Research Article

Cooperation and Conflict
1–30

DOI: 10.1177/00108367241288073
journals.sagepub.com/home/cac
Sage

Trust in artificial intelligence: Producing ontological security through governmental visions

Stefka Schmid, Bao-Chau Pham and Anna-Katharina Ferl

Abstract

With developments in artificial intelligence (AI) widely framed as security concern in both military and civilian realms, governments have turned their attention to regulating and governing AI. In a study of United States (US), Chinese, and European Union (EU) AI documents, we go beyond instrumental understandings of AI as a technological capability, which serves states' self-interests and the maintenance of their (supra)national security. Our specific interest lies in how AI policies tap into both problem-solving approaches and affective registers to achieve both physical and ontological securities. We find that in governmental visions, AI is perceived as a capability that enhances societal and geopolitical interests while its risks are framed as manageable. This echoes strands within human–computer interaction that draw on human-centered perceptions of technology and assumptions about human–AI relationships of trust. Despite different cultural and institutional settings, the visions of future AI development are shaped by this (shared) understanding of human–AI interaction, offering common ground in the navigation of innovation policies.

Keywords

innovation, ontological security, technology, trust, vision

Introduction

The development of artificial intelligence (AI) has been widely considered a security issue, with governments around the world deeming AI urgent enough to be addressed through policy and regulation (Johnson, 2019). This becomes most obvious when considering governments' ambitions to increasingly develop and utilize AI in the military realm, and specifically, the attention research and development (R&D) of autonomous weapon systems has garnered (Sauer, 2021). However, security is also of concern when

Corresponding author:
Stefka Schmid, Department of Computer Science, Technical University of Darmstadt, Pankratiusstraße 2, 64289 Darmstadt, Germany.
Email: schmid@peasec.tu-darmstadt.de

it comes to civilian applications, for example as part of critical infrastructures or in law enforcement (Fischer and Wenger, 2021). In this context, scholars have been speaking of a "race to regulate" AI (Smuha, 2021). While the United States (US) and China are often considered the main competitors over technological developments of AI capabilities (Pecotic, 2019), the European Union (EU) has made efforts to live up to its (self-appointed) role as a norm entrepreneur, proposing "Trustworthy AI" (European Commission (EC), 2019: 1) and approving the comprehensive AI Act in early 2024. As the implications of AI development and application are inherently uncertain and will only materialize in the future, there is increasing interest in prospective technology assessment (Grunwald, 2018) and the anticipatory nature of policies in preventive arms control (Prem, 2022). Similarly, scholarly work has compared existing AI policies, in particular those put forward by the US, China, and the EU, and how they address security issues surrounding AI (Cath et al., 2018; Roberts et al., 2021a, 2023).

Some scholars (Bächle and Bareis, 2022; Bareis and Katzenbach, 2022; Ferl, 2024), for example, have critically investigated AI policy documents to understand governmental visions of technologies and sociotechnical imaginaries and how these visions shape AI policies and regulation. This includes how these visions function as stabilizers of power relations. Focusing on such visions allows us to unpack fundamental understandings of hopes and fears that are embedded in and enact AI policies. We add to this existing body of work, which predominantly highlights differences in governmental AI visions and policies, by arguing that US, Chinese, and EU AI policies indeed share significant similarities in their AI visions. In our analysis, we lay out how these visions function as a common frame of reference and how they situate and reinforce governmental actors' self-understanding and self-stabilization as security providers in a globally competitive environment. In doing so, our work contributes to the critical and interdisciplinary study of AI policies by drawing on international relations (IR), science and technology studies (STS), and human–computer interaction (HCI).

More concretely, our work proposes an interpretation of what these (supra)national visions of AI produce by considering their relationship with ontological security. Thereby, we rely on insights from HCI as a guide for analyzing policy documents that frequently refer to characteristics relating to usability and human-centered design. With an interest in governmental actors' efforts to produce ontological security, we, therefore, ask:

1. What kind of visions of AI are enacted in governmental AI policies?
2. How do these visions foster ontological security?

Drawing on ontological security and accounting for dimensions that go beyond instrumental understandings of AI as merely technological tools enables us to shift the focus beyond strategic capabilities or the implications of (lethal) autonomous weapons systems, which remain one of the more prominently researched areas in security studies and AI. This is in line with more recent scholarship that considers the legal, normative, and ethical implications around AI in international security and that has been critical of techno-deterministic accounts of (security) policy and technology (in particular AI) (Leese and Hoijtink, 2019). Adding to these studies, we aim to broaden the research agenda to explore visions of not only military but also civilian applications of AI in

policy that are relevant to the identity–security nexus. We argue that governmental visions regarding civilian AI applications similarly reflect how state actors navigate through and produce the reality of the international system. Building on our analysis, we argue that governmental AI visions also serve the purposes of producing ontological security, in that they entail and aggregate collective hopes (and anticipatory skepticism) as well as emotional accounts of the society–technology relationship.

Further contributing to the burgeoning literature on AI and international security, which has explored underlying narratives, utopias, or imaginaries that give shape to AI policies, we unpack how visions of AI in governmental policies draw heavily on insights from the research field of HCI, in particular the ideas of human-centered and problem-solving design. This goes along with an increasing focus on HCI and warfare (Leese, 2019). We show empirically how these visions refer to HCI as a constitutive knowledge base, which allows for the production of ontological security in AI governmental policies.

Our analysis of AI policy documents published by China, the EU, and the US suggests that governmental policies by these three competitors share similarities in viewing AI as a capability, which can be harnessed for national purposes. Albeit actor-specific differences in their AI visions, China, the US, and the EU share (1) an instrumental understanding of technology as a problem-solving tool which serves proclaimed self-interests. Following from this rationalist perspective, AI is considered a "technological fix" (Katzenbach, 2021: 1) which allows for the delivery of different societal promises while (2) anticipated risks are framed as manageable for (human) actors. In this regard, anticipatory risk management that addresses issues of reliability, explainability, and human control helps to activate "basic trust" and mediate potential fears. Drawing on insights from HCI, we show how these problem-solving-oriented perspectives of human–AI interaction, that underlie human-centered design, influence AI policies. Taken together with the view of AI as a useful tool, governmental visions (3) reflect an understanding of AI as a military capability.

In the following, we map out AI and international security as our research focus (see "AI and international security" section). We then outline ontological security (see "Ontological security" subsection), emotions, and trust as important components to the role of science in the production of (ontological) security (see "Emotions, trust in science, and security" subsection) and concepts in HCI (see "Human–AI interaction" subsection) as our conceptual framing. This is followed by an overview of our case selection, data collection, and analysis (see "Methodology" section). Subsequently, we present the findings from our analysis of AI visions in Chinese, US, and EU AI policy documents (see "Visions of AI" section 5). Based on the results, we discuss implications for the identity–security nexus (see "Discussion" section) and offer a conclusion (see "Conclusion" section).

AI and international security

AI as a general-purpose technology plays an ever more important role in both civilian and military spheres. AI is understood to solve problems independently (from human interaction). Especially AI applications based on machine learning (ML) algorithms have

become widely used in areas where vast amounts of data and information have to be assessed and filtered for humans to make time-sensitive decisions. Current developments in AI applications rely on ML that can focus on specific tasks with a narrow scope, including object recognition, or drawing inferences from large data sets to make predictions (Reuter-Oppermann and Buxmann, 2022). In the field of international security, AI has mostly been discussed as enabling autonomous functions in weapon systems. Autonomy in weapon systems is broadly understood as conducting tasks of the mission cycle (such as target identification, selection, or the application of force) without any human intervention (Boulanin et al., 2021). Nevertheless, human interaction takes place in the context of training ML models. Improved algorithmic performance, broad access to digitized data as well as availability of opensource toolkits and libraries, has contributed to the expansion and rapid development of novel AI-ML applications (Reuter-Oppermann and Buxmann, 2022).

While different technology readiness levels due to different application contexts can be noted (Schmid, 2023)—military applications necessitate a high degree of accuracy and require specific training conditions considering low latency, inaccessibility of locations—this mainly applies to AI implemented to conduct tasks of the mission cycle. Besides more recent developments in the field of autonomous weapon systems, such as increased duration of flight or swarms (Longpre et al., 2022), AI is regularly used to conduct more organizational tasks, for example, in logistics or training (Grand-Clément, 2023).

Given these technological developments, AI has increasingly become a security concern. In strategic studies, the military use of AI applications is often perceived as enabling the military to "fight at machine speed" (Horowitz, 2019: 4) or to enhance battlefield awareness through better-informed human decision-making and improved command and control (Altmann and Sauer, 2017; Sauer, 2022). These anticipated benefits of AI applications in the military realm drive much of the current technological development and procurement decisions. Others have critically examined the "moral implications of this integration of nonhuman logics and systems into existing processes of military violence" (Renic and Schwarz, 2023: 322) and have highlighted fundamental questions around the dehumanization of warfare (Asaro, 2012; Schwarz, 2021). These ethical and moral concerns also raise questions about who bears accountability and responsibility for the use of AI systems if things go wrong (Garcia, 2016). This raises legal questions about the application of International Humanitarian Law (IHL) in relation to military uses of AI (Brehm, 2017; Garcia, 2024; Heyns, 2016). Taken together, AI has increasingly become a topic of discussion in international security with the debates around fundamental technical, ethical, legal, and security-related questions remaining largely open and contested (Bode et al., 2024).

This article builds on scholarship in IR and STS-informed security studies that understand AI as more than merely a hard power capability and bounded technology but instead considers its sociotechnical aspects. Through this, we (1) refocus the investigation not only onto military uses of AI but also civilian applications, and (2) consider AI policies and governance as constitutive of international security.

While much discussion in the literature on AI and international security focuses on the military use of AI, in this article we take into account that developments in the civilian field play an equally significant role in advancing AI. In contrast to earlier technological

developments, where military technologies spun-off into civil technologies, AI and ML applications nowadays are mostly developed by private tech companies that then diffuse into the military realm (Fischer, 2022; Verbruggen, 2019). Therefore, we cannot only discuss developments in the military domain as a separate field but need to consider the dual-use[1] quality of AI, including spill-overs from the civilian to the military sphere (Schmid et al., 2022). While the military depends on civilian innovation in the field of AI, countries engage with it differently. The US, for instance, has a longer tradition of civil–military relations where the Pentagon solicits tech companies to develop specific projects (e.g., Project Maven, see Suchman (2020)). In China, the official policy of "military–civil fusion" leads to closer ties between the civilian AI sector and the military (Carrozza et al., 2022: 10). The EU is much more stratified in its civil–military relation—not least since EU defense research initiatives have only recently taken off (Martins and Mawdsley, 2021). The path-breaking EU AI Act of 2024 for example explicitly "does not apply to AI systems that are exclusively for military, defense or national security purposes, regardless of the type of entity carrying out those activities" (EC, 2023), reflecting the intergovernmental nature of EU defense policy while leaving the regulation of AI dual-use applications to member states (Carrozza et al., 2022: 29).

