NATIONAL RESEARCH UNIVERSITY
HIGHER SCHOOL OF ECONOMICS

Aleksei Turobov

ARTIFICIAL INTELLIGENCE AND SECURITY: TRANSFORMATION AND CONSISTENCY

BASIC RESEARCH PROGRAM
WORKING PAPERS

SERIES: POLITICAL SCIENCE
WP BRP 88/PS/2022

Aleksei Turobov[1]

ARTIFICIAL INTELLIGENCE AND SECURITY: TRANSFORMATION AND CONSISTENCY

What are the dynamics of using artificial intelligence technology in national security? The expansion of the security domains required a rethinking of the "who" and "what" could become a security issue. Digitisation and the rapid development of ICT have led to changes in security. One of the most popular and rapidly developing digital technologies is artificial intelligence (AI). Despite the attention to this technology, little research focuses on HOW technology is entering the security sphere and WHAT changes are taking place in the state's security system. Existing work on artificial intelligence technology in security studies highlights the opportunities, potential risks, and benefits of using the technology. However, it remains unclear what changes are occurring in countries' national security with AI. This study aims to address a specific problem - to offer a well-grounded understanding (and illustration of the dynamics) of the infiltration of artificial intelligence technologies into national security. An empirical model has been developed, tested, and verified for studying changes and assessing the state security system. An empirical model of Security Consistency was implemented in five countries from a comparative perspective (time coverage from 1996 to 2019). The results show that the capabilities of national security systems broaden with the implementation of advanced digital technologies. And these dynamics are non-linear. Despite theoretical concerns, governments have time to assess the risks and benefits of technologies in proportion to the internal features of the security system and, based on this assessment, work on implementation. This research demonstrates that countries lack both "alarming" and exaggerated fears and leniency on technological changes in national security systems.

Keywords: Security, Security studies, Artificial intelligence, Threat evaluation, AI capability, France, Finland, Germany, Sweden, USA.

JEL Classification: E10, E17, F52, H56, O33,

[1] Lecturer, School of Politics and Governance. Research Fellow, Faculty of Social Science. HSE University (Moscow, Russia). Email: aturobov@hse.ru

Introduction

The academic and applied research results indicate specific changes in security under the influence of digital technologies. One of the most popular and rapidly developing digital technologies is artificial intelligence (AI). Yet, despite the attention to this technology, little research focuses on HOW technology is entering the security sphere and WHAT changes are taking place in the state's security system itself.

Security studies operate on the concepts of threats and resources/capabilities to eliminate/stop/overcome threats. In this vein, digital technologies can act as both security threats and security tools. Questions arise: how do states define and perceive security threats of digital technologies (for example, artificial intelligence)? What are the dynamics of the use of artificial intelligence in security provision?

Payne [Payne, 2018], drawing parallels between artificial intelligence technology and nuclear weapons, points to significant strategic and scientific uncertainty about the military implications of technology. The ongoing changes in the military activity require new strategic decisions: "... the rapid advances of AI that seeks to optimize human goals is the beginning to transform military activity, and demands new strategic thought" [Payne, 2018: 30].

Some studies demonstrate fears that AI-assisted expansion of conventional weapons exacerbates the risk of unintended escalation caused by the convergence of nuclear and strategic non-nuclear weapons. Also, the increasing speed of war undermines strategic stability and increases the risk of nuclear confrontation [Johnson, 2020]. It is indicated that advanced capabilities such as autonomous drone swarms and AI-controlled hypersonic weapons will destabilise the strategic level of the conflict and will significantly increase the speed and pace of combat operations [Johnson, 2020: 29-30]. From a pragmatic point of view, AI is integrated into the performance of combat missions to improve knowledge about the operational situation, the enemy's capabilities, and the speed and accuracy of offensive and defensive weapons [Davis, 2019]. A special issue of the journal Security Dialogue for 2017 is wholly devoted to the issue of AI, leading readers to the idea of the growing uncertainty of the future in the field of security. It also raises decision-making problems in the security field in the face of even more significant uncertainty [Amoore, Raley, 2017].

Research on the application of sophisticated algorithms (AI) in a separate element of national security - crime prevention [Lum, Isaac 2016; Wang & Zhao 2016; Ensign et.al 2017; Seo et al. 2018; Zhang et al, 2021; etc.] - demonstrates a large number of problems at the level of (1) the reliability of the results, (2) the quality of forecasts/prediction, (3) the difficulties with the application, (4) the quality of the data, (4) the bias of algorithms, (5) the admissibility of the application of algorithms, in the context of the rights and freedoms of citizens, et cetera.

This study aims to address a specific problem - to offer a well-grounded understanding (and illustration of the dynamics) of the infiltration of artificial intelligence technologies into national security.

At the same time, there is no clear understanding of which analytical toolkit (method/approach) allows tracing such changes (directly dynamics) with the possibility of evaluating the consequences for the security system. Of course, expert interviews provide some "ground" for research, but attempts to "objectively" look at the security system through quantitative methodology are rare. At the instrumental level, this work seeks to provide some evidence of *how* are the dynamics of changes in the field of security by constructing an empirical model for studying changes and assessing the state security system under the influence of artificial intelligence technology.

The developed empirical model is evaluative and aims to form a concept for assessing a security system (not only evaluating threats but also identifying capabilities)

and creating an empirical approach to measuring such a concept of the state's security sphere.

Thus, at the theoretical level of problematisation, the work contributes to the subject area of research into the process of changes in security under the influence of digital technologies. At the methodological level, it offers an approach to measuring such changes using an empirical model.

The following section briefly reviews existing work on security studies and artificial intelligence technology. The second section presents hypotheses, while the third section discusses empirical strategy. The fourth section presents the results of the analysis. The final section concludes.

Literature review

Security is political on two levels. The first is insecurity issues, which are necessarily a product of political action [Weldes et al., 1999] since neither security nor insecurity is a "natural" state of affairs. Political actors promote discourses of security or insecurity and feed appropriate public emotions through rhetorical "speech acts" - asserting vulnerability, promising security - and thereby shape the political landscape [Wirls, 1992]. The second level is directly the development and implementation of security policies. In addition to the obvious questions about the legislature's activities, political technologies, and mechanisms, it is worth noting the topic of mobilisation that is "slipping away" from research agendas. It's not only about mobilising the population in the face of real/imaginary threats to security from political actors but also more complex processes of mobilising resources (to ensure security) and mobilising elites (to support and maintain the chosen course of national security) [Krebs, 2018: 9]. Krebs argues that not all security preferences are equally accounted for in defining "national interests," as political institutions empower actors and direct aggregation of preferences differently [Krebs 2018:2]. Political institutions also influence decision-makers' available resources and what policy instruments they find attractive. Leaders gravitate towards those policy instruments over which they have more control, which explains many security decisions.

The field of security is evolving. The content of the concepts "security" and "threat" is developing. Moreover, the borders/spheres of national security are also being transformed. Traditionally, security has been viewed in terms of war and peace [Wright, 1942; Mead Earle, 1944; Bayley, 1975; Paret, 1986; Baldwin, 1995; Parker, 1996]. Only those threats directly in the plane of military readiness/capabilities of states were perceived as threats to security. That is why the study of war and peace is a separate area of academic research. Initially, research was focused on the triad: State Power - War - Strategy.

After the First World War, the concept of collective security developed intensively [Kennedy, 1987], although the practice of collective security (within the framework of narrow alliances) is a much earlier phenomenon. Collective security is not the same as international security. Collective security applies only to those actors (states) that have assumed the responsibility of maintaining a certain level of security (under international agreements and treaties). After World War II [Wolfers, 1952; Buzan, 1991; Baldwin, 1997], the security concept has constantly been evolving as the very fact of the world war has demonstrated the inconsistency of the previously existing approaches to ensuring security. Thus, collective security (with the creation of an appropriate system) is becoming the dominant concept and practice. The concept of "security dilemma" acquired rapid development (especially during the Cold War), and the specificity of "control and deterrence" was formed.

However, the early 1990s brought a new conceptualisation of security. The Copenhagen School of Security proposed a sectoral approach to security [Buzan 1991;

Buzan and Waever 2003; Brauch et al. 2008; Floyd and Matthew 2013; Hanlon and Christie 2016; Neack 2017]. And Busan presented the securitisation theory [Buzan et al. 1998; Balzacq 2011]. These two concepts expanded the classical understanding of the security sectors from the military sphere and territorial sovereignty to the economic, environmental, social, and political sectors. Despite the criticism and objective shortcomings (for example, by the Paris School of Security), the expansion of the sectors continues. For example, information security [Nance and Straub, 1988; Alter and Sherer, 2004; Deshmukh, 2004; Bulgurcu et al., 2010; Whitman and Mattord, 2011] does not just stand out as a different sector but permeates all security sectors. Likewise, cybersecurity stands out, which is actively infiltrating all security sectors.

The terrorist attacks of September 11, 2001, also contributed to expanding the security research agenda. The events themselves and the US response have generated unprecedented attention to the growing number of non-state actors and their impact on the international security environment [Enders and Sandler 2006; Hoffman 2006; Pape 2006; Sageman 2008; Cronin 2009]. New discussion spaces have also emerged about the "war on terror" in the relationship between the state and civil society, including in liberal democracies [e.g., Aradau and van Munster 2007; Bigo et al. 2015; Jarvis and Lister 2015] and an increased focus on technology development.

The expansion of the security domains required a rethinking of the "who" and "what" could become a security issue. In addition, digitalisation and the rapid development of ICT have led to the emergence of new, inventive and aggressive ways to monitor, predict and/or neutralise potential security threats [Hendershot and Mutimer 2018]. These "new" security practices and processes (for example, widespread use of biometrics [see Muller 2008], surveillance [see Bell 2006], drones and targeted assassinations [see Grayson 2016], algorithmic security [see Amoore and Raley, 2017]) opened theoretical and empirical possibilities, and posed several new problems.

This brief overview demonstrates the transformation in security research. The very "the notion of security expanded greatly as a consequence of transformations in policy" [Schlag et al, 2015:12] under various factors, including the technological ones. Moreover, the environment can lead to a "…transformation of what it means to secure" [Schlag et al, 2015:152]. At the same time, there is "…a link between the transformation of security policies and the development of security studies" [Schlag et al, 2015:233]. Thus, *awareness of the expansion of the content of "security" and the presence of significant changes - "transformations" - in politics serve as the starting point of this study*.

Thus, I propose the following logic: digital technologies infiltrate all security sectors (by analogy with the information sphere) and act as independent elements of threats and tools to overcome them. Artificial intelligence, as a type of digital technology, is no exception.

There are various concepts of understanding AI, but in a broad sense, this type of digital technology is defined as intelligent systems with the ability to think and learn [Russel, Norvig, 2010]. It is a heterogeneous set of tools, methods, and specific algorithms [Jarrahi, 2018]. Artificial intelligence is also defined as a system that can independently interpret external data and learn from them to achieve particular results through flexible adaptation [Kaplan, Haenlein, 2019]. Various applications and methods - from neural networks (and deep machine learning models) to speech and/or image recognition and genetic algorithms, natural language processing and machine vision - are united by the umbrella concept of AI technologies [Reis, Santo, Melao, 2019].

Some studies indicate that artificial intelligence does not exist (e.g. Galanos, 2018 or Edwards, 1997). However, considering AI (1) as a separate category of intelligence, which differs from "natural" intelligence and (2) that intelligence is an inherent property of physical subjects (for example, people) or objects (for example, robots) - it is

demonstrated that neither the first nor the second is impossible within the framework of the usual social philosophy since intelligence is a systemic phenomenon, and not a property of a separate unit [Longino, 2014]. Within the framework of research in the social sciences, an "umbrella" concept of AI technologies is quite acceptable, which defines a set of approaches, tools, and algorithms since the main "refraction" of AI is social impact and socio-political effects. In other words, for the study of social and political processes, it does not matter how "artificial" or "natural" the intellect and what it is [Vallverdu, 2017] if we consider the phenomena through understanding the intellect as the primary phenomenon with people and machines as its agents [Galanos, 2018:362].

Despite the difficulty in understanding and content-rich AI technology, states have pretty "successfully" entered a new "arms race" concerning AI technologies [Horowitz, 2018; Sharre, 2019; Brose, 2019]. Although AI technology is believed to be the key to economic growth, national security, and strategic advantages, the competition between countries to dominate AI is getting fierce [Fatima et al., 2021].

The UN report Militarizing Artificial Intelligence (2019) indicates that AI is not a single technology but a collection of theories, methods, technologies and applications to stimulate and expand human intelligence. The prism of practical military application and influence on the security system and the strategy of states and international stability is considered separately.

For this research, artificial intelligence technology is understood as *algorithmic and computer systems (including software and/or hardware) that, while learning, can solve complex problems, make predictions or perform tasks requiring human perception, learn, plan, communicate or perform a physical act, necessarily in the security domain or, directly, in the military sphere*.

Based on such a definition, we can trace the main discussions about artificial intelligence in security. Some argue that AI will disrupt the balance of power, become a critical part of future weapon systems, and provide new opportunities for adversarial attacks [Horowitz et al., 2020]. Others demonstrate that AI-enabled systems are widely used by national security agencies for intelligence, surveillance, reconnaissance, logistics and cybersecurity operations [Hoadley, 2019]

The scientific and technological agenda also draw much attention. Especially the high potential in the automation of the scientific process itself (including the military one) with the help of AI [Briscoe, Fairbakks, 2020]. This type of advancement stands to upend the fundamentals of technological development. At the same time, R&D efforts on autonomous weapon systems have already been tested in the context of drones that can target and hit enemy radar installations [Simonite, 2019].

The implications of military applications of AI in ways that reduce the risk that states' uncertainty about changes in military technology undermine international security and stability [Horowitz et al, 2020]. The world's leading developed countries consider AI development a major strategy to enhance national competitiveness, protect national security and see it as a strategic technology that will play a leading role in the future [Allen, 2019]. Arguably, states defend their advances in AI for political, social, and economic reasons and security [Aradau & Blanke, 2017]. The Offense-Defense theory notes that advances in artificial intelligence will escalate the number of weapons platforms and the number of software vulnerabilities that they can discover [Garfinkel & Dafoe 2019].

