I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17635

The Impact of Learning about AI Advancements on Trust

Milena Nikolova
Marco Angrisani

JANUARY 2025

I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17635

The Impact of Learning about AI Advancements on Trust

Milena Nikolova
University of Groningen, Brookings Institution and IZA

Marco Angrisani
University of Southern California Dornsife and USC Center for Economic and Social Research

JANUARY 2025

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany
Phone: +49-228-3894-0
Email: publications@iza.org
www.iza.org

ABSTRACT

The Impact of Learning about AI Advancements on Trust*

Can people develop trust in Artificial Intelligence (AI) by learning about its developments? We conducted a survey experiment in a nationally representative panel survey in the United States (N = 1,491) to study whether exposure to news about AI influences trust differently than learning about non-AI scientific advancements. The results show that people trust AI advancements less than non-AI scientific developments, with significant variations across domains. The mistrust of AI is the smallest in medicine, a high-stakes domain, and largest in the area of personal relationships. The key mediators are context-specific: fear is the most critical mediator for linguistics, excitement for medicine, and societal benefit for dating. Personality traits do not affect trust differences in the linguistics domain. In medicine, mistrust of AI is higher among respondents with high agreeableness and neuroticism scores. In personal relationships, mistrust of AI is strongest among individuals with high openness, conscientiousness, and agreeableness. Furthermore, mistrust of AI advancements is higher among women than men, as well as among older, White, and US-born individuals. Our results have implications for tailored communication strategies about AI advancements in the Fourth Industrial Revolution.

JEL Classification: C91, D83, O33, Z10

Keywords: Randomized Controlled Trial (RCT), survey experiment, Artificial Intelligence (AI), trust, United States

Corresponding author:
Milena Nikolova
University of Groningen
Nettelbosje 2
9747 AE Groningen
The Netherlands
E-mail: m.v.nikolova@rug.nl

* The authors would like to thank David Wollenberg for excellent research assistance and Ioana Matei for help with literature review tasks. Nikolova acknowledges financial support from the Institute for Humane Studies (IHS) at George Mason University (Grant Numbers: IHS017788 and IHS017786). All errors are the authors'.

1. Introduction

Artificial Intelligence (AI) refers to the continuously evolving field of computational technologies that emulate human intelligence. These technologies can autonomously learn from data, adapt their behavior, and make decisions (Berente, Gu, Recker, & Santhanam, 2021; Raj & Seamans, 2019). AI already outperforms humans in tasks with well-defined parameters, such as image classification, language processing, and coding, but remains less effective than humans in complex activities requiring abstract reasoning, commonsense decision-making, or long-term planning (HAI, 2024; Nurski, 2024).

Although human progress and innovation are longstanding (Mokyr, Vickers, & Ziebarth, 2015), the current AI wave is unprecedented in its speed and scope. This rapid development has sparked significant public attention and discussion. A notable example is the launch of ChatGPT in November 2022, an advanced AI-based large language model that can have human-like conversations (Haque, Dharmadasa, Sworna, Rajapakse, & Ahmad, 2022; Karanouh, 2023; Leiter et al., 2024). ChatGPT took the world by surprise and initially created a lot of excitement and praise about its capabilities, which later morphed into worry about its inaccuracies, the spread of misinformation, and job loss (Leiter et al., 2024).

The public discourse about ChatGPT illustrates how information about technological advancements influences attitudes toward such developments. News about technological and scientific advancements can foster literacy about and familiarity with new developments, which, in turn, can shape trust toward them (Brossard, 2013; Nguyen & Hekman, 2024). But do these reactions differ when people learn about AI versus scientific advancements and, if so, how do these differences shape people's trust in AI versus other scientific developments? As AI continues to advance, understanding public trust in AI—and how it compares to trust in traditional scientific progress—becomes essential for fostering acceptance and coexistence with this rapidly evolving field.

We are the first to investigate how individuals form trust in AI versus non-AI scientific advancements across three societal domains—linguistics, medicine, and personal relationships—as well as how key factors shape these differences. Using a pre-registered survey experiment[1] (N= 1,491) implemented in the Understanding America Study (UAS), a probability-based Internet panel representative of the U.S. adult population (Kapteyn, Angrisani, Darling, & Gutsche, 2024), we compare how learning about AI versus non-AI scientific progress impacts trust. For example, individuals may perceive an AI breakthrough in linguistics differently than a non-AI scientific discovery in the same domain because of AI's novelty (Glikson & Woolley, 2020), or because AI and non-AI developments trigger different expectations about their societal usefulness. Additionally, we study how socio-demographic factors and personality traits influence differential trust in AI vs. non-AI progress.

We conceptualize trust in AI and non-AI advancements as situational: trust depends on the immediate learning about an innovation in a specific domain (i.e., linguistics, medicine, or personal relationships). Our approach highlights the processes of trust formation in technology, which underpins technology adoption (Ghazizadeh, Lee, & Boyle, 2012; Glikson & Woolley, 2020).

Our findings demonstrate that trust in AI advancements is lower than trust in non-AI scientific advancements, with significant variation across domains. Trust in AI

[1] The experiment is registered under RCT ID AEARCTR-0012977 on the American Economic Association (AEA) registry for randomized controlled trials on February 9, 2024.

advancements is higher in medicine, a high-stakes domain, than in low-stakes areas like dating and linguistics, where mistrust is more pronounced. We explore the role of four channels – excitement, fear, understanding of the advancement, and its perceived benefit – to explain the differential trust patterns between AI and non-AI advancements. The key mediators vary by context: in linguistics, fear is the most significant mediator; in medicine, it is excitement; and in dating, it is the perceived societal benefit of the innovation. Furthermore, we find that the trust consequences of exposure to information about AI advancements versus non-AI advancements in the linguistics domain do not depend on personality traits. In medicine, mistrust of AI compared to non-AI developments is concentrated among individuals with high agreeableness and neuroticism, while for dating, mistrust of AI is higher among those with more pronounced openness, conscientiousness, and agreeableness. The differential mistrust of AI is concentrated among women and older individuals, meanwhile. These results underscore the importance of assessing the role of communicating information about (non-)AI advancements for shaping public attitudes and trust in these developments.