Many studies on AI and IR have compared and unpacked AI policies (Bareis and Katzenbach, 2022). Complementing existing work that has focused on autonomous weapon systems and questions of identity (Bode et al., 2023; McCarthy, 2021; Nadibaidze, 2022), our work approaches AI innovation policies as an arena in which ontological (in) securities are produced with respect to both civilian and military contexts of application. As noted by Pham and Davies (2024) in their study of EU policy documents, infrastructures such as AI policies are not static but constitute affective processes that allow for identity-formulating ontological experiments.

We build on and contribute to this research by engaging explicitly with insights from HCI (see "Human–AI interaction" subsection) as a so-far overlooked scholarship in the literature at the intersection of AI and international security. HCI allows us to take into account human-centered design and usability characteristics and how they inform how governmental actors make sense of AI and security. We refer to HCI to inform our theoretical argument that these shape understandings and policies of AI which play into the production and stabilization of ontological security (see "Ontological security" subsection).

Conceptual framework

Ontological security

In drawing on the concept of ontological security, we approach AI policies and the visions of the technology therein as a subject that is not only relevant to physical security but also to states' security of the self as well as stabilization efforts regarding their self-perception and positioning in the international system (Mitzen, 2006; Zarakol, 2017). We seek to show how state identities are performed, (re)constructed, and (re)negotiated in light of ongoing AI developments and in parallel with security concerns (Lupovici, 2022). Ontological security refers to *security-as-being* (Kinnvall and Mitzen, 2017) as opposed to *security-as-survival* (Rumelili, 2015). At the core, ontological security refers to the

state in which an actor feels secure in themselves. A stable self is, for instance, held up through biographical continuity, the maintained performance of routines as well as "basic trust," reflecting "optimism that things generally work out in the end" (Browning and Joenniemi, 2016: 35). Although routinizing may help to counter "anxieties" and dreadfulness, a state of ontological security does allow for change in behavior and identity, as the agent is able to adapt and show reflexivity (Peoples and Vaughan-Williams, 2020). For the individual, this means that they can "rely on a social normality" in which their everyday life is predictable, their relations can be trusted and their position in life is perceived as stable (Croft, 2012). One example where ontological security is anchored in is the idea of home, where both a physical and emotional permanence and continuity are thought to be provided (Kinnvall, 2004). Different studies have since focused on ontological security at the national level and considered communities' everyday (in)securities (Browning, 2018). Ontological security is here maintained through mundane but performative acts that reproduce a sense of national belonging. The focus shifts away from exogenous physical threats, such as death, bodily harm, or tanks, to more "psychological ramifications of security discourse" (deRaismes Combes, 2017: 128). At the same time, the production of ontological security necessitates a stable self which often means that other identities need to be inferiorized and securitized (Bilgin, 2010; Habib, 2018; Noble, 2005). While the concept of ontological security has been used for analysis of state behavior at critical junctures (Ejdus, 2018) or under the conditions of an anarchic international system (Peoples and Vaughan-Williams, 2020), it has so far rarely been used to research state responses to emerging digital technologies (Lupovici, 2022). Most recently, Nadibaidze (2024) has investigated Russia's narratives around AI technologies through the lenses of ontological security and status-seeking literature in IR, thereby making visible how states mobilize visions and narratives around technologies to "deal with the constant uncertainty about recognition of their self-perceived identity" (p. 1). We build on the idea that states seek the maintenance of their ontological security through the activation of "basic trust" and routinizing behavior, which gives us insights into "how we construct and perform certain identities" (deRaismes Combes, 2017: 128). It allows us to ask why, on an international level, states introduce policy, such as AI strategies, in the first place. Lupovici (2022: 2) suggested that states introduce programs and adopt strategies for ambiguous and elusive topics such as digital sovereignty not with concrete policy aims in mind—which in themselves often times seem ineffective or difficult to implement—but as a way of seeking ontological security. Thus, ontological security, in the context of AI governance, may not only be targeted at the public audience, whose trust or technology acceptance is desired. It may also be self-interested, self-reflected security sought by governmental actors in a highly competitive and geopoliticized global environment. In this context, we find that institutional trust in science is foundational to governmental actors in their efforts to produce ontological security. Both affective and rationalizing behaviors play into the generation of trust in AI that is essential for technology adoption.

Emotions, trust in science, and security

IR scholarship has noted an increased interest in "emotion and hot cognition" (Kertzer and Tingley, 2018), which is reflected in key debates around emotions and the *affective*

turn in IR and the social sciences (Åhäll, 2018). In an effort to measure emotions and to connect to works of ontological security, scholars (Bilgin, 2010; Kertzer and Tingley, 2018) emphasized positively associated terms to endorse affirming emotions and thus a "community's way of life" (Koschut, 2018) or sense of self and belonging. Consequently, negatively ascribed emotions such as anger or anxiety have been identified to reflect outsider positions, low status, or unstable identities (Kertzer and Tingley, 2018). Bridging both rationalist and emotion-oriented approaches, scholars focusing on public/institutional trust and identity have noted the entanglement of instrumental and affective logics that come into play in the creation of trust. Instead of building on a rationalist approach that understands trust as "encapsulated interest" (Hardin, 1991, as cited in Ruokonen, 2013) and emphasizes the rational assessment of others' "goodwill and reliability" (Bilgic et al., 2019), a synthesized perspective notes the subjective and emotional nature of trust as an "affective attitude" (Jones, 1996, as cited in Ruokonen, 2013). This is exemplified by understanding trust of one actor having an "optimistic attitude about the goodwill of B [. . .] [and] B's competence when it comes to B's expected behavior [. . .]" (Ruokonen, 2013). Institutions also receive impersonal trust, including trust in their rules and norms (Muringani and Noll, 2021). This also applies to science as an institution and its technological artifacts (Hartley, 2021; Townley and Garfield, 2013). With regard to the latter, conceptualizations of interpersonal trust have been extended due to digital technologies mimicking human behavior (Muringani and Noll, 2021). Critical of overtrust in AI, scholars (Grodzinsky et al., 2011; Taddeo, 2017, as cited in Schmid et al., 2022) have differentiated between technology and human actors and argued for relationships of reliance instead of trust among human and nonhuman agents. Others (Nissenbaum, 2001; Winfield and Jirotka, 2018, as cited in Schmid et al., 2022) have emphasized the importance of institutions and regulations in building up trust that is essential for technology acceptance and adoption.

The enhancing character of trust is also noted by Bilgic (2014), who emphasizes the transformative effect on actors' self-interests and identities as security dilemmas can be transcended.[2] Thus, generalized trust in science and technology may form a mechanism through which science and technology produce ontological security. However, physical insecurity can be increased through particularized distrust. Our work follows others who have analytically distinguished between a scientific and political sphere in the context of professionalized politics in a "reflexive modernity" (Beck, 2016; Lidskog and Sundqvist, 2015). While STS-inspired works (Elbe and Buckland-Merrett, 2019; Fischer and Wenger, 2021) have focused on the co-production of (applied) science and security, we are particularly interested in how HCI as a scientific knowledge base helps to produce "common sense" understandings of human interaction with trustworthy AI and thereby stabilizes governmental Selves.

Human–AI interaction

To gain a better understanding of how ontological security and future visions are performed through AI policies, our analysis draws on HCI. This proves suitable as political institutions themselves have identified computer scientific research, such as explainable AI or human–AI interaction research, as crucial to the successful implementation of

"Trustworthy AI" (EC, 2019) and its integration into "warfighting and defense sectors" (Vorm, 2020). At the same time, scholarly work has aimed at introducing "high-level" AI design principles into academic debates and translating these principles into technical design requirements (Mäntymäki et al., 2023). Assuming design practices and underlying ideas (Suchman, 2007, cited in Leese, 2019) and values (Friedman et al., 2017) to structure (prospective) human–technology interaction, it is useful to approach AI policies through a HCI lens. Recently, HCI has started to connect with the AI tech community (Li et al., 2020; Liao et al., 2019; Xu et al., 2021) and included a focus on quality characteristics increasing usability and deriving design implications from empirical findings. Generally, technology is considered usable when it can be used "effectively, efficiently and with satisfaction" (ISO, 2018: 1) (i.e., ensuring a good user experience (UX)) in relation to specified goals and the context of use (Bevan et al., 2016).

Characteristics such as explainability have been identified as crucial to an AI system's usability and trustworthiness (Xu et al., 2021). This is deemed important as current military AI applications are frequently regarded as unreliable (Vorm, 2020). Precision, accuracy and robustness are design requirements which aim to ensure reliability (Reuter-Oppermann and Buxmann, 2022). Often used interchangeably, explainability and interpretability are characteristics that offer users explanations to understand AI-enabled decision-making. Interpretability "refers to a passive characteristic of a model" and "the level at which a given model makes sense for a human observer" (Barredo Arrieta et al., 2020: 6, as cited in Alexander, 2021). Thus, interpretability of a system deals with how decisions are understood by human users and relies on transparency. Conversely, explainability is considered the "active characteristic of a model [. . .] with the intent of clarifying or detailing its internal functions" (Barredo Arrieta et al., 2020: 7, as cited in Alexander, 2021). While explainable AI is often criticized for its "black box" nature (Rudin, 2019: 1), it is still positively associated with trustworthy AI systems across academic and political institutions (Bach et al., 2022). Trustworthiness of technology is associated with "ability, benevolence, and integrity" (Baughan et al., 2023) and thus combines what is defined as "cognitive trust" ("knowledge-driven") and "affective trust" ("motivated by emotion") in "behavioral trust" that "refers to the willingness to take action based on the judgment or information provided by the AI system" (Chen and Sundar, 2023). Crucial for technology adoption, more recently, human–AI interaction has particularly necessitated studies on trust (Bansal et al., 2023). Trust between humans and technology is also particularized in terms of which concrete implementations are seen as trustworthy as well as in terms of accessibility. Corbett and Le Dantec (2021) focused on the latter in their interest in "trust work" and identify, together with other third wave, critical studies of HCI (Erete et al., 2023; Frauenberger, 2016), individualistic design as perpetuating power structures. Identified design characteristics are proclaimed to not only contribute to better usability, but also create a more enjoyable UX. Focusing on the latter, HCI's emphasis on human-centered technology design and user satisfaction becomes even more apparent. A human-centered perspective is usually prevalent in HCI (Bach et al., 2022), implying a hierarchical human–technology relationship. A shift in human perception on technology from a tool to a collaborative agent can also be observed (Saßmannshausen et al., 2023). And yet, HCI is fundamentally human-centered in that it is interested in designing technological solutions that can be

purposeful to humans. While human-centered design is guided by a normative claim of developing "legitimate" technology (Dourish, 2019), it also may sugar-coat governmental policies (Frey and Schaupp, 2020). With HCI positioning itself by noting design to be "human-made," it reflects science in a "reflexive modernity," that is responsible (and in control of) creating and managing risks (Beck, 2016). In light of identifiable and much-debated risks of AI, the discipline offers a more positive and self-confident perspective of human control and agency, stating: "to ensure that human–AI collaborations do more good than harm, it is vital that we understand, measure, and shape human–AI trust and reliance" (Bansal et al., 2023). In contrast to apocalyptic depictions of future AI, this systematized, empirically oriented perspective offers ontological security.