Turning to national strategies on artificial intelligence, it is worth noting that only nine plans mention national defence and security [Fatima et al., 2020]: China[2], Chzech

[2] Next Generation Artificial Intelligence Development Plan. Next Generation Artificial Intelligence Development Plan. September 2017

Republic[3], Finland[4], Germany[5], Italy[6], Korea[7], Qatar[8], Spain[9], USA[10]. Six countries mentioned the defence industry in their strategic AI plans. The defence sector was seen as a critical collaborator in advancing the deployment of AI-enabled solutions in all facets of national security operations [Fatima et al., 2020]. AI systems can be a threat to security in two broad ways: (1) intentional use of destructive AI (e.g., autonomous weapons) and (2) unintentional malfunctioning in AI systems (in autonomous cars etc.) that could damage humans, properties and natural resources. Twenty-one plans recognised that AI systems can cause harm and the need to carefully consider the malicious use of the technology [Fatima et al., 2020].

Drawing together the above discussion points to some major remaining questions about the impact of artificial intelligence technology on the security sector. First, although much work has been done to assess the prospects and risks of using AI technology, it remains unclear how this technology is applied. Secondly, the focus of existing works excludes quantitative analysis of the dynamics of infiltration into the field of security. Finally, much of the work on applying AI technology in the field of security focuses on individual cases or general phenomena. However, comparative analysis can significantly expand existing discussions and reveal general trends among different states.

In this paper, I attempt to contribute to existing work by tackling some of these unanswered questions both empirically and theoretically.

Theory and Hypotheses

As discussed in the previous section, current work in security studies and technology applications in security suggests that governments are wary of new technologies. Based on the logic of the empirical model (the ratio of the threat evaluation indicator and the digital capabilities of responding to threats), it can be assumed that the threat indicator will exceed the capability indicator in country models. The justification is the specificity of digital technologies, the ambiguity of their application and their rapid development. Analyzing and evaluating digital threats (and assessing "traditional" threats using digital technologies) will outpace existing threat countermeasures. Testing the hypothesis will reveal official "attitudes" toward transformations in the sphere of

[3] National Artificial Intelligence Strategy of the Czech Republic. Ministry of Industry and Trade of the Czech Republic. May 2019

[4] Finland's Age of Artificial Intelligence. Ministry of Economic Affairs and Employment, Finland. December 2017

[5] AI Strategy. Federal Ministry of Education and Research, the Federal Ministry for Economic Affairs and Energy, and the Federal Ministry of Labour and Social Affairs, Germany. November 2018
Germany's plan noted that "The use of AI-based technologies and systems will have implications for the armed forces and is therefore an important issue to be taken into account for the future of the Bundeswehr. As in other fields of application, the Federal Government will undertake a comprehensive analysis of the benefits and risks involved" [Germany AI Plan 2018, p. 31]

[6] National Strategy on AI. Italian Ministry of the Economic Development. August 2018

[7] Mid- to Long-Term Master Plan Intelligent Information Society. Government of the Republic of Korea. December 2016

[8] National Artificial Intelligence Strategy for Qatar. Qatar Center for Artificial Intelligence. February 2019

[9] Spanish Strategy for RDI in Artificial Intelligence. Ministry of Science, Innovation and Universities, Spain. March 2019

[10] The National AI Research and Development Strategic Plan. National Science and Technology Council, USA. October 2016.
USA's plan said that "Machine learning agents can process large amounts of intelligence data and identify relevant patterns-of-life from adversaries with rapidly changing tactics. These agents can also provide protection to critical infrastructure and major economic sectors that are vulnerable to attack. Digital defense systems can significantly reduce battlefield risks and casualties" [USA AI Plan, 2016, p. 11]

security initiated by digital technologies. The explanatory mechanism will allow to understand how states determine the role and place of technologies (threat evaluation will show the fears of states and the potential of capabilities - how states can respond to threats). If the hypothesis is confirmed, we will see an increase in security concerns since governments are paying more attention to the potential of threats (both directly from digital technologies and in issues where technology should serve as a tool for eliminating threats). Conversely, if the hypothesis is refuted, governments successfully integrate technologies by a balanced assessment of threats. In this case, we demonstrate that states have adapted to modern changes and designed their security systems so that the response capabilities exceed the evaluated threats. More specifically, the hypothesis is formulated as follows:

> **Hypothesis 1.** Threat evaluation will exceed capabilities of threat response

The temporal coverage of the empirical model makes it possible to formulate hypotheses in terms of the dynamics of change. The organisational principle of contemporary national security is expressed in the following maxim "maximisation of military power through the use of technology" [Farrell 2002: 67]. Accordingly, it is logical to assume that the use of digital technologies in security occurs before a meaningful discussion in society about these digital technologies. If they are genuinely committed to maximising opportunity, governments must start making policy decisions and adopting technologies faster and earlier than societies are included in these discussions. The primary public debate about applying artificial intelligence technology in security began around 2010 in a discussion of concerns associated with an "artificial intelligence arms race"[11]. Hypothesis testing will provide an understanding of the immediate dynamics. It will allow us to assess governments' " inclusion " in harnessing the potential of technological changes in security systems. In other words, we can reveal how proactive governments are. Do we see the intention of the state to apply technologies (given the duration of the technological cycle from the moment of decision-making to direct practical implementation) at a faster pace before society pays attention to the significant risks and opportunities associated with digital technologies. If the hypothesis is confirmed, we can argue that states seek to maximise power and security through new technologies. Thus, observed changes (and, more broadly, transformations) are supported and even initiated by governments. However, there may be several explanatory mechanisms if the hypothesis is refuted. The primary mechanism will be that states fear threats and risks more than they perceive the potential for technologies' applications. Therefore, the authorities will seek first to obtain an evidence base of contingent benefits and only then begin to introduce and apply technologies. An alternative explanation would be a redistribution of government's focus, i.e. maximisation can and does occur, but through other technologies or strategies and tactics. Thus, the second hypothesis of the study is formulated as follows:

> **Hypothesis 2.** The dynamics of the use of artificial intelligence technology in the field of security will manifest itself in the period 2008-2010.

[11] Artificial Intelligence arms race. For example, see the article on the popular Wikipedia resource: https://en.wikipedia.org/wiki/Artificial_intelligence_arms_race (accessed: 11.09.2021)

There is a consensus in security studies that different states (governments) do not assess threats similarly. States can evaluate and interpret the same threat differently [Wolfers, 1952:151]. Although one of the leading papers on this topic (the study of Wolfers) refers to "traditional" approaches to security and can be considered "outdated", the issue of uncertainty of threats is very relevant. Despite existing discussions about the correct and proper understanding of security, the perception of threats (and their content) is both a scientific and a practical problem. In other words, in this hypothesis, we rely on the substantive component of the argument that states define and evaluate threats differently. The empirical model of this study allows us to analyse the evaluation and perception of threats by each state separately and evaluate such perceptions. However, despite differences in threat assessment, there is security ambiguity regarding digital technologies, creating unnecessary fears. Therefore, we assume that the evaluations of threats will be homogeneous in the analysed countries, and heightened concerns will characterise them. By testing this hypothesis, we seek to identify the existence of uncertainty in the assessment of threats, complicated by the specifics of digital technologies. On the one hand, identifying differences in threat assessment will confirm the empirically grounded theoretical assumption about differences among states in the evaluation and perception of threats. On the other hand, we will be able to demonstrate whether there is (or is not) a particular specificity related to digital technologies. If the hypothesis is confirmed, we will refute the theoretical notion of uncertainty in threat assessment and, in doing so, demonstrate the unique attitude of states towards digital technologies. This will significantly expand the discussion about the transformational effect of digital technologies. If the hypothesis refutes, we will confirm the existing theory about the uncertainty of threat evaluation and demonstrate that governments' perceptions of digital technologies do not differ significantly from other tools. More specifically, the third hypothesis is formulated as follows:

> **Hypothesis 3.** The perception of threats and the type of threats will reach maximum values in all analysed countries.

Comparative analysis will also test the assumption that states are highly alert to digital technologies' risks and challenges. We assume that by 2018-2019 in all analysed countries as leaders in the field of technology, high values of the indicators of the empirical model will be observed. These will indicate that states (1) are on high alert to risks and threats arising from digital technologies and (2) are rapidly introducing digital technologies into security to meet current challenges. When testing this hypothesis, we will be able to assert the existence of changes (and, more broadly, transformations) and find out the directions of these changes. The fast pace of technology adoption and a high assessment of risks and challenges will demonstrate the adaptability and readiness of governments to contemporary security challenges associated with digitalisation. More specifically, the fourth hypothesis is formulated as follows:

> **Hypothesis 4.** By 2018-2019 the indicators of the empirical model of all analysed countries will approach their maximum values, which indicates a high readiness of states to face the risks and challenges posed by AI technologies.

Empirical Design

Hypothesis testing was carried out in five countries[12]: the USA, Sweden, Germany, Finland and France.

In forming an empirical model, our choice of indicators and parameters is determined by emphasising what is measurable [Fioramonti, Kononykhina, 2015:476]. Working with indicators, in turn, is subject to the need to strike a balance between completeness and availability of data [Fioramonti, Kononykhina, 2015:477].

The Security Consistency index is derived from the difference in threat metrics (a measure of threat evaluation) and the ability of AI technologies to respond (a measure of AI capabilities) to such threats. This can be represented in the form of a block diagram.

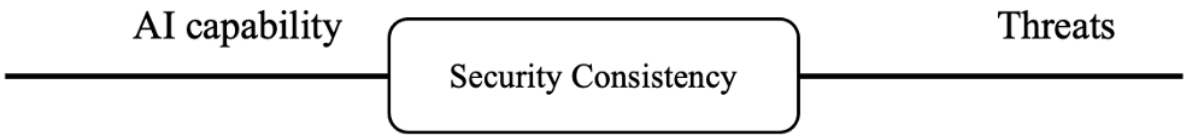

Fig.1 Security Consistency index block diagram

The Security Consistency index reflects how a state can assess threats (an indicator of threats) and whether the state has the necessary capabilities to repel them (an indicator of capabilities). Thus, formula (1) for calculating the Security Consistency index has the following form:

$$Security\ Stability = AI\ Capability\ -\ Threats\ Evaluation \quad (1)$$

The AI capability parameter is calculated in the logic of composite indexes. Based on the conceptual framework for understanding artificial intelligence and the goals of creating the indicator, four areas have been identified (technological area, economic area, governance, social area) with the distribution of weights. The formula for calculating the AI capability indicator (2) is as follows (a detailed description of the calculation is presented in *Appendix 1*):

$$AIcapability = \frac{0.25 \cdot UT + 0.25 \cdot MF + 0.3 \cdot (SC + LA) + 0.2 \cdot (JO + RS)}{4} \quad (2)$$

[12] Two independent network analyses were carried out to determine the pool of countries to focus on the study. The first network analysis focused on digitalisation in competition between states (networks were built in the logic of Affiliation and Co-Affiliation). In this logic, a network of countries was built in which ties are represented by indicators denoting widely used technologies.
The second network analysis considers countries as nodes that share common links expressed in terms of technology trade. A specific country acts as a node, and an indicator of trade with other countries "lies" on edge, which will form a connection. At the same time, communication in networks is directed - from the country-seller to the country-buyer and based on data from the World Trade Organization. Thus, the countries that demonstrate the leading positions according to the results of two network analyses were used for the study.
The results of the second network analysis are presented in the publication: Turobov A. Opportunities for Transplantation of Political Institutions in Technology Trade: Results of Network Analysis in a Comparative Perspective // Vestnik TGU. Filosofiya. Sotsiologiya. Politologiya (RU) – Tomsk State University Journal of Philosophy, Sociology and Political Science – 2022 – V.65 – p.310-327 (Туробов А.В. Возможности трансплантации политических институтов при торговле технологиями: результаты сетевого анализа в сравнительной перспективе // Вестник Томского государственного университета. Философия. Социология. Политология – 2022–№65 - С. 310-327.)

The formula for calculating the AI capability parameter represents the *UT* indicator (as an indicator of the application/use of AI in government and the military sphere), which belongs to the Technological area. The *MF* (Military AI Funding) metric refers to the Economic area. *SC* (an indicator of state-owned AI companies) and *LA* (an indicator of whether there is legal authorisation to use AI technology for military purposes) belong to the area of Governance. And also, two indicators are related to the Social sphere: *JO* (an indicator of employment in the field of AI) and *RS* (an indicator of start-ups in the field of AI).

The design of **the Threat evaluation parameter** is based on the approaches of a highly specialised subject field of research on weapons and military threats - Threat Evaluation and Weapon Assignment - TEWA (for example, [Cocelli, Arkin, 2017; Johansson, Falkman, 2008; Naeem, Masood, 2010; Naseem et al. , 2017; Kumar, Tripathi, 2016]). This study's logic requires evaluating threats and searching for a balance between these threats and the defended objects and security sectors. The formula for calculating threats (3) for research purposes is presented below (a detailed description of the calculation is presented in *Appendix 2*):

$$TE = \frac{PV \cdot (TP + TT)}{DA \cdot TF} \quad (3)$$

The Threat evaluation formula contains indicators of significance (protection value - *PV*), threat perception (*TP*), threat nature/type of threat(*TT*), number of protected objects (*DA*) and threat factors (*TF*).

In other words, the logical content of the formula can be represented as follows: the Threat evaluation indicator (*TE*) is the *relation* the sum of the threat perception indicator (*TP* - how threats are presented at the level of regulatory legal acts - an act of political will) and the indicator of the nature/type of threat (*TT* - how threats are presented in reports, relevant literature) multiplied by the Protection Value (*PV* - as decision-makers and political actors assess the importance/significance of threats) *to* the indicator of the number of protected objects (*DA* - objects/assets to which threats are directed) by the indicator of threat factors (*TF* - assessment and ranking of security sectors to which threats belong).

The empirical model for each selected country was implemented according to a unified protocol. In the first step, the relevant time coverage for each country was determined. When determining the time coverage, the following were taken into account:

(1) bare normative legal acts in the field of security (national security strategies, laws, decrees, doctrines, etc., defining the state security system);

(2) the primary regulatory documents of the information technology sphere, with a focus on digitalisation, algorithmisation, automation of politics and public administration (taking into account the concept of e-government, up to regulation of specific types of digital technologies);

(3) time coverage with multiple governments/administrations for dynamics.