We contribute to the literature in two distinct ways. First, to the best of our knowledge, this study is the first to empirically investigate whether *learning about AI vs. non-AI scientific advancements* influences trust. Unlike prior work (see Glickson & Woolley, 2020 for an overview), we provide empirical evidence of how learning about AI advancements—through the presentation of information—affects trust formation differently across domains, offering a deeper understanding of how trust forms situationally within specific contexts. As such, we offer new insights into the processes through which learning shapes trust differently for AI and non-AI scientific advancements.

Second, this study builds on existing research on trust in technology (Glikson & Woolley, 2020; Hoff & Bashir, 2015) by focusing on how contextual factors, such as high-stakes versus low-stakes applications, emotional responses (fear and excitement), and perceived societal benefits shape situational trust in AI. While extant work has primarily explored the role of task-based factors (e.g., the nature or complexity of tasks AI performs) and functional factors (e.g., reliability and transparency), our study shifts the focus to the contextual and emotional aspects of situational trust of AI and non-AI advancements.

2. Theoretical Background

2.1. Trust

Interpersonal trust facilitates economic exchange by reducing uncertainty, fostering cooperation, and minimizing transaction costs (Arrow, 1972). Therefore, trust is a key marker of the quality of the social fabric and contributes to economic growth and social prosperity (Akçomak & Ter Weel, 2009; Algan & Cahuc, 2010; Knack & Keefer, 1997).

Individuals can extend trust to non-human entities, such as institutions, central banks, machines, and products (Algan, 2018; Mcknight, Carter, Thatcher, & Clay, 2011; Roth, Nowak-Lehmann D, & Otter, 2022; Sætra, 2024), which can have important behavioral consequences. For example, trust in AI is key for public acceptance and willingness to engage with the technology (Glikson & Woolley, 2020; Kreps, George, Lushenko, & Rao, 2023).

Key factors influencing trust in machines include their reliability (consistent and accurate performance over time), transparency (the ability of the system to clearly explain its processes and decisions), the user's understanding of the system's capabilities, the machine's anthropomorphic features (e.g., voice or human-like behavior), and situational factors (e.g., the complexity of the task or the user's stress level) (Glikson & Woolley, 2020; Henrique & Santos Jr, 2024; Hoff & Bashir, 2015; Mcknight et al., 2011; Sætra, 2024). Moreover, trust in AI is *cognitive* or *emotional* and depends on the representation of AI itself (e.g., robot vs. virtual bot) and AI's capabilities and degree of sophistication (Glikson & Woolley, 2020; Riley & Dixon, 2024). Emotional trust, which is not well explored in the literature, stems from affective and social factors like anthropomorphism and immediacy behaviors that foster a sense of connection or comfort with the AI (Glikson & Woolley, 2020).

The scarce body of literature on the determinants of trust in science suggests that trust in science may stem from similar sources as trust in technology. Specifically, trust in science depends on positive beliefs about science's capacity to provide reliable knowledge and contribute to societal well-being (Wintterlin et al., 2022). It increases when scientists demonstrate expertise, integrity, and benevolence (i.e., acting in the interest of others and following ethical principles), while it diminishes when scientists are viewed as "elites" disconnected from ordinary people (Wintterlin et al., 2022). Effective communication based on non-technical language that resonates with diverse audiences further enhances public trust in science (Henkel, Jacob, & Perrey, 2023). Finally, individuals are more likely to trust scientific findings that align with their societal values (Drummond & Fischhoff, 2017).

Trust in AI is conceptually distinct from trust in machines and automation (Glikson & Woolley, 2020) and trust in science (Alvarado, 2023), though no paper to date has formally explored the differences.[2] AI differs from traditional automation because it learns and adapts, making its decisions harder to understand (Castelvecchi, 2016), unlike traditional automation, which handles simple, repetitive tasks that are easy to program (Glikson & Woolley, 2020). Trust in AI likely changes over time, shaped by transparency, perceived intelligence, and human-like traits, unlike trust in traditional automation, which tends to be static.

Furthermore, trust in science encompasses interpersonal and relational trust based on the integrity, expertise, and accountability of scientists and scientific institutions (Alvarado, 2023). In contrast, trust in AI is only "epistemic" in that it is based on the machine's technical performance and reliability to perform the given task (Alvarado, 2023). Additionally, while scientific processes emphasize openness and peer review to foster trust, AI systems often operate as "black boxes" (Castelvecchi, 2016) whereby humans cannot understand how the system made a particular decision or reached a certain outcome.

2.2. Trust in AI and non-AI advancements across domains

It is a priori unclear whether trust in AI is higher or lower than trust in non-AI scientific developments. On the one hand, non-AI scientific progress is more long-standing, while AI innovations are more recent, raising questions about their ethical implementation, potential misuse, and destructive power. People may also view scientific advancements as products of *human* effort and research, whereas they may perceive AI advancements as impersonal. These factors could lead to lower trust in AI compared to traditional scientific progress. On the other hand, AI advancements may inspire greater

[2] Montag, Becker, and Li (2024) demonstrate that there is a small to moderate overlap between trust in humans and trust in AI, suggesting that the two concepts are distinct.

trust due to their potential to solve complex problems that humans cannot tackle on their own.

Plausibly, trust in AI and non-AI advancements may vary across societal domains, though it is not a priori obvious which contexts may invoke higher trust in AI relative to non-AI technologies.

In low-stakes domains, such as linguistics, entertainment, or dating, trust in AI systems may be higher compared to trust in non-AI advancements. This is because the consequences of potential failures are generally limited and less impactful, and the benefits, such as convenience or fun, may outweigh the costs of occasional errors. At the same time, in the context of dating, individuals may mistrust AI in its ability to ensure meaningful connections.

In high-stakes domains such as healthcare, trust in AI developments may be lower compared to trust in non-AI advancements because the AI systems are not yet well-regulated or fine-tuned. In the medical domain, non-AI technologies, instruments, and devices, are subject to regulations from the Food and Drug Administration (FDA) ensuring their predictability and safety, making them potentially more trust-worthy. At the same time, the potential consequences of mistakes within the AI healthcare context, such as incorrect diagnoses or surgical errors, carry significant risks for patients and society. For instance, bias in healthcare AI systems can worsen health inequalities by relying on historical data that reflect existing inequalities, leading to less accurate diagnoses for underrepresented populations and perpetuating unequal outcomes (Larrazabal, Nieto, Peterson, Milone, & Ferrante, 2020; Obermeyer, Powers, Vogeli, & Mullainathan, 2019; Seyyed-Kalantari, Zhang, McDermott, Chen, & Ghassemi, 2021). However, AI's perceived ability to rapidly process vast amounts of data, detect patterns, and generate actionable insights may, in some cases, lead individuals to trust AI more than human decision-makers.