Drawing on the discipline of HCI, we are able to show how problem-solving-oriented perspectives of human–AI interaction, that underlie human-centered design, influence AI policies and visions. The policy documents contribute to visions of AI that on one hand evoke hopes for a secure future, while on the other mitigate fears and risks through the upholding of design requirements and performance tests of AI models, including accuracy, predictability, and explainability of decisions.

Methodology

Case selection

R&D in the AI sector is deeply entangled, especially through business collaborations between US and Chinese tech firms. Competition and collaboration co-exist at the same time, and the dual-use nature of the technology has consequences for the military competition between the two states. China engages very directly in the civil–military integration of the dual-use nature of AI technology through their "military–civil fusion" strategy (Kania, 2019: 1). This policy has contributed to US perceptions of China as a strategic competitor in the field of AI innovation and in turn led to a renewed focus of the US government on military AI developments. In contrast, the EU traditionally tends more toward rule and norm making. In the case of AI, they have positioned themselves as a major player in regulating technology and developing ethical guidelines (Carrozza et al., 2022). Considering other countries are also heavily investing in AI innovation, the AI strategies put forward by the chosen actors here seem (1) most consequential on a multilateral scale, therefore most impactful for international security concerns, (2) will inevitably inform the strategies of their allies and opponents, and (3) for pragmatic reasons, are available in English.

In analyzing these policy documents, we are interested in the sociotechnical dimension through the visions they perform, meaning we are interested in what such policies say and do, that is, allocating institutional resources (Bareis and Katzenbach, 2022). This is in line with scholarship that engages with the role of discourses and visions in institutionalizing and materializing (digital) technology and local practices (Ferl, 2024; Mager and Katzenbach, 2021). Visions are at the same time representative of the hopes and fears of governmental actors as they are performative (and at times prescriptive) in making a certain sociotechnical pathway intelligible and desirable (see Jasanoff and Kim, 2015). Analysis of governmental visions also connects to existing studies of technology assessment that have considered visions in science or industries (Grunwald, 2018).

Data collection

The starting point for our data collection was the OECD overview of AI policies and a list of initiatives sorted according to countries (https://www.oecd.ai/dashboards?selectedTab=countries). As a first step, we collected all documents published between 2015 and 2023 by executive government offices and relevant ministries (e.g., the Chinese Ministry of Foreign Affairs (MFA), Ministry of Education (MoE), and Ministry of Industry and Information Technology (MIIT). In the 2010s, developments in generative AI helped to revive AI R&D which were followed by increased regulatory efforts (Héder, 2020). While specific AI policies and guidelines, such as the US "Preparing for the Future of Artificial Intelligence" (NSTC, 2016) were published in 2016 (Cath et al., 2018), we followed existing studies in considering the Chinese 10-year plan "Made in China 2025" (Roberts et al., 2021b) and thus chose 2015 as the starting point of our sample. We then reduced the text corpus, cross-checking with other scholarly work and think tank publications to identify the most influential documents for each actor's AI strategy while striving for diversity in publication date and authorship. Except for texts by ministries of defense or education, we did not include publications specific to policy fields such as environmental issues, public health, or public services. Our sample consists of 31 documents (see Table 1), with some Chinese documents translated into English by think tanks such as the Center for Security and Emerging Technology (CSET). Although regional initiatives such as local governments supporting "national championships" (Roberts et al., 2021b) or local regulations aiming at Silicon Valley inventions (Washington Post, 2024) are relevant to state–market relationships that contribute to AI R&D, we refrained from extending our sample in this regard and focused on (supra-)state-level policies. Most policies target the broad application of AI either focusing primarily on civilian industries or on both civilian and defense sectors, but we also included policies directly relating to AI R&D for military purposes of each governmental actor.

Data analysis

For the qualitative analysis, we developed categories in a deductive-inductive approach and re-evaluated after initial coding. Building on debates of human–computer and human–AI interaction (Xu et al., 2021) and STS-adjacent literature on IR and technology (Bareis and Katzenbach, 2022), the coding scheme ultimately consisted of two main categories that aimed at understandings of *human–AI interaction* and *AI visions* (see Figures 1 and 2). The material was coded by two researchers. Each researcher coded an equal share of selected publications in MAXQDA Plus 2022. Subsequently, the authors discussed the initial impressions and ambiguities. The content analysis in a qualitative coding software helped to structure the research process and systematize our findings.

Visions of AI

AI as securing people's essential needs

One key aspect that transpires in our analysis is the way in which AI is framed as a potential solution to securing people's essential needs. In this regard, AI is presented as a

Table 1. Selected policy documents of the People's Republic of China, the European Union, and the United States of America. Publications reflect governmental positions, published by executive.

	Actor	Publication	Type	Year
China	State Council (translated by CSET)	Made in China 2025	broad	2015
	PRC Ministry of Science and Technology (MOST; translated by CSET)	The "13th Five-Year" Special Plan for S&T Military–Civil Fusion Development	broad	2016
	State Council (translated by New America)	A Next Generation Artificial Intelligence Development Plan	broad	2017
	Ministry of Industry and Information Technology (MIIT)	Three-Year Action Plan for Promoting Development of a New Generation Artificial Intelligence Industry (2018–2020)	broad	2017
	Ministry of Education	Notice of the Ministry of Education on Issuing the Artificial Intelligence Innovation Action Plan for Institutions of Higher Education	broad	2018
	China Institute for Science and Technology Policy at Tsinghua University	China AI Development Report 2018	broad	2018
	The Big Data Security Standards Special Working Group (translated by CSET)	Artificial Intelligence Security Standardization White Paper (2019 Edition)	broad	2019
	China Academy for Information and Communications Technology (CAICT; translated by DigiChina)	Translation: Artificial Intelligence Security White Paper [excerpts]	broad	2019
	Ministry of Foreign Affairs	Position Paper of the People's Republic of China on Regulating Military Applications of Artificial Intelligence	military	2021
	The China Academy of Information and Communications Technology (CAICT; translated by CSET)	Artificial Intelligence White Paper	broad	2022
	Ministry of Foreign Affairs	Position Paper of the People's Republic of China on Strengthening Ethical Governance of Artificial Intelligence	broad	2022
EU	European Commission	Communication from the Commission to the European Parliament, the European Council, the European Economic and Social Committee and the Committee of the Regions	broad	2018
	European Commission	Liability for Artificial Intelligence and other emerging digital technologies	broad	2019
	High-level Expert Group on Artificial Intelligence (HLEG)	Ethics Guidelines for Trustworthy AI	broad	2019

(Continued)

Table 1. (Continued)

	Actor	Publication	Type	Year
	European Union Agency for Cybersecurity (ENISA)	AI Cybersecurity Challenges Threat Landscape for Artificial Intelligence	broad	2020
	European Commission	White Paper On Artificial Intelligence—A European approach to excellence and trust	broad	2020
	European Commission	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	broad	2021
	European Commission	Communication: Fostering a European approach to artificial intelligence	broad	2021
	European Commission	Coordinated Plan on Artificial Intelligence 2021 Review	broad	2021
	European Defence Agency	Joint quest for future defense applications	military	2021
The United States	Executive Office of the President	Preparing for the Future of Artificial Intelligence	broad	2016
	Department of Defense	Summary of the 2018 Department of Defense Artificial Intelligence Strategy Harnessing AI to Advance Our Security and Prosperity	military	2018
	Executive Office of the President	Summary of the 2018 White House Summit on Artificial Intelligence for American Industry	broad	2018
	The President	Executive Order 13859 of February 11, 2019 Maintaining American Leadership in Artificial Intelligence	broad	2019
	Executive Office of the President	The National Artificial Intelligence Research and Development Strategic Plan: 2019 Update	broad	2019
	The White House (Office of Science and Technology Policy)	American Artificial Intelligence Initiative: Year One Annual Report	broad	2020
	Executive Office of the President	Guidance for Regulation of Artificial Intelligence Applications	broad	2020
	Department of Defense	DoD AI Education Strategy	military	2020
	National Security Commission on Artificial Intelligence	Final Report	broad	2021
	Department of Defense	Implementing Responsible Artificial Intelligence in the Department of Defense	military	2021
	Department of Defense	DoD Directive 3000.09 Autonomy in Weapon Systems	military	2023

