Artificial Intelligence and Political Trust

DANICA FINK-HAFNER AND KATARINA KAIŠIĆ

sciendo

Politics in Central Europe (ISSN 1801-3422)
Vol. 21, No. 3
DOI: 10.2478/pce-2025-0014

Abstract: *This article is based on understanding political trust as a relational concept. In the frame of liberal democratic thought, the article refers to the relationship between citizens and political institutions, and the role played by political trust as the crucial glue that holds it together, thereby enabling the democratic political system to function successfully. Based on a narrative literature review, we have developed a conceptual model that illustrates the complex relationship between trust in artificial intelligence (AI), political trust and the broader context in which these relationships are co-constructed by various actors.*

Keywords: ***political trust, artificial intelligence (AI), democracy, AI governance***

Introduction

Issues involved in the relationship between artificial intelligence (AI) and political trust have been overlooked for a long time (Robles & Mallinson 2023a) but have appeared on the public and research agenda recently for multiple reasons. Among these issues are increasing concerns about the negative impacts of AI technologies on human rights and security (Gillespie et al. 2023: 7), the inclusion of such technologies in public administration (Kleizen et al. 2023: 10) and the impacts of biases in AI-supported policymaking and implementation (Valle-Cruz et al. 2020: 5). This article focuses on the relationship between trust in AI and political trust, as both have faced significant challenges (Nie 2024). The central research question is: What factors influence the relationship between trust in AI and political trust? The aim is to develop a conceptual model, grounded in a literature review, that identifies the factors affecting trust in AI and political trust, and the relation between the two.

While public trust and political trust are often confused in public discourse – especially since in literature, the former is sometimes used synonymously with the latter – this article emphasises the need to distinguish between them. Our understanding of political trust is relational, reflecting the connection between citizens, political institutions and democracy – a concept that draws on Easton's (1975: 437) view of political support as citizens' perceptions of the outputs and performance of political authorities. Easton's definition of political support is also widely used in empirical social science research through indicators of trust in specific political institutions, while we use the term 'public trust' in relation to AI in terms of citizens' trust in new technologies grouped under the term 'AI'. This differs from political trust.

Recent literature on AI has shown a growing but delayed interest in the relationship between AI and democracy (Nie 2024); however, systematic research on general or specific connections between them is still lacking. Existing studies either note some links between trust in AI and political trust (see, for example, Kreps & Kriner 2023; Nie 2024) or highlight contextual factors affecting trust in AI (e.g. Afroogh et al. 2024; Li et al. 2024). Furthermore, most empirical research focuses on individual citizens' trust in AI and political institutions (e.g. Nie 2024; Zuiderwijk et al. 2021), while other works also recognise other actors that influence this relationship (e.g. Ryan 2020). The fragmentation of literature has made it difficult to fully grasp the complex interplay of various factors impacting the relationship between trust in AI and political trust.

This study aims to address the existing gap in literature by organising ideas around the complex relationship between trust in AI and political trust, and the broader context in which this relationship is co-constructed by various actors. We do this by developing a conceptual model based on a literature review related to AI, findings from literature on political trust and expert judgments drawn from long-term academic experience. At this stage, the model provides a foundation for more systematic empirical research, but also further theoretical and conceptual developments in the field.

This article begins with the theoretical and methodological framework, where we explain how we found the literature included in our research and define the two main concepts from the title (AI and political trust). We follow this with a literature review divided into subsections, with each subsection covering a particular segment of the studied relationships. These segments are then synthesised into a model. We conclude with suggestions for further research.

Theoretical and methodological framework

Trust and trustworthiness

Trust is a relational concept. In psychology, it entails the intention of a truster (A) to accept vulnerability based upon positive expectations of the intentions or behaviour (X) of a trustee. However, in some other disciplines, context is also exposed as a factor impacting this relationship (Pillulta 2005: 406).

In research, trust is usually viewed as a quality of an individual. On the other hand, trustworthiness is defined as 'an informal social contract where principals authorise agents to act on their behalf in the expectation that the agent will fulfil their responsibilities with competency, integrity, and impartiality despite conditions of risk and uncertainty' (Norris 2022a: 3).

Defining AI and trust in AI

There is no universal definition of AI. For the purpose of this article, we use a more recent maximalist definition of AI by Rai, Constantinides and Sarker (2019: iii) as 'the ability of a machine to perform cognitive functions that we associate with human minds, such as perceiving, reasoning, learning, interacting with the environment, problem solving, decision-making, and even demonstrating creativity'.

In defining a trustworthy AI, an increasingly long list of criteria has been developed that a particular AI system needs to fulfil to earn trustworthiness. In 2019, three main criteria seem to have prevailed in the policy arena: AI needs to be lawful, ethical and robust (European Commission 2019). In 2022, researchers published a more detailed list of criteria: robustness, generalisation, explainability, transparency, reproducibility, fairness, privacy preservation and accountability (Li et al. 2022). These criteria also resonate with more recent lists of criteria published to guide enterprises, which include accountability, explainability, fairness, interpretability and transparency, privacy, reliability, robustness and security (Gomstyn, Jonker & McGrath 2024). Nevertheless, critics have exposed a tendency for regulators to narrowly understand 'trustworthiness' in terms of the 'acceptability' of the risks associated with AI (Laux, Wachter & Mittelstadt 2024).

Defining political trust

In the literature on AI, the term 'public trust' is often used. Usually, it is not conceptualised in more detail; however, when looking at elements related to public trust, we can say that it is a synonym for the political science term 'political trust'.

Political scientists have distinguished political trust from legitimacy. The concept of the latter refers to the whole permanent political system, while political trust refers to the rulers in power (Dogan 1992: 121). The concept of political trust has survived despite its fuzziness and elusiveness (Carstens 2023: 298), as well as criticism that it has been theoretically dubious (Hooghe & Zmerli 2011: 1–2). It has been continuously used in empirical research as an important indicator of the relationship between citizens, political institutions and democracy.

Easton's (Easton 1975) understanding of political support is still valued as a definition of political trust. Easton stressed the uniqueness of the relationship between political support and citizens' feelings about obtaining it from what they see as the outputs and performance of political authorities (Easton 1975: 437). The citizens' perceptions may or may not be correct in some objective sense (Easton 1975: 438).