All in all, trust differences between AI and non-AI advancements likely vary depending on the context and domain. We test empirically in which situations (if any) trust in AI surpasses trust in non-AI.

2.3. Meditators

Understanding of the development, perceptions of its societal benefit, and emotional responses may mediate differences in trust between AI and non-AI advancements. The complexity and opacity of AI systems could potentially require greater effort to comprehend, which may lead to skepticism and reduced trust, particularly among individuals with limited technological literacy (Riley & Dixon, 2024). In contrast, non-AI advancements might be perceived as more intuitive and based on familiar scientific principles, making them easier to trust.

Moreover, perceived societal benefit could also explain differences in trust. People may view AI as transformative, particularly in high-stakes domains like healthcare, where it appears capable of solving complex societal challenges and outperforming humans or other methods in diagnosing challenging conditions (Cao et al., 2023; Cid et al., 2024; Schopf et al., 2024). Conversely, individuals may view non-AI advancements, as more incremental but potentially more predictable and reliable than AI.

Finally, emotional responses might further shape trust differences caused by learning about AI vs. non-AI advances (Freeman, Alhoori, & Shahzad, 2020; Glikson & Woolley, 2020). For example, AI advancements may evoke stronger fear due to ethical concerns, loss of human agency when machines take over, potential misuse, or worries about job displacement, while non-AI advancements, being more familiar, could seem less threatening. Conversely, AI might generate greater excitement for its innovative capabilities, whereas non-AI advancements may be viewed as less groundbreaking and less exciting.

2.4. Personality traits

Personality traits can influence how individuals initially react to and process information (Baumert & Schmitt, 2012; Matthews, 2008). Yet, there is little theoretical or empirical research guiding the assessment of how personality traits can differentially impact situational trust when receiving information about AI and non-AI advancements. Therefore, multiple scenarios are possible. For example, individuals with high openness to new experiences may view AI as innovative and forward-thinking, which aligns with their preference for exploring new and unconventional ideas, thereby fostering greater excitement and trust in AI advancements compared to traditional, more familiar non-AI technologies. High neuroticism might lead to skepticism or anxiety about AI given its complexity and opacity, leading to lower trust in AI compared to non-AI technologies that are perceived as more stable and predictable. Consequently, it is an empirical question whether and how personality traits play a differential role in shaping trust in AI compared to non-AI scientific advancements. We take on this question in our data collection and analysis.

3. Methodology

Following Jackson et al. (2023), we employ a survey experiment based on the randomized provision of information with a between-subjects design to assess whether learning about AI versus scientific breakthroughs influences trust. Specifically, we expose study participants to information about either AI or scientific advancements in linguistics, medicine, and dating, and compare differences in trust by type of advancement and across domains. We also investigate whether understanding of the advancement, perceptions of its societal benefit, as well as emotional responses and personality traits play a role in the relationships we uncover. We focus on the domains of linguistics, medicine, and personal relationships, as these areas capture different low- and high-stakes aspects of everyday life.

We conducted the experiment with the respondents of the Understanding America Study (UAS), a nationally representative probability-based panel of U.S. residents aged 18 and older. The UAS was established in 2014 and has been managed since then by the Center for Economic and Social Research at the University of Southern California (Kapteyn et al., 2024). The UAS fielded the survey experiment (in English and Spanish) between March 6, 2024, and April 8, 2024. Respondents completed the survey using a computer, tablet, or smartphone and were compensated using the UAS standard rate of $2 for 3 minutes of interview time (the average response time was approximately 4 minutes). We invited 2,143 UAS members to participate in the experiment. Of them, 1,501 completed the survey yielding a 70% response rate.[3]

The experiment proceeded in three steps, which is common in the literature (Fuster & Zafar, 2023). First, before the information treatment, respondents in each condition reported their level of awareness of AI (or science) developments on a 5-point Likert

[3] Table 2 shows that the demographic balance between experimental conditions ensured by ex-ante randomization was preserved in the final sample of participants (suggesting no differential non-response across demographic groups).

scale, with the responses ranging from very limited awareness (rarely seeking or coming across information on AI/scientific developments) to an expert level (deep understanding of AI/scientific advancements, possibly because of professional or educational background). Second, participants received information about either AI (treated group) or science (control group) advancements in linguistics, medicine, and dating. Each respondent read three excerpts—one for each domain (Table 1 and Appendix A). The difference between the treated and control groups was that the treated group read about AI advancements and the control group read about different non-AI scientific advancements in the same domain. We used one-paragraph texts from reputable science and news websites (e.g., Science Daily, BBC, Deutsche Welle).

Table 1: Summary of Information Treatments

Domain	Category	Description
AI	Linguistics	Discusses GPT-3.5 and GPT-4, their capabilities in language translation, text generation, and other linguistic tasks. Source: BBC Science Focus, Link: https://www.sciencefocus.com/future-technology/gpt-3
AI	Medicine	Introduces an AI algorithm for cervical cancer detection through image analysis. Source: Science Daily, Link https://www.sciencedaily.com/releases/2019/01/190110164701.htm
AI	Dating	Covers AI chatbots like Replika and their role in digital relationships. Source: Deutsche Welle, Link: https://www.dw.com/en/ai-love-why-romance-with-a-chatbot-is-complicated/a-66238378
Science	Linguistics	Examines how babies learn language through rhythm rather than phonetic information. Source: Science Daily, Link: https://www.sciencedaily.com/releases/2023/12/231201123644.htm

Science	Medicine	Details the use of mRNA therapeutics for ovarian cancer treatment in preclinical studies.[4] Source: Science Daily, link: https://www.sciencedaily.com/releases/2024/01/240124132806.htm
Science	Dating	Explores the role of oxytocin and dopamine in romantic love and their connection to brain pathways. Source: Science Daily, Link: https://www.sciencedaily.com/releases/2024/01/240109004404.htm

Third, after reading each excerpt, participants in both the treated and control conditions rated the following items on a 5-point scale, where 1 = Completely Disagree and 5 = Completely Agree:

1. "I trust this innovation/advance based on the information provided."
2. "This development has the potential to benefit humanity."
3. "I have a good understanding of this innovation/advance."
4. "Reading about this makes me feel afraid."
5. "Reading about this development makes me feel excited."