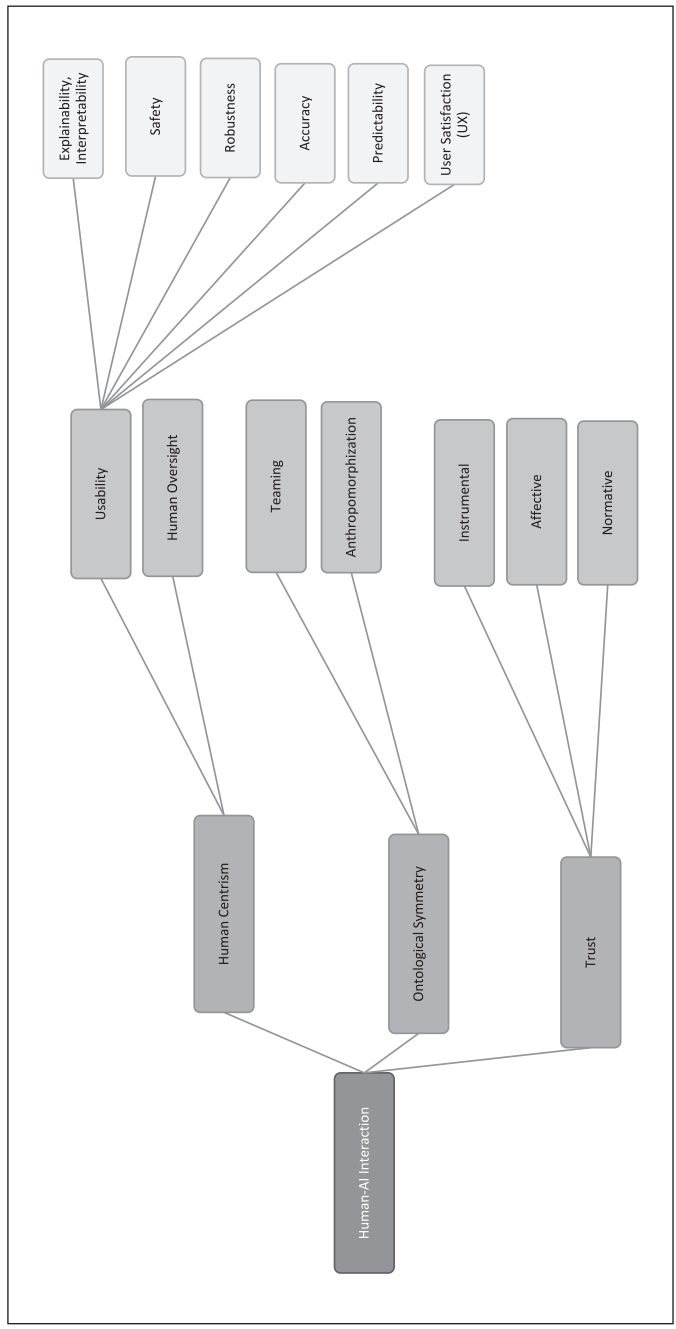

Figure 1. First category of the coding scheme, including codes and sub-codes that aim to capture understandings of human–AI interaction.

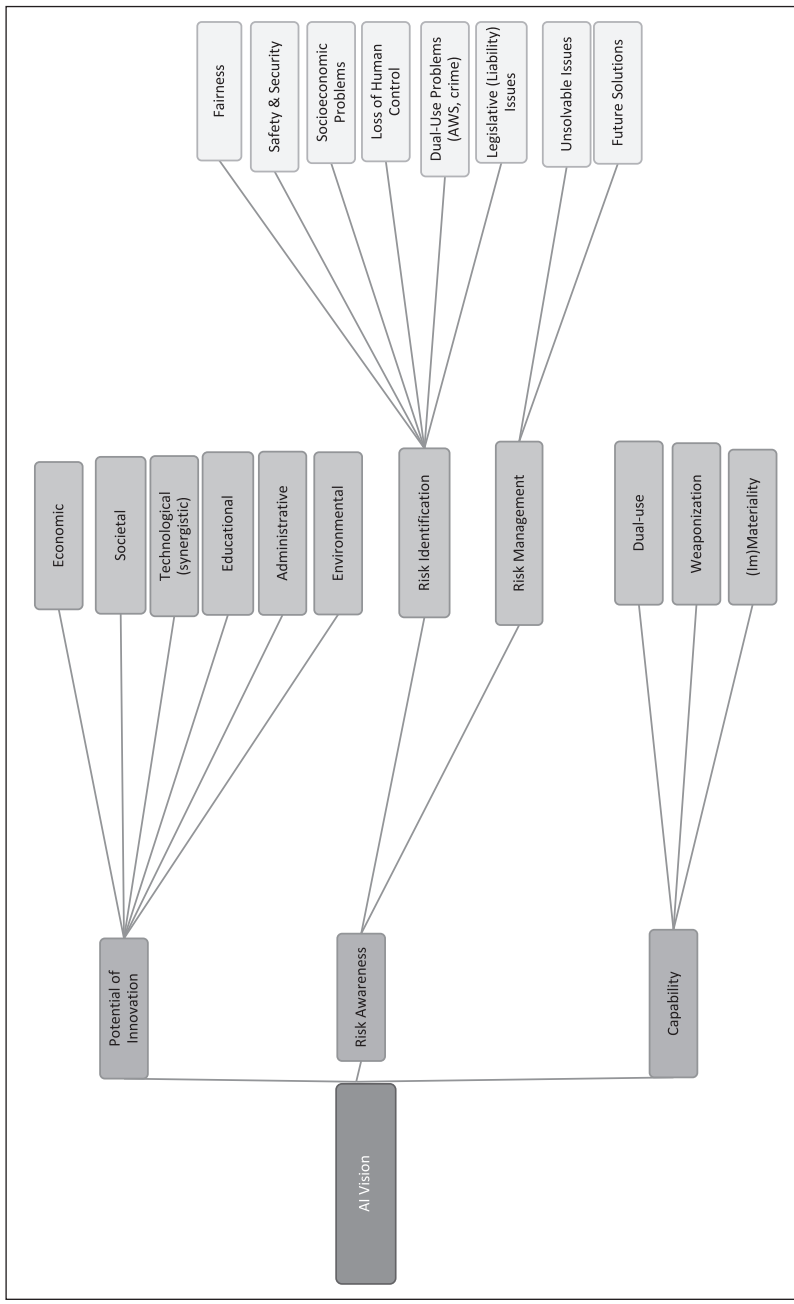

Figure 2. Second category of the coding scheme, including codes and sub-codes that aim to capture visions of AI R&D.

technological fix-all, or as "the main" (Digichina, 2017: 8) and "important" (CSET, 2022: 7) driving force for a whole host of solutions to (global) problems (EC, 2021a). In fact, in the documents we find that AI is envisioned not only as fulfilling societal needs but also as instrumental to improving people's quality of life. AI is repeatedly referred to in the context of people's everyday work lives. The goal for AI would be to allow people to "live a richer and more colorful life, as they are *liberated* from manual labor and even conventional intellectual labor, devote more energy to creative activities for fuller development of humankind and human society" (China Institute for Science and Technology Policy at Tsinghua University (CISTP), 2018: 96, emph. added). Within all this, so the documents argue, is it crucial that humans are seen at the center. AI and AI-enabled technologies are not to displace, replace, or harm humans but should support humans. One of the policy documents, for example, states that "[t]he ultimate purpose of developing AI is not to replace humans but make humans smarter" (CISTP, 2018: 97). Rather than taking over human tasks, AI is better imagined as a "super assistant" (CISTP, 2018: 62), being in a hierarchical relationship with human agents to whom AI offers its service. While socioeconomic transformation requires adaptation by humans, they are imagined to be prepared for it through education (CISTP, 2018: 97; EC, 2021a: 10). Our analysis is in line with the idea of the human–AI relationship prevalent in HCI scholarship as outlined above, where technology is seen as something that can be utilized for the purpose of a specific goal, and serves the function of making things more "effective" and "efficient." Rather than replacing human labor, AI is intended to make work less "repetitive or dangerous" (CISTP, 2018: 97) and instead more fulfilling. This is envisioned by developing AI that is "transparent" and provides "reasonable guarantees on the safety, security, robustness, and resiliency" (The White House Office of Science and Technology Policy (WHST), 2020: 6). While the requirement of transparency reflects the increasing relevance of UXs in human–AI interaction and thus a focus on satisfaction of user needs, the latter characteristics emphasize awareness of the importance of flawless and uninterrupted performance of AI systems. Fundamental safety and security requirements, which make systems reliable and guarantees integrity, are rationalized as a necessary condition for the production of "trustworthy AI" (EC, 2019: 1). Another argument given as to why humans should be placed at the center is that AI is wrongly mistrusted or being met with disproportionate skepticism. Without the wider public acceptance and uptake of AI, the logic of the documents goes, it is difficult to achieve and fully take advantage of the promises that AI holds. As the WHST (2020), for example, posits,

> [n]ot using AI technologies because of perceived or potential harms, however, could be just as problematic, depriving individuals — or the Nation — of the significant benefits that AI technologies could bring. Fully realizing the potential of AI, therefore, requires public trust and confidence in these technologies. (p. 19)

Were AI not to be adopted, a loss and, in fact, additional costs to society as a whole, ranging from compromised national security to reduced standards of living, could be incurred (Department of Defense (DoD), 2019: 5). Here, we can see both emotional and rational dimensions of trust playing into each other. The message that is transported in the documents is that there are no alternatives to putting trust in governmental

institutions, which offer both physical and emotional permanences and thus a home to citizens. While governmental policies reflect an effort to offer ontological security to the public, they produce ontological security for the governmental actors themselves in the competitive global landscape. Through personalizing "the Nation" as a representation of society, ontological security of its citizens contributes to the stabilization of governmental selves. While Chinese, EU, and US documents may reflect different self-understandings of state–society relationships, they all target citizens' everyday lives as an issue relevant to their AI approaches "on the global stage" (EC, 2021a: 9). Not only might individuals' "basic trust" be activated through the assurance of institutions and "human-centric, sustainable, secure, inclusive, accessible and trustworthy" (EC, 2021a: 9) AI development, but governmental actors might also relate to these issues when conducting their routine work in formulating policies directed at both domestic and international audiences. Comparing AI development to historical cases such as electrical power, the US hypes AI as "not even comparable" but, at the same time, refers to Thomas Edison's "humility" in his self-description of his discoveries. Today, it is the National Security Commission on Artificial Intelligence (NSCAI) (2021) that humbly acknowledges how much remains to be discovered "but still know[s] enough about AI today" (p. 7). Referring to national history, governments are able to represent a successful, naturally appearing continuance such as US conduct of R&D without any "affective dissonance" (Wood and Cox, 2021). Similarly reflecting authentic values, in the production of Chinese ontological security "amid rapid AI development," "*harmonious* and positive" human–AI interaction serves as an important point of reference and can best be promoted through benefiting "all regions, all industries and all groups *equally*" (CISTP, 2018: 99; emph. added). Both "com[ing] up with methods" and imaginations of a "beautiful future" (CISTP, 2018: 99) are important to governmental actors' visions. Technology governance can be understood as an important policy field from which these actors can draw self-awareness. This is possible through both instrumental perspectives on learning and applying knowledge as well as emotional reflections on technological development.