In the 1970s, researchers found that political trust as an affective orientation towards government can be traced on various levels, from high trust to high distrust or political cynicism (Miller 1974: 952). In addition, a healthy mistrust has also been considered an important factor in a healthy democracy (Carstens 2023). Nevertheless, political trust has been valued as an important source of liberal democracy. At this level of analysis, political trust refers to citizens' assessments of the core institutions of the polity and entails a positive evaluation of the most relevant attributes that make each political institution trustworthy, such as credibility, fairness, competence, transparency in its policymaking and openness to competing views (Zmerli 2014).

More recently, researchers have pointed to political trust as an indicator of political legitimacy, and empirical research has revealed that political trust is more or less as stable as some other attitudes, such as those exhibited towards immigration and redistribution (Devine & Valgarðsson 2024). Although a decline in political trust might signal a legitimacy crisis, this is not always the case (van der Meer & van Erkel 2024), and since political trust can have both positive and negative impacts on the development of democracy, it has been suggested that healthy political scepticism or sceptical trust (Norris 2022b) is supportive of democracy.

For our research, we also drew on literature on political trust to identify its key factors (see the section 'Factors impacting political trust').

Methodological approach

This article is based on a qualitative narrative literature review conducted between July and December 2024. We opted for a narrative literature review (Arksey & O'Malley 2005) due to a lack of systematic focus in the existing literature on the relationship between political trust and AI, and because of the aim to explore the topic more broadly. The qualitative approach has allowed

us to capture the complexity of perceptions, emotions and contextual factors, which quantitative methods often overlook.

While this type of review does not adhere to specific guidelines, we used a semi-structured approach to gather the literature, consisting of the following steps. First, we used Google Scholar and Cobiss+ (a digital catalogue of Slovenian libraries) as the two main search engines for finding literature. Units were selected by using the following English keywords in our searches: 'Artificial Intelligence', 'Artificial Intelligence and Trust', 'Trust in Artificial Intelligence', 'Artificial Intelligence and Political Trust' and 'Artificial Intelligence and Institutional Trust'.

Additionally, we used the snowball effect method for identifying additional sources on lists of references in the previously found units. To ensure that only credible sources were used for analysis, we included in our search results only academic articles, books, book chapters, and selected conference proceedings and expert reports. We mainly relied on peer-reviewed sources, and titles, abstracts and keywords from these sources were reviewed to determine the relevance and comprehensiveness of the uncovered literature. Some sources were excluded due to the inaccessibility of the full texts, and the final selection was focused on literature addressing the intersection between AI and political trust.

Findings from the literature review were synthesised and organised according to the commonalities found in the literature in the form of a research report (Kaišić & Fink-Hafner 2025), which served as the foundation for building the novel conceptual model from the bottom up. Given the complexity and interdependence of factors shaping both trust in AI technology and political trust, the conceptual model offers a clear and structured way to capture these dynamics and guide future research.

Building a model of the relationship between trust in AI and political trust

Trust in AI and factors affecting trust in AI technology

Trust in AI technology is not a constant; rather, it can be built and destroyed, because many factors co-determine such trust (Table 1).

Demographic factors. Demographic elements, such as gender, age, education and managerial roles, play a significant role in shaping individuals' perception of AI, as highlighted by Li et al. (2024). With regard to age, younger generations tend to show greater trust in AI globally, although the opposite is true in some countries, such as China and South Korea, where older populations demonstrate higher trust levels. Gender differences in trust are quite minimal, according to a worldwide survey conducted by Gillespie et al. (2023), apart from in the United States, Singapore and South Korea, where notable gaps were

Table 1: Factors affecting trust in AI technology

Factors	Variables in more detail
Demographic factors	Gender, age, education, managerial roles
Factors related to the user	Personality Context Norms, values, ideology Users' attitudes towards AI technology Users' perceptions of machine–human relationships in the case of AI Reviews of AI from other users Confidence in institutional safeguards
Technology-based factors	***Basic AI technological qualities:*** – robustness, accuracy, reliability, trialability – transparency and explainability of AI and AI outcomes – trustworthiness of AI – usability, including competence, functionality, performance, helpfulness, reputation ***Characteristics of AI with ethically burdened social impact:*** – privacy protection – fairness – accountability ***Human face of AI technology:*** – anthropomorphism and warmth – AI personality
Contextual factors	**Social context:** level of economic development, sociocultural factors including values and norms **Organisational context:** team characteristics, task risk **Social representation of AI:** – particularly AI representation in mass media – image of AI in sci-fi books and movies – government interference – regulation of media for accurate AI representation
Factors of control over technology and its use	**Technological/technical:** – verification tools **Governing:** – governance – regulators – auditors

Source: Authors

observed. Education is also a crucial factor, as individuals with university-level qualifications tend to trust AI more. Similarly, those in managerial roles exhibit higher trust in AI, reflecting how work positions can influence someone's perception of AI technology (Gillespie et al. 2023).

User-related factors. Based on a literature review, we revealed six user-related subgroups of factors. The first refers to factors linked to individuals' personalities (including psychological traits), which co-determine individuals' trust in

AI. There is also a general willingness to trust AI technology, and hedonic motivation is driven by the enjoyment of using AI (which further enhances trust). Furthermore, the image and perception of AI also influence how users connect emotionally to it and how they evaluate its trustworthiness (Duenser & Douglas 2023; Li et al. 2024).

Second, trust in AI is also shaped by individuals' past experiences and cultural norms, which define and guide users' attitudes towards AI (Duenser & Douglas 2023). According to Kleizen et al. (2023), these factors – along with pre-existing values, ideologies and political orientations – can vary widely among individuals. For instance, people may hold conflicting views on issues such as privacy versus safety, which can significantly affect how they perceive and trust AI systems.

