Hazra, Sanchaita; Serra-Garcia, Marta

Working Paper

Understanding Trust in AI as an Information Source: Cross-Country Evidence

CESifo Working Paper, No. 11954

Provided in Cooperation with:
Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Hazra, Sanchaita; Serra-Garcia, Marta (2025) : Understanding Trust in AI as an Information Source: Cross-Country Evidence, CESifo Working Paper, No. 11954, CESifo GmbH, Munich

This Version is available at:
https://hdl.handle.net/10419/322516

CESifo WORKING PAPERS

11954
2025

June 2025

Understanding Trust in AI as an Information Source: Cross-Country Evidence

Sanchaita Hazra, Marta Serra-Garcia

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.ifo.de/en/cesifo/publications/cesifo-working-papers
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.ifo.de/en/cesifo/publications/cesifo-working-papers

Understanding Trust in AI as an Information Source: Cross-Country Evidence

Sanchaita Hazra and Marta Serra-Garcia*

This version: June 2025

Abstract

LLMs are emerging as information sources that influence organizational knowledge, though trust in them varies. This paper combines data from a large-scale experiment and the World Values Survey (WVS) to examine the determinants of trust in LLMs. The experiment measures trust in LLM-generated answers to policy-relevant questions among over 2,900 participants across 11 countries. Trust in the LLM is significantly lower in high-income countries—especially among individuals with right-leaning political views and lower educational attainment—compared to low- and middle-income countries. Using large-scale data on trust from the WVS, we show that patterns of trust in the LLM differ from those in generalized trust but closely align with trust in traditional information sources. These findings highlight that comparing trust in LLMs to other forms of societal trust can deepen our understanding of the potential societal impacts of AI.

JEL Classification: D83, D91, C72, C91.

Keywords: Information, generative AI, accuracy, trust, experiment.

*Hazra: Department of Economics, University of Utah, Salt Lake City, UT, United States (e-mail: sanchaita.hazra@utah.edu); Serra-Garcia: Rady School of Management, UC San Diego, La Jolla, CA, United States, and CESifo (email: mserragarcia@ucsd.edu). We thank Yan Chen, Mircea Epure, Oliver Hauser, Dorothea Kübler, Michel Marechal, Gael Le Mens, and Theo Offerman for their excellent feedback. This research was conducted under IRB #810990. Wanting Zhou and Sky Chen provided excellent research assistance.

1 Introduction

The adoption of new technologies, such as (generative) AI, is critical to how individuals and organizations make decisions in a variety of domains. Generative AI, in the form of LLMs, complements human intelligence (Bubeck et al., 2023; Jones, 2023) in two main ways: First, it helps individuals *complete tasks*, such as writing (e.g., Noy and Shang, 2023), idea generation (e.g., Doshi and Hauser, 2023), and programming (e.g., Peng et al., 2023); and second, it *provides information* to individuals, substituting search from standard web engines (Rowlands, 2025).

While there is growing evidence of the role of AI in human task completion, the impact of AI as an information source is less well understood. The main challenge in using LLMs as an information source is that, although LLMs attempt to answer questions accurately, they unintentionally make mistakes (Ji et al., 2023; Narayanan and Kapoor, 2024). Understanding how individuals differ in their trust in LLMs and their ability to detect errors is essential for explaining patterns of information acquisition via LLMs. Their trust can affect decision-making over time and shape the accumulation of societal knowledge (e.g., Del Rio-Chanona et al., 2023).

In this paper, we provide novel evidence of individuals' trust in LLMs as information sources by combining experimental data with a widely used cross-national dataset that provides comprehensive measures of trust — the World Values Survey (WVS) — across 11 countries. We first use the experiment to elicit a behavioral measure of trust in LLMs, and document within- and across-country differences. We then complement the analysis with WVS measures of generalized trust and trust in different organizations in society, to gain a deeper understanding of the meaning of trust in LLMs.