AI as a manageable issue

Understanding and framing AI as a potential solution to securing people's needs also means that AI as a technology needs to be manageable so that associated risks can be mitigated. While AI is predominantly seen as a beneficial technology by all three governmental actors, when its limitations are assessed, they are often depicted as problems that can easily be solved, often through technology itself. AI as a manageable issue means that risks can be identified *a priori*. Their risk management is feasible with the establishment of specific design requirements and techniques to solve "technical hurdles that require more R&D" (WHST, 2020: 12). Typical of a risk society (Beck, 2016), reflexive political and scientific institutions are continuously needed to tackle future challenges in R&D, building on "basic trust." Especially in EU policy papers, we find a high level of risk awareness, with most documents transparent about the problems that could come with AI. These risks range from data protection and privacy concerns (EC, 2020: 11), technical robustness and explainability of AI (European Union Agency for Cybersecurity (ENISA), 2020: 27), to "manipulation, deception, herding and conditioning, all of which

may threaten individual autonomy" (EC, 2019: 16). In contrast to China or the US, the EU (2021c) highlights the need for human control and oversight over AI systems as a risk mitigation measure so that "high-risk AI" can be constituted as trustworthy:

> High-risk AI systems should be designed and developed in such a way that natural persons can oversee their functioning. For this purpose, appropriate human oversight measures should be identified by the provider of the system before its placing on the market or putting into service. (p. 31)

While all three actors mention the ethical implications of AI, they deal with them in different ways. The EU is very upfront about ethical issues, while the US does not specifically develop a framework for dealing with potential ethical issues arising out of this technology. In China's case, developing ethical AI seems to be explicitly linked to managing the technical risks of AI and connecting ethical questions to AI security (MFA, 2022: 8). Although levels of risk awareness or the focus on different kinds of risks differ between China, the US, and the EU, they all three emphasize trustworthy AI as a risk mitigation strategy in their policies, albeit not explicitly labeling it this way. However, understandings of how to achieve trustworthiness differ. For instance, China focuses on technical flaws of AI systems that have an impact on IT security, such as "fragility and vulnerability" and "complexity and uncertainty" (CSET, 2022: 22) that need to be solved to guarantee the predictability of AI. R&D is necessary to overcome these potential risks, which again reflects an instrumental understanding of the human–AI relationship or "cognitive trust" (Chen and Sundar, 2023) based on expectations regarding AI to be reliable. The US, also relying on training and testing (DoD, 2021: 2), understands trustworthiness to arise out of the "development of explainability mechanisms" (WHST, 2020: 12) for human operators to understand (and thus be less frustrated by) AI decisions. Here, the affective dimension of trust (Townley and Garfield, 2013) becomes apparent, focusing on AI's "benevolence" (Baughan et al., 2023) and explainability as a crucial usability characteristic. The EU (2020) also associates trustworthiness with the need for "technically robust and accurate" (p. 21) AI systems and highlights explainability as one way to mitigate risk as well as the central role of human agency for countering potential technical issues by enabling users "to make informed autonomous decisions" (EC, 2019: 18). This includes the ability to override or challenge the system's decisions which in turn requires knowledge and training. With its risk-awareness approach, the EU may be more oriented toward guaranteeing secure and reliable human–AI interactions and reflect a more rationalized view on trustworthy AI. Yet, the EU focus' on human autonomy and risk aversion is idealistic as well and matches the trajectory of EU identity-formation processes built on regulation and protection of civil liberties. Despite the evaluation of risks appearing non-affirming or negative, particularized trust with regard to manageable technology is enabled and creates a common and differentiated ground for identity among EU member states. There are only few instances, in which awareness of the inherent problematic uncertainty of AI decision-making is revealed and, in these cases, explainable AI (which is believed to be realizable in the future) is posited as the solution to the problem:

> An explanation as to why a model has generated a particular output or decision (and what combination of input factors contributed to that) is not always possible. These cases are referred

> to as "*black box*" *algorithms* and *require special attention*. In those circumstances, other explicability measures (e.g., traceability, auditability, and transparent communication on system capabilities) *may be required, provided* that the system as a whole respects fundamental rights. (EC, 2019: 13; emph. added)

Reproducing a vision of the manageability of AI can become inherently problematic when the focus is predominantly on "measuring and evaluating AI technologies through standards and benchmarks" (WHST, 2020: 12) because it is perceived to be controllable. Unlike visions of radical uncertainty (Katzenstein, 2022), framing AI in terms of risks also makes it manageable and finally resolvable. Reflexive statements on the limitations and conditionality of explainable AI, as posed by the EU, indicate that actors balance between activating "basic trust" through suggestion of (well-established) operative design characteristics and reflecting limited knowledge, which could put them in a less ontologically secure position.

AI as a (military) capability

Alongside representing AI as a tool to secure people's needs as well as a manageable issue, our analysis supports the argument that governmental actors envision AI as a capability, which can be used for geopolitical purposes. The increasing geopoliticization of technology is apparent in the visions of AI that are laid out in the documents, while the indeterminacy of technological development opens up a wide scope of imaginable action. For example, the US points out that:

> AI is also the quintessential "dual-use" technology. The ability of a machine to perceive, evaluate, and act more quickly and accurately than a human represents a competitive advantage in any field—civilian or military. AI technologies will be a source of enormous power for the companies and countries that harness them. (NSCAI, 2021: 9)

Notions of AI as a useful technology to achieve relative advantages and position national interests in the global arena are also made by China, emphasizing the necessity "to build China's first-mover advantage" (Digichina, 2017: 2). Furthermore, the European Defence Agency (EDA) refers to " the most important areas of AI for Europe's strategic autonomy" (EDA, 2021: 34). We also find that the beneficial capabilities of AI to one's geopolitical standing are regularly associated with a human-centric perspective of technology. For example, the EU (2021b) focuses on "creating EU global leadership in human-centric AI" (p. 3) and notes that "[c]ountries around the world are choosing to use AI as a means of technical advancement due to its utility" (EC, 2021a: 5). Fitting with the US Silicon Valley tradition of creative, iterative, and disruptive R& D, US policy papers depart from other publications in their focus on human–machine teaming. Introducing this concept into the policy arena reflects an idea of anthropomorphized AI (especially compared with more traditional EU and Chinese approaches to technology as a tool). Complementing human work through teaming may take place "as a side effect of AI development, or a system might be developed specifically" (NSCAI, 2021: 9) and is seen as highly important:

> Mastering human-AI collaboration and teaming is a foundational element for future application of AI. Synergy between humans and AI holds the promise of a whole greater than the sum of its parts. (NSCAI, 2021: 34)

US readiness to adopt a more systemic (and less individualized) approach to human–AI interaction shows that ontological security does not necessarily build on the absence of change. Comfortable in its positioning as a leading country in innovation, the US formulates a closer, more collaborative human–AI relationship to which progress in AI adoption is tied. While this may indicate a shift to a more interactional and less instrumental view, the US vision does not deviate from other governmental actors' policies in their relationship with science. Ultimately, theories are to be developed for application, indicating instrumental impersonal trust toward science, by which "[f]or better insights, intelligence agencies will need to develop innovative approaches to human–machine teaming that use AI" (NSCAI, 2021: 10). Chinese policy also reflects a functional understanding of science, relating to "making breakthroughs in human-centric human–machine fusion theories (. . .) to support AI-driven industrial development" (CISTP, 2018: 70). However, in contrast to the EU, these two actors explicitly focus on the innovative character of interactional approaches. In doing so, they do not need to sacrifice ontological security. Not only does AI allow for the generation of hopes regarding the prosperity in different spheres of life (see "AI as securing people's essential needs" subsection), it is envisioned to be particularly useful—and a much-needed capacity—to military operations (NSCAI, 2021: 11). While the US is eager to translate "AI research into military capabilities," China pushes for a more synergistic approach to AI R&D investments. The promotion of "military–civilian sharing and joint use" of resources to achieve "deep military–civil integration" (Digichina, 2017: 13) allows China to "seize the initiative in the new round of international science and technology competition" (Digichina, 2017: 2). With China and the US most explicitly proclaiming the importance of AI to military leadership, they both stabilize their positions as competing superpowers, albeit in different ways. Facing a highly competitive and (in terms of security) challenging international arena (Digichina, 2017: 13; EC, 2021a: 4), governmental visions differ in their claims—with the EU (2021b: 2) in particular positioning itself as a norm entrepreneur and China pushing for synergistic technological superiority (Digichina, 2017: 2). However, visions by all three governmental actors label AI as a dual-use technology (Digichina, 2017: 13; NSCAI, 2021: 7), emphasizing potential technology use in military application contexts. Formulating a vision of military applications of AI as a necessity to maintain or secure military advantages may not lead to positive effects by the public. Still, the governmental actors themselves routinely conduct their institutional work of guaranteeing international security as well as regulating the labor force market through the generation of incentives and training for AI personnel. Offering a material and countable representation of AI, workforce is seen as "capability" (National Science and Technology Council (NSTC), 2019: 4) which serves as a common point of reference in governmental security policies. At the same time, already normalized views on the military environment as another economic industry (that now, with AI adoption, potentially imply less dangers to humans) help governmental actors to produce ontological security in their relationship with domestic audiences. The intersecting modes of

conducting security and economic politics have been noted by the governmental actors themselves, who point out that AI is strategically important to the "crossroads of geopolitics, commercial stakes and security concerns" (EC, 2021a: 4).

Discussion

Reproducing identities in a geopoliticized playing field

The reproduction of actor identities as providers of security in various spheres of life (see "AI as securing people's essential needs" subsection) contributes to "basic trust" through routinization (Mitzen, 2006). Still, in this dynamic policy field, reflexive behavior is possible (Browning and Joenniemi, 2016) and may also include shifts in the (formerly) hierarchical understanding of the human–AI relationship. Adapting to more symmetric understandings of human–AI relationships may even be perceived as necessary by governmental actors (e.g., the US as outlined above) in their efforts to stabilize their competitive position and self-understanding as providers of security. This rings true particular in emergent political-economic environments (Lupovici, 2022), such as AI governance where the understanding of what subjectivity or governmental actors' selfhoods entail (Browning and Joenniemi, 2016) has only begun to be negotiated. Our analysis has highlighted what the three actors share in their AI visions and has also allowed us to assess differences in their identities as instrumental to their positioning in the global economic competition. Deviating from each other, *othering* becomes possible (Habib, 2018; Noble, 2005). In contrast to perceptions of AI as a global common or flawed technology, the presented AI visions offer ontological security to governmental actors in that they allow them to position themselves as regulating actors. This regulation takes place both with regard to international security and national prosperity as well as in the governmental actors' interactions with transnational (industrial and scientific) actors and international organizations. As noted by Bilgic (2014), trust can lead to structural change in identities. However, while AI as the referent object to the production of trust is perceived as a capability in economic and geopolitical settings, there is no substantial change in perceiving other governments as potential threat actors or competitors.