This anonymized survey experiment data, along with basic socio-demographics, are publicly available for download as study UAS610 (available at https://uasdata.usc.edu/survey/UAS+610). We merged the most recent participants' data about employment, health status, cognitive ability, and personality traits with the experiment survey dataset using the UAS Comprehensive File (available at https://uasdata.usc.edu/page/Comprehensive+File), which compiles data from all UAS core surveys, administered to the entire panel every two years. We dropped 28 individuals who did not provide answers to one of the key variables (trust, fear, excitement,

[4] Due to an omission, the source was not provided to the experiment participants.

understanding, and perceived benefit), resulting in an analysis sample of 1,491 individuals.

Table 2 details the distribution of respondent characteristics by treatment status. The two groups are similar across socio-demographic characteristics (age, gender, foreign-born status, race, marital status, education, income) and prior knowledge of AI/science. The only exception is working status, whereby respondents in the treatment group (AI condition) are 5.4 percentage points less likely to be working than the control group. Importantly, we see that there are no differences in the treated and control groups based on prior knowledge, suggesting that the effects we find are truly due to learning about the advancement in the context of the experiment.[5]

Table 2: Summary statistics of the analysis sample, by treatment status

Variable	AI (n)	AI (%)	Science (n)	Science (%)	Difference (AI - Science)	Test Statistic (df)	p-value
Age	744	52.37 (Mean)	747	51.40 (Mean)	0.97	t(1489) = -1.18	0.238
Female	744	61.8	747	60.1	1.7	z = -0.68	0.496
Born in US	744	90.7	747	89.4	1.3	z = -0.84	0.401
Currently Working	743	51.3	746	56.7	5.4	z = 2.10	0.036*
Awareness of AI/Science						$\chi^2(4) = 6.09$	0.192
1 (Lowest Awareness)	222	29.8	210	28.1	1.7		
2	337	45.3	312	41.8	3.5		
3	136	18.3	156	20.9	-2.6		
4	39	5.2	56	7.5	-2.3		

[5] We found no moderation effects of the treatment on trust across domains based on respondents' prior knowledge, indicating that the observed effects stem from learning about the advancement at that specific moment. These results are available upon request.

5 (Expert Awareness)	10	1.3	13	1.7	-0.4		
Race							
White	496	66.7	491	65.7	1		
Black	98	13.2	89	11.9	1.3	$\chi^2(3) = 1.44$	0.696
Others	76	10.2	82	11	-0.8		
Hispanic	74	9.9	85	11.4	-1.5		
Marital Status							
Married	402	54.1	408	54.6	-0.5		
Separated/Divorced/Widowed	166	22.3	160	21.4	0.9	$\chi^2(2) = 0.19$	0.91
Never Married	175	23.6	179	24	-0.4		
Education Level							
High School or Less	154	20.7	160	21.4	-0.7		0.708
Some College	246	33.1	232	31.1	2	$\chi^2(2) = 0.69$	
Bachelor or More	344	46.2	355	47.5	-1.3		
Household Income							
<$35,000	200	27	189	25.4	1.6		
$35,000-$74,999	213	28.7	206	27.7	1	$\chi^2(2) = 1.02$	0.601
>=$75,000	328	44.3	348	46.9	-2.6		
Census Region							
Northeast	126	16.9	135	18.1	-1.2		
Midwest	205	27.5	192	25.7	1.8	$\chi^2(3) = 3.32$	0.344
South	286	38.4	269	36	2.4		
West	127	17.2	151	20.2	-3		
Urbanicity							
Rural	112	15.1	87	11.7	3.4		
Mixed	358	48.1	379	50.8	-2.7	$\chi^2(2) = 3.80$	0.149
Urban	274	36.8	280	37.5	-0.7		

In principle, because the information treatment is randomized across respondents, we can directly estimate the impact of learning about AI on trust by comparing the mean trust outcomes of the treatment and control groups. Nevertheless, we also provide results with basic characteristics (age, biological sex, race) and additional characteristics (prior awareness, income, marital status, education, working status, census region, and urbanicity) added as controls.

Using Ordinary Least Squares (OLS) estimators, we test whether exposure to AI vs. non-AI advancements differentially affects trust. We also conduct mediation analysis to explore the mechanisms driving the observed effects, such as the perceived understanding of the technology, its potential to benefit humanity, and the emotions triggered by the innovation. Finally, we explore whether the results we find depend on the respondent's personality traits.

6. Results

Table 3 displays the findings regarding the impact of the information treatment on trust in each domain. For each domain, we present three specifications, starting with the impact of the treatment on trust, then adding basic control variables, and finally including additional controls. The inclusion of controls slightly reduces the magnitude of the coefficient estimates but does not change the overall patterns we observe.

The main message of Table 3 is that across all domains, learning about advancements in AI instigates less trust than learning about scientific advancements in the same area. This suggests that individuals are more likely to trust traditional scientific progress than novel but perhaps more unfamiliar AI-based innovations. Specifically, exposure to information about AI advancements decreases trust by 75% of a standard deviation in linguistics, 23% of a standard deviation in medicine, and 122% of a standard deviation in dating, compared to non-AI advancements, based on models (2), (5), and

(8), respectively.[6] The trust differences between AI and non-AI developments are the largest in the dating domain, followed by linguistics, and the smallest in medicine. In dating, the use of AI may be seen as impersonal or unintuitive, whereas in medicine, the potential gains in terms of improved diagnostics or life-saving interventions may make individuals optimistic about their application.