The experimental data stems from a large-scale pre-registered experiment that included participants from 11 countries: Australia, Canada, Chile, India, Kenya, Mexico, New Zealand, Spain, South Africa, the UK, and the US. Nine of these countries ranked among the top 50 countries in global traffic to the LLM (GPT-4o), including three in the top 3 (India, the US, and the UK) at the time of the study (Similarweb, 2024). Each participant ($N = 2,928$) was presented with 20 out of 100 policy-relevant questions and was

incentivized to evaluate whether the LLM's answer is accurate. The questions covered topics such as the gender wage gap, racial differences in financial markets, and macroeconomic statistics— topics individuals might look up through a search engine or ask directly to an LLM.

Across all countries, individuals' ability to spot incorrect responses is very limited. But, their trust in the LLM — defined as their belief that the LLM provided an accurate response – differs substantially across countries. In low- and middle-income countries, participants overly trust the LLM's accuracy. By contrast, individuals in high-income countries display lower trust in the LLM. The US stands out because participants showed excessive skepticism. Within countries, patterns of trust in the LLM differ significantly by age and education: individuals who are older and more educated trust LLMs more. Men trust the LLM more. Political preferences matter, but they do so differentially by country: In high-income countries, individuals with right-leaning political views trust the LLM less, while in low- and middle-income countries, individuals with right-leaning political views trust the LLM more.

These patterns of trust in the LLM as an information source across and within countries raise the question: What does trust in the LLM mean? The experimental data could reflect trust that is specific to the task in which the LLMs are used, or it could reflect a broader form of societal trust. To address this question, we use the World Value Survey (WVS). The WVS includes a measure of "generalized trust", defined as the degree to which individuals trust others in society, often interpreted as a form of social capital. Such trust has been shown to be a key determinant of economic and financial development (e.g., Guiso et al., 2004). In addition, the WVS includes measures of trust in different organizations in society: trust in traditional information sources (the press and TV) and trust in political organizations (government and political parties). These measures have been collected for the same 11 countries that are part of our experiment, including over 76,000 individuals.

The WVS data reveals that the cross-country patterns of trust in the LLM are not correlated with cross-country differences in generalized trust. For instance, while the country with the highest trust in LLMs is Kenya, it is the country with the lowest generalized trust – trust towards others.

By contrast, trust in LLMs is significantly correlated with trust in traditional informa-

tion sources. The countries with higher trust in the LLM also exhibit higher trust in the press and TV. For example, Kenya is the second highest country in trust in the press, close to its top ranking in trust in the LLM.

Not only are the cross-country patterns of trust in LLMs and trust in traditional information sources similar, but within-country differences in trust are also consistent with each other. The data show that individuals with right-leaning political views trust the press and TV less in high-income countries, reflecting the patterns of trust in LLMs. Similarly, individuals with high education in high-income countries trust both the LLM and the press significantly more than individuals with low education.

These findings provide new insights and a methodological contribution to the growing literature on human-AI interaction. First, we uncover new patterns of trust by focusing on the role of LLMs as information sources and providing a comparison across 11 countries. By including a cross-country comparison, we can speak to the use of this technology (LLMs) in a wide range of cultures, where it is already in use, and document the importance of considering the societal context, when examining whether individuals overly trust or distrust new technologies, such as LLMs.[1]

Second, we provide a new methodology to further understand human-AI interaction. A wide range of studies have documented the complicated interaction between humans and new technologies, such as algorithms, LLMs and other forms of AI (e.g., Bao et al., 2024; Xu et al, 2024; Wang et al., 2024; Caplin et al., 2025; Serra-Garcia and Gneezy, 2025). Some studies find algorithmic appreciation (e.g., Logg et al., 2019), suggesting that individuals are willing to trust new technologies, while others have shown algorithm aversion (e.g., Dietvorst et al., 2019; Dargnies et al., 2024). By combining experimental data with the WVS, we are able to better interpret why individuals show low levels of trust in LLMs in some societies

[1] Anecdotal evidence suggests increasing awareness of the role of LLMs as search engines. For example, in a Reddit thread with over 4,900 net positive votes, the topic is: "Anyone else basically done with Google search in favor of ChatGPT?". https://www.reddit.com/r/ChatGPT/comments/13ik8wh/ anyone_else_basically_done_with_google_search_in/, consulted October 18, 2024 (Reddit, 2024a). Similarly, a second Reddit thread posed the question "When you need a question answered or problems solved, which do you use first: an LLM (ChatGPT or not) or a search engine?" In the responses, the modal answer was "an LLM first" compared to "a search engine first." https://www.reddit.com/r/OpenAI/comments/18imawm/when_you_need_a_question_answered_or_problems/, consulted October 18, 2024.

but not in others. The analyses suggest that individuals may perceive LLMs as they perceive traditional information sources. Their trust in LLMs goes beyond the specific experimental task, reflecting trust in information sources in society more broadly, while at the same time being distinct from generalized (interpersonal) trust.