After forming the timetable, the data collection for each model indicator began. A prerequisite for data collection was registering national legal acts and official statistics, international reports, databases, etc., for the analysed country. As a result, a data set was formed for each country, where a note about the data source was made for each

indicator's value. The data is available in the open GitHub repository at a stable link in separate files for each country[13].

To build, test and validate the model, all calculations were carried out in a free program environment for statistical analyses and graphics R v.4.0.5[14].

Results

Comparing country models in terms of AI technology capabilities allows us to trace the dynamics of how governments assess technology capabilities in the security sector. A graphical comparison of the models is shown below in Figure 2.

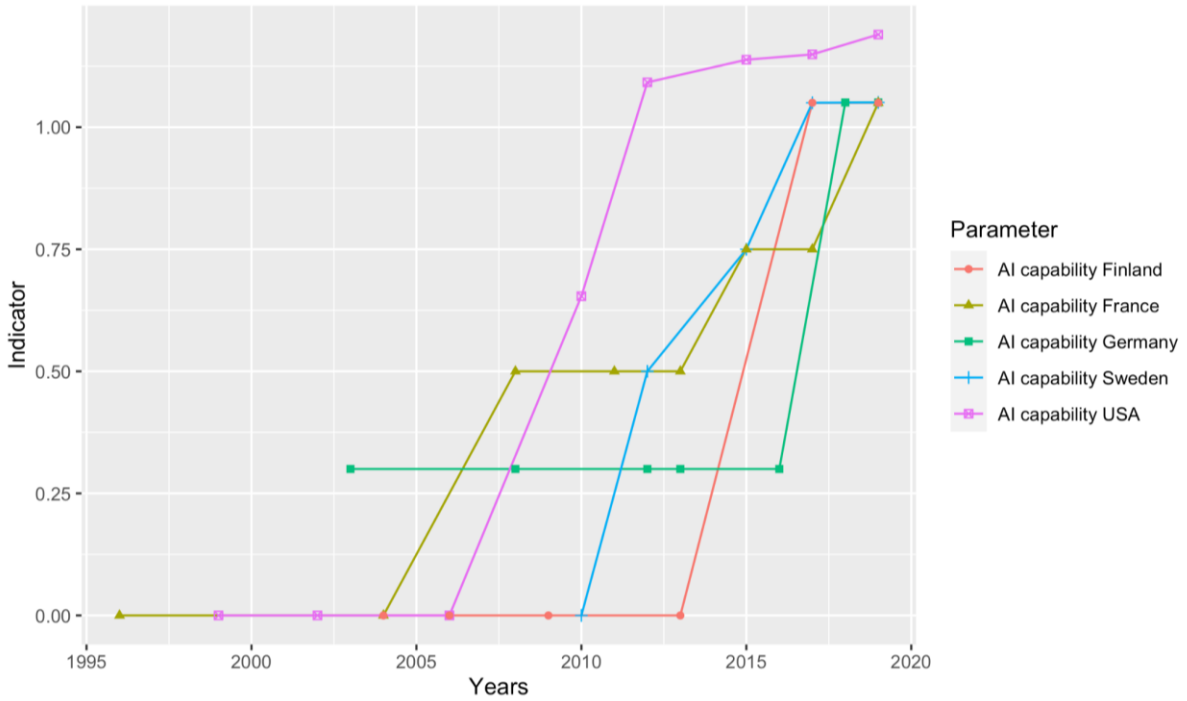

Fig. 2 Comparison chart of country models in terms of capabilities of artificial intelligence technology

First of all, attention is drawn to the "leadership" of the US model in this parameter since 2010 and the similarity of the dynamics of the US and Swedish models. These results of the US model confirm the existing discussion about starting in the 2010 Artificial Intelligence Arms Race. Moreover, we see circumstantial evidence that the US government was preparing for this race in advance and, directly to 2010, began to show consistently high rates.

From 2003 to 2004, the assessment of capabilities was at zero, despite the theoretical developments that already existed, practical tests, and experiments with artificial intelligence technology. The above may indicate the technology's immaturity and the acquisition of meaningful content closer to the mid-2000s.

The results of the German model stand out firmly from the rest, firstly, because it has the earliest effects on this parameter (2003). Secondly, it does not start from a zero position, while, until 2016, the results of the German model for this parameter show a plateau. Although such a result is difficult to interpret unambiguously, the most likely

[13] Repository with data for each country: https://github.com/AlTurobov/PhD_SecurityConsistency (stable link)
A separate file corresponds to each country, and an additional link with comments and descriptions of sources is indicated for each country.

[14] Mode of access: https://www.r-project.org (accessed: 24.10.2021)

explanation lies in the specifics of the country's national security due to the historical context (including the lack of national security strategies) and the government's political will.

However, one of the most surprising results is that all countries, except for the US model, will show the same result by 2019. With a difference of 2 years (starting from 2017), all country models come to a single point. Perhaps this result illustrates a situation where all countries reach a uniform peak in developing and absorbing the available technology. An alternative explanation may be the nature of the international interaction of these countries within the framework of the standard rules of the European Union and international agreements. It can be criticised from existing approaches to understanding national security and the militarisation of technology. This dynamic is subject to further research, both in terms of content and at the level of expansion of the time.

A comparative analysis of country models in terms of threat evaluation allows you to track how governments of countries define threats associated with a technological factor in security. A graphical comparison of the models is shown below in Figure 3.

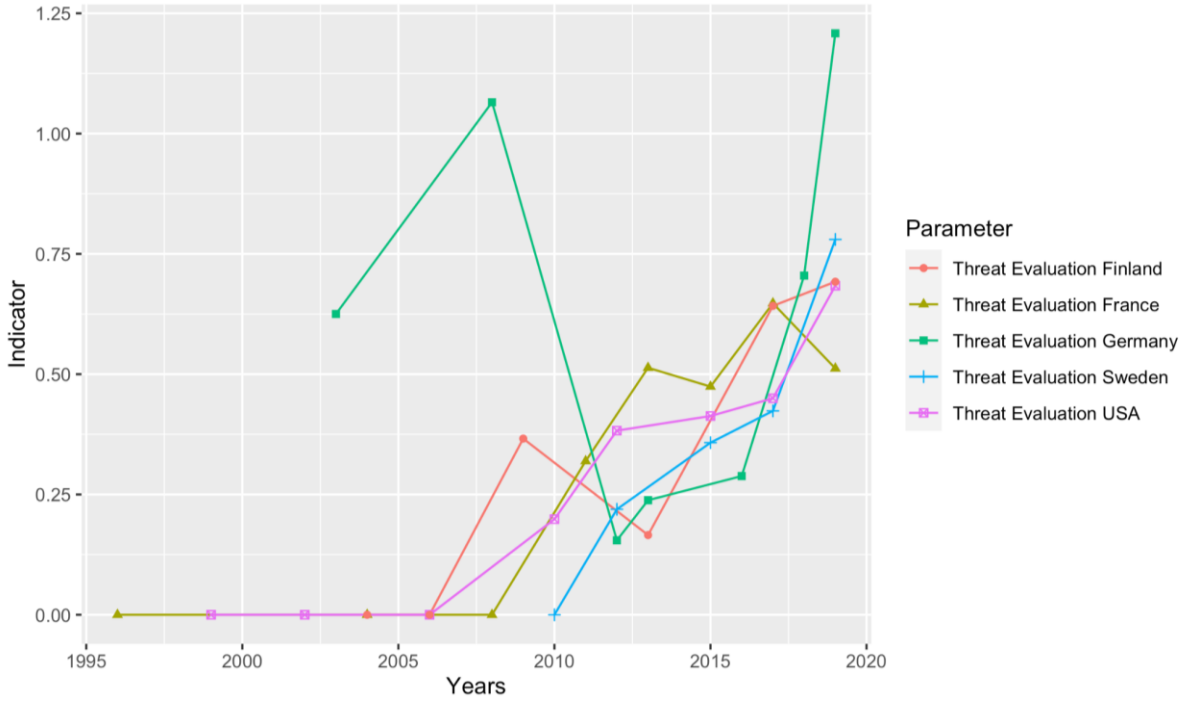

Fig. 3 Comparison chart of country models by the threat evaluation parameter

Threat evaluation results for other models of countries demonstrate similar dynamics. Thus, the starting point for the Finnish and US models is 2006, for France - in 2008, and for Sweden - in 2010. The similarity of dynamics is also reflected in progressive growth, except for the Finnish model, which shows a decline by 2013 and later again progressive increase, and the model of France, in which we can observe two recessions (2015 and 2019). It is noteworthy that only the French model declines by 2019 when all other models show a substantial increase in the parameter.

Attention is drawn to the similarity of the dynamics of the threat evaluation in the United States and Sweden models. Moreover, we see similar patterns of dynamics and an almost identical result in the last period - in 2019. In addition, we observe the exact similarity in the dynamics of the AI capability parameter. A possible explanation may lie in technology diffusion and technological ties between the two countries, including solid trade channels between the US and Sweden. An alternative explanation could be the

influence of security experts visiting experts in Sweden, both in terms of bilateral cooperation and in the direction of international cooperation, including military cooperation with the United States and cooperation with the NATO alliance (as an Enhanced Partner).

The logic of the Security Consistency index reflects the readiness of the state to respond to threats. Readiness is manifested by considering how the state evaluates threats (Threat Evaluation parameter) and how the state determines the ability to stop/overcome threats (AI capability parameter). Thus, a high Security Consistency index means that the state's security system is ready to adequately respond to the threats defined (identified) by the government. A comparative analysis of country models according to the developed index of consistency of the security system is shown below in Figure 4.

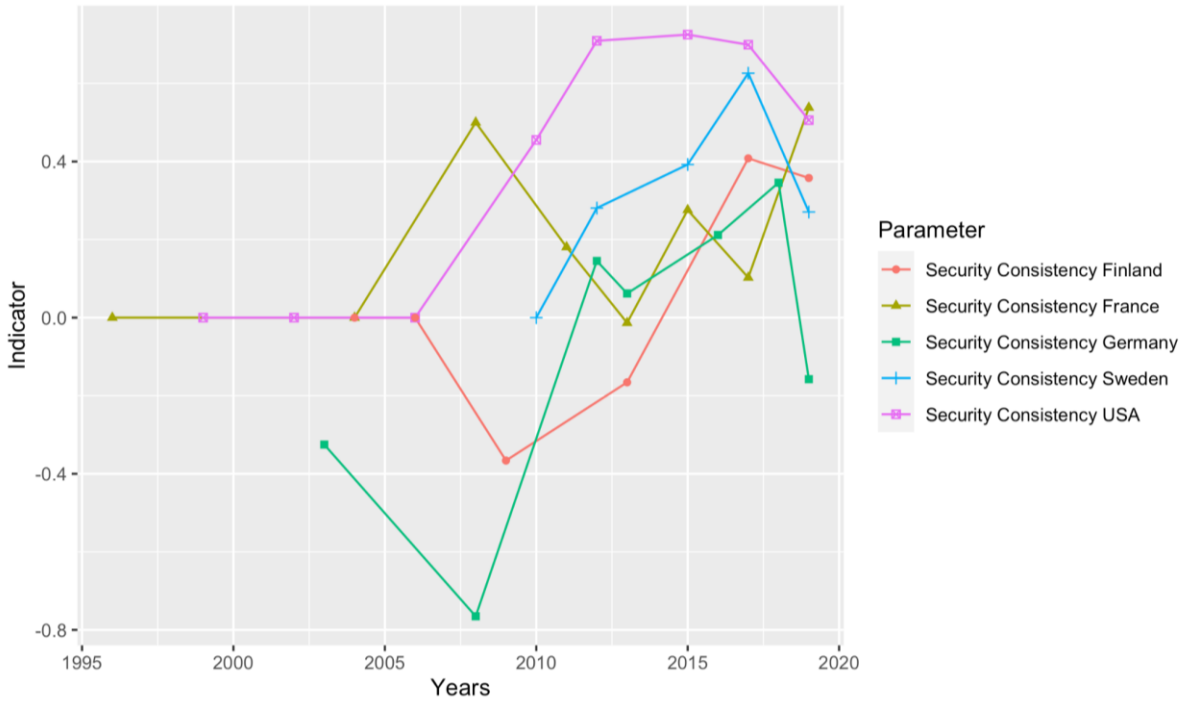

Fig. 4 Comparison chart of country models by the security consistency index

The model comparison results reflect several exciting findings. Firstly, the leader in this index is the model of France, which by 2019 will be ahead of the US model. The lead is not significant but remarkable because the overall dynamics of the US model have been leading since 2010 and only in 2019 yielded to the French model.

Second, the two models show a negative Security Consistency index. The German model started in 2003 from a negative value with a fall in 2006. Only by 2012 is it moving towards a positive trend. However, by 2019 it still drops to negative values. Based on the model's logic, this result means that Germany pays more attention to threat evaluation, i.e. the evaluation of threats is superior to determine the possibilities of stopping/overcoming them. The explanatory mechanism behind this result may be that the German government is very apprehensive about introducing technology to security. Therefore, it pays more attention to risks than to potential opportunities. Accordingly, the unfavourable indicators of the Finnish model follow a similar logic. The difference between these two models lies in the final (as of 2019) result - the Finnish model continues to show positive dynamics, unlike the German model.

Thirdly, the wavy trend in all models is remarkable (although it is weakest in the US model). Finally, we see that the index of security system coherence in all analysed countries does not develop linearly with a constant increase/fall but has pronounced peaks and "attenuation". The undulating nature is consistent with the existing literature on the heterogeneity of perceptions of threats by different states. However, it is complemented by the impressive results of the technological specificity of this study.

Fourth, the results of the models quite clearly reflect the emerging interest on the part of governments in artificial intelligence technology in the field of provision. We can observe the absence of any interest and development (zero indicators) in all models. However, since 2003-2004 and 2006, all models of countries, one way or another, have begun to demonstrate significant changes in national security systems. At the same time, it is noticeable that countries react differently both to the assessment of threats and directly to opportunities. Because of this, the consistency of security systems differs in all countries. The only exception is the similarity of the dynamics of the US and Sweden.

The results for each country are available in *Appendix 3*.

Model verification

A series of simulation analyses [e.g. Marquardt 2020] is carried out to verify and test the model. The simulation logic aims to test the model's sensitivity [Gelman et al. 2020: chapter 5]. What if some of the data does not correspond to reality, is false, or is completely missing? How much will it affect the results? Testing the model through simulation aims to answer these questions.