Table 3: The impact of AI information treatment on trust, by domain

	(1) Linguistics	(2) Linguistics	(3) Linguistics	(4) Medicine	(5) Medicine	(6) Medicine	(7) Dating	(8) Dating	(9) Dating
Treatment	-0.757***	-0.747***	-0.726***	-0.228***	-0.227***	-0.204***	-1.232***	-1.224***	-1.220***
	(0.048)	(0.047)	(0.047)	(0.051)	(0.051)	(0.050)	(0.041)	(0.040)	(0.040)
Basic controls	N	Y	Y	N	Y	Y	N	Y	Y
Additional controls	N	N	Y	N	N	Y	N	N	Y
Observations	1,491	1,491	1,481	1,491	1,491	1,481	1,491	1,491	1,481
R^2	0.143	0.173	0.215	0.013	0.040	0.098	0.380	0.399	0.417

Notes: Robust standard errors in parentheses.*** p<0.01, ** p<0.05, * p<0.1. The dependent variable is based on the statement: "I trust this innovation/advance based on the information provided" on a 5-point Likert scale and standardized to have a mean of 0 and a standard deviation of 1. Models (1), (4), and (7) include no control variables. Models (2), (5), and (8) include basic characteristics (age, biological sex, race), and Models (3), (6) and (9) include additional characteristics (prior awareness, income, marital status, education, working status, census region, and urbanicity).

We next explore whether the understanding of the advancement, its perceived societal benefit, and the emotions associated with learning about it explain the differential trust in AI/non-AI scientific development in each domain (Table 4). To this end, we performed standard mediation analysis using the products of coefficients approach (Preacher & Hayes, 2008), with bootstrapped standard errors (5,000 replications), using Stata's -sureg- and the nlcom commands.

[6] We focus on models (2), (5), and (8) which include the set of exogenous controls because in models (3), (6), and (9) we include several additional controls, but at the cost of losing observations. Ultimately, the results with the full set or a restricted set of controls differ only slightly from each other.

Table 4 displays the indirect effects (via the four mediators: understanding, usefulness, fear, and excitement), the total indirect effect, the direct effect of the treatment on trust, and the total effect (combined effect via the mediators and the direct effect of the treatment on trust). The results reveal that the importance of mediators varies across societal areas. The mediators play their most prominent role in medicine, where they explain about 90% of the total effect (0.20 standard deviations out of a total 0.23 standard-deviation decrease in trust). In contrast, the four mediators account for about 64% of the total effect in dating, and 46% of the total effect in linguistics.

Furthermore, we observe several notable patterns related to how specific mediators contribute across domains. In linguistics, fear (-0.120, SE = 0.024, corresponding to 35% of the indirect effect) and excitement (-0.095, SE = 0.017, corresponding to 28% of the indirect effect) are the strongest mediators, while in medicine and dating, perceptions of societal benefit dominate the indirect effects. In medicine, excitement accounts for more than half (54%) of the indirect effect, while in dating, the perceived societal benefit accounts for approximately 43% of the indirect effect.

The role of understanding of the advancement varies, having a moderate effect in linguistics (-0.051, SE = 0.013, corresponding to 15% of the indirect effect) and dating (-0.124, SE=0.017, corresponding to 16% of the indirect effect), but a minimal influence in medicine (-0.003, SE = 0.012, corresponding to 1.5% of the indirect effect). These patterns suggest that the mechanisms through which learning about AI vs. non-AI advancements affect trust are context-dependent, with different factors driving trust across different societal areas.

Table 4: Mediation Analysis

	Linguistics		Medicine		Dating	
Effect	Value	Standard Error (SE)	Value	Standard Error (SE)	Value	Standard Error (SE)
Indirect Effect via Understanding	-0.051	0.013	-0.003	0.012	-0.124	0.017
Indirect Effect via Societal Benefit	-0.076	0.018	-0.075	0.018	-0.335	0.036
Indirect Effect via Fear	-0.120	0.024	-0.017	0.011	-0.126	0.025
Indirect Effect via Excitement	-0.095	0.017	-0.110	0.020	-0.195	0.031
Total Indirect Effect	-0.342	0.036	-0.205	0.039	-0.780	0.038
Direct Effect	-0.405	0.041	-0.022	0.035	-0.444	0.040
Total Effect	-0.747	0.048	-0.227	0.050	-1.224	0.040

Notes: The table represents the mediation analysis results (product of coefficients method), applied with the -sureg- command and with bootstrapped standard errors with 5000 replications. The total effects correspond to the total effects from models (2), (5), and (8) respectively from Table 2. All estimations include a set of basic control variables. N=1,491.

We also explored the moderating role of the Big Five personality traits by dichotomizing the scores into high (above the median, coded as 1) and low (below the median, coded as 0). The results in Table 5 reveal that personality traits moderate the effect of exposure to AI versus non-AI scientific advancements to varying degrees, depending on the domain. In linguistics, personality traits do not significantly moderate the effect of the treatment on trust. In other words, in linguistics, the mistrust towards AI innovations compared with non-AI advancements is similar across people with different personality traits.

In the medicine domain, the negative effects of mistrust of AI are concentrated among individuals with above-median agreeableness and neuroticism scores. In the dating context, the negative effect of the treatment on trust is amplified for individuals with above-median openness and conscientiousness scores.

All in all, rather than identifying universal personality traits that consistently moderate the relationship between information about AI/non-AI advancements and trust, the findings reveal that the effects of personality are highly context-specific, varying significantly across different contexts.

The direct effect of personality traits on trust demonstrates that more agreeable and open individuals are likely to trust advancements of all kinds (both AI and non-AI) in the medicine and personal relationships domains, while more extroverted individuals are marginally more likely to trust advances in linguistics. None of the other personality traits is associated with trust in our setting.

Finally, we conducted a post-hoc heterogeneity analysis to explore whether the treatment effects we identify differ based on respondents' characteristics in terms of biological sex, age, race, and immigrant status. Table 6 details that the mistrust of AI advancements is higher among women than men, and among older, White, and US-born individuals. This suggests that addressing concerns about AI advancements may require tailored communication strategies to build trust among these specific demographic groups.