With these new insights, this paper contributes and complements existing work on the impacts of LLM in society. Several studies have shown that LLMs can be used to generate content (write messages), complementing our focus on using LLMs to retrieve content (answer questions). This work has shown that individuals have a very limited ability to detect content generated with LLMs, both if it is true or false (e.g., Kreps et al., 2022; Chen and Shu, 2023; Feuerriegel et al., 2023; Greevink et al., 2023; Spitale et al., 2023; Goldstein et al., 2024). Consistent with these and prior findings on individuals' ability to detect false information (Bond and DePaulo, 2006; Belot and van de Ven, 2017; Serra-Garcia and Gneezy, 2021), our experiments find a limited ability to detect mistakes made by the LLM, which were incorrect answers to factual and policy-relevant questions. But, they display significant overconfidence in the ability to detect mistakes, suggesting that individuals do not have well-calibrated beliefs when using LLMs as information sources.

Additionally, the findings in this paper complement existing work using large-scale surveys of trust across societies. Generalized trust has been shown to matter for economic and financial decision-making of individuals (e.g., Knack and Keefer, 1997; Karlan et al., 2009; Tabellini, 2010; Gorodnichenko and Roland, 2011; Alesina et al., 2013). Societies with higher levels of trust, for example, also show higher levels of stock market participation (e.g., Guiso et al., 2004). In this paper, we show that societies that have higher levels of trust in information sources exhibit such trust across both traditional and new sources: the press, TV and also LLMs. These results underscore that understanding trust in AI and regulating AI use will require specific consideration of the task and the society involved.

2 Data Sources

2.1 Experimental Design

The core experimental task required participants to assess whether answers provided by an LLM (GPT-4o) to policy-relevant factual questions were correct. Each participant evaluated 20 such responses, and one response was randomly selected for bonus payment. Participants received $2 if they correctly identified whether the LLM's answer was accurate.

The LLM as an Information Source. We submitted 100 policy-relevant questions to GPT-4o, evenly divided between U.S.-specific (50) and international (50) topics. For each, we recorded a one-shot response from the model.

We systematically selected questions relating to policy-relevant topics that have an objectively correct answer. We first identified the most discussed issues for a respective country from the Internet (e.g., from recent articles in top media outlets such as New York Times, Washington Post, Forbes). To obtain factually correct answers, we chose questions that have factual data available from major government websites and leading research institutes (such as Pew Research). Topics included the economy, migration, health, crime and politics. For instance, U.S.–specific questions addressed Black homeownership rates and the gender wage gap across age and racial groups. The questions, answers, and their respective sources are available Online Appendix F.

We introduced variation in how the questions were asked to the LLM, by including a source to use in the search for the answer or not providing a source. This variation allows us to examine whether LLM accuracy changes based on the type of prompt, and whether individuals' perceptions of accuracy are affected. In both cases, we ask the same question and ask the LLM to cite appropriate sources that it used to generate its responses. If the LLM was provided a source, it was told it could consult it, but it was not restricted to only using the provided source. The inclusion of a source did not significantly affect LLM accuracy (70% accuracy without a source, and 78% accuracy with a source, p-value= 0.202).

To evaluate the accuracy of LLM responses, three raters independently assessed each answer. Discrepancies were discussed until a consensus was reached. Detailed rater instruc-

tions are provided in the Online Appendix E.