The third subgroup of factors includes users' overall acceptance and confidence in the technology, as individuals' familiarity with the way in which AI works enhances trust (Gillespie et al. 2023; Kleizen et al. 2023). More precisely, it concerns users' knowledge, expertise and understanding of the design, applications and limitations (benefits and risks) of AI (Gillespie et al. 2023; Kleizen et al. 2023). It is about a sense of control over AI's decision-making, self-efficacy or confidence in using it effectively, all of which are linked to a better understanding of AI (Gillespie et al. 2023; Kleizen et al. 2023). A clear comprehension of how AI is designed, how it functions, and the benefits and risks of AI systems make users more confident in how the technology will perform (Banavar 2016; Kleizen et al. 2023). A sense of control over the AI decision-making process can enhance trust, as users have a feeling that they can influence actions taken by AI (Kleizen et al. 2023; Li et al. 2024).

The fourth subgroup of user-related factors includes the influence of other users of AI systems on how individuals perceive and trust AI (Siau & Wang 2018). The fifth subgroup tackles issues concerning the relationship between humans and machines, bearing in mind that users' perception of AI as a supportive tool rather than a replacement for human expertise fosters trust in AI (Li et al. 2024). The sixth and final subgroup concerns users' confidence in institutional safeguards (i.e. regulatory frameworks, oversight mechanisms), which enhances users' trust (Gillespie et al. 2023).

Technology-based factors

Technological characteristics of AI also matter in shaping individuals' trust in it, and there are three sets of factors that relate specifically to technology. The first includes the perception of the main technological qualities of AI, including its robustness, accuracy, functionality, reliability, trialability and explainability (Siau & Wang 2018; Kleizen et al. 2023; Afroogh et al. 2024). Trialability allows users to engage with AI, fostering trust through first-hand experience (Siau & Wang 2018), while the competence (ability to fulfil AI's

functional claims) and performance (actually fulfilling AI's functional claims) of AI validate its ability to meet expectations (Li et al. 2024). Transparency and explainability enhance trust by making AI processes clearer, enabling users to understand its decision-making process (Siau & Wang 2018; Li et al. 2024). All the above-listed characteristics of AI ensure consistency in its performance and also that errors are minimised (Kleizen et al. 2023; Afroogh et al. 2024; Li et al. 2024). Its reputation depends on the experience with multiple AI technological characteristics, which enhances its trustworthiness (Afroogh et al. 2024).

Second, several factors are rooted in the characteristics of AI that have ethically burdened impacts on society, particularly privacy protection, fairness and accountability (Li et al. 2024). Ensuring privacy protection and fairness in using AI through data security can enhance trust in such technology, while accountability emphasises clear roles and mechanisms for addressing AI malfunctions (Emaminejad, North & Akhavian 2022; Li et al. 2024).

The third set of factors comprises human-like factors related to AI traits, including anthropomorphism, warmth and personality (Li et al. 2024). These traits enhance trust in AI by encouraging emotional connections, creating positive perceptions of AI intentions and ensuring relatable and/or supportive interactions with such systems (Glikson & Williams Wolley 2020; Lockey et al. 2021).

Contextual factors

The context of the development and use of AI matters in terms of trusting such technology. Perceptions of and trust in AI are influenced by levels of economic development and sociocultural factors, including values and norms within a particular society (Afroogh et al. 2024; Li et al. 2024). Individuals who perceive the use of AI as socially acceptable are also more likely to express positive attitudes towards it (Kleizen et al. 2023; Li et al. 2024).

On the organisational level, team characteristics and task risk also contribute to the whole dynamic of trust in AI technology (Afroogh et al. 2024).

Finally, social representation of AI is also a relevant factor regarding trust in such technology. Media, in a broad sense, portrays AI in a particular way, and not only mass media but also science fiction books and movies can contribute to either fear of or fascination with AI (Siau & Wang 2018; Lockey et al. 2021). Since media can spread both accurate and inaccurate information on AI technology, proper media regulation is vital (Li et al. 2024).

Factors of control over technology and its use

With the rapid development and spread of multiple AI technologies, issues concerning biases and misuse have evolved. AI regulation has become essential for mitigating public risks and impacting trust in such technologies (Li et al. 2024).

Based on experiences with state intervention lagging behind real-life processes, authors increasingly call for proactive regulation, including methods for certifying, explaining and auditing AI systems. National governance of AI is not sufficient; rather, global governance needs to be developed together with verification tools used by regulators and auditors (Siau & Wang 2018; Butcher & Beridze 2019), as these are vital for fostering trust both domestically and internationally.

The reviewed literature has pointed out the rapidly evolving global regulation of AI using different approaches. The European Union (EU) has been highlighted as a leader in establishing a comprehensive legal framework with its EU Artificial Intelligence Act, which emphasises trustworthy AI based on principles of legality, ethics and robustness. This includes key pillars such as technical robustness and safety, transparency, human agency and oversight, privacy and data governance, non-discrimination, accountability and societal well-being (Neuwirth 2023: 10). The EU model has been exposed for building on proactive and precautionary principles, applying a risk-based classification of AI systems (e.g. high-risk versus low-risk) and setting explicit legal obligations for high-risk AI systems (Gillespie et al. 2023: 70).

In contrast, the United States (US) has been characterised as adopting a more decentralised and sector-specific approach, often focused on guidelines and principles rather than binding laws (Lockey et al. 2021). For example, the US Defense Advanced Research Project Agency's explainable AI initiative reflects the emphasis on technological transparency and control, aimed at making AI systems understandable and manageable, particularly in military contexts (Butcher & Beridze 2019). US governance focuses more on remaining flexible and promoting innovation, with less emphasis on prescriptive regulation, relying instead on initiatives with specific agencies and private sector self-regulation (Butcher & Beridze 2019; Li et al. 2024).

Besides the US and EU, countries such as Canada, the United Kingdom and China are also beginning to approach AI regulation, although in a less regulatory way (Butcher & Beridze 2019; Li et al. 2024).

Negative and positive impacts of various factors on trust in AI technology

Here, previously presented findings are summarised into two sets of factors impacting trust in AI technology. The first includes those with a positive impact on trust in AI technology, and the second covers those factors that negatively impact trust in such technology.

Factors with a negative impact. Unsupportive social context matters. Pre-existing negative opinions, beliefs and attitudes, and a negative cultural influence on attitudes towards AI technology constitute an important basis for

further shaping attitudes toward this technology. More narrowly, a general mistrust of technology, the perception of AI as a danger and a lack of trust in AI systems together constitute an overall technological distrust.