Trustworthy AI visions based on science

As we have argued and analytically shown above, governmental visions of AI, that refer to HCI as a constitutive knowledge base, allow for the production of ontological security in AI governmental policies. This becomes apparent across many of the policy documents, where design characteristics, which are aimed at increasing usability and security, are repeatedly presented as requirements to realize effective and trustworthy human–AI interaction. Thus, the policy documents contribute to visions of AI that, on one hand, evoke hopes for a secure future, while, on the other hand, mitigate fears and risks through the upholding of design requirements and performance tests of AI models, including accuracy, predictability, and explainability of decisions. Integrating security-by-design approaches and human-centered approaches into policies helps to build instrumental trust and plays into the affective dimension of (basic) trust in institutions and technology.

Policy documents allow both institutional actors and public readership to feel at "home" (Kinnvall, 2004) when it comes to AI governance. As a design-oriented discipline, HCI offers a positivist and problem-solving perspective. This also assumes interaction design to imply different socio-technological affordances which make the realization of behavioral patterns more or less probable (Frauenberger, 2016) and should be considered for user-centered design (Abe, 2021). Apart from critical scholarship[3] endeavors (Erete et al., 2023), the discipline carries a less "skeptical" perspective on technology, relying on the human ability to steer the design process and analysis based on mid-range, solution-oriented frameworks. This fosters implicit notions of trust in technology. However, considering the opacity of AI, it has become necessary to investigate the human–AI relationship to formulate design requirements for trustworthy AI (Xu et al., 2021). Our analysis of AI policy documents has shown how emphasizing a functional understanding of science as a provider of applicable knowledge reflects the important role of traditional HCI (Dourish, 2019) for sociopolitical risk management (see "AI as a manageable issue" subsection).

STS-inspired works of IR have focused on "processes of hybridization" in epistemic communities (Lidskog and Sundqvist, 2015) or crisis situations, such as epidemics, causing disruption and tighter coupling of academia and political institutions (Elbe and Buckland-Merrett, 2019). In contrast, societal discourses on AI and military technology adoption have indirectly and broadly diffused into and influenced scientific discourses of HCI or the definition of which empirical problems should be solved by the information systems community. Interested in how ontological security is pursued through governmental policies, further research can complement our work by focusing on other channels of communication and knowledge transfer, such as military-funded research projects, and how (in)security constitutes science and technology. Here, visions guided by perceived insecurity (regarding future challenges) may be identifiable, although one may also assume problem-solving-oriented research, that is "second wave" HCI, that reflects an ontologically secure positioning, to share an optimistic vision of trustworthy AI.

Multiple (in)securities

As noted by Mitzen (2006), "routines that perpetuate physical insecurity can provide ontological security" (p. 14). This makes it difficult to envision alternative scenarios instead of security dilemmas, in which AI is perceived as a dual-use technology (Lupovici, 2021) or a military capability (see "AI as a (military) capability" subsection). While ontological security might be successfully produced, physical insecurities appear in the context of a competitive environment. Although governmental policies reassure audiences that any identified risk can or will be solved, the diverse list of risks identified by different actors makes it harder to provide ontological security to domestic audiences. Anxiety is produced by EU risk assessment, defining a variety of potential threats and harmful ways of AI use, complemented by options to deal with them (Brinker et al., 2023). The performativity of AI policies becomes apparent, with risk assessment communications shifting from framing technology adoption as an issue of "security-as-survival" (Rumelili, 2015: 2) to "security-as-being" (Kinnvall and Mitzen, 2017: 2). In comparison, Chinese documents, and their relatively small focus on risks as well as less

emotionally laden perspective on technological "disruption," might offer the easiest access to envisioning a harmonious future in which everyday life goes by smoothly. Still, more research is needed on the identity–(in)security nexus and how visions of "black box" and explainable AI (Rudin, 2019: 1) play into this relationship.

Conclusion

AI policies are shaped by and at the same time enact visions of the technology they govern. Offering a perspective on the production of ontological security in this context that is often labeled disruptive and as defined by different regional AI policies, our study sheds light on how governmental actors continue to present themselves as providers of security and regulators of technology. As our analysis shows, differences in visions relate to governmental actors' identities formed by ideals and innovation cultures but also (self-perception of) competitive leverage. This is indicated by a relatively dynamic understanding of human–AI teaming, which allows for change. Furthermore, state-science relationships are also embedded differently in domestic societies, necessitating varying investments of "trust work" (Corbett and Le Dantec, 2021) regarding diverse combinations of affective and instrumental notions of trustworthy AI.

Overall, in Chinese, US, and EU AI policy, visions are characterized by instrumental perceptions of AI as a multi-beneficial tool that can be applied to bring societal prosperity. Here, the affective dimension of trust is also enforced through reliance on human-centered design that ensures security and a satisfactory UX. Following from the instrumental understanding of AI as a technology that can be controlled, it is also framed as a military capability. Such commonalities in AI visions are backed by references to human-centered design. Understanding HCI as relevant knowledge base contributing to AI visions, we are able to identify human-centered design and usability as common points of reference across policy documents. Thus, actors seek ontological security by navigating innovation policies in a competitive and dynamic environment; at the same time, this may lead to a less physically secure world.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This contribution was supported by the German Federal Ministry of Education and Research (BMBF) as part of TraCe "Regional Research Center Transformations of Political Violence'" (01UG2203E). The authors would like to thank the anonymous reviewers for their feedback as well as Josefine Süpke and Désirée Hoppe for their assistance in the research process.

ORCID iD

Stefka Schmid https://orcid.org/0000-0003-0948-8893

Notes

1. Dual-use technologies can be understood to be used in both civilian and military application contexts, as technologies that can also be part of a weapon system (Forge, 2010) or as technologies that can be used for both beneficial and malicious purposes (Brundage et al., 2018).
2. Contrasting positive depictions, Wille and Martill (2023) has focused on the *loss* of trust in IR and noted the change in actors' perceptions of future possibilities of cooperation and conflict.
3. Here, we follow prior work (Cox, 1981) that has established the difference between traditional (problem-solving-oriented) and critical science in IR, with the latter acknowledging contradictions and the societal embeddedness of science.

References

Abe N (2021) Human–machine interaction and design methods. In: Elliott A (ed.) *The Routledge Social Science Handbook of AI*. Milton Park: Routledge,pp. 138–154.

Åhäll L (2018) Affect as methodology: Feminism and the politics of emotion. *International Political Sociology* 12(1): 36–52.

Alexander F (2021) *Workshop report for next-gen AI for proliferation detection: Accelerating the development and use of explainability methods to design AI systems suitable for nonproliferation mission applications*. Report. Washington, DC: Brookhaven National Laboratory & US Department of Energy.

Altmann J and Sauer F (2017) Autonomous weapon systems and strategic stability. *Survival* 59(5): 117–142.

Asaro P (2012) On banning autonomous weapon systems: Human rights, automation, and the dehumanization of lethal decision-making. *International Review of the Red Cross* 94(886): 687–709.

Bach TA, Khan A, Hallock H, et al. (2022) A systematic literature review of user trust in AI-enabled systems: An HCI perspective. *International Journal of Human–Computer Interaction* 40(5): 1–16.

Bächle T and Bareis J (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: 20.

Bansal G, Smith-Renner AM, Buçinca Z, et al. (2023) Workshop on trust and reliance in AI-human teams (trait). In: *CHI EA' 23: Extended abstracts of the 20 CHI conference on human factors in computing systems*, New Orleans, LA, 29 April–5 May, pp. 1–6. New York: ACM.

Bareis J and Katzenbach C (2022) Talking AI into being: The narratives and imaginaries of national AI strategies and their performative politics. *Science, Technology, & Human Values* 47(5): 855–881.

Barredo Arrieta A, Díaz-Rodríguez N, Del Ser J, et al (2020) Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion* 58: 82–115.

Baughan A, Wang X, Liu A, et al. (2023) A mixed-methods approach to understanding user trust after voice assistant failures. In: *CHI' 23: Proceedings of the 20 CHI conference on human factors in computing systems*, Hamburg, 23–28 April, pp. 1–16. New York: ACM.

Beck U (2016) *Risikogesellschaft: Auf Dem Weg in Eine Andere Moderne*. Berlin: Suhrkamp Verlag.

Bevan N, Carter J, Earthy J, et al. (2016) New ISO standards for usability, usability reports and usability measures. In: Kurosu M (ed.) *Human-Computer Interaction. Theory, Design, Development and Practice*. Cham: Springer,pp. 268–278.

Bilgic A (2014) Trust in world politics: Converting "identity" into a source of security through trust-learning. *Australian Journal of International Affairs* 68(1): 36–51.

Bilgic A, Gjørv GH and Wilcock C (2019) Trust, distrust, and security: An untrustworthy immigrant in a trusting community. *Political Psychology* 40(6): 1283–1296.

Bilgin P (2010) Identity/security. In: Burgess JP (ed.) *The Routledge Handbook of New Security Studies*. Milton Park: Routledge,pp. 81–89.

Bode I, Huelss H, Nadibaidze A, et al. (2023) Prospects for the global governance of autonomous weapons: Comparing Chinese, Russian, and US practices. *Ethics and Information Technology* 25(1): 5.

Bode I, Huelss H, Nadibaidze A, et al. (2024) Algorithmic warfare: Taking stock of a research programme. *Global Society* 38(1): 1–23.

Boulanin V, Goussac N and Bruun L (2021) Autonomous weapon systems and international humanitarian law: Identifying limits and the required type and degree of human–machine interaction. Available at: https://www.sipri.org/publications/2021/policy-reports/autonomous-weapon-systems-and-international-humanitarian-law-identifying-limits-and-required-type (accessed 23 May 2024).

Brehm M (2017) *Defending the Boundary: Constraints and Requirements on the Use of Autonomous Weapon Systems Under International Humanitarian and Human Rights Law*. Geneva: Geneva Academy Briefing.

Brinker N, Skalt R and Pleil H (2023) Objective assessment of reasonable machines? Role and limitations of risk management in the European AI regulation efforts. Available at: https://doi.org/10.13140/RG.2.2.15528.34567 (accessed 16 October 2024).