Table 5: The impact of AI information treatment on trust, with personality traits interactions

	(1)	(2)	(3)
	Trust (linguistics)	Trust (medicine)	Trust (dating)
Treatment	-0.569***	0.025	-0.953***
	(0.136)	(0.145)	(0.110)
High Agreeableness	0.066	0.212***	0.161***
	(0.070)	(0.078)	(0.058)
High Conscientiousness	-0.008	0.036	0.026
	(0.069)	(0.080)	(0.056)
High Extroversion	0.117*	-0.028	0.063
	(0.066)	(0.074)	(0.054)
High Neuroticism	-0.074	0.020	0.019
	(0.070)	(0.078)	(0.056)
High Openness	0.105	0.155**	0.145***
	(0.065)	(0.075)	(0.054)
Treatment*High Agreeableness	-0.144	-0.249**	-0.174*
	(0.109)	(0.117)	(0.093)
Treatment*High Conscientiousness	-0.059	-0.103	-0.149*
	(0.107)	(0.116)	(0.089)
Treatment*High Extroversion	-0.102	-0.024	-0.083
	(0.102)	(0.108)	(0.084)
Treatment*High Neuroticism	-0.006	-0.191*	0.015
	(0.108)	(0.116)	(0.090)
Treatment*High Openness	-0.066	0.067	-0.171**
	(0.101)	(0.109)	(0.085)
Observations	1,360	1,360	1,360
R-squared	0.191	0.059	0.430

Notes: Robust standard errors in parentheses.*** p<0.01, ** p<0.05, * p<0.1. The dependent variable is based on the statement: "I trust this innovation/advance based on the information provided" on a 5-point Likert scale and standardized to have a mean of 0 and a standard deviation of 1. All models include basic characteristics (age, biological sex, race). The personality traits are binary variables indicating whether the respondent scores above the median for each trait.

Table 6: The impact of AI information treatment on trust, heterogeneity analysis

	(1)	(2)	(3)	(4)	(5)	(6)
	Variable=Female			Variable= Age Over 60		
	Linguistics	Medicine	Dating	Medicine	Medicine	Medicine
Treatment	-0.541***	-0.033	-1.124***	-0.575***	-0.187***	-1.097***
	(0.080)	(0.082)	(0.066)	(0.058)	(0.065)	(0.052)
Variable	0.116*	-0.007	-0.006	0.042	0.064	0.057
	(0.064)	(0.073)	(0.054)	(0.066)	(0.073)	(0.055)
Treatment*Variable	-0.338***	-0.318***	-0.165**	-0.463***	-0.093	-0.342***
	(0.099)	(0.105)	(0.083)	(0.099)	(0.105)	(0.082)
Observations	1,491	1,491	1,491	1,491	1,491	1,491
R2	0.179	0.046	0.400	0.185	0.037	0.403

	(7)	(8)	(9)	(10)	(11)	(12)
	Variable=White			Variable= US-Born		
	Linguistics	Medicine	Dating	Medicine	Medicine	Medicine
Treatment	-0.583***	-0.283***	-1.023***	-0.360**	-0.184	-0.989***
	(0.082)	(0.089)	(0.071)	(0.149)	(0.165)	(0.131)
Variable	0.077	0.095	0.049	-0.166	-0.268**	-0.162*
	(0.066)	(0.077)	(0.057)	(0.118)	(0.129)	(0.097)
Treatment*Variable	-0.245**	0.082	-0.302***	-0.428***	-0.049	-0.259*
	(0.100)	(0.108)	(0.086)	(0.157)	(0.174)	(0.137)
Observations	1,491	1,491	1,491	1,491	1,491	1,491
R2	0.171	0.037	0.403	0.171	0.034	0.397

Notes: Robust standard errors in parentheses.*** p<0.01, ** p<0.05, * p<0.1. The dependent variable is based on the statement: "I trust this innovation/advance based on the information provided" on a 5-point Likert scale and standardized to have a mean of 0 and a standard deviation of 1. All models include basic characteristics (age, biological sex, race). The personality traits are standardized to have a mean of 0 and a standard deviation of 1.

7. Conclusion

This study contributes to the burgeoning literature on trust in AI (Glikson & Woolley, 2020) by providing the first experimental evidence that trust in AI advancements is lower than trust in non-AI scientific advancements. Furthermore, we document that this differential trust in AI vs. non-AI progress is context-specific. Trust in AI is higher in medicine, a high-stakes domain, compared to lower-stakes societal areas such as linguistics and dating. Emotional responses, particularly fear, and excitement, along with perceived societal benefits, play a central role in mediating the relationship between exposure to information about AI and non-AI advancements and trust. Furthermore, the empirical evidence highlights different patterns across contexts. Fear plays the biggest role in linguistics, excitement is most influential in medicine, and societal benefit is the key mediator in dating. Furthermore, personality traits do not influence differences in trust when comparing exposure to information about AI advancements versus non-AI advancements in the linguistics domain. In medicine, the mistrust of AI compared with non-AI developments is primarily concentrated among individuals with above-median agreeableness and neuroticism scores. The mistrust of AI in personal relationships is higher among individuals with high levels of openness, conscientiousness, and agreeableness.

While the study provides causal insights on trust in AI vs. non-AI innovations, it relies on a single survey experiment, which may not capture the full complexity of real-world interactions with AI systems. Future research should explore longitudinal designs and field experiments to assess how trust evolves over time and in practical applications and how information about AI translates into the actual adoption of AI. Additionally, examining cultural and policy differences in trust formation could provide valuable insights into how AI acceptance varies globally. Finally, AI is a constantly evolving technological frontier, requiring novel research into how specific AI-powered innovations – e.g., industrial robots equipped with AI capabilities, self-driving cars, or large language models – shape trust over time, depending on their applications and contexts.

Our findings underscore the complexity of trust in AI, emphasizing the importance of considering both contextual and individual factors when communicating about AI (and non-AI) scientific advancements. Ensuring transparency in AI communication and aligning its development with democratic values is essential for fostering equitable and inclusive adoption (Boix, 2022). Implementing clear communication strategies and trust-building mechanisms could reduce techno-anxiety, address public concerns, and ensure that AI development reflects societal and democratic principles.

References

Akçomak, I Semih, & Ter Weel, Bas. (2009). Social capital, innovation and growth: Evidence from Europe. *European Economic Review, 53*(5), 544-567.

Algan, Yann. (2018). Trust and social capital. In J. Stiglitz, J. Fitoussi, & M. Durand (Eds.), *For good measure: Advancing research on Well-being metrics beyond GDP* (pp. 283-320).

Algan, Yann, & Cahuc, Pierre. (2010). Inherited Trust and Growth. *The American Economic Review, 100*(5), 2060-2092.

Alvarado, Ramón. (2023). What kind of trust does AI deserve, if any? *AI and Ethics, 3*(4), 1169-1183.