Overall, the LLM answered 74% of the questions correctly. In the context of our questions, the source of incorrect responses of the LLM can be attributed to failures in knowledge retrieval and comprehension (Ji et al., 2024). Since the questions we asked are factual and the sources we used are reputable, we minimized the likelihood that the LLM error is based training on erroneous knowledge (Lin et al., 2022) or strategically driven wish to use propaganda. Because we relied on one-shot responses, our design does not capture stochastic variability in the model's outputs (Farquhar et al., 2024). Consistent with this type of incorrect responses, mistakes by the LLM are more common when the answer is provided in a figure, rather than in text or table (in one or more online sources). When errors occurred, they typically involved either misreporting a statistic from the correct source or fabricating a number not found in any cited material. For detailed information on the LLM's accuracy, see Online Appendix Tables A1 and A2.

Additional Measures. After providing their evaluation of LLM answers, we elicited two measures of participants' confidence. First, we elicited absolute confidence by asking participants how many of the 20 questions they believed they had evaluated correctly. Participants received a $1 bonus if their self-assessment matched their actual performance.

Next, we measured relative confidence by asking participants: "Relative to others who completed the same study as you, how do you think you did?" They selected a quartile within the overall performance distribution and earned a $1 bonus if their answer was accurate.

Experimental Procedures. We pre-registered a target sample of 3,000 observations from 11 countries.[2] We aimed to gather 300 observations each, or as many as possible up to this number, from the following countries: the United States, the United Kingdom, Canada, Australia, New Zealand, South Africa, Spain, and Mexico. Additionally, we intended to collect 200 observations each, or as many as possible up to this number, from India, Chile, and Kenya.

Of these 11 countries, 9 ranked among the top 50 globally in terms of online traffic

[2]The preregistration (blinded for peer-review) can be found at https://aspredicted.org/9b8r-t5wd.pdf.

to GPT models between July and September 2024, according to Similarweb (2024). Their rankings were as follows: India (1), United States (2), United Kingdom (3), Canada (6), Mexico (8), Australia (14), Spain (15), South Africa (20), and Chile (34).

Overall, the data include 2,928 participants, recruited on Prolific Academic in the summer of 2024. The fraction of female participants ranged from 30 to 71%, with average ages ranging from 27 to 40 years old, across countries. Their socio-demographic characteristics are summarized in Online Appendix Table B.1.

The survey was implemented using Qualtrics, with settings configured to prevent participants from copying and pasting text—limiting their ability to easily consult external sources. The survey also included captcha verification and the IP addresses were used to verify the location of the participant. At the end of the experiment, participants were asked whether they had looked up any answers online. Overall, 29% reported checking at least some of the LLM-generated answers. However, we find no evidence that this behavior was associated with greater accuracy in their evaluations (t-test, p-value= 0.237, Online Appendix Table B.9).

A potential concern with the experiment is that the findings may reflect participants' lack of effort rather than meaningful evaluations. We address this concern with two pieces of evidence. First, the majority of participants spent more time than the 15 minutes allocated for the task. The median completion time was 20.8 minutes, suggesting that most participants did not rush through their evaluations of the 20 LLM-generated answers. Second, as discussed in the Results Section, participants' actual accuracy was significantly correlated with their self-assessed accuracy across all countries. This relationship indicates that participants were able to distinguish between instances where they performed well versus poorly, suggesting that they were attentive during the task.

Another concern with respect to the comparison across countries is that there might be differential selection into participation in the experiment. A first step to address this concern is to control for participants' sociodemographic characteristics throughout (age, gender, education and political views), as we do. An additional step is to include the WVS data, and compare the results regarding trust across countries, because sampling for the WVS is representative of all people in the age 18 and older residing within private households in

each country. The similarity we document between the experimental data with patterns of trust (in traditional information sources) in the WVS provides evidence to suggest that cross-country comparisons within the experiment are not invalid.

2.2 Survey Measures of Trust: World Values Survey

We use trust measures from the World Values Survey (WVS; Inglehart et al., 2022). These measures have been widely used to study how trust relates to a range of important outcomes. In economics, for example, trust has been linked to key market outcomes, including stock market participation (Guiso et al., 2004) and economic growth (Knack and Keefer, 1997; Gorodnichenko and Roland, 2011). This data is particularly useful to understand the meaning of trust in LLMs, as measured in our experiment, by enabling the comparision between patterns of trust in the WVS (across and within countries) and those observed experimentally.