Browning CS (2018) Brexit, existential anxiety and ontological (in)security. *European Security* 27(3): 336–355.

Browning CS and Joenniemi P (2016) Ontological security, self-articulation and the securitization of identity. *Cooperation and Conflict* 52(1): 31–47.

Brundage M, Avin S, Clark J, et al. (2018) The malicious use of artificial intelligence: Forecasting, prevention, and mitigation. Available at: https://arxiv.org/pdf/1802.07228

Carrozza I, Marsh N and Reichberg GM (2022) Dual-use AI technology in China, the US and the EU: Strategic implications for the balance of power. Report, PRIO paper. Oslo: PRIO.

Cath C, Wachter S, Mittelstadt B, et al. (2018) Artificial intelligence and the "good society": The US, EU, and UK approach. *Science and Engineering Ethics* 24: 505–528.

Center for Security and Emerging Technology (CSET) (2022) Translation artificial intelligence white paper (2022). Available at: https://cset.georgetown.edu/publication/artificial-intelligence-white-paper-2022/ (accessed 15 August 2023).

Chen C and Sundar SS (2023) Is this AI trained on credible data? The effects of labeling quality and performance bias on user trust. In: *CHI' 23: Proceedings of the 20 CHI conference on human factors in computing systems*, Hamburg, 23–28 April, pp. 1–11. New York: ACM.

China Institute for Science and Technology Policy at Tsinghua University (CISTP) (2018) China AI development report 2018. Available at: https://indianstrategicknowledgeonline.com/web/China_AI_development_report_2018.pdf (accessed 15 August 2023).

Corbett E and Le Dantec C (2021) Designing civic technology with trust. In: *CHI' 21: Proceedings of the 20 CHI conference on human factors in computing systems*, Yokohama, Japan, 8–13 May, pp. 1–17. New York: ACM.

Cox RW (1981) Social forces, states and world orders: Beyond international relations theory. *Millennium* 10(2): 126–155.

Croft S (2012) *Securitizing Islam: Identity and the Search for Security*. Cambridge: Cambridge University Press.

Department of Defense (DoD) (2019) Summary of the 2018 Department of Defense Artificial Intelligence Strategy: Harnessing AI to Advance Our Security and Prosperity. Available

at: https://media.defense.gov/2019/Feb/12/2002088963/-1/-1/1/SUMMARY-OF-DOD-AI-STRATEGY.PDF (accessed 15 April 2022).

Department of Defense (DoD) (2021) Implementing responsible artificial intelligence in the department of defense. Available at: https://media.defense.gov/2021/May/27/2002730593/-1/-1/0/IMPLEMENTING-RESPONSIBLE-ARTIFICIAL-INTELLIGENCE-IN-THE-DEPARTMENT-OF-DEFENSE.PDF (accessed 15 August 2023).

deRaismes Combes M (2017) Encountering the stranger: Ontological security and the Boston marathon bombing. *Cooperation and Conflict* 52(1): 126–143.

Digichina (2017) Full translation: China's "new generation artificial intelligence development plan" (2017). Available at: https://digichina.stanford.edu/work/full-translation-chinas-new-generation-artificial-intelligence-development-plan-2017/ (accessed 15 August 2023).

Dourish P (2019) User experience as legitimacy trap. *Interactions* 26(6): 46–49.

Ejdus F (2018) Critical situations, fundamental questions and ontological insecurity in world politics. *Journal of International Relations and Development* 21(4): 883–908.

Elbe S and Buckland-Merrett G (2019) Entangled security: Science, co-production, and intra-active insecurity. *European Journal of International Security* 4(2): 123–141.

Erete S, Rankin Y, Thomas J, et al. (2023) A method to the madness: Applying an intersectional analysis of structural oppression and power in HCI and design. *ACM Transactions on Computer-Human Interaction* 30(2): 1–45.

European Commission (EC) (2019) Ethics guidelines for trustworthy AI. Available at: https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1 (accessed 15 April 2022).

European Commission (EC) (2020) White paper: On artificial intelligence: A European approach to excellence and trust. Available at: https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en (accessed 15 April 2022).

European Commission (EC) (2021a) 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: Fostering a European approach to artificial intelligence. Available at: https://eur-lex.europa.eu/resource.html?uri=cellar:01ff45fa-a375-11eb-9585-01aa75ed71a1.0001.02/DOC_1\&format=PDF (accessed 27 May 2022).

European Commission (EC) (2021b) Coordinated plan on Artificial Intelligence 2021 review. Available at: https://digital-strategy.ec.europa.eu/en/library/coordinated-plan-artificial-intelligence-2021-review (accessed 15 August 2023).

European Commission (EC) (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. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX\%3A52021PC0206 (accessed 15 April 2022).

European Commission (EC) (2023) Artificial intelligence—questions and answers. Available at: https://ec.europa.eu/commission/presscorner/detail/en/qanda_21_1683 (accessed 14 May 2024)

European Defence Agency (EDA) (2021) Artificial intelligence: Joint quest for future defence applications. Available at: https://eda.europa.eu/news-and-events/news/2020/08/25/artificial-intelligence-joint-quest-for-future-defence-applications (accessed 15 August 2023).

European Union Agency for Cybersecurity (ENISA) (2020) AI cybersecurity challenges: Threat landscape for artificial intelligence. Available at: https://www.enisa.europa.eu/publications/artificial-intelligence-cybersecurity-challenges (accessed 15 April 2022).

Ferl AK (2024) Imagining meaningful human control: Autonomous weapons and the (de-)legitimisation of future warfare. *Global Society* 38(1): 139–155.

Fischer SC (2022) Military AI applications: A cross-country comparison of emerging capabilities. In: Reinhold T and Schörnig N (eds) *Armament, Arms Control and Artificial Intelligence: The Janus-Faced Nature of Machine Learning in the Military Realm*. Cham: Springer,pp. 39–56.

Fischer SC and Wenger A (2021) Artificial intelligence, forward-looking governance and the future of security. *Swiss Political Science Review* 27(1): 170–179.

Forge J (2010) A note on the definition of "dual use." *Science and Engineering Ethics* 16: 111–118.

Frauenberger C (2016) Critical realist HCI. In: *CHI EA' 16: Proceedings of the 20 CHI conference extended abstracts on human factors in computing systems*, San Jose, CA, 7–12 May, pp. 341–351. New York: ACM.

Frey P and Schaupp S (2020) Futures of digital industry: Techno-managerial or techno-political utopia? *Behemoth-A Journal on Civilisation* 13(1): 98–108.

Friedman B, Hendry DG and Borning A (2017) A survey of value sensitive design methods. *Delft: Now Foundations and Trends*. Available at: https://www.nowpublishers.com/article/Details/HCI-015

Garcia D (2016) Future arms, technologies, and international law: Preventive security governance. *European Journal of International Security* 1: 94–111.

Garcia D (2024) Algorithms and decision-making in military artificial intelligence. *Global Society* 38: 24–33.

Grand-Clément S (2023) *Artificial intelligence beyond weapons: Application and impact of AI in the military domain*. Report. Geneva: UNIDIR.

Grodzinsky FS, Miller KW and Wolf MJ (2011) Developing artificial agents worthy of trust: "would you buy a used car from this artificial agent?" *Ethics and Information Technology* 13: 17–27.

Grunwald A (2018) *Technology Assessment in Practice and Theory*. Milton Park: Routledge.

Habib MAR (2018)Hegel on identity and difference. In: Habib R (ed.) *Hegel and the Foundations of Literary Theory*. Cambridge: Cambridge University Press,pp. 72–86.

Hardin R (1991) Trusting persons, trusting institutions. In: Zeckhauser R (ed.) *Strategy and Choice*. Cambridge, MA: MIT Press,pp. 185–209.

Hartley K (2021) Public trust and political legitimacy in the smart city: A reckoning for technocracy. *Science, Technology, & Human Values* 46(6): 1286–1315.

Héder M (2020) A criticism of AI ethics guidelines. *Információs Társadalom* 20: 57.

Heyns C (2016) Autonomous weapons systems: Living a dignified life and dying a dignified death. In: Bhuta N, Liu HY and Beck S (eds) *Autonomous Weapons Systems. Law, Ethics, Policy*. Cambridge: Cambridge University Press,pp. 3–20.

Horowitz MC (2019) When speed kills: Lethal autonomous weapon systems, deterrence and stability. *Journal of Strategic Studies* 42(6): 764–788.

ISO (2018) ISO 9241-11:2018: Ergonomics of human-system interaction — part 11: Usability: Definitions and concepts. Available at: https://www.iso.org/standard/63500.html

Jasanoff S and Kim SH (2015) *Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power*. Chicago, IL: University of Chicago Press.

Johnson J (2019) Artificial intelligence & future warfare: Implications for international security. *Defense & Security Analysis* 35: 1–23.

Jones K (1996) Trust as an affective attitude. *Ethics* 107(1): 4–25.

Kania EB (2019) In military-civil fusion, China is learning lessons from the United States and starting to innovate. Available at: https://www.cnas.org/publications/commentary/in-military-civil-fusion-china-is-learning-lessons-from-the-united-states-and-starting-to-innovate (accessed 18 July 2023).

Katzenbach C (2021) "AI will fix this"–The technical, discursive, and political turn to AI in governing communication. *Big Data & Society* 8(2): 20539517211046182.

Katzenstein PJ (2022) Worldviews in world politics. In: Katzenstein PJ (ed.) *Uncertainty and Its Discontents. Worldviews in World Politics*. Cambridge: Cambridge University Press,pp. 1–69.

Kertzer JD and Tingley D (2018) Political psychology in international relations: Beyond the paradigms. *Annual Review of Political Science* 21(1): 319–339.

Kinnvall C (2004) Globalization and religious nationalism: Self, identity, and the search for ontological security. *Political Psychology* 25(5): 741–767.

Kinnvall C and Mitzen J (2017) An introduction to the special issue: Ontological securities in world politics. *Cooperation and Conflict* 52(1): 3–11.

Koschut S (2018) Appropriately upset? a methodological framework for tracing the emotion norms of the transatlantic security community. *Politics and Governance* 6: 125.

Leese M (2019) Configuring warfare: Automation, control, agency. In: Hoijtink M and Leese M (eds) *Technology and Agency in International Relations*. Milton Park: Routledge,pp. 42–65.

Leese M and Hoijtink M (2019) How (not) to talk about technology: International relations and the question of agency. In: Hoijtink M and Leese M (eds) *Technology and Agency in International Relations*. Milton Park: Routledge,pp. 1–19.