The simulation was implemented independently and in stages for each indicator of the model[15], i.e. for each indicator, a question was asked separately about the validity of the data. This was implemented as follows: (1) fake data[16] was created for each indicator (10,000 simulations); (2) the model was reproduced with all the original indicators, but instead of the tested indicator, values with false data were randomly used; it is important that the model was necessarily built using the minimum value of false data (i.e., the permissible minimum) and separately with the maximum value of false data (similarly: the permissible maximum); (3) as a result, after each simulation, a new final result of the model was obtained, which was compared with the primary one.

Thus, each indicator from the model was tested. When comparing (including with the minimum and maximum allowable values for each specific indicator), a conclusion was made about the stability of the model to the data of a particular indicator. It should be noted that when comparing, the decision on the stability of the model was made relative to the distribution of the modelled (taking into account simulations) final indicator within one standard deviation. In other words, if the distribution was within one standard deviation, a decision was made about the stability of the model to the variability of values in the data of a particular indicator.

As a result of simulations with fake data, it was found (full information on the simulation results is presented in *Appendix 4*):

1. The **model is resistant** to indicators:

- Military fundings (*MF*) from the AI Capabilities parameter (economic sphere);
- Threat perception (*TP*) from the Threat Evaluation parameter.

2. The **model is sensitive** to indicators:

[15] The simulation was carried out on the US model [see. Turobov and Mironyuk, 2021]

[16] The creation of false data was based on the mean and standard deviation of the data used in the model. Various variations of working with false data can be found on the site Statistical modeling, causal inference and social science
Mode of access: https://statmodeling.stat.columbia.edu/?s=%22fake+data%22 (accessed: 29.09.2021)

- the technological sphere from the AI Capabilities parameter (*UT* - use of technology), but only towards the maximum values;
- Social sphere from the parameter of AI Capabilities (*JO* (job openings) + *RS* (startups)), but only towards the maximum values;
- Threat factors (*TF*) from the Threat Evaluation parameter to both minimum and maximum values.

3. Indicators that are not subject to testing because their minimum and maximum accepted values were present in the primary model for different years:

- Governance environment (*SC* (state companies) + *LA* (legal authorization)) from the AI Capabilities parameter;
- Threat Type (*TT*), Protected Objects/Assets (*DA*), and Protection Value (*PV*) from the Threat Evaluation parameter.

The model demonstrates stability to most indicators. However, special attention should be paid to the threat factors (*TF*) indicator data sources from the Threat Evaluation parameter. This indicator can give a significant error when building a model. Therefore, when working with data on countries, the sources for each indicator underwent additional verification. A similar but less threatening situation exists with the Technological and Social indicators from the AI Capabilities parameter, where the model is sensitive only to the maximum values of the indicators. In addition to these indicators, overestimation (exaggeration) in the available sources' data was also checked.

Conclusion and discussion

On the theoretical level, this study presents a description of *how* digital technologies (by the example of AI technology) find their way into the sphere of security and *which* changes are taking place in the security system of a state. The results show a maximisation of national security capabilities using advanced digital technologies (however, this process does not occur evenly due to states' institutional features).

In security studies, there is a debate about the boundary between security and development [Gheciu and Wohlforth 2018b]. After September 11, 2001, the controversy outlined the prospect of politically narrowing security concerns to development opportunities. An alternative point of view was expressed by researchers who called for a reformulation of the agenda to recognise the fundamental connection between the two areas [Baranyi, 2008; Newman 2010; Tschirgi et al. 2010]. Academic debate about security now goes to two extremes: (1) we are seeing a deepening link between security and development, or (2) development is being "securitised", subordinating it to a paramilitary security agenda. Indirectly, the results demonstrate that states actively apply the achievements of big (in the context of national and international markets) technology companies in the sphere of security. Thus, the trend of subordination and securitisation of development to security is apparent.

The dynamics of changes in all analysed countries' models are not uniform but have several similar stages. By 2018-2019 the indicator of AI capabilities is stabilising, while the indicator of threat evaluation by this period shows rapid growth. The security consistency index demonstrates differences among states regarding the beginning of the infiltration of digital technologies into the sphere of security (an increase from 2003 to 2006). While some countries are actively building up potential for transforming their national security systems (for example, the United States and Sweden), other countries demonstrate the opposite trend of increased fears and risks of changes (for example, Germany and Finland).

At the second level of problematisation - in the instrumental (methodological) plane - an empirical model of changes and evaluation of the state security system under the influence of digital technologies has been developed, tested and validated (by the

example of the AI technology). The model aims to offer means of assessing the security system (it does not only evaluate threats but also identifies capabilities) and to create an empirical approach to measuring the sphere of security of a state.

The first hypothesis has been **partially refuted**. Despite the specificity, ambiguity, and rapid development of digital technologies, particularly artificial intelligence technology, most states have time, at least at the institutional level, to determine and consolidate the ability to respond to such threats. In other words, we observe that states, for security provision at the initial stage, assess capabilities of responding to threats more extensively (in comparison with the threat evaluation). However, this state of affairs does not apply to all states uniformly. For example, Germany's model is the only one where, by 2019, threat evaluation scores exceed its AI Capability index scores. This is the reason why we can only partially refute the first hypothesis.

A meaningful interpretation of the results of the first hypothesis testing allows to determine official "attitudes" toward changes in the sphere of security initiated by digital technologies. States define the roles and significance of technologies in proportion to their potential capabilities. Governments are successful in integrating technology when evaluating threats in a balanced manner. We demonstrate that states have adapted to changes and designed security systems so that capabilities exceed the perceived threats. With a broader interpretation of the results, it can be assumed that states are in complete control (or try to be in total control) of changes and transformations in the sphere of security. In other words, this is not a "spontaneous transformation", not a one-dimensional reaction to emerging challenges. Successful transformations are made possible by systematic and strategic political activities to assess capabilities vis-à-vis threats. Every state does not take this approach. In some states (for example, Germany), there is an increase in security concerns because governments are paying more attention to the potential for threats associated with AI technology.

The second hypothesis is inspired by the organisational principle of modern security forces of "maximising military power through the use of technology". The results of the models **partly refute** the assumption that the use of digital technologies by states in the sphere of security occurs before there is a public discussion about these types of digital technologies. Despite similar dynamics, the German model shows the proof of application of digital technologies only by 2016, and the models of France and Finland - by 2013. Thus, we cannot state that in all countries, proofs of the application of digital technologies in the sphere of security precede relevant public discussions. Although the US and Swedish models support the hypothesis, these results do not apply evenly to all countries. A possible explanatory mechanism for such an unexpected result lies in the specifics of artificial intelligence technology, which is not originally a military/security technology. It has first developed for commercial (civilian) purposes.

Ambiguous results do not allow us to unequivocally conclude that states strive to maximise power and security by introducing technology. A potential explanation for the variation in country models may be that some states fear threats and risks associated with the technology more than they see benefits from technology application. Therefore, they will first try to obtain a solid evidence base of benefits and only after that governments begin to introduce and apply technologies. Also, variation can be explained by the existing opportunities for technological development and the human capital (presence or absence of highly qualified personnel to develop and implement technologies).

The third hypothesis is constructed around the consensus that states can perceive the same threat differently. We wanted to test whether there is substantial uncertainty about digital technologies creating unnecessary fears. Therefore, the perception of threats will reach maximum values in all analysed countries. The results of the analysis

demonstrate that the hypothesis **is confirmed**. However, *this hypothesis is partially confirmed from the standpoint of empirical accuracy*. This indicator may fluctuate in different periods, even within the same country. Thus, the perception of threats by states is not linear. Nevertheless, maximum scores for this indicator are observed in the mid-2010s with a subsequent decrease. This finding confirms that states define and assess threats differently. We have confirmed the existing consensus on the differences in the definition and evaluation of threats among states and have also clarified it directly regarding digital technologies. Moreover, we observe that the same threats are defined and evaluated differently in different periods. This does not mean that the perception of the same threat should "decrease" over time.

We find significant specifics concerning digital technologies and expand the discussion about the transformational effects of technologies. Moreover, we demonstrate that such transformations are not linear. Instead, states adapt to threats and try to assess the emerging potential for reacting to them proportionately. In turn, digital technologies act both as a tool (reinforcing change) and as a factor that raises fears. This is a significant result, which allows an extended interpretation to argue that the transformation can be associated with positive changes and the emergence of reasonable worries.

The fourth hypothesis has been formulated in the logic of states' readiness for risks and challenges associated with digital technologies. The hypothesis has been **partially confirmed**. We expected to observe a roughly uniform trend in all models of the analysed countries by 2018-2019. The indicator of AI capabilities levels off and remains stable at the same level over several years. States begin to apply the technology with an adequate understanding of the technology's capabilities and limitations of applicability. We expected the threat evaluation indicator to grow rapidly by 2018-2019, which should have indicated a high assessment of risks and threats by states in recent years as more knowledge on the impact of technologies and associated risks became available. We expected the security consistency index to increase from 2006 up to 2010 with approximately similar dynamic and to increase sharply from 2016 up to 2017 because during this period, firstly, the states could observe, in fact, the first real consequences of the use of technologies. Secondly, there was an increase in the politicisation and securitisation of digital technologies.

However, the results do not always reflect these expectations. They are valid for the countries of the USA and Sweden. For these countries, the demonstrated logic fits into the government's approach to changes in national security systems under the influence of AI technologies. Still, it does not fully apply to Germany, France, and Finland. Indeed, in all the countries analysed, the parameter of the AI capabilities reaches a similar maximum value by 2019. Still, the differences between the countries' models are significant for other parameters and their dynamics.

First, the threat evaluation in all country models is not linear, achieving high indicators by 2019. Thus, the model of France, for example, shows a decline in threat evaluation from 2017 to 2019. Second, the Security Consistency index does not necessarily indicate a gradual, systematic growth. Its dynamics are very undulating and unique for each country, except for the US and Sweden. This reinforces existing discussions about the heterogeneity of countries' national security systems and demonstrates new findings on the diversity of government responses to technological challenges in national security. Thirdly, contrary to theoretical expectations, a sharp jump in the security consistency indicator in 2016-2017 was observed only in the French model, when other models declined. This can be explained by the complexity of artificial intelligence technology, where there was more encouraging information about the benefits and potential in earlier times. Over time, potential expectations were replaced by a pragmatic understanding of the limitations of the technology and increasing risks

regarding the application. An alternative explanation may lie in the increasing securitisation of technology and the shift in attitudes towards AI from public discussion to information restriction (secrecy).

It can be argued that the dynamics of the application of AI technology are not linear but undulating and reflects both national features of the conceptualisation of national security and the internal specifics of states with a particular type of digital technology. Despite theoretical concerns, governments have time to assess the risks and benefits of technologies in proportion to the internal features of the security system and, based on this assessment, work on implementation. In other words, the results demonstrate that countries lack both "alarming" and exaggerated fears and leniency on technological changes in national security systems.

At the same time, there are general trends in the countries both in time (for example, similar terms for the adoption of national strategies and programs on artificial intelligence) and in the dynamics of determining the capabilities of AI - approaching the maximum values by 2019. The above may indicate either the functioning of the system of international cooperation both in the field of security and digital technologies or the development and strengthening of competition between countries regarding AI technology. However, the actual mechanism is likely to lie in these explanations that countries continue to increase international cooperation while increasing international competition.

Approbation of the empirical model for evaluating the state security system based on the measurement of security consistency has also demonstrated some limitations. The logic of security consistency indicates that the lower the index, the more "coherent" the state's security system is. In other words, based on the mathematical logic of the difference (we subtract the threat parameter from the capability parameter), the more capabilities a state has and the more threats a state defines for itself, the higher the probability that the consistency indicator tends to zero. However, it is necessary to consider the difficulties in threat evaluation at the level of interpretation and argumentation. For example, the state may "not see" the threat (intentionally or accidentally), but it will exist in reality.

On the contrary, the state may unnecessarily politicise and securitise entire sectors (and actors in these sectors). As a result, the model will show a high indicator of threat evaluation, but these threats will not have an actual embodiment. Similar fluctuations can occur with the indicator of AI capabilities (when the state "exaggerates" the capabilities, etc.). Also, the lack of data for some countries can be a significant limitation. Same, it should be noted that at the moment, it is difficult to talk about the existence of threshold values that would be optimal for the security consistency index.

The proposed model does not purposefully consider military applications of AI technology, including since information about such applications may be confidential (for example, to obtain and (or) maintain an advantage in an actual or potential conflict). Any model that claims to reflect the natural world's complexity is an inevitable simplification of reality. The parameters and indicators used in the model describe not so much reality ("in fact") as its reflection (and understanding), the elements of which are contained in official documents.

Further directions of research using the theoretical and instrumental (methodological) developments of this study may relate both to the topic of data-driven, namely the inclusion of new data (regarding various types of digital technologies) in the subject field of security research, and within the framework of the comparative paradigm of political science.

First, it is possible to expand the study's boundaries and look at the links between the dynamics of national security systems relative to digital technologies and regime

variations. In other words, one of the following directions may be a comparative analysis of models of countries with different regime markers. For example, this is not just about comparing "autocracy" against "democracies". Considering a larger regime spectrum, at least based on the V-Dem typology, consider four countries with different regime characteristics (liberal democracy, electoral democracy, electoral autocracy, closed autocracy). Secondly, applying the model to different types of digital technologies will require the development of a system model that will evaluate the security system not to one type of technology but in an interconnection of different types. Such a model will allow us to look at the dynamics of countries' security systems more comprehensively and in general linkage with the implementation of digitalisation and automation strategies. Thirdly, over time, the additional model building will be required to track changes after 2019. Also, an exciting direction maybe constructs already predictive models based on the existing ones. Predictive models will expand our understanding of the transformational impact of technology and allow for a more informed approach to both practical planning for technology impact and improved policy impact.

References

Allen G.C. (2019) Understanding China's AI Strategy: Clues to Chinese Strategic Thinking on Artificial Intelligence and National Security.Center for a New American Security, Washington, DC

Alter S. Sherer S.A. (2004) A general, but readily adaptable model of information system risk. Communications of the association for information systems. Vol.14, Article 1, p.1-28.

Amoore L., Raley R. (2017) Securing with algorithms: knowledge, decision, sovereignty. Security dialogue. Vol. 48, Iss. 1, p. 3–10.