Additionally, there are unacceptable characteristics of AI technology that work against trusting it. Among them are characteristics that raise privacy and security concerns, the autonomy of AI itself, AI bias and erratic behaviour, the non-transparency of AI, issues related to the explainability and interpretability of AI, misperformance of AI and non-compatibility with human values, and the lack of accountability.

Another negative set of factors includes a scarcity of information and education, which contributes to the lack of understanding of how AI works and limited citizen knowledge of ethical AI. As already mentioned, biased information with negative connotations, along with negative media and press influence (e.g. media reporting, fake news and misinformation, news media attention on negative aspects of AI) may represent a negative set of factors on its own. Additionally, the spread of AI technology and its impact on job security raise negative economic concerns related to the AI technological revolution.

The way in which AI is governed also matters. The exclusivist governance and lack of open dialogue, along with clear and transparent information sharing, lead to the exclusion of citizens from developing AI policy, and because of citizens' perception that their policy concerns are not heard, distrust in AI tends to increase. Similarly, poor regulation of AI, including a lack of ethical guidelines and hard-to-comprehend policies, is a significant negative factor.

Factors with a positive impact. As mentioned previously, contextual factors are important. Supportive pre-existing opinions, beliefs and attitudes, as well as a supportive cultural influence, help in developing trust in technology in general and AI technology in particular.

Supportive trust is as complex as distrust, and an overall trust in technology, confidence in government and technological organisations, and trust in AI engineers positively impact trust in AI. Technological adequacy also supports trust in AI, including key factors such as the safety, good performance, transparency, and explainability and interpretability of AI, compatibility with human values, and a belief in having control over AI alongside appropriate safeguards. As expected, a positive impact is also found in cases of supportive information and education variables, and where accountability is established in relation to the development, use and misuse of AI.

Furthermore, inclusive governance supports trust in AI, and open dialogue, transparency and clarity of information sharing, citizens' inclusion in AI policy development and addressing citizens' policy concerns are all important factors. Acceptable regulation of AI means understandable and ethical AI regulation, which is committed to societal values. Human oversight and the ability to challenge AI-related policy decisions also need to be available after adopting such

policies. Economic factors matter as well, as when people perceive economic benefits, they tend to put more trust in AI technology.

The interfering variables – actors. In the reviewed literature, actors are not systematically addressed as factors in shaping trust in AI technology. Researchers have mentioned various actors when tackling a variety of issues so far, and both AI engineers and producers have been exposed when mentioning technological issues. AI users develop their attitude toward AI technology based on their experience with it. As a rule, governments and citizens are revealed as important actors when it comes to making and implementing particular public policies related to AI governance.

Factors impacting political trust

Various factors impact political trust, and in political science literature, micro factors are often distinguished from macro ones.

Micro factors. Researchers have focused predominantly on micro factors of political trust, i.e. those that shape individuals' personal characteristics and perceptions. Among the most cited are personal characteristics (individual psychological or biological characteristics) and socio-economic factors (an individual's position in the labour market in connection with their education and skills, experiences of unemployment, financial distress and the welfare system) (Carstens 2023).

Macro factors. Macro factors can be understood broadly as a context (Zmerli & Hooghe eds. 2011; Eder, Mochmann & Quandt 2015; Martini & Quaranta 2020). More specifically listed factors often include corruption, macroeconomic performance, the inclusiveness of institutions, socio-economic inequalities among people and regions, and divisions among economic winners and losers, as well as cultural and social norms and values (Newton 2015; Uslaner 2015; Dodsworth & Cheeseman 2020). However, it is important that political trust is related to perceptions of these macro factors and not necessarily their objective characteristics (Carstens 2023), particularly when it comes to the government's responsiveness in times of various crises and natural disasters. Also, poor quality of public services at the local and national level, as well as low inclusivity of welfare policies, negatively impact political trust (Newton 2001; Carstens 2023).

Furthermore, there are also political factors in a narrower sense that impact political trust, such as democratic development and stability, the rule of law, effective institutional checks and balances, belief in democracy, satisfaction with how democracy works (Newton 2001) and the transparency of governments' decision-making (Robles & Mallinson 2023a: 11). Very specific factors are also linked to politicians – namely, their competence (Uslaner 2015), trustworthiness (Newton 2001), predictability, intrinsic commitment, competence and responsiveness (Winsvold et al. 2024). What is particularly interesting in

relation to the (mis)use of AI technology is the media's coverage of individual politicians' conduct (particularly scandals) and the media's framing of political events, which may shift people's trust (Devine & Valgarðsson 2024: 482).

Additional factors. Researchers have found that various kinds of trust are interconnected (Newton 2001), with some factors supporting political trust and others damaging it. Nevertheless, political trust is not only a dependent but also an independent variable (Carstens 2023).

Impacts of trust in AI on political trust

As shown in previous sections, researchers have pointed out many factors that co-shape people's trust in AI; however, as trust in AI is formed, it does impact political trust. With a dynamic change of trust in AI, its impact on political trust may also change. The factors that impact trust in AI can be both positively and negatively attributed (Table 2).

Table 2: Impacts of trust in AI on political trust

Impact of trust in AI on political trust	Variables related to the impact of trust in AI technology on political trust	Interfering variables: characteristics and behaviours of actors
Negative	Distrust in AI technology per se AI eroding trust in media, representation, accountability, government AI destabilising democratic societies AI complicating political engagement and participation	Citizens Users Stakeholders Governments Media
Positive	AI trustworthiness Transparency in the use of AI in the public sector Platforms designed to boost trust in AI Effective institutionalised human oversight of AI	

Source: Authors

Negative impacts of (dis)trust in AI on political trust. In literature, the negative impacts of (dis)trust in AI on political trust dominate over revealing the positive impacts of such trust, due to a significant distrust of AI and challenges preventing the development of (positive) trust in such technology. In this context, researchers expose in particular AI's problematic practices of reinforcing biases, infringing on privacy and fairness, spreading misinformation, contributing to job displacement and enabling malicious uses among the risks of AI failures (Gillespie et al. 2023). Kreps and Kriner (2023) also find that AI has an indirect negative impact on political trust, which it generates by complicating political engagement and participation, damaging trust in media, representation, accountability and government, and by destabilising democratic societies

(Kreps & Kriner 2023; Nie 2024: 2). The unclear accountability and the risks of AI failures (Zuiderwijk, Chen & Salem 2021) are additional factors that support the above-described causal relationships.