Arrow, Kenneth J. (1972). Gifts and exchanges. *Philosophy & Public Affairs*, 343-362.

Baumert, Anna, & Schmitt, Manfred. (2012). Personality and Information Processing. *European Journal of Personality, 26*(2), 87-89. doi:10.1002/per.1850

Berente, Nicholas, Gu, Bin, Recker, Jan, & Santhanam, Radhika. (2021). Managing artificial intelligence. *MIS Quarterly, 45*(3), 1433-1450.

Boix, Carles. (2022). AI and the economic and informational foundations of democracy. In Justin B. Bullock (Ed.), *The Oxford Handbook of AI Governance*.

Brossard, Dominique. (2013). New media landscapes and the science information consumer. *Proceedings of the National Academy of Sciences, 110*(supplement_3), 14096-14101.

Cao, Kai, Xia, Yingda, Yao, Jiawen, Han, Xu, Lambert, Lukas, Zhang, Tingting, . . . Lu, Jianping. (2023). Large-scale pancreatic cancer detection via non-contrast CT and deep learning. *Nature Medicine, 29*(12), 3033-3043. doi:10.1038/s41591-023-02640-w

Castelvecchi, Davide. (2016). Can we open the black box of AI? *Nature News, 538*(7623), 20.

Chen, Qi, Yuan, Yufei, Feng, Yuqiang, & Archer, Norm. (2021). A decision paradox: benefit vs risk and trust vs distrust for online dating adoption vs non-adoption. *Internet Research, 31*(1), 341-375. doi:10.1108/INTR-07-2019-0304

Cid, Y. D., Macpherson, M., Gervais-Andre, L., Zhu, Y., Franco, G., Santeramo, R., . . . Montana, G. (2024). Development and validation of open-source deep neural networks for comprehensive chest x-ray reading: a retrospective, multicentre study. *Lancet Digit Health, 6*(1), e44-e57. doi:10.1016/s2589-7500(23)00218-2

Drummond, Caitlin, & Fischhoff, Baruch. (2017). Individuals with greater science literacy and education have more polarized beliefs on controversial science topics. *Proceedings of the National Academy of Sciences, 114*(36), 9587-9592.

Freeman, Cole, Alhoori, Hamed, & Shahzad, Murtuza. (2020). Measuring the diversity of Facebook reactions to research. *Proceedings of the ACM on Human-Computer Interaction, 4*(GROUP), 1-17.

Fuster, Andreas, & Zafar, Basit. (2023). Chapter 4 - Survey experiments on economic expectations. In Rüdiger Bachmann, Giorgio Topa, & Wilbert van der Klaauw (Eds.), *Handbook of Economic Expectations* (pp. 107-130): Academic Press.

Ghazizadeh, Mahtab, Lee, John D., & Boyle, Linda Ng. (2012). Extending the Technology Acceptance Model to assess automation. *Cognition, Technology & Work, 14*(1), 39-49. doi:10.1007/s10111-011-0194-3

Glikson, Ella, & Woolley, Anita Williams. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals, 14*(2), 627-660.

HAI. (2024). *Stanford Institute for Human-Centered Artificial Intelligence (HAI) AI index report 2024*. Retrieved from Stanford: https://aiindex.stanford.edu/

Haque, Mubin Ul, Dharmadasa, Isuru, Sworna, Zarrin Tasnim, Rajapakse, Roshan Namal, & Ahmad, Hussain. (2022). "I think this is the most disruptive technology": Exploring Sentiments of ChatGPT Early Adopters using Twitter Data. *arXiv preprint arXiv:2212.05856*.

Henkel, Maria, Jacob, Armin, & Perrey, Lennart. (2023). *What Shapes Our Trust in Scientific Information? A Review of Factors Influencing Perceived Scientificness and Credibility.* Paper presented at the European Conference on Information Literacy.

Henrique, Bruno Miranda, & Santos Jr, Eugene. (2024). Trust in artificial intelligence: Literature review and main path analysis. *Computers in Human Behavior: Artificial Humans*, 100043.

Hoff, Kevin Anthony, & Bashir, Masooda. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human factors, 57*(3), 407-434.

Jackson, Joshua Conrad, Yam, Kai Chi, Tang, Pok Man, Sibley, Chris G, Waytz, Adam, Jackson, Joshua Conrad, . . . Waytz, Adam. (2023). Exposure to Automation Explains Religious Declines. *Proceedings of the National Academy of Sciences of the United States of America, 120*(34), e2304748120.

Kapteyn, Arie, Angrisani, Marco, Darling, Jill, & Gutsche, Tania. (2024). The Understanding America Study (UAS). *BMJ open, 14*(10), e088183.

Karanouh, Maya. (2023). Mapping ChatGPT in Mainstream Media to Unravel Jobs and Diversity Challenges: Early Quantitative Insights through Sentiment Analysis and Word Frequency Analysis. *arXiv preprint arXiv:2305.18340*.

Knack, Stephen, & Keefer, Philip. (1997). Does Social Capital Have an Economic Payoff? A Cross-Country Investigation. *The Quarterly Journal of Economics, 112*(4), 1251-1288. doi:10.1162/003355300555475

Kreps, Sarah, George, Julie, Lushenko, Paul, & Rao, Adi. (2023). Exploring the artificial intelligence “Trust paradox”: Evidence from a survey experiment in the United States. *PloS One, 18*(7), e0288109.

Larrazabal, Agostina J, Nieto, Nicolás, Peterson, Victoria, Milone, Diego H, & Ferrante, Enzo. (2020). Gender imbalance in medical imaging datasets produces biased classifiers for computer-aided diagnosis. *Proceedings of the National Academy of Sciences, 117*(23), 12592-12594.

Leiter, Christoph, Zhang, Ran, Chen, Yanran, Belouadi, Jonas, Larionov, Daniil, Fresen, Vivian, & Eger, Steffen. (2024). ChatGPT: A meta-analysis after 2.5 months. *Machine Learning with Applications, 16*, 100541. doi:https://doi.org/10.1016/j.mlwa.2024.100541

Matthews, Gerald. (2008). Personality and information processing: A cognitive-adaptive theory. *Handbook of personality theory and assessment, 1*, 56-79.