The most widely used WVS measure of trust is the generalized trust question. Participants are asked: "Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people?", with two answer options: "Most people can be trusted" or "Need to be very careful." We code a participant as showing generalized trust if they chose the first answer option.

Participants are also asked how much they trust various organizations. They are asked "Could you tell me how much confidence you have in them: is it a great deal of confidence, quite a lot of confidence, not very much confidence or none at all?" We consider two organizations that are important information providers, the press and TV. We also consider two organizations that are important for policy-making and implementation, the government and political parties, because they central to the topics of the questions and answers considered in the experiment.[3] We code participants as trusting an organization if they report having either "a great deal" or "quite a lot" of confidence in it. While "confidence" may differ conceptually from "trust", it is elicited within the same section of the WVS (Wave 7), and

[3]The full list of organizations included in the two most recent waves of the survey includes: churches, armed forces, press, television, labor unions, police, courts, government, political parties, civil service, universities, elections, major companies, banks, environmental organizations, women's organizations, charitable organizations, trade unions, regional organizations (e.g., EU, NAFTA), and the UN.

we thus treat it as a proxy for trust.

The WVS includes trust measures for thousands of individuals across the 11 countries included in our experiment. For example, there are 85,096 observations of generalized trust across these countries, including 17,601 responses from the most recent wave (Wave 7, 2017–2022). The set of countries included in each survey wave varies slightly. In the most recent wave, trust measures were collected in 9 out of the 11 countries in the experiment, and Spain and South Africa were not included. However, Spain and South Africa were included in a previous wave (2010-2014), which did not cover Canada, Kenya and the UK. We include all available waves of data and account for wave-level variation using wave fixed effects in all analyses. The number of observations by wave and country is reported in Online Appendix C.