Positive impacts of trust in AI on political trust. When people trust AI, it is expected that this will have a positive impact on political trust, with positive attributes of such technology including the transparency of AI algorithms and AI outcomes (Gillespie et al. 2023; Kleizen et al. 2023). The lack of some negatively evaluated factors may work in a similar fashion, such as AI's lack of infringement of privacy and fairness, and not spreading misinformation, endangering jobs or enabling malicious misuses (Gillespie et al. 2023).

AI's trustworthiness is a crucial element in being able to trust it (Zuiderwijk, Chen & Salem 2021), and algorithmic transparency has been particularly important in using AI in the public sector (Robles & Mallinson 2023a: 12). Institutionally supported trust in AI is helpful, as are accountability, consent and intermediaries (e.g. platforms, human oversight of AI), which enhance trust in AI (Gillis, Laux & Mittelstadt 2024).

While looking at the positive and negative impacts of trust in AI on political trust, it should not be overlooked that such technologies are not (yet) fully autonomous and that they differ among themselves in terms of levels of autonomy. Researchers highlight the important role of various actors involved in the development, production, spread and use of AI technologies, as the characteristics and behaviours of these actors impact both trust and distrust in AI, as well as affect political trust (Omrani et al. 2022).

With growing awareness of ethical issues related to AI, authors increasingly point out characteristics and behaviours of a wide range of actors impacting the (non-)ethical use of AI, particularly among citizens, users, stakeholders, government and media (Zuiderwijk, Chen & Salem 2021; Kleizen et al. 2023). There are particularly high expectations that the government should act in favour of both trust in AI and political trust by introducing regulations designed to prevent breaches of trust (Butcher & Beridze 2019; Zuiderwijk, Chen & Salem 2021; Kleizen et al. 2023). On the one hand, governments are pressured by the technology industry, which offers AI solutions for enhanced delivery of public goods and services; on the other hand, governments need to ensure that citizens are satisfied with the delivery of such public goods and services. A positive public attitude is of critical importance for introducing and expanding the use of AI in governance, especially in the public sector (Zuiderwijk, Chen & Salem 2021; Gutierrez Gaviria 2022; Wilson 2022). Such attitudes are not a given; therefore, positive experiences with the use of AI are important (Kleizen et al. 2023).

Characteristics of technology may also contribute to positive experiences with AI, one example being platforms that are designed to boost trust in AI (Gillis, Laux & Mittelstadt 2024: 187). However, human oversight of AI supported by effective institutions and their activities are very important for enhancing

trust in such technology (Gillis, Laux & Mittelstadt 2024: 187), and indirectly political trust as well.

Impacts of political trust on trust in AI technology

Research on the response to a particular innovation has revealed that acceptance of such an innovation and the policies related to it may be largely determined by socio-interactional factors (Oudshoorn & Pinch 2003; Silverstone & Hirsch 1992; Veen et al. 2010: 811). However, there has been a delay in studying the relationship between political trust and public trust in technology (Robles & Mallinson 2023b).

Trust is vital for public acceptance of AI (Robles & Mallinson 2023b), as well as being essential for its societal adoption (Kreps et al. 2023). Gillespie et al. (2023) warned that without public trust, the adoption of AI technologies would be limited and that this would restrict the opportunities for AI to have positive societal and economic impacts.

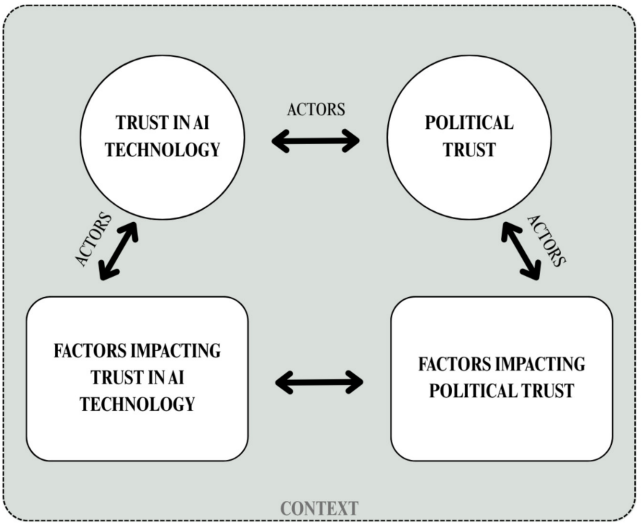
Governments' experimentation with AI impacts trust in AI (Zuiderwijk, Chen & Salem 2021: 1), and it should be borne in mind that increasing political trust can improve people's attitudes towards new technologies in general, and AI in particular (Wilson & van der Velden 2022). The government needs to invest in building both political trust and trust in AI (Flynn, Ricci & Bellaby 2012; Ahmed, Wahed & Thompson 2023), which could be achieved by developing and implementing strategies and procedures for managing AI risks (Robles & Mallinson 2023b), encompassing the government's inclusiveness (citizen involvement in policymaking and public service delivery), transparency, accountability, enhancing public debate (Tambotoh et al. 2017; Criado & Gil-Garcia 2019; Kemper & Kolkman 2019; Wilson 2022), ensuring ethical AI regulation in their implementation (Criado & Gil-Garcia 2019), and educating people about the benefits and drawbacks of AI (Ahmed, Wahed & Thompson 2023).

Political trust can impact trust in AI technology both negatively and positively. When political trust is low, it is not realistic to expect people to trust the government's use of AI in policymaking and policy implementation processes; this is because perceptions are critical for legitimacy. More precisely, perceptions of the government's bias will inevitably undermine the perceived legitimacy of its use of AI (Robles & Mallinson 2023a: 12). In contrast, when citizens trust the government, they exhibit more trust in the government's AI policy and the instrumentalisation of AI in the public sector.