Aradau C., Blanke T. (2017) Governing others: Anomaly and the algorithmic subject of security. European Journal of International Security, 3 (1), pp. 1-21

Aradau Claudia and Rens van Munster (2007) Governing Terrorism through Risk: Taking Precautions, (Un)knowing the Future. European Journal of International Relations, 13 (1): 89–115.

Baldwin D.A. (1995) Security studies and the end of the Cold War. World politics. Vol. 45, P.117-141

Baldwin D.A. (1997) The concept of security. Review of international studies. Vol. 23, P. 5-26.

Balzacq T. (2011) (ed.) Securitization theory. London : Routledge. 272p.

Baranyi Stephen (2008) The Paradoxes of Peacebuilding Post-9/11. Vancouver: UBC Press.

Bayley D.H. (1975) The Police and political development in Europe. In Tilly C., Ardant G. (eds.). The Formation of National States in Western Europe. Princeton, NJ: Princeton university press, p. 328–339.

Bell C. (2006) Surveillance Strategies and Populations at Risk: Biopolitical Governance in Canada's National Security Policy. Security Dialogue, 37(2): 147–65.

Bigo, Didier, Evelien Brouwer, Sergio Carrera, Elspeth Guild, Emmanuel-Pierre Guittet, Julien Jeandesboz, Francesco Ragazzi, and Amandine Scherrer (2015) The EU Counter- Terrorism Policy Responses to the Attacks in Paris: Towards an EU Security and Liberty Agenda. CEPS Liberty and Security in Europe (81). http://www.ceps.eu/system/files/LSE81Counterterrorism.pdf

Brauch H. G., Spring O. Ú., Mesjasz C., Grin J., Dunay P., Behera N. C., Chourou B., Kameri-Mbote P., Liotta P. H. (2008) (eds.). Globalization and environmental challenges: rreconceptualizing security in the 21st century: Vol. 3. Berlin: Springer Science, Business Media, 1141p.

Briscoe E., Fairbanks J. (2020) Artificial Scientific Intelligence and its Impact on National Security and Foreign Policy, Orbis, Volume 64, Issue 4, p.544-554, https://doi.org/10.1016/j.orbis.2020.08.004.

Brose C. (2019) The new revolution in military affairs: War's sci-fi future. Foreign affairs. Vol. 98, N 3. Mode of access: https://www.foreignaffairs.com/articles/2019- 04-16/new-revolution-military-affairs (accessed: 15.09.2021).

Bulgurcu B., Cavusoglu H., Benbasat I. (2010) Information security policy compliance: an empirical study of rationality-based beliefs and information security awareness. MIS Quarterly. №34, p. 523–548.

Buzan B., Wæver O. (2003) Regions and powers. Cambridge University Press. p.598, DOI: https://doi.org/10.1017/CBO9780511491252

Buzan B., Waever O., Wilde, J. de. (1998) Security: a new framework for analysis. London: Lynne Rienner. 239p.

Buzan B. (1991) People, states and fear. An agenda for international security studies in the Post-Cold War Era. 318p.

Cocelli M., Arkin E. (2017) A threat evaluation model for small-scale naval platforms with limited capability. EEE Symposium Series on Computational Intelligence (SSCI 2016). p.1–8.

Cronin Audrey Kurth (2009) How Terrorism Ends: Understanding the Decline and Demise of Terrorist Campaigns. Princeton, NJ: Princeton University Press

Davis, Zachary (2019) Artificial intelligence on the battlefield: implication for deterrence and surprise. Institute for national strategic security. p. 114–131.

Deshmuhk, A. (2004) A Framework for Online Internal Controls. AMCIS August. 2004, p. 4471-4479.

Edwards P.N. (1997) The closed world: computers and the politics of discourse in Cold War America. Cambridge : MIT press, p. 468.

Enders Walter and Sandler Todd (2006) The Political Economy of Terrorism. New York: Cambridge University Press.

Ensign D., Friedler S.A., Neville S., Scheidegger C., Venkatasubramanian S. (2017) Runaway Feedback Loops in Predictive Policing // eprint arXiv. p. 1–12. http://arxiv.org/abs/ 1706.09847

Farrell Theo (2002) Constructivist Security Studies: Portrait of a Research Program. International Studies Review , Vol. 4, No. 1 p. 49- 72

Fatima S., Desouza K. C., Dawson G. S. (2020) National strategic artificial intelligence plans: A multi-dimensional analysis, Economic Analysis and Policy, Volume 67, p.178-194, https://doi.org/10.1016/j.eap.2020.07.008.

Fatima S., Desouza K. C., Denford J. S., Dawson G. S. (2021) What explains governments interest in artificial intelligence? A signaling theory approach. Economic Analysis and Policy, Volume 71, p.238-254, https://doi.org/10.1016/j.eap.2021.05.001.

Fioramonti L., Kononykhina O. (2015) Measuring the Enabling Environment of Civil Society: A Global Capability Index. Voluntas.,Vol. 26, p.466-487.

Floyd R., Matthew R. A. (2013) Environmental security: approaches and issues. In environmental security: approaches and issues. 322p.

Galanos V. (2018) Artificial intelligence does not exist: Lessons from shared cognition and the opposition to the nature/nurture divide. IFIP advances in information and communi- cation technology. p.359–373.

Garfinkel B, Dafoe A. (2019) How does the offense–defense balance scale? J. Strategic Stud. 42:736–63

Gelman A., Hill J., Vehtari A. (2020) Regression and Other Stories (Analytical Methods for Social Research). Cambridge University Press; 1st edition (July 23, 2020), p.552. ISBN-10: 1107676517

Gheciu Alexandra and Wohlforth William (2018b) The Future of Security Studies. The Oxford Handbook of International Security. DOI: 10.1093/oxfordhb/9780198777854.013.1

Hanlon R.J., Christie K. (2016) Freedom from fear, freedom from want: an introduction to human security. In Freedom from Fear, Freedom from Want. 289p.

Hendershot C and Mutimer D. (2018) Critical Security Studies. The Oxford Handbook of International Security Edited by Alexandra Gheciu and William C. Wohlforth

Hoadley D.S. (2019) Artificial Intelligence and National Security.Congressional Research Service

Hoffman Bruce (2006) Inside Terrorism. New York: Columbia University Press.

Horowitz M., Kahn L., Ruhl C., Cummings M., Lin-Greenberg E., Sharre P., Slayton R. (2020) Policy Roundtable: Artificial Intelligence and International Security. Texas National Security Review. https://tnsr.org/roundtable/policy-roundtable-artificial-intelligence-and-international-security/.

Horowitz M.C., Kahn L., Mahoney C. (2020) The Future of Military Applications of Artificial Intelligence: A Role for Confidence-Building Measures?. Orbis, Volume 64, Issue 4, p.528-543, https://doi.org/10.1016/j.orbis.2020.08.003.

Horowitz, M.C. (2018) Artificial intelligence, international competition, and the balance of power. Texas national security review. Vol. 1, p.37–57.

Jarrahi M.H. (2018) Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. Business horizons. Vol. 61, Iss. 4, P. 577–586.

Jarvis Lee and Lister Michael (eds.) (2015) Critical Perspectives on Counter-Terror-ism. London: Routledge.

Johansson F., Falkman G.A. (2008) Comparison between two approaches to threat evaluation in an air defense scenario. Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics). Vol. 5285, p.110–121. DOI: https://doi.org/10.1007/978-3-540-88269-5_11

Johnson J.S. (2020) Artificial Intelligence: A Threat to strategic stability. Strategic studies quarterly. Vol. 14, p. 16–39.

Kaplan A., Haenlein M. (2019) Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. Business Horizons. Vol. 62, Iss. 1, p.15–25.

Kennedy D. (1987) The move to institutions. Cardoso law review. №8(5), p. 841–988.

Krebs R.R. (2018) The Politics of National Security. The Oxford Handbook of International Security. Edited by Alexandra Gheciu and William C. Wohlforth. DOI: 10.1093/oxfordhb/9780198777854.013.42

Kumar S., Tripathi B.K. (2016) Modelling of threat evaluation for dynamic targets using Bayesian network approach. Procedia technology., Vol. 24, p.1268–1275.

Longino H. (2014) Individuals or populations? In: Cartwright N., Montuschi E. (eds). Philosophy of social science: an introduction. Oxford : Oxford university press, p. 102–120.

Lum K., Isaac W. (2016) To predict and serve? Significance, Vol.13, Issue 5, p. 14-19.

Marquardt K.L. (2020) How and how much does expert error matter? Implications for quantitative peace research. Journal of Peace Research. No57(6):692-700. doi:10.1177/0022343320959121

Mead Earle E. (1944) (ed.). Makers of Modern Strategy: Military Thought from Machiavelli to Hitler. Princeton, NJ: Princeton university press, p.951

Muller B. (2008) Securing the Political Imagination: Popular Culture, the Security Disposi- tif and the Biometric State. Security Dialogue, 39(2–3): 199–220.

Naeem H., Masood A. (2010) An optimal dynamic threat evaluation and weapon scheduling technique. Knowledge-based systems. Vol. 23, Iss. 4, p.337–342.

Nance William D. (1988) Straub Detmar W. An Investigation into the Use and Usefulness of Security software in Detecting Computer Abuse. ICIS 1988 Proceedings. 36. http://aisel.aisnet.org/icis1988/36

Naseem A., Shah S.T. H., Khan S.A., Malik A.W. (2017) Decision support system for optimum decision-making process in threat evaluation and weapon assignment: Current status, challenges and future directions. Annual reviews in control. Vol. 43, p.169–187.

Neack L. (2017) National, International, and Human Security: A Comparative Introduction. Lanham, MD: Rowman & Littlefield, p.236.

Newman Edward (2010) Peacebuilding as Security in "Failing" and Conflict-Prone States. Journal of Intervention and Statebuilding, 4: 305–22.

Pape Robert (2006) Dying to Win: The Strategic Logic of Suicide Terrorism. New York: Random House.

Paret P. (1986) (ed.). Makers of Modern Strategy from Machiavelli to the Nuclear Age. Princeton, NJ: Princeton University Press, 951 p.

Parker G. (1996) The Military Revolution: Military Innovation and the Rise of the West. 1500–1800: 2nd ed. Cambridge: Cambridge University Press, 292 p.

Payne K. (2018) Artificial intelligence: A revolution in strategic affairs? Survival, Vol.60, p.7-32.

Reis J., Santo P.E., Melão N (2019) Artificial intelligence in government services: a systematic literature review. Advances in intelligent systems and computing. p.241–252.

Russell S., Norvig P (2010) Artificial intelligence a modern approach. New Jersey : Prentice Hall, 1152 p.

Sageman Marc (2008) Leaderless Jihad: Terror Networks in the Twenty-First Century. Philadelphia, PA: University of Pennsylvania Press.

Schlag G., Junk J., Daase C. Transformations of security studies: dialogues, diversity and discipline. - Routledge. 2015. - P.250

Seo S., Chan H., Brantingham P.J., Leap J., Vayanos P., Tambe M., & Liu Y. (2018) Partially Generative Neural Networks for Gang Crime Classification with Partial Information // ACM Conference on Artificial Intelligence, Ethics, and Society. https://doi.org/10.1145/3278721.3278758

Sharre P. (2019) Killer apps: The real dangers of an AI arms race. Foreign Affairs. Mode of access: https://www.foreignaffairs.com/articles/2019-04-16/killer-apps (accessed: 20.09.2021).

Simonite T. (2019) When robots can decide whether you live or die. Wired. https://www.wired.com/story/robots-decide-you-live-die/

Tschirgi Necla, Lund M. and Mancini F. (eds.) (2010) Security and Development: Search-ing for Critical Connections. Boulder, CO: Lynne Rienner.

Turobov A.V., Mironyuk M.G. (2021) Empirical model for analysis of the dynamics of algorithmization (artificial intelligence technology) in the field of secu- rity by the example of the USA. Political science (RU). N 3, P. 72–111. DOI: http://www.doi.org/10.31249/poln/2021.03.04

Vallverdu J. (2017) The emotional nature of post-cognitive singularities. In: Callaghan V. et al. (eds). The Technological singularity, the frontiers collection. Germany : Springer-Verlag, Heidelberg, p. 193–208.

Wang L., Zhao J.S. (2016) Contemporary police strategies of crime control in U.S. and China: a comparative study // Crime, Law and Social Change. Vol. 66. No.5. p. 525–537. https://doi.org/10.1007/s10611-016-9641-7

Weldes Jutta, Mark Laffey, Hugh Gusterson, and Raymond Duvall (1999) Cultures of Insecurity: States, Communities, and the Production of Danger. Minneapolis, MN:

Uni-versity of Minnesota Press

Whitman M. E., Mattord H. J. (2011) Principles of Information Security Fourth Edition. Learning. 656 p.

Wirls Daniel (1992) Buildup: The Politics of Defense in the Reagan Era. Ithaca, NY: Cor-nell University Press.

Wolfers A. (1952) “National Security” as an Ambiguous Symbol. Political Science Quarterly. №67(4), p. 481–502.

Wright Q. (1942) A Study of War. Chicago: University of Chicago Press, 466 p.

Zhang Yi, Wu Mengjia, Tian George Yijun, Zhang Guangquan, Lu Jie (2021) Ethics and privacy of artificial intelligence: Understandings from bibliometrics. Knowledge-Based Systems, Volume 222, 106994, https://doi.org/10.1016/j.knosys.2021.106994

Appendix

Appendix 1

AI capability indicator

Measuring and assessing capabilities is a relevant area in political science (from the compilation of indices of state capacity and ending with the study of organizational capabilities). All measurements of capability in the literature are presented either by various regression models to identify the relationships between indicators and their impact on capabilities or by expert interviews (for more details, see [Grant, Verona, 2015], which analyses the primary empirical studies and their problem areas in the organizational capabilities). A separate area of research since 2010 is the creation of composite indices "capabilities" at the national and international levels (for example, Global capability index, Composite Index of National Capability).