Li Y, Kumar R, Lasecki WS, et al (2020) Artificial intelligence for HCI: A modern approach. In: *CHI EA' 20: Extended abstracts of the CHI conference on human factors in computing systems*, Honolulu, HI, 25–30 April, pp. 1–8. New York: ACM.

Liao QV, Wang YC, Bickmore T, et al. (2019) Human-agent communication: Connecting research and development in HCI and AI. In: *CSCW' 19: Conference companion publication of the 20 on computer supported cooperative work and social computing*, Austin, TX, 9–13 November, pp. 122–126. New York: ACM.

Lidskog R and Sundqvist G (2015) When does science matter? International relations meets science and technology studies. *Global Environmental Politics* 15(1): 1–20.

Longpre S, Storm M and Shah R (2022) Lethal autonomous weapons systems & artificial intelligence: Trends, challenges, and policies. *MIT Science Policy Review* 3(1): 47–56.

Lupovici A (2021) The dual-use security dilemma and the social construction of insecurity. *Contemporary Security Policy* 42(3): 257–285.

Lupovici A (2022) Ontological security, cyber technology, and states' responses. *European Journal of International Relations* 29: 153–178.

Mager A and Katzenbach C (2021) Future imaginaries in the making and governing of digital technology: Multiple, contested, commodified. *New Media & Society* 23(2): 223–236.

Mäntymäki M, Minkkinen M, Zimmer MP, et al. (2023) Designing an AI governance framework: From research-based premises to meta-requirements. In: *Thirty-first European Conference on Information Systems (ECIS 2023)*, pp. 1–18. Atlanta, GA: Association for Information Systems.

Martins BO and Mawdsley J (2021) Sociotechnical imaginaries of EU defence: The past and the future in the European defence fund. *Journal of Common Market Studies* 59(6): 1458–1474.

McCarthy DR (2021) Imagining the security of innovation: Technological innovation, national security, and the American way of life. *Critical Studies on Security* 9(3): 196–211.

Ministry of Foreign Affairs (MFA) (2022) Position paper of the People's Republic of China on strengthening ethical governance of artificial intelligence (AI). Available at: https://www.fmprc.gov.cn/eng/wjdt_665385/wjzcs/202211/t20221117_10976730.html (accessed 15 August 2023).

Mitzen J (2006) Ontological security in world politics: State identity and the security dilemma. *European Journal of International Relations* 12(3): 341–370.

Muringani J and Noll J (2021) Societal security and trust in digital societies: A socio-technical perspective. In: *14th CMI international conference—critical ICT infrastructures and platforms*, Copenhagen, 25–26 November, pp. 1–7. New York: IEEE.

Nadibaidze A (2022) Great power identity in Russia's position on autonomous weapons systems. *Contemporary Security Policy* 43(3): 407–435.

Nadibaidze A (2024) Technology in the quest for status: The Russian leadership's artificial intelligence narrative. *Journal of International Relations and Development* 27(2): 117–142.

National Science and Technology Council (NSTC) (2016) Preparing for the future of artificial intelligence. Available at: https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf (accessed 15 April 2022).

National Science and Technology Council (NSTC) (2019) The national artificial intelligence research and development strategic plan: 2019 update. Available at: https://www.nitrd.gov/pubs/National-AI-RD-Strategy-2019.pdf (accessed 15 April 2022).

National Security Commission on Artificial Intelligence (NSCAI) (2021) Final report national security commission on artificial intelligence. Available at: https://www.nscai.gov/wp-content/uploads/2021/03/Full-Report-Digital-1.pdf (accessed 15 April 2022).

Nissenbaum H (2001) Securing trust online: Wisdom or oxymoron. *Boston University Law Review* 81(3): 635–664.

Noble G (2005) The discomfort of strangers: Racism, incivility and ontological security in a relaxed and comfortable nation. *Journal of Intercultural Studies* 26(1–2): 107–120.

Pecotic A (2019) Whoever predicts the future will win the AI arms race. Available at: https://foreignpolicy.com/2019/03/05/whoever-predicts-the-future-correctly-will-win-the-ai-arms-race-russia-china (accessed 26 January 2023).

Peoples C and Vaughan-Williams N (2020) *Critical Security Studies: An Introduction*. Milton Park: Routledge.

Pham BC and Davies S (2024) Policy as infrastructure: Enacting artificial intelligence and making Europe. In: Klimburg-Witjes N and Trauttmansdorff P (eds) *Technopolitics and the Making of Europe*. Milton Park: Routledge,pp. 125–141.

Prem B (2022) Governing through anticipatory norms: How UNIDIR constructs knowledge about autonomous weapons systems. *Global Society* 36(2): 261–280.

Renic N and Schwarz E (2023) Crimes of dispassion: Autonomous weapons and the moral challenge of systematic killing. *Ethics & International Affairs* 37(3): 321–343.

Reuter-Oppermann M and Buxmann P (2022) Introduction into artificial intelligence and machine learning. In: Reinhold T and Schörnig N (eds) *Armament, Arms Control and Artificial Intelligence: The Janus-Faced Nature of Machine Learning in the Military Realm*. Cham: Springer,pp. 11–26.

Roberts H, Cowls J, Hine E, et al. (2021a) Achieving a "good AI society": Comparing the aims and progress of the EU and the US. *Science and Engineering Ethics* 27: 68.

Roberts H, Cowls J, Hine E, et al. (2023) Governing artificial intelligence in China and the European Union: Comparing aims and promoting ethical outcomes. *The Information Society* 39(2): 79–97.

Roberts H, Cowls J, Morley J, et al. (2021b) The Chinese approach to artificial intelligence: An analysis of policy, ethics, and regulation. *AI & Society* 36: 59–77.

Rudin C (2019) Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence* 1: 206–215.

Rumelili B (2015) Identity and desecuritisation: The pitfalls of conflating ontological and physical security. *Journal of International Relations and Development* 18(1): 52–74.

Ruokonen F (2013) Trust, trustworthiness, and responsibility. In: Mäkelä P and Townley C (eds) *Trust* (Value inquiry book series). Rodopi: Brill,pp. 1–14.

Saßmannshausen T, Burggräf P, Hassenzahl M, et al. (2023) Human trust in otherware—A systematic literature review bringing all antecedents together. *Ergonomics* 66(7): 976–998.
Sauer F (2021) Lethal autonomous weapons systems. In: Elliott A (ed.) *The Routledge Social Science Handbook of AI*. Milton Park: Routledge,pp. 237–250.
Sauer F (2022) The military rationale for AI. In: Reinhold T and Schörnig N (eds) *Armament, Arms Control and Artificial Intelligence: The Janus-faced Nature of Machine Learning in the Military Realm*. Cham: Springer, pp. 27–38.
Schmid S (2023) *Safe and Secure? Visions of Military Human-Computer Interaction in: Mensch und Computer—Workshopband*. Bonn: GI.
Schmid S, Riebe T and Reuter C (2022) Dual-use and trustworthy? a mixed methods analysis of AI diffusion between civilian and defense R&D. *Science and Engineering Ethics* 28(1): 21–23.
Schwarz E (2021) Autonomous weapons systems, artificial intelligence, and the problem of meaningful human control. *The Philosophical Journal of Conflict and Violence* 5: 53–72.
Smuha NA (2021) From a "race to AI" to a "race to AI regulation": Regulatory competition for artificial intelligence. *Law, Innovation and Technology* 13(1): 57–84.
Suchman L (2007) *Human-Machine Reconfigurations: Plans and Situated Actions*. Cambridge: Cambridge University Press.
Suchman L (2020) Algorithmic warfare and the reinvention of accuracy. *Critical Studies on Security* 8: 175–187.
Taddeo M (2017) Trusting digital technologies correctly. *Minds and Machines* 27: 565–568.
The White House Office of Science and Technology Policy (WHST) (2020) American artificial intelligence initiative: Year one annual report. Available at: https://www.nitrd.gov/nitrdgroups/images/c/c1/American-AI-Initiative-One-Year-Annual-Report.pdf (accessed 15 April 2022).
Townley C and Garfield JL (2013) Public trust. In: Townley C and Maleka P (eds) *Trust: Analytic and Applied Perspectives*. Rodopi: Brill, pp. 95–107.
Verbruggen M (2019) The role of civilian innovation in the development of lethal autonomous weapon systems. *Global Policy* 10(3): 338–342.
Vorm ES (2020) Computer-centered humans: Why human-AI interaction research will be critical to successful AI integration in the DoD. *IEEE Intelligent Systems* 35(4): 112–116.
Washington Post (2024) In big tech's backyard, California lawmaker unveils land-mark AI bill. Available at: https://www.washingtonpost.com/technology/2024/02/08/california-legislation-artificial-intelligence-regulation/ (accessed 26 March 2024).
Wille T and Martill B (2023) Trust and calculation in international negotiations: How trust was lost after Brexit. *International Affairs* 99(6): 2405–2422.
Winfield AF and Jirotka M (2018) Ethical governance is essential to building trust in robotics and artificial intelligence systems. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376(2133): 20180085.
Wood S and Cox L (2021) Status, imitation, and affective dissonance in international relations. *International Relations* 35(4): 634–656.
Xu W, Dainoff MJ, Ge L, et al. (2021) From human-computer interaction to human-AI interaction: New challenges and opportunities for enabling human-centered AI. arXiv [preprint]. DOI: 10.48550/arXiv.2105.05424
Zarakol A (2017) States and ontological security: A historical rethinking. *Cooperation and Conflict* 52(1): 48–68.

Author biographies

Stefka Schmid is a research associate at the Chair of Science and Technology for Peace and Security (PEASEC) at the Department of Computer Science, TU Darmstadt. Her research interests include innovation policies as a subject of critical security studies, science and technology in peace and

conflict research, and human–computer interaction in crisis scenarios. Her doctoral research focuses on the collective use of technology in the context of global security politics.

Bao-Chau Pham is a PhD candidate at the Department of Science and Technology Studies, University of Vienna. Her doctoral research explores the co-production of artificial intelligence, society, and security, with a particular focus on discourses around AI policy in the European Union.

Anna-Katharina Ferl is doctoral researcher at the Peace Research Institute Frankfurt (PRIF) and member of the research group Emerging Disruptive Technologies. She focuses on emerging military technologies, especially artificial intelligence and autonomy in weapons systems, arms control and regulation policies, and international security. Her PhD dissertation analyzes how knowledge about technological developments is generated, and what role this knowledge and expertise plays in arms control processes concerning autonomous weapon systems.