A model of the relationship between trust in AI and political trust

We have developed a model of the relationship between trust in AI and political trust in an inductive, bottom-up manner. As shown in Figure 1, the relationship between trust in AI and political trust may be direct and two-way, but also indirect.

Figure 1: Model of the relationship between trust in AI and political trust in the frame of a broader context



Source: Authors

Based on the AI-related literature review, the complexity of variables and relationships includes the following: 1) the revealed factors impacting trust in AI (Table 1); 2) the revealed factors impacting political trust (section under the subtitle 'Factors impacting political trust'); 3) the reverse impact of trust in AI as an independent variable on factors of trust in AI (Table 2); 4) the reverse impact of political trust as an independent variable on factors of political trust (section under the subtitle 'Impacts of political trust in AI technology'); and 5) the overall contextual factors relevant for both trust in AI and political trust (multiple factors including various types of social trust, socio-economic factors, cultural factors, various crises) (see Table 1 and the section under the subtitle 'Factors impacting political trust'). Characteristics of actors and their behaviour co-shape the causal relationships (see Table 1, Table 2 and the section under the subtitle 'Factors impacting political trust').

From the literature on political trust, additional variables and relationships need to be taken into account in the model. Due to limited space, we only mentioned those that are part of the context and, at the same time, impact each other. These are as follows: socio-economic stakes, market relations, relations between generalised trust and particular trusts, reciprocity issues, social (including political) divisions and conflicts, power relations, available public spaces, social capital (Warren ed. 1999), accountability (Sztompka 2022), (multiple) levels of government (Kappler et al. 2024), models of governing (van der Meer 2017), international factors such as global interconnectedness (Fisher 2012) and (management of) international crises (Weinberg 2022).

Conclusions

In this article, we have focused on the mutual impact of trust in AI technology on political trust and the impact of political trust on trust in AI technology. The reviewed literature that connects AI and political trust has been limited by the scope of English language and the cited sources. Nevertheless, we have been able to reveal a complex set of actors and other factors that impact on trust in AI, on political trust, and the relationship between the two. A disciplinary inclusive research approach has further allowed us to develop the presented novel research model. We hope it will spark further theoretical developments, empirical research in the form of case studies and cross-country studies as well as disciplinary (political science) and interdisciplinary research.

Further research is needed to incorporate literature in other languages and regions – particularly in China and other parts of Asia where the development of AI has been especially dynamic. More research could provide additional insights for our model by examining AI governance in authoritarian regimes and established democracies, but also in backsliding democracies.

It should be taken into account that the current accelerated technological development may significantly interfere with the present global trend of declining democracy in a context that favours authoritarianism (including unprecedented social inequality gaps in particular states and on a global scale). It is also important to systematically take into account political science literature that finds that political trust can have positive and negative impacts on the development of democracy.

From the policy relevance perspective, the reviewed literature supports a thesis that at this stage there are still people (citizens, users), stakeholders (developers, companies, other actors) and governments who have the upper hand in the development and use of AI technology. Therefore, they are responsible for the development and enforcement of ethical AI technology and its use. In order to support socially responsible and ethical policies in this field, researchers need to focus on comparing different governance models (including frameworks for AI oversight in democratic, democratic backsliding and non-democratic settings), and on identifying their sources, consequences and characteristics of trust in AI governance. Such research should not overlook the relevance of context, particularly rapidly increasing social inequalities and political power inequalities.

Acknowledgements

This research was funded by the Slovenian Research and Innovation Agency (ARIS) Grant No. P5–0136.

Bibliography

Afroogh, S., Akbari, A., Malone, E., Kargar, M. & Alambeigi, H. (2024): Trust in AI: Progress, Challenges, and Future Directions. *Humanities and Social Sciences Communications*, 11(1), 1–32.

Ahmed, N., Wahed, M. & Thompson, N. C. (2023): The Growing Influence of Industry in AI Research. *Science*, 379, 884–886.

Arksey, H. & O'Malley, L. (2005): Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology*, 8(1), 19–32.

Banavar, G. (2016): Learning to Trust Artificial Intelligence Systems: Accountability, Compliance and Ethics in the Age of Smart Machines. *IBM*, September, <accessed online: https://www.ibm.com/watson/assets/pdfs/Learning_to_trust_AI_systems.pdf>.

Butcher, J. & Beridze, I. (2019): What is the State of Artificial Intelligence Governance Globally? *The RUSI Journal,* 164(5–6), 88–96.

Carstens, J. (2023): Unpacking Political Trust: A Review of Conceptualisations, Determinants, and Consequences. *French Politics*, 21, 295–314.

Criado, J. I. & Gil-Garcia, R. J. (2019): Creating Public Value Through Smart Technologies and Strategies: From Digital Services to Artificial Intelligence and Beyond. *International Journal of Public Sector Management*, 32(5), 438–450.

Devine, D. & Valgarðsson, V. O. (2023): Stability and Change in Political Trust: Evidence and Implications from Six Panel Studies. *European Journal of Political Research*, 63(2), 478–497.

Dodsworth, S. & Cheeseman, N. (2020): Political Trust: The Glue that keeps Democracies Together. University of Birmingham & Westminster Foundation for Democracy, May, <accessed online: https://www.agora-parl.org/sites/default/files/agora-documents/Political%20trust%20-%20The%20glue%20hat%20keeps%20democracies%20together.pdf >.

Dogan, M. (1992): Conceptions of Legitimacy. In: Hawkesworth, Mary and Kogan, M. (eds.): *Encyclopedia of Government and Politics*. London and New York: Routledge, 116–128.

Duenser, A. & Douglas, D. M. (2023): Whom to Trust, How and Why: Untangling Artificial Intelligence Ethics Principles, Trustworthiness, and Trust. *IEEE Intelligent Systems,* 38(6), 19–38.

Easton, D. (1975): A Re-Assessment of the Concept of Political Support. *British Journal of Political Science*, 5(4), 435–457.

Eder, C., Mochmann, I. C. & Quandt, M. (eds.) (2014): *Political Trust and Disenchantment with Politics: International Perspectives*. Leiden: Brill.