Within the framework of this study, the methodological approach of measuring the Global Capability Index [Fioramonti, Kononukhina, 2015] and the Composite Index of National Capability, which considers the costs and financing of the military sphere, is taken as a basis. This approach forms a single indicator for the country, the sum of indicators of specific areas divided by the number of these areas. For example, the Global capability index is formed based on measurements of three areas (dimensions):

- Socio-economic environment: education, equality and gender equality, digital participation, communication technology infrastructure;
- Socio-cultural environment: trust, social tolerance, participation in collective actions, etc.;
- Governance environment: individual and collective opportunities for social and political activity, the rule of law, political dialogue, the legal framework of civil associations and organizations, etc.

Considering an extensive array of parameters in each area, the final calculation of the capability indicator is the sum of indicators by area, considering the coefficients relative to the number of each indicator, divided into three areas.

A similar approach is used when measuring the Composite Index of National Capability. However, only six parameters are calculated (and not three, as with the Global Capability Index). They are summed up and divided by the number of parameters.

Based on the conceptual framework for understanding Artificial Intelligence and the goals of creating an AI capability indicator, I identified four areas:

1. Technological: considers the technological aspects of Artificial Intelligence, namely the use of technology in public administration and the field of security. It is a reflection of the technological capabilities of the technology;
2. Economic environment: considers the funding of AI technology in the context of the overall military budget. I understand that the security sector can have various funding sources, including classified budget items, funding from other articles and sections, etc. I'm forced to rely on public data on financial support for technology directly in the military budget. The financial and economic capabilities of the technology are demonstrated, without funding and economic incentives, the development of technology and especially its application in the field of security is unlikely;
3. Governance environment: considers the number of state-owned companies associated with AI technology and the existence of legal sanctions for the use of AI in the military sphere. Reflects the readiness of the state to develop technology and apply the capabilities of technology;
4. Social environment: considers the employment of the population in the spheres and areas of development and application of AI technology and the number of startups focusing on technology. Reflects the involvement of the public and the ability of the state

to mobilize highly qualified personnel.

The indicated areas with indicators and their brief description of the calculation are shown in Table 1.

Table 1.

Dimension	Indicator	Brief characteristic
Technological (0.25)	UT (use of technology)	Application of AI in government and military sphere (yes / no)
Economic environment (0.25)	MF (military funding)	AI military spending / General military spending
Governance environment (0.3)	SC (state companies)	State-owned AI companies (yes / no)
	LA (legal authorization)	Legal authorization for the use of AI in the military sphere (yes / no)
Social environment (0.2)	JO (job openings)	Employment - Job openings / workforce
	RS (related startups)	Artificial intelligence related startups

Aggregation of these areas is the final stage in forming the AI capability indicator, at which these areas are "weighed". Technological and Economic are given a weight of "0.25" out of 1, and the Governance area - "0.3" due to the socio-political importance of this area in terms of security. For similar reasons, the Social sphere weights of "0.2", despite the importance of civil society and public reaction, in the sphere of ensuring security, the population "knows only what the state considers it permissible to know". In other words, the role of society in matters of technology capability in security will be the least significant next to the rest of the areas.

As a result, the formula (2) for calculating the AI capability indicator is as follows:

$$AIcapability = \frac{0.25 \cdot UT + 0.25 \cdot MF + 0.3 \cdot (SC + LA) + 0.2 \cdot (JO + RS)}{4} \qquad (2)$$

where the first UT indicator (an indicator of the application/use of AI in public administration and military sphere) - refers to the Technological sphere; one indicator relates to the Economic sphere: MF - indicator of AI funding in the military sphere; two indicators relate to the area of Governance: SC - an indicator of state-owned companies in the field of AI, LA - an indicator of the existence of legal sanction for the use of AI technology for military purposes; two indicators relate to the Social sphere: JO is the indicator of employment in the field of AI, RS is the indicator of startups in AI.

Appendix 2

Threat evaluation indicator

Threat evaluation is devoted to a separate section of specialized literature, often united by a single direction - Threat Evaluation and Weapon Assignment (TEWA) [for example, Cocelli, Arkin, 2017; Johansson, Falkman, 2008; Naeem, Masood, 2010; Naseem et al., 2017; Kumar, Tripathi, 2016]. TEWA is considered the main component of the Air Defense system (ADS). Recently, the most common are models based on Bayesian networks [Kumar, Tripathi, 2016], fuzzy logic / fuzzy inference rules [Naeem, Masood, 2010; Johansson, Falkman, 2008], and decision support systems [Naseem et al., 2017].

TEWA models based on Bayesian networks allow overcoming uncertainties (incompleteness of information about objects; lack of information about the state of infrastructure; probability and / or randomness in the control of a specific weapon, etc.) in modeling. In the Bayesian approach, the variables of the TEWA model contain probability limits or probability distributions, which allows evaluating threats even in the event of incomplete data.

In turn, models based on fuzzy logic rules (the concept of fuzzy sets) are built according to the principle of membership functions. In the theory of crisp sets, the members x of the universal set X are either members or not members of the set $A \subseteq X$. Thus, the values assigned to x fall within the range, indicating the degree of membership of the element in the (fuzzy) set in question. Larger values indicate a higher degree of membership, while lower values indicate a lower degree. In other words, it can be difficult for a specific context to define clear boundaries (measures / parameters) of a variable; therefore, a membership function is used, which calculates the indicators of a variable that are similar in terms of the degree of membership. Two critical remarks: (1) membership grades in the rules of fuzzy sets do not apply to the assessment of probability [Johansson, Falkman, 2008], (2) there is no unity in the scientific, academic environment regarding the rules of fuzzy logic themselves - some researchers, in principle, do not recognize this approach, considering it too abstract.

The use of decision-making systems in TEWA models allows considering the indicators of geographic information systems (GIS mapping of vulnerable assets), supplementing the model with forecasting methods, distributing and assessing the "cost-effective purpose of weapons" [Naseem et al., 2017:169]. Thus, the decision-making system allows expanding the list of parameters in the model and evaluating additional factors (for example, economic feasibility) when assessing threats. TEWA models themselves with a decision-making system built based on machine learning (the most popular models with a decision tree; more advanced models are based on deep learning, such as the Tactical Air Combat Decision Support System), game theory, and dynamic Bayesian networks.

TEWA models imply a complete threat response and countermeasures system, which is beyond the scope of this study. Thus, based on available research on threat evaluation, including a large-scale meta-analysis (156 publications of TEWA studies from 1975 to 2016) conducted by Nasim, Shah, Khan, and others, an empirical model for calculating threats has been developed.

Threat evaluation is often represented by two [Naeem, Masood, 2010] or three [Naseem et al., 2017] stages. The two-stage model involves (1) assessing and ranking the threat, and correspondingly (2) weapon assignment. The three-stage model consists of (1) threat perception evaluation, (2) threat index calculation, and the corresponding (3) weapon assignment. In addition, each threat assessment model must consider the threat's correspondence to the protected object/asset (defended asset).

Each step includes calculating specific characteristics [Naseem et al., 2017]. So, at the stage of threat perception, the critical parameters of a specific type of weapon (for

example, a cruise missile) are calculated: speed, height, radar cross-section / effective surface of scattering of radar waves (Radar cross-section), manoeuvrability, dive angle, attack approach, etc. The calculation of the threat index includes the characteristics of speed, direction, altitude, threat of manoeuvring, distance from vulnerable points, threat of lethality from the knowledge base, etc. defensive weapons and contains parameters:

1. The threat is assigned to the weapon based on the threat index.
2. The threat with the highest threat index (TI) is assigned first.
3. The threat is assigned to the weapon with the highest probability of being killed.

Generally, the TEWA model is an assessment and ranking of a threat according to the characteristics listed above and a calculation of the correspondence of the assessed threat to the protected objects/assets, followed by the definition of weapons to ensure the security of the objects and neutralize the threat.

For this study, Threat Perception metrics (TP) and Threat Type metrics (TT) are conceptualized as follows (all characteristics are proxy metrics), as shown in Table 2.

Table 2.

Stage	Characteristic
Threat perception (based on strategies and legal acts)	1. Education (Ed) - attackers will face fewer obstacles when exploiting AI vulnerabilities without educating and retraining the population to keep pace with technological change and different types of threats[17]. An educated population will also reduce unintentional mistakes. Thus, any state that indicates the threats to AI, at the policy level, adapts the educational system to increase educational awareness of AI technology. In this study, the education indicator will be calculated based on the indicators of the country's results in the PISA rating exclusively in mathematics since mathematics education is fundamental to the development and application of AI technology. The calculation is the ratio of the country's rating in a particular year to the highest possible rating indicator[18]. 2. Regulation (Reg) - the perception of government threats is reflected in the regulatory domain. The proportion of regulatory legal acts that secure (to one degree or another) threats from the artificial intelligence technology will be calculated as follows: the number of legal acts with a mention of AI technology per the total number of legal acts per year. 3. domestic patents (DP) - determination of the perception of threats from the state is impossible without scientific and technical research and development. To calculate this indicator, only domestic patents on the subject of "Artificial Intelligence" will be taken into account. The indicator will represent the proportion of the number of internal patents by topic to the total number of patents in the year under review.
The type of threat	This indicator was intended to demonstrate specific (and computed)

[17] "AI Using Standards of Mitigate Risk" Public-private analytic exchange program 2018. US Department of Homeland Security & Office of the Director of National Intelligence United States of America. Official website of the Department of Homeland Security. – Mode of access: https://www.dhs.gov/sites/default/files/publications/2018_AEP_Artificial_Intelligence.pdf (accessed: 20.04.2021).

[18] In each test subject, there is theoretically no minimum or maximum score in PISA; rather, the results are scaled to fit approximately normal distributions, with means for OECD countries around 500 score points and standard deviations around 100 score points. About two-thirds of students across OECD countries score between 400 and 600 points. Less than 2% of students, on average across OECD countries, reach scores above 700 points, and at most a handful of students in the PISA sample for any country reach scores above 800 points. – Mode of access: http://www.oecd.org/pisa/pisafaq/ (accessed: 20.04.2021).

(based on research, indices and indicators of reports)	metrics of the type of threats. Unfortunately, there are currently no robust statistics for directly relevant AI technologies. Almost all reports and studies are “doctrinal” in nature. Namely, they indicate some threat or AI has a specific threat characteristic, but there is no statistical or mathematical expression. In this connection, the decision was made to construct a numerical binary indicator: 0 - the type of the threat was absent in the country in the year under review; 1 - the type of the threat was present (recorded / documented) in the country in the year under review, according to the following list of types of AI threats[19]: 1.1. Threats to critical infrastructure; 1.2. Cyberthreats / Cyberattacks using AI (AI expands the vectors of threats vulnerable to cyberattacks by detecting and exploiting weaknesses in the system); 1.3. Disinformation companies (including Deepfakes); 1.4. Violation of human rights (meaning, bias algorithms, personal data violations, threats to biometric data, etc.). According to the specified list for each specific year, an indicator from "0" to "4" will be formed. If some type of threat was recorded in the studied year - "1" is set, if not - "0".

The calculation of the threat perception index (TP) is the sum of the indicators of Education (Ed), Regulation (Reg), and Domestic patents (DP). Therefore, the indicator of the type of threats (TT) will take a value from "0" to "4".

The proposed threat assessment approach will be based on the Threat Perception (TP) assessment and the Threat Type (TT) assessment to Defence Assets (DA). Threat Perception (TP) for Artificial Intelligence technology is based on analyzing countries' national security strategies and security regulations. Threat type (TT), in turn, is typologies from the relevant theoretical overview, expert assessments, indicators, and indices from international and national reports.

Defenced Objects/Assets (*DA*) - what, in essence, the threats of AI technology are aimed at. Given the specifics of the focus of this study, from a practical point of view, it is impossible to calculate the total number of protected assets/objects. Therefore, all areas mentioned in the national legal acts of the country as protected will be taken into account (numerical binary indicator). An exemplary list[20] is shown in Table 3.

[19] The list was formed on the basis of an analysis of relevant national legal acts and reports (for example, National Security Commission on Artificial Intelligence, Interim Report, 2019 // The National Security Commission on Artificial Intelligence (NSCAI) - Mode of access: https://www.nscai.gov / reports (accessed: 05.05.2021); Artificial Intelligence and UK National Security: Policy Considerations, from RUSI, 2020 // RUSI - Mode of access: https://rusi.org/publication/occasional-papers/artificial-intelligence-and-uk-national-security-policy-considerations (accessed: 05/05/2021); etc.),
as well as international reports (for example, Attacking Artificial Intelligence: AI's Security Vulnerability and What Policymakers Can Do About It, from Harvard Kennedy School. Belfer Center for Science and International Affairs, 2019 // Belfer Center, Harvard Kennedy School - Mode of access: https://www.belfercenter.org/sites/default/files/2019-08/AttackingAI/AttackingAI.pdf (accessed: 05/05/2021))

[20] An example list is illustrated in the section on AI from the National Security Strategy of the United States of America. December 2017. - Mode of access: https://www.whitehouse.gov/wp-content/uploads/2017/12/NSS-Final-12-18-2017-0905.pdf (accessed: 19.04.2021).

Table 3.

The area determined by the normative legal act	Numerical indicator (definition in regulatory legal acts)
Intelligence	0 - absent; 1 - present.
State-owned companies data	0 - absent; 1 - present.
Personal data	0 - absent; 1 - present.
Self-driving cars	0 - absent; 1 - present.
Autonomous weapons	0 - absent; 1 - present.
…	…

A separate element of the calculation is the indicator of the significance / value of assets / objects. Kumar and Tripathi point to the high role of accounting in the model for the "protection value assigned by the decision maker" and "lies between 0 and 1" [Kumar, Tripathi, 2016:1270]. Such an indicator is necessary to consider the priority distribution of threats from political actors and decision-makers in the field of security. Therefore, the proposed model will also have a Protection Value (PV) and take a value from "0" to "1", but with a distribution of weights. The characteristics of the assessment are shown in Table 4.