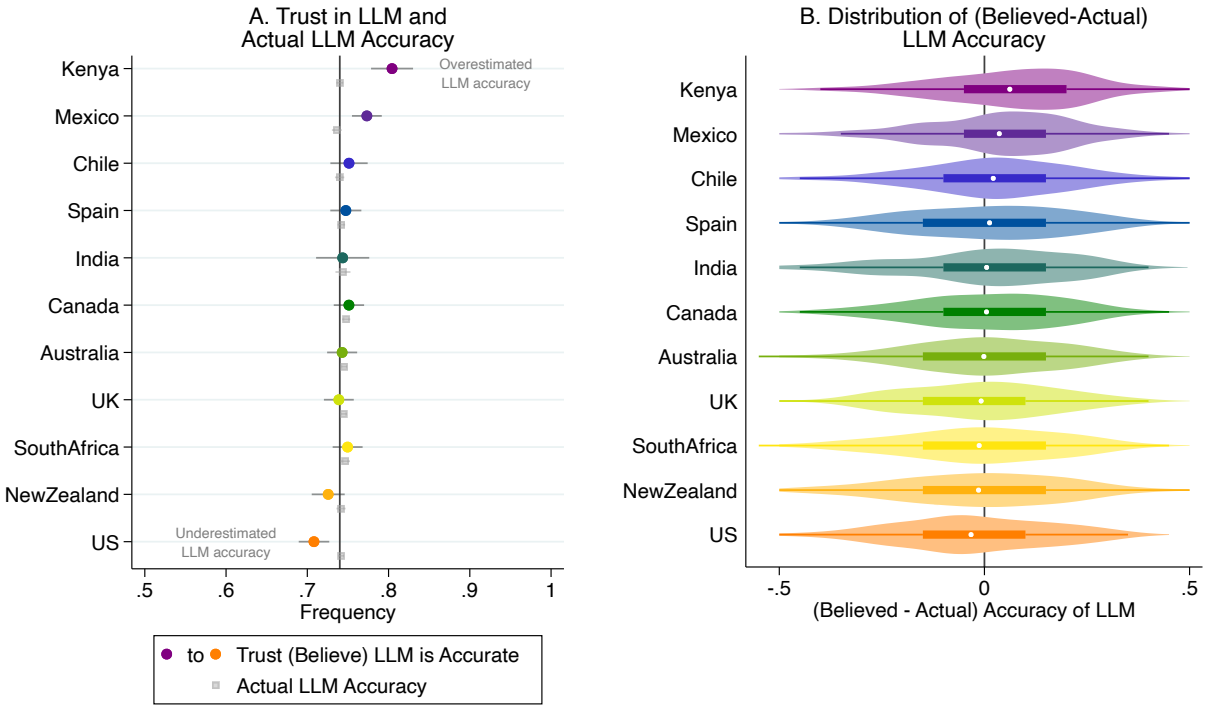
3 Results

3.1 Cross-country differences in trust in the LLM

We begin by examining how much individuals trust LLM-generated answers, measured as the percentage of cases (out of 20) in which participants rated the LLM as correct. Across all countries, and consistent with the wisdom of the crowds (e.g., Surowiecki, 2005; Lee and Lee, 2017), individuals are equally likely to overestimate or underestimate LLM accuracy (Fig. 1, Panel A; $t-$test, $p-$value$= 0.28$). This pattern holds both in the aggregate and in eight of the eleven countries. Three countries deviate from this pattern. In Kenya and Mexico, participants exhibit excessive trust in the LLM, overestimating its accuracy. In contrast, participants in the United States display significant distrust, consistently underestimating the LLM's accuracy (Table 1).

At the individual level, there is significant heterogeneity in trust toward the LLM (Fig. 1, Panel B). Across all countries, 64% of individuals overestimate or underestimate the LLM's accuracy by more than 10 percentage points, indicating that many individuals misperceive how accurate LLM-generated answers are. When comparing low- and middle-income countries (Chile, Kenya, India, Mexico, and South Africa) with high income countries (Australia,

Figure 1: Human Beliefs about the Accuracy of LLM Output



Notes: Panel A shows the average believed LLM accuracy and the actual LLM accuracy, based on regression-adjusted estimates for country-level beliefs, including fixed effects for question (100 questions in total) and presence of a source in it. The vertical gray line indicates the average accuracy of the LLM over all questions (74%). The bars underlying each symbol are 95% confidence intervals, from covariate-adjusted means that included socio-demographic characteristics and fixed effects for each question-answer and for whether the LLM was provided a source. Panel B shows the distribution of the difference between individual level believed accuracy and actual accuracy for the 20 LLM responses that each individual evaluated. The white circle in each row indicates the mean difference, while the box inside the density plot indicates the 25th and 75th percentiles. The vertical line indicates the observations for which there is no difference between believed and actual LLM accuracy.

Canada, New Zealand, Spain, UK and US) we find significantly higher trust in the LLM's accuracy in low- and middle-income countries. Their average perceived accuracy is 76.2%, compared to 73.7% in high-income countries ($t-$test, $p-$value< 0.001). This latter estimate is marginally below the actual accuracy.

Result 1. *Trust in the LLM is aligned with its actual accuracy in a majority of the countries. Across three countries there are significant differences: In the US individuals overly distrust the LLM, while individuals from Kenya and Mexico overly trust it.*

Table 1: Human Trust and Actual LLM Accuracy

	(1) Trust LLM Belief is accurate = 1		(2) LLM Is accurate= 1	
Canada	0.008	(0.013)	0.003	(0.003)
Chile	0.008	(0.015)	-0.006*	(0.003)
India	0.001	(0.019)	-0.001	(0.005)
Kenya	0.061***	(0.016)	-0.005*	(0.003)
Mexico	0.030**	(0.013)	-0.009**	(0.004)
NewZealand	-0.017	(0.014)	-0.004	(0.003)
SouthAfrica	0.007	(0.013)	0.001	(0.004)
Spain	0.005	(0.014)	-0.003	(0.003)
UK	-0.004	(0.013)	-0.001	(0.003)
US	-0.035***	(0.013)	-0.004	(0.003)
Female	-0.017***	(0.006)		
Age	0.001**	(0.000)		
High Education	0.015**	(0.007)		
Politically leaning right (standardized)	-0.000	(0.003)		
Constant	0.687***	(0.023)	0.962***	(0.023)
Observations	58,560		58,560	
Clusters	2,928		100	