Emaminejad, N., North, A. M. & Akhavian, R. (2022, May 24): *Trust in AI and Implications for the AEC Research: A Literature Analysis*. ASCE International Conference on Computing in Civil Engineering, May 24, <accessed online: https://ascelibrary.org/doi/10.1061/9780784483893.037>.

European Commission (2019): Ethics Guidelines for Trustworthy AI. *European Commission*, April, <accessed online: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.

Fisher, J. A. V. (2012): Globalization and Political Trust. *MPRA Paper*, 37763, 1–63.

Flynn, R., Ricci, M. & Bellaby, P. (2012): Ambiguity, Complexity and Uncertainty Surrounding the Hazards of Hydrogen and Public Views of Emergent Risks. *Journal of Risk Research*, 15, 373–387.

Gillespie, N., Lockey, S., Curtis, C., Pool, J. & Akbari, A. (2023): Trust in Artificial Intelligence: A Global Study. *The University of Queensland & KPMG Australia*, February, <accessed online: https://assets.kpmg.com/content/dam/kpmg/au/pdf/2023/trust-in-ai-global-insights-2023.pdf>.

Gillis, R., Laux, J. & Mittelstadt, B. (2024): Trust and Trustworthiness in Artificial Intelligence. In: Paul, Regine, Carmel, E. & Cobbe, J. (eds.), *Handbook on Artificial Intelligence and Public Policy*. Cheltenham: Edward Elgar, 181–193.

Glikson, E. & Williams Woolley, A. (2020): Human Trust in Artificial Intelligence: Review of Empirical Research. *Academy of Management Annals,* 14(2), 627–660.

Gomstyn, A., Jonker, A. & McGrath, A. (2024): What is trustworthy AI? *IBM*, October, <accessed online: https://www.ibm.com/think/topics/trustworthy-ai#:~:text=Trustworthy%20AI%20refers%20to%20artificial, among%20stakeholders%20and%20end%20users>.

Gutierrez Gaviria, C. I. (2022): The Role of Artificial Intelligence in Pushing the Boundaries of U.S. Regulation: A Systematic Review. *Santa Clara High Technology Law Journal*, 38(2), 123.

Hooghe, M. & Zmerli, S. (2011): Introduction: The Context of Trust. In: Zmerli, Sonja & Hooghe, M. (eds.): *Political Trust: Why Context Matters*. Colchester: ECPR Press, 1–12.

Kaišić, K. & Fink-Hafner, D. (2025): *Raziskovalno poročilo 2: Politično zaupanje in umetna inteligenca: pregled literature [Research Report 2: Political Trust and Artificial Intelligence: Literature Review]*. Ljubljana: Center za politološke raziskave.

Kappler, M., Schomaker, R., Guashino, E. & Verhoest. (2024): How do Elite Core Actors Assess Trust in National and EU Authorities? The Varying Role of Generalised Trust at Different Governmental Levels. *Comparative European Politics*, 1–32.

Kemper, J. & Kolkman, D. (2019): Transparent to Whom? No Algorithmic Accountability Without a Critical Audience. *Information, Communication & Society*, 22(14), 2081–2096.

Kleizen, B., Van Dooren, W., Verhoest, K. & Tan, E. (2023): Do Citizens Trust Trustworthy Artificial Intelligence? Experimental Evidence on the Limits of Ethical AI Measures in Government. *Government Information Quarterly,* 40(4), 1–21.

Kreps, S. & Kriner, D. (2023): How AI Threatens Democracy. *Journal of Democracy,* 34(4), 122–131.

Kreps, S., George, J., Lushenko, P. & Rao, A. (2023): Exploring the Artificial Intelligence 'Trust Paradox`: Evidence from a Survey Experiment in the United States. *PLos One*, 18(7), e0288109.

Laux, J., Wachter, S. & Mittelstadt, B. (2024): Trustworthy Artificial Intelligence and the European Union AI Act: On the Conflation of Trustworthiness and Acceptability of Risk. *Regulation & Governance*, 18, 3–32.

Li, B., Qi, P., Liu, B., Di, S., Liu, J., Pei, J., Yi, J. & Zhou, B. (2022): Trustworthy AI: From Principles to Practices. *ACM Computing Surveys*, 55(9), 1–56.

Li, Y., Wu, B., Huang, Y. & Luan, S. (2024): Developing Trustworthy Artificial Intelligence: Insights from Research on Interpersonal, Human-Automation, and Human-AI Trust. *Frontiers in Psychology,* 15, 1–13.

Lockey, S., Gillespie, N., Holm, D. & Someh, I. A. (2021, January 5): *A Review of Trust in Artificial Intelligence: Challenges, Vulnerabilities and Future Directions*. Proceeding of the 54th Hawaii International Conference on System Science, January 5, <accessed online: https://www.researchgate.net/publication/349157208_A_Review_of_Trust_in_Artificial_Intelligence_Challenges_Vulnerabilities_and_Future_Directions>.

Martini, S. & Quaranta, M. (2020): *Citizens and Democracy in Europe: Contexts, Changes and Political Support*. Cham: Palgrave Macmillan.

Miller, A. H. (1974): Political Issues and Trust in Government: 1964–1970. *American Political Science Review*, 68(3), 951–972.

Neuwirth, R. J. (2023): *The EU Artificial Intelligence Act: Regulating Subliminal AI Systems*. London: Routledge.

Newton, K. (2015): Trust and Political Disenchantment. In: Eder, Christina, Mochmann, I. C. & Quandt, M. (eds.): *Political Trust and Disenchantment with Politics*. Leiden: Brill, 19–30.

Nie, M. (2024): Artificial Intelligence: The Biggest Threat to Democracy Today? *Proceedings of the AAAI Symposium Series*, 3(1), 376–379.

Norris, P. (2022a): Two Faces of Trust. In: Norris, Pippa (ed.): *Praise of Skepticism: Trust but Verify*. Oxford: Oxford University Press, 3–25.

Norris, P. (2022b): The General Theory of Skeptical Trust. In: Norris, Pippa (ed.): *Praise of Skepticism: Trust but Verify*. Oxford: Oxford University Press, 26–51.