Table 4

Protection Value	Numerical indicator	Weigh t	Characteristic
AI technology in the National Security Strategy	0 - absent; 1 - present.	0.3	The consolidation of technology in the national security strategy is the highest "recognition" by the state of the importance of both the technology itself and the potential threats.
Separate government agency on AI issues	0 - absent; 1 - present	0.3	If within the framework of the system of state bodies, a particular body dedicated to AI technology has been created in the structure of security, we can assert that the state determines the high priority of this technology.
National AI Strategy	0 - absent; 1 - present	0.2	The national strategy on AI, although not directly related to the field of security, however, due to the specifics of the technology itself, to a certain extent will also regulate security issues.
Formulated definition of AI in the national strategy for digital transformation	0 - absent; 1 - present	0.1	The absence of separate regulation (or, at least, the doctrinal consolidation of intention in the form of a national strategy) with a focus on AI reflects not so strong "interest" on the part of the state.
Government-owned (or	0 - absent;	0.1	Accounting for technology companies reflects the

government-affiliated) military and/or security AI technology companies	1 - present		state's own (not foreign) technical capabilities (computing power, software, etc.) to identify technology threats.
Maximum value of the indicator			**1**

The most critical point in computing threats is taking factors and individual categories into account. For example, when creating the International Security Index[21], the PIR Center points out the priority of military factors over others (political factors, terrorism, man-made and natural factors, economic factors)[22]. Moreover, it is supposed to rank within each group in global, regional, and local safety factors. In the proposed approach, these factors are also considered. Based on the general theoretical framework of the study - the sectoral approach to the analysis of the security sphere proposed by the Copenhagen School, the indicator - Threat Factor (TF) is introduced. The Threat Factor in this model is presented as the ratio of each country's score to the overall score for Barry Buzan's five security sectors.

Table 5

Security Sector	**Description [Buzan, Waever, Wilde, 1998]**	**Numerical indicator**
Political	Threats to sovereignty, attacks on legitimacy and authority	0 - absence; 1 - local; 2- regional; 3 - international
Military	All military issues are defined as security threats (except for peacekeeping purposes and disaster relief)	0 - absence; 1 - local; 2- regional; 3 - international
Economical	Threats to the economic stability of the state and to some aspects of the economic system (for example, the banking sector)	0 - absence; 1 - local; 2- regional; 3 - international
Ecological	All environmental issues on the territory of the national borders of the state, including global international climate challenges related to the state (global warming, pollution, the ozone layer, etc.)	0 - absence; 1 - local; 2- regional; 3 - international
Societal	Issues of collective identity (linguistic, cultural, religious, etc.) and the balance of identity in the state (for example, the ratio of different cultures and multiculturalism)	0 - absence; 1 - local; 2- regional; 3 - international

Threat Factors (TF) are calculated as the proportion of the sum of the assessment

[21] International Security Index (iSi). Description and calculation methodology // PIR Center. - Mode of access: http://www.pircenter.org/media/content/files/9/13462438640.pdf (accessed: 04/21/2021).

[22] ".. The general political or economic crisis can be somehow overcome, the consequences of even a global environmental catastrophe, including those caused by the actions of terrorists, can be neutralized, albeit not completely ... As for a global nuclear war, this phenomenon can be considered completely irreversible and "lethal" for all mankind "(p. 5) - Mode of access: http://www.pircenter.org/media/content/files/9/13462438640.pdf (accessed: 04/21/2021).

scale to the maximum number of assessments of security sectors.

The general formula for Threats evaluation (3) for research purposes is presented below:

$$TE = \frac{PV \cdot (TP + TT)}{DA \cdot TF} \quad (3)$$

Where *PV* is an indicator of significance - protection value (PV∈ [0; 1]); *TP* is the threat perception indicator; *TT* is an indicator of the nature / type of threat (TT ∈ [0; 4]); *DA* is an indicator of the number of protected objects (DA = {DA1, DA2,… DAn}); *TF* is an indicator of threat factors.

In other words, the logical content of the formula can be represented as follows: the Threat evaluation indicator (*TE*) is the *relation* the sum of the threat perception indicator (*TP* - how threats are presented at the level of regulatory legal acts - an act of political will) and the indicator of the nature/type of threat (*TT* - how threats are presented in reports, relevant literature) multiplied by the Protection Value (*PV* - as decision-makers and political actors assess the importance/significance of threats) *to* the indicator of the number of protected objects (*DA* - objects/assets to which threats are directed) by the indicator of threat factors (*TF* - assessment and ranking of security sectors to which threats belong).

Appendix 3

USA

The definition and justification of periods were based on the US national security strategy and federal regulations on using / implementing technologies for more than 20 years. Therefore, eight-time periods are subject to analysis, namely, the years: 1999, 2002, 2006, 2010, 2012, 2015, 2017, 2019. The analysed periods cover

1. all the critical dates for defining and regulating the US national security strategy (national strategies often determine the main vectors of the direction of the security sphere),
2. the main regulatory documents in the field of technological development (starting from the concept of e-government, ending directly with regulation AI technologies),
3. the last four Administrations, which allows you to track the immediate dynamics of changes.

After calculating the model, the following indicators were obtained for the AI capability and Threat Evaluation parameters and the Security Consistency index. The results are presented in Table 6.

Table 6

Year	AI capability	Threat Evaluation	Security Consistency
1999	0	0	0
2002	0	0	0
2006	0	0	0
2010	0.6538471962	0.1986746297	0.4551725665
2012	1.091858052	0.3826546118	0.7092034399
2015	1.138222821	0.4130319641	0.7251908565
2017	1.149086336	0.4498542608	0.6992320752
2019	1.189734347	0.6837136475	0.5060206997

The results are presented in graphical form in Figure 5. The graph allows us to track the dynamics of changes in the Security Consistency index and each AI Capabilities and Threat Evaluation indicator over the years. It is noteworthy how the “interest” of the state and the definition of threats change with the development of the technology itself, as well as the rapid growth of AI capabilities (in the years of maximum progress in computing power and the emergence of new algorithms) more and more new “discoveries” in AI technologies are no longer revolutionary, but evolutionary. Interestingly, before 2008 the US government seemed to have ignored artificial intelligence technology in security, and after that, rapid growth began both in threats evaluation and in identifying capabilities.

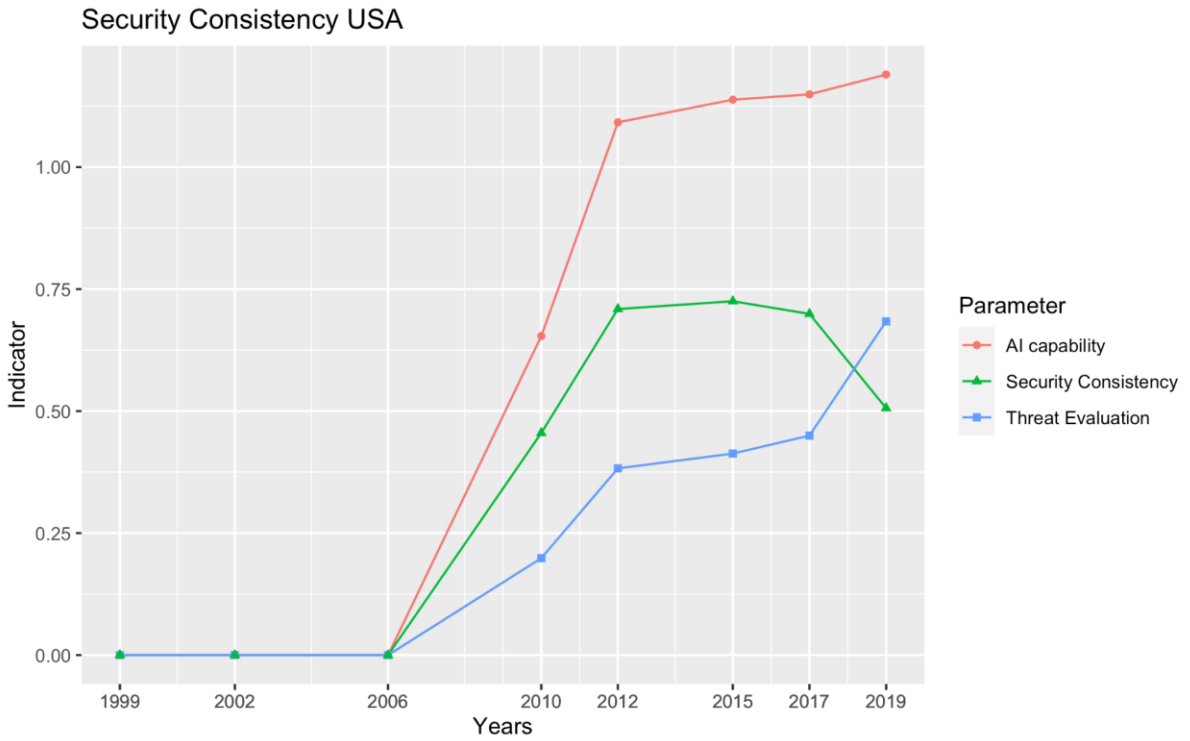

Fig.5 Graph of Security Consistency, AI Capabilities, Threat Evaluation USA (1999 to 2019)

Sweden

In the case of the Swedish analysis, the definition and justification of the time intervals to be taken into account in the model cover ten years. However, the starting point of the analysis was 2010 because it was this year that one of the country's first strategies for introducing digital technologies in the healthcare sector at the national level (National eHealth) was adopted.

The specified period covers:

1. The critical dates for determining and regulating the national security strategy (including cyber security)
2. The primary regulatory documents in the field of technological development (starting from the concept of the e-health system, ending directly with the regulation of AI technologies)
3. The last four electoral cycles of elections of the leader of the Government of Sweden make it possible to trace the dynamics of changes.

After computing the model, the following parameters were obtained for the AI capability, threat evaluation, and security consistency index. The results are presented in Table 7.

Table 7

Year	AI capability	Threat Evaluation	Security Consistency
2010	0	0	0
2012	0.5	0.2192023174	0.2807976826
2015	0.75	0.3576949926	0.3923050074
2017	1.05	0.4236833866	0.6263166134
2019	1.050734363	0.779946551	0.2707878116

The results are graphically presented in Figure 6. The analysis results demonstrate the gradual development and interest of the Swedish government in both the issues of artificial intelligence technology and its potential in security and the steady increase in threat evaluation. However, in 2017 assessment of AI capabilities has reached a kind of plateau, but the evaluation of threats, on the contrary, has intensified. This may indicate an increase in the meaningful understanding of risks on algorithmic systems.

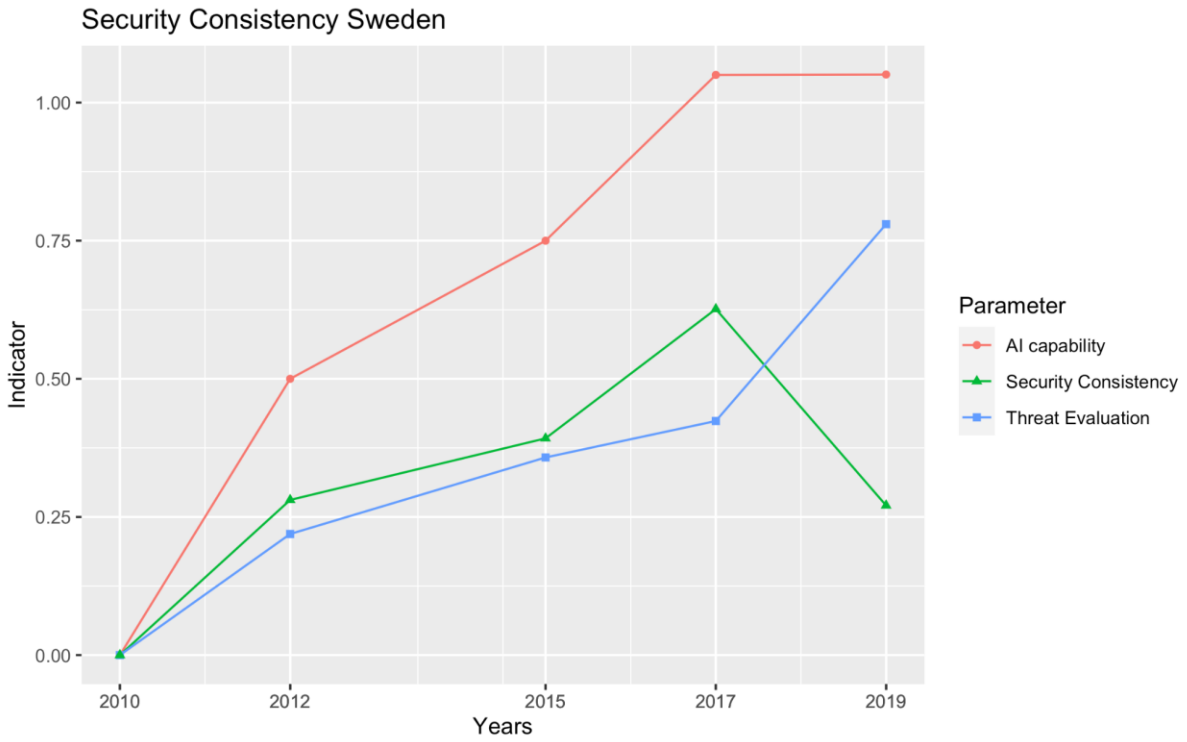

Fig.6 Graph of Security Consistency, AI Capabilities, Threat Evaluation Sweden (2010 to 2019)

Germany

The analysis of Germany began in 2003 when the Social Code (book 12, paragraph 64j) first mentioned digital care applications. The period is 15 years.

Formally, the indicated period of 15 years allows covering only two administrations (considering the "political survival" of Frau Merkel). However, in terms of content, we can trace the discussion about the need for a security strategy in Germany and the development of ICT and the digital component.

After computing the model, the following parameters were obtained in terms of AI capability, Threat Evaluation, and Security Consistency. The results are presented in Table 8.

Table 8

Year	AI capability	Threat Evaluation	Security Consistency
2003	0.3	0.6251324876	-0.3251324876
2008	0.3	1.065037768	-0.7650377679
2012	0.3	0.1547313253	0.1452686747
2013	0.3	0.2380659193	0.06193408073
2016	0.3	0.2882713103	0.2117286897
2018	1.050663548	0.7049233076	0.3457402402
2019	1.050762188	1.208400727	-0.1576385386

The results are graphically presented in Figure 7. The results of the German analysis do not show a gradual increase, as in Sweden and the United States. Noteworthy is the plateau in the definition of AI capabilities until 2016, followed by a rapid leap. In turn, the threat evaluation by the German government can be characterized in waves: with peaks in 2008, a subsequent decline, and rapid development since 2016 and to this day. This dynamic could mean a shift in the government's focus on national security and the role of digital technologies in it over time. Interestingly, Germany is one of the few countries that conceptualizes artificial intelligence technology in the environmental sector field, though starting from 2020.

