekenkiant DI technologijų diegimui Europoje“ (Brattberg et al. 2020). Disertacijoje teigiama, kad jau atliktuose ES požiūrio tyrimuose daugiausia dėmesio skiriama ekonominiams ir reguliavimo klausimams bei svarstymams apie ES rinkos konkurencingumą, nepaisant ES formuojamoje DI politikoje įtvirtintos saugumo sampratos. Disertacija įtraukia šią perspektyvą ir parodo ne tik jos aktualumą, bet ir tai, kaip siekis apsaugoti pamatines žmogaus teises bei demokratines vertybes yra pristatomas kaip išskirtinai europietiškas, tačiau kartu ir universaliai taikytinas požiūris, tapęs neatskiriama ir integralia ES formuojamos DI politikos dalimi. Tai skatina kompleksiškiau diskutuoti apie tai, kokie yra ES prioritetai ir kaip jie nusako ne tik ES poziciją, bet ir jos požiūrį į DI, neapsiribojant vien tik bendrosios rinkos ir konkurencingumo logika.

ABOUT THE AUTHOR

Justinas Lingevičius holds a Bachelor's degree in Political Science (2014) and a Master's degree in International Relations and Diplomacy (Magna Cum Laude, 2016) from the Institute of International Relations and Political Science (IIRPS), Vilnius University.

During his doctoral studies he published four peer-reviewed articles and a book review based on his PhD research. He presented at 17 international conferences and workshops, including co-convening an early career scholars' workshop on the geopolitics of AI governance, hosted by the European International Studies Association (EISA). He also participated in the European Consortium for Political Research (ECPR) Summer School of the Standing Group on International Relations and continues to be a member of the ECPR Standing Group on Knowledge Politics and Policies.

Justinas was a Fellow of the Charlemagne Prize Academy (2021–2022, Germany), conducting research on military AI in the European Union. He completed a doctoral internship at the University of Antwerp (Belgium) and taught a Master's course in International Relations Theory at the IIRPS, Vilnius University, in 2024. Since March 2024, he has served as a member of the Early Career Research Development Group of EISA. He currently works as a research assistant on the project "In Search for Recognition: Lithuanian Foreign Policy 2015–2024" at Vilnius University.

In 2023, the Research Council of Lithuania awarded him a scholarship for academic excellence during his doctoral studies.

PUBLICATIONS

Publications based on PhD research:

Lingevičius, Justinas. 2025. "The ethics of artificial intelligence in defence." *International Affairs* 101(4): 1528–1529. https://doi.org/10.1093/ia/iiaf124.

Lingevicius, Justinas. 2025. "Frontiers of digital sovereignty: drones and military security in the Baltic States." *Journal of Contemporary European Studies*, 1–121. https://doi.org/10.1080/14782804.2025.2525142.

Lingevicius, Justinas. 2024. "Transformation, insecurity, and uncontrolled automation: frames of military AI in the EU AI strategic discourse." *Critical Military Studies* 11(2): 175–96. https://doi.org/10.1080/23337486.2024.2387890.

Lingevicius, Justinas. 2023. "Military artificial intelligence as power: consideration for European Union actorness." *Ethics and Information Technology* 25(19). https://doi.org/10.1007/s10676-023-09684-z.

Jakniūnaitė, Dovilė, and Justinas Lingevičius. 2021. "Digital geopolitical competition in the age of artificial intelligence: the visions of the United States, China, and the European Union." *Darbai ir dienos* 76: 75–97. https://doi.org/10.7220/2335-8769.76.5.

Other publications:

Lingevičius, Justinas. 2022. "What These Walls Saw and Heard? Reconstructing the Traces of Oppressive Structures." *Politologija* 106(2): 165–173. https://doi.org/10.15388/Polit.2022.106.5.

Lingevičius, Justinas. 2020. "Identity Discourse within a Geopolitical Crisis: The Case of Lithuania." *POLITIKON: The IAPSS Journal of Political Science* 44: 26–43. https://doi.org/10.22151/politikon.44.2.

Lingevičius, Justinas. 2018. "Being a Small State: Discussion on the Role of Size." *Baltic Journal of Political Science* 7(8): 73–91. https://doi.org/10.15388/BJPS.2018.7-8.5.

Lingevičius, Justinas. 2016. "Kaip kalbėti apie mažas valstybes? Mažumo reikšmių analizė." *Politologija* 2(82): 32–74. https://doi.org/10.15388/Polit.2016.2.10104.

Lingevičius, Justinas. 2015. “Identity Tensions: The Case of Michail Golovatov’s Release.” *Lithuanian Foreign Policy Review* 34: 87–108. doi.org/10.1515/lfpr-2016-0004.

Lingevicius, Justinas. 2015. “Lost in Self-Identification? In Search of NATO’s Identity.” *Politikon: IAPSS Political Science Journal* 27: 103–124. https://doi.org/10.22151/politikon.27.5.

Lingevičius, Justinas. 2015. „Lietuvos tapatybė saugumo ir užsienio politikoje 1991-1994 metais: „grįžimo“ keliai ir įtampos”. *Politologija* 3(79): 127–162. https://doi.org/10.15388/Polit.2015.3.8431.

Lingevičius, Justinas. 2015. “Tapatybinės įtampos: Michailo Golovatovo paleidimo atvejis.” In *Ambicingas dešimtmetis: Lietuvos užsienio politika 2004–2014*, edited by Dovilė Jakniūnaitė, 75–98. Vilnius: Vilniaus universiteto leidykla.

ACKNOWLEDGMENTS

This work would not have been possible without the help, inspiration and encouragement I received from many people over the past five years while writing this thesis.

I am immensely grateful to my supervisor, Prof. Dr. Dovilė Jakniūnaitė, for being both an academic role model and for giving me the intellectual freedom and trust not only during this research process but also throughout earlier stages of my academic journey. Her rational guidance during moments of self-doubt and her drive for excellence motivated me to persevere and continuously improve this thesis through many iterations.

I would also like to thank Prof. Dr. Florian Rabitz and Prof. Dr. Ramūnas Vilpišauskas for reading earlier drafts and providing thoughtful feedback that significantly improved the quality of the work. I am also sincerely grateful to the members of the defence jury for their valuable comments, which I will carry forward in whatever form this research takes in the future.

My gratitude extends to the Institute of International Relations and Political Science, its leadership, and colleagues, where I felt privileged to shape my studies in ways that aligned with my interests and goals. I also appreciate the entire PhD cohort, especially fellow Rasa, Simona, and Neringa, for celebrating small successes and sharing our frustrations, along with academic jokes that made this journey less lonely. I thank my friends for regularly checking in on me. Their interest and encouragement helped me recognize the importance of my work.

This process would have been very different without the opportunity to meet, receive feedback from, and engage with researchers from diverse academic backgrounds at international conferences, workshops, and beyond. These exchanges not only helped shape the direction of my research but also fostered professional connections – and friendships – that made me feel part of a broader academic community. For all of this, I remain truly indebted.

Finally, and most importantly, I dedicate this thesis to my mother and grandmother. As the first graduate in our family, I have never faced doubt in my academic choices – only support, trust, and genuine care. I share this achievement with them as a reflection of my heartfelt gratitude.

NOTES

Vilniaus universiteto leidykla
Saulėtekio al. 9, III rūmai, LT-10222 Vilnius
El. p. info@leidykla.vu.lt, www.leidykla.vu.lt
bookshop.vu.lt, journals.vu.lt
Tiražas 15 egz.