Omrani, N., Rivieccio, G., Fiore, U., Schiavone, F. & Garcia Agreda, S. (2022): To Trust or not to Trust? An Assessment of Trust in AI-based Systems: Concerns, Ethics and Contexts. *Technological Forecasting & Social Change,* 181, 1–10.

Oudshoorn, N. & Pinch, T. (2003): How Users and Non-Users Matter. In: Oudshoorn, Nelly & Pinch, T. (eds.): *How Users Matter: The Co-Construction of Technologies and Users*. Cambridge: MIT Press, 4–22.

Pillulta, M. M. (2005): Trust. In: Nicholson, Nigel, Audia, P. G. & Pillutla, M. M. (eds.): *The Blackwell Encyclopedia of Management: Organizational Behavior*. Malden, Oxford and Carlton: Blackwell Publishers Ltd., 406–409.

Rai, A., Constantinides, P. & Sarker, S. (2019): Next Generation Digital Platforms: Towards Human-AI Hybrids. *MIS Quarterly*, 43(1), iii–ix.

Robles, P. & Mallinson, D. J. (2023a): Artificial Intelligence Technology, Public Trust, and Effective Governance. *Review of Policy Research*, 1–18.

Robles, P. & Mallinson, D. J. (2023b): Catching up with AI: Pushing Toward a Cohesive Governance Framework. *Politics & Policy*, 51(3), 355–372.

Ryan, M. (2020): In AI We Trust: Ethics, Artificial Intelligence, and Reliability. *Science and Engineering Ethics*, 26, 2749–2767.

Siau, K. & Wang, W. (2018): Building Trust in Artificial Intelligence, Machine Learning, and Robotics. *Cutter Business Technology Journal,* 31(2), 47–53.

Silverstone, R. & Hirsch, E. (eds.) (1992): *Consuming Technologies: Media and Information in Domestic Spaces*. London and New York: Routledge.

Sztompka, P. (2022): Trust. In: Harris, Phil, Bitonti, A., Fleisher, C. S. & Binderkrantz, A. S. (eds.): *The Palgrave Encyclopedia of Interest Groups, Lobbying and Public Affairs*. Cham: Palgrave Macmillan, 1428–1432.

Tambotoh, J. J. C., Supangkat, S. H., Kosala, R., Ranti, B., Prabowo, H., Isa, S. M. & Pudjianto, B. W. (2017, November 15–17): *Conceptual Model for Creating Effective Public Value through Key Practices in Information Technology Governance Mechanisms*. International Conference on Information Management and Technology, November 15–17, <accessed online: https://www.researchgate.net/publication/322873473_A_conceptual_model_for_creating_effective_public_value_through_key_practices_in_information_technology_governance_mechanisms>.

Uslaner, E. M. (2015): Trust, Political. In: Wright, James D. (ed.): *International Encyclopedia of Social & Behavioral Sciences*. Oxford: Elsevier, 658–663.

Valle-Cruz, D., Criado, J. I., Sandoval-Almazan, R. & Ruvalcaba-Gomez, E. (2020): Assessing the public policy-cycle framework in the age of artificial intelligence: From agenda-setting to policy evaluation. *Government Information Quarterly*, 37(4), 1–12.

van der Meer, T. W. G. (2017): Political Trust and the "Crisis of Democracy". In: Thompson, William R. (ed.): *Oxford Research Encyclopedia of Politics*. Oxford: Oxford University Press, 1–22.

van der Meer, T. W. G. & van Erkel, P. F. A. (2024): Moving beyond the Political Trust Crisis Debate: Residual Analyses to Understand Trends in Political Trust. *European Journal of Political Research*, 63(3), 1240–1257.

Veen, M., Gremmen, B., te Molder, H., & van Woerkum, C. (2010): Emergent Technologies against the Background of Everyday Life: Discursive Psychology as a Technology Assessment Tool. *Public Understanding of Science*, 20(6), 810–825.

Warren, M. E. (ed.) (1999): *Democracy and trust*. Cambridge, New York: Cambridge University Press.

Weinberg, J. (2022): Can Political Trust Help to Explain Elite Policy Support and Public Behaviour in Times of Crisis? Evidence from the United Kingdom at the Height of the 2020 Coronavirus Pandemic. *Political Studies*, 70(3), 655–679.

Wilson, C. (2022): Public Engagement and AI: A Values Analysis of National Strategies. *Government Information Quarterly*, 39(1), 101652.

Wilson, C. & van der Velden, M. (2022): Sustainable AI: An Integrated Model to Guide Public Sector Decision-Making. *Technology in Society*, 68, 101926.

Winsvold, M., Haugsgjerd, A., Saglie, J. & Segaard, S. B. (2023): What Makes People Trust or Distrust Politicians? Insights from Open-Ended Survey Questions. *West European Politics*, 47(4), 759–783.

Zmerli, S. (2014): Political Trust. In: Michalos, Alex C. (ed.): *Encyclopedia of Quality of Life and Well-Being Research*. Dordrecht: Springer, 1–4.

Zmerli, S. & Hooghe, M. (eds.) (2011): *Political Trust: Why Context Matters*. Milton Park: ECPR Press.

Zuiderwijk, A., Chen, Y. & Salem, F. (2021): Implications of the Use of Artificial Intelligence in Public Governance: A Systematic Literature Review and a Research Agenda. *Government Information Quarterly,* 38(3), 1–9.

Danica Fink-Hafner *is a professor at the Faculty of Social Sciences, University of Ljubljana, Slovenia. Her main research fields are artificial intelligence, political parties and democracy. Among her recent publications is an open-access book Party System Changes and Challenges to Democracy: Slovenia in a Comparative Perspective (Springer Nature, 2024). E-mail: danica.fink-hafner@fdv.uni-lj.si; ORCID: 0000-0002-4455-9223.*

Katarina Kaišić *is a master's student at the Faculty of Social Sciences, University of Ljubljana, Slovenia. Her main research fields are artificial intelligence and public policy process, and artificial intelligence and governance. E-mail: katarinakaisic1@gmail.com; ORCID: 0009-0005-9410-4701.*