Mcknight, D Harrison, Carter, Michelle, Thatcher, Jason Bennett, & Clay, Paul F. (2011). Trust in a specific technology: An investigation of its components and measures. *ACM Transactions on Management Information Systems (TMIS), 2*(2), 1-25.

Mokyr, Joel, Vickers, Chris, & Ziebarth, Nicolas L. (2015). The History of Technological Anxiety and the Future of Economic Growth: Is This Time Different? *Journal of Economic Perspectives, 29*(3), 31-50. doi:10.1257/jep.29.3.31

Montag, Christian, Becker, Benjamin, & Li, Benjamin J. (2024). On trust in humans and trust in artificial intelligence: a study with samples from Singapore and Germany extending recent research. *Computers in Human Behavior: Artificial Humans*, 100070.

Nguyen, Dennis, & Hekman, Erik. (2024). The news framing of artificial intelligence: a critical exploration of how media discourses make sense of automation. *AI & society, 39*(2), 437-451.

Nurski, Laura. (2024). *AI AT WORK: Why there's more to it than task automation*. Retrieved from Brussels: https://www.ceps.eu/ceps-publications/ai-at-work/

Obermeyer, Ziad, Powers, Brian, Vogeli, Christine, & Mullainathan, Sendhil. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science, 366*(6464), 447-453.

Preacher, Kristopher J, & Hayes, Andrew F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior research methods, 40*(3), 879-891.

Raj, Manav, & Seamans, Robert. (2019). Primer on artificial intelligence and robotics. *Journal of Organization Design, 8*(1), 11. doi:10.1186/s41469-019-0050-0

Riley, Breagin K, & Dixon, Andrea. (2024). Emotional and Cognitive Trust in Artificial Intelligence: A Framework for Identifying Research Opportunities. *Current Opinion in Psychology*, 101833.

Roth, Felix, Nowak-Lehmann D, Felicitas, & Otter, Thomas. (2022). Has the financial crisis shattered citizens' trust in national and European governmental institutions? Evidence from the EU member states, 1999–2010. In *Public Support for the Euro: Essays on Labor Productivity, Monetary Economics, and Political Economy, Vol. 2* (pp. 187-217): Springer.

Sætra, Henrik Skaug. (2024). A Machine's ethos? An inquiry into artificial ethos and trust. *Computers in Human Behavior, 153*, 108108.

Schopf, Cody M., Ramwala, Ojas A., Lowry, Kathryn P., Hofvind, Solveig, Marinovich, M. Luke, Houssami, Nehmat, . . . Lee, Christoph I. (2024). Artificial Intelligence-Driven Mammography-Based Future Breast Cancer Risk Prediction: A Systematic Review. *Journal of the American College of Radiology, 21*(2), 319-328. doi:https://doi.org/10.1016/j.jacr.2023.10.018

Seyyed-Kalantari, Laleh, Zhang, Haoran, McDermott, Matthew BA, Chen, Irene Y, & Ghassemi, Marzyeh. (2021). Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. *Nature medicine, 27*(12), 2176-2182.

Wintterlin, Florian, Hendriks, Friederike, Mede, Niels G, Bromme, Rainer, Metag, Julia, & Schäfer, Mike S. (2022). Predicting public trust in science: The role of basic orientations toward science, perceived trustworthiness of scientists, and experiences with science. *Frontiers in Communication, 6*, 822757.

Appendix A: Full Text of the Vignettes

AI Vignettes

Linguistics (AI condition):

"GPT-3 (Generative Pretrained Transformer 3), GPT-3.5, and GPT-4 are state-of-the-art language processing AI models developed by OpenAI. They are capable of generating human-like text and have a wide range of applications, including language translation, language modeling, and generating text for applications such as chatbots. GPT-3.5 gives a user the ability to give a trained AI a wide range of worded prompts. These can be questions, requests for a piece of writing on a topic of your choosing, or a huge number of other worded requests. This language-processing AI model is a program able to understand human language as it is spoken and written, allowing it to understand the worded information it is fed, and what to output in response. ChatGPT has a very wide range of abilities, from writing poems about sentient farts to explaining quantum mechanics in simple terms or writing full-length research papers and articles."

Medicine (AI condition):

"Researchers have developed an AI algorithm designed to improve cervical cancer detection through image analysis. This AI system is capable of identifying abnormal cell growth with greater accuracy and speed than traditional diagnostic methods, potentially offering earlier intervention for at-risk patients. In clinical trials, the AI tool demonstrated a significant reduction in missed diagnoses compared to conventional approaches, highlighting its potential to transform cancer screening processes."

Dating (AI condition):

"AI chatbots, such as Replika, are designed to provide companionship and support through personalized conversations. These chatbots use natural language processing algorithms to engage users in meaningful dialogue and adapt to their preferences over time. While they are not intended to replace human relationships, such AI systems are being increasingly used to alleviate loneliness and provide emotional support."

Science Vignettes

Linguistics (Science condition):

"A recent study shows that babies learn language by focusing on rhythm rather than phonetic details in the early stages of development. Researchers found that infants are highly attuned to the rhythmic patterns of speech, which helps them segment words and phrases even before they understand their meanings. This discovery sheds light on the foundational processes of human language acquisition."

Medicine (Science condition):

"Results of a preclinical study offer hope for new treatment options in the medium term. Ovarian cancer is often very aggressive and responds poorly to the therapies currently available. A recent study offers hope that this could change in the medium term. The researchers used an mRNA therapeutic. With its help, the tumor cells produced a protein again that prevents their uncontrolled proliferation or induces cell death. The mRNA therapeutic successfully combated cancerous cells and tumors in vitro as well as metastases in mice."

Dating (Science condition):

"Love is blind, the saying goes, and thanks to a new study we are now a step closer to understanding why. Researchers have measured how a part of the brain is responsible for putting our loved one on a pedestal in that first flush of romance. It is well known that romantic love changes the brain, releasing the so-called love hormone oxytocin, responsible for the euphoria we feel when falling in love. It turns out that when we are in love, our brain reacts differently. It makes the object of our affections the center of our lives. The new study shows that romantic love is linked to changes in behavior as well as emotion. The way that loved ones take on special importance, however, is due to oxytocin combining with dopamine, a chemical that our brain releases during romantic love. Essentially, love activates pathways in the brain associated with positive feelings."