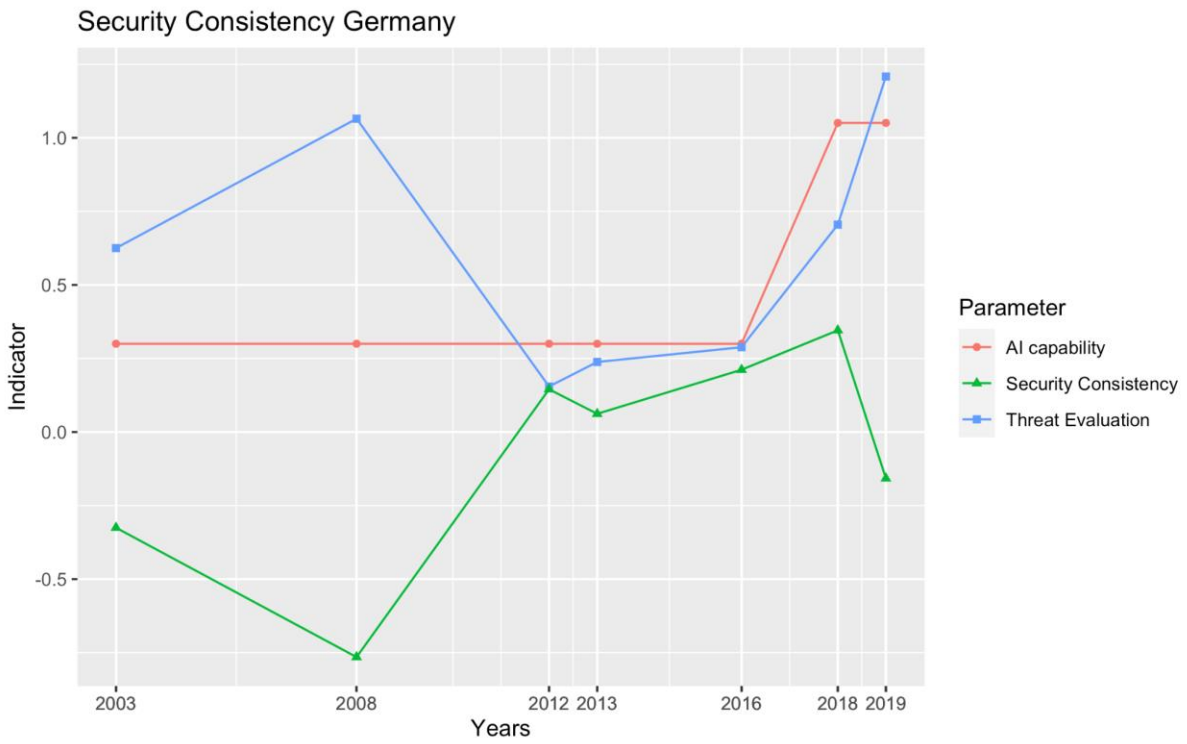

Fig.7 Graph of Security Consistency, AI Capabilities, Threat Evaluation Germany (2003 to 2019)

Finland

The time limits for the analysis of Finland have been defined since 2004, namely, since the adoption of two relevant legal acts. The first is dedicated to creating a secure information society in the country, and the second is a specific program of education,

training, and research to support that very information society. As a result, a 15-year time period is subject to coverage.

The specified period covers the last three administrations of the Finnish government, allows you to trace the dynamics of the introduction of ICT and digital technologies, and reflects the main changes in the country's national security system.

AI capability, Threat Evaluation, and Security Consistency scores are shown in Table 9.

Table 9

Year	AI capability	Threat Evaluation	Security Consistency
2004	0	0	0
2006	0	0	0
2009	0	0.3661503444	-0.3661503444
2013	0	0.1656428571	-0.1656428571
2017	1.05	0.6419444445	0.4080555556
2019	1.050000023	0.6921428571	0.3578571658

The results are graphically presented in Figure 8.

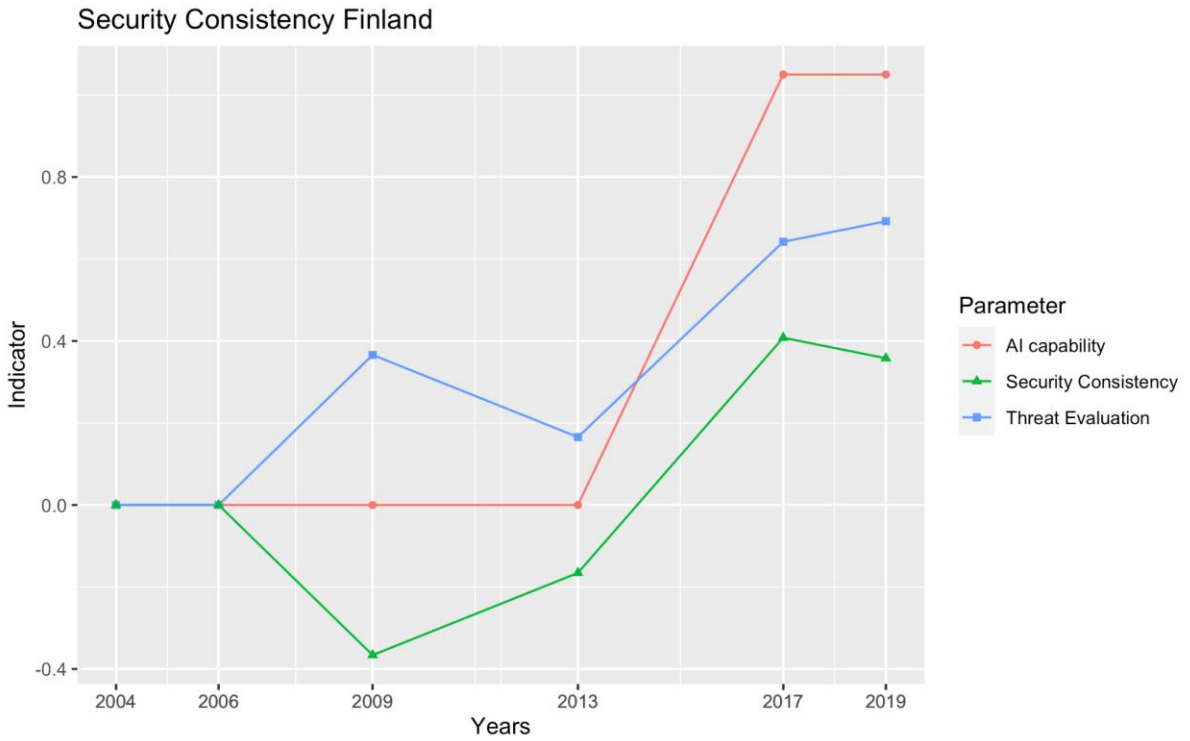

Fig.8 Graph of Security Consistency, AI Capabilities, Threat Evaluation Finland (2004 to 2019)

Finland shows a more similar pattern to Germany, which differs from the stable growth of the US and Sweden. Interestingly, until 2013 the Finnish government identified more threats than capabilities, and only after 2013 did the capabilities of AI technology show an increase. Threat evaluation by the Finnish government comes in waves, as with Germany, which could mean a shift in the government's focus on technology in national security over time, with the first peak in 2009 and the second in the present day. The

definition of capabilities is more uniform: a long plateau until 2013, followed by an increase until 2017. To date, we are seeing another plateau in the definition of opportunities from the Finnish government.

France

The analysis of France starts from 1996, which is the earliest period of all analyzed countries. Such an early date is based on France's significant national plan for digitization (National digitization plan), which prepared the basis for the subsequent transformations of the country's politics and public administration. The specifics of French security legislation is that the national security strategy/plan is presented in the form of White Papers, which cover not only security issues but also the military sphere, issues of international peace, and others. The total time coverage is 23 years.

Thus, such a long period allows us to analyze the work of the last four administrations of the President of France, taking into account both the specifics of national security policy and the impact of digitalization.

After calculating the model, the following indicators were obtained for the AI capability and Threat Evaluation parameters and the Security Consistency index. The results are presented in Table 10.

Table 10

Year	AI capability	Threat Evaluation	Security Consistency
1996	0	0	0
2004	0	0	0
2008	0.5	0	0.5
2011	0.5	0.3192920419	0.1807079581
2013	0.5	0.5133615182	-0.01336151819
2015	0.75	0.4742145771	0.2757854229
2017	0.75	0.6473350761	0.1026649239
2019	1.050000443	0.5119223978	0.5380780452

The results are presented graphically in Figure 9. The results of the analysis of France show the most "unstable" dynamics. Namely, we can observe many peaks and declines:

1. Notably, not a single parameter in France acquires negative values, which is similar to the United States and Sweden and different from Germany and Finland.
2. Before 2004 The focus of the French government was neither on evaluation threats nor on identifying capabilities of AI technology. In turn, initially, the French government began to determine the capabilities of technology, and only in 2008 paid attention to threat evaluation. This pattern continues to this day.
3. The assessment of threats in France itself has a pronounced wave-like nature with peaks in 2013 and 2017. At the same time, the definition of capabilities is associated with reaching a plateau, but today we see a clear interest on the part of the state in the potential of technologies.

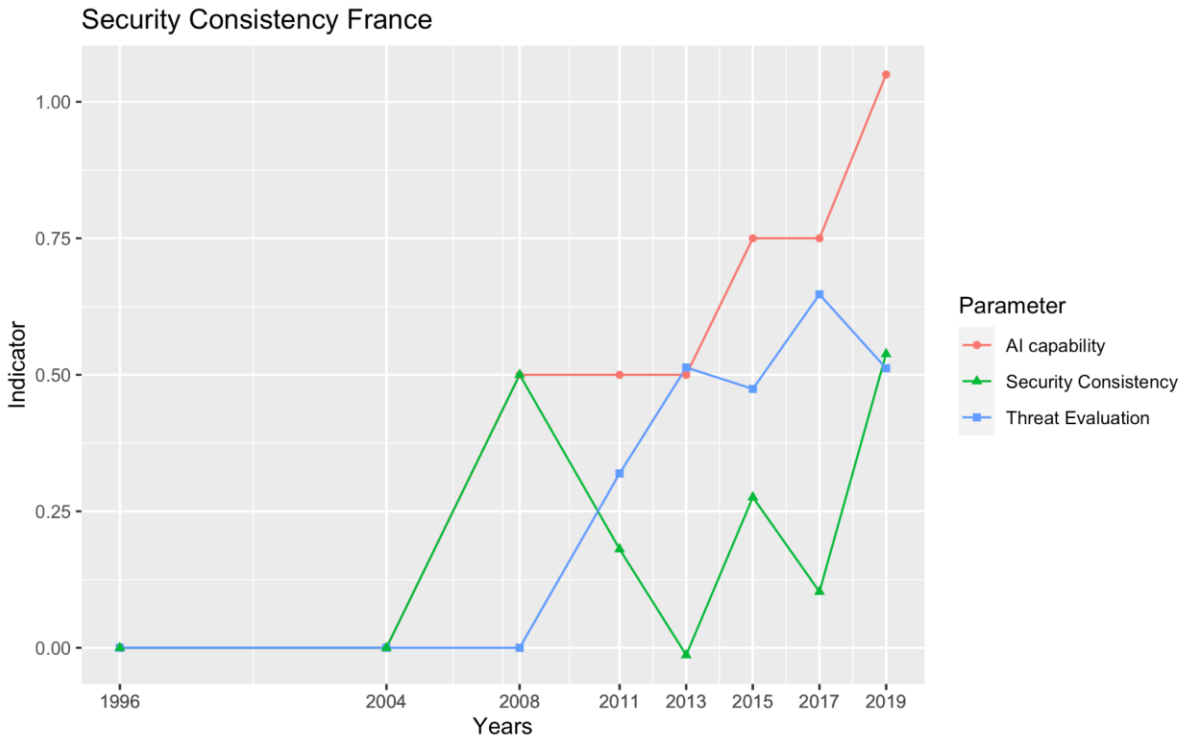

Fig.9 Graph of Security Consistency, AI Capabilities, Threat Evaluation France (1996 to 2019)

Interestingly, France is one of the few countries that pay special attention to the environmental sphere (there is a similarity with Germany) and directly raises human rights violations when using algorithms.

Appendix 4

Simulation and fake data

Indicator	Fake data of indicator	Standard deviation of the model	Result of the simulation	
MF (Military funding) Economic environment, AI Capabilities	Min: 0 Max: 0.9255269	0.7071068	Min: 0.6218285 Max: 0.8532102 (within 1sd)	**The model is stable**
UT (use of technology) + Test + AA (algorithm accuracy) Technological environment, AI Capabilities	Min: 0 Max: 5.450846	0.7071068	Min: 0.02811615 Max: 1.390828	**The model is sensitive to the indicator (towards max)**
The Governance environment (SC (state companies) + LA (legal authorization)) is not subject to testing because the minimum indicator (0) and the maximum indicator (2) are present in the model for different years				
JO (job openings) + RS (startups) Social environment , AI Capabilities	Min: 0 Max: 2.712146	0.7071068	Min: 0.5116695 Max: 1.054099	**The model is sensitive to the indicator (towards max)**
Threat Perception (TP) Threat Evaluation	Min: 0 Max: 1.976549	0.7071068	Min: 0.8982343 Max: 0.5276314	**The model is stable**
Threat Type indicator (TT) is not subject to testing because the minimum indicator (0) and the maximum indicator (4) are present in the model for different years				
Protected Object/Assets (DA) is not subject to testing because the minimum indicator (0) and the maximum indicator (8) are present in the model for different years				
The indicator Protection Value (PV) is not subject to testing because is in the range between 0 and 1 and is present in the model for different years				
Threat Factor (TF) Threat Evaluation	Min: 0.4444445 Max: 1.3695864	0.7071068	Min: 0.007321798 Max: 1.115742	**The model is sensitive to the indicator (to both min and max)**

Author

Aleksei Turobov
HSE University (Moscow, Russia). Lecturer, School of Politics and Governance. Research Fellow, Faculty of Social Science.
Email: aturobov@hse.ru

Acknowledgments: The author is grateful to colleagues at HSE University for their astute observations and criticism: A. S. Akhremenko, I. II Marques and K. L. Marquardt. For useful comments, help and support, the author thanks his supervisor M. G. Mironyuk.