Notes: This table displays the estimated coefficients from linear regressions on likelihood that the human beliefs the LLM answered correctly (column (1)), on the likelihood that the LLM is correct (column(2)). The regressions include fixed effects for question (100 questions in total), and presence of a source in it. Each country name represents an indicator variable for that country, where the omitted country is Australia. Robust standard errors, clustered at the participant level next to column (1) and at the question level next to column (2), in parentheses. * p<.10; ** p<.05; *** p<.01

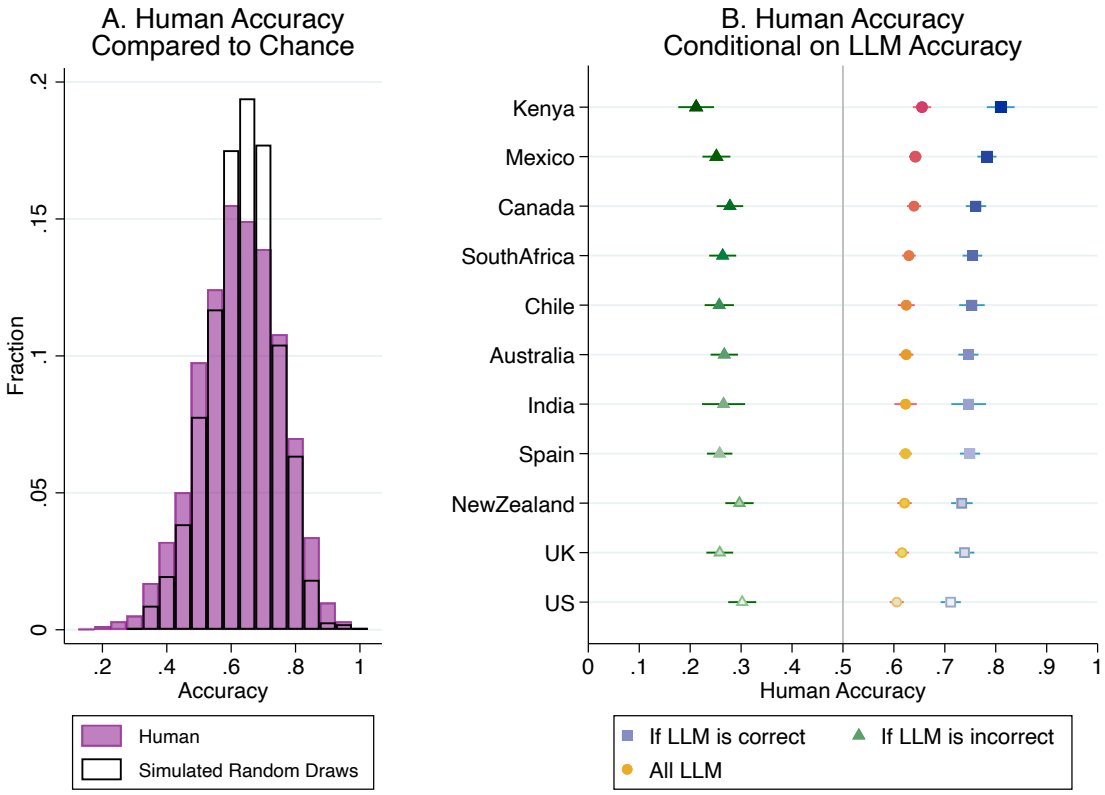
3.2 Accuracy and Confidence

Next, we examine individuals' ability to spot mistakes by the LLM. We answer the question, when confronted with a specific accurate or an inaccurate LLM answer, are they able to distinguish correct and incorrect answers? We measure how often the participant accurately (or correctly) rates the LLM output as correct or incorrect.

Individuals' ability to accurately rate the LLM's answers is limited. They are correct in 62.7% of the cases. While this rate is higher than 50%, the appropriate benchmark is how often an individual would spot mistakes if they knew the accuracy of the LLM (74%) and they would randomly rate an answer as correct with a probability of 0.74. We simulate the accuracy of such a random data generator, with 58,640 draws that follow a probability of 0.74 of rating an LLM statement as correct. In that case, the accuracy (by chance) would

be 61.6%. Although the accuracy by chance is statistically significant lower than that of participants, 62.7% ($t-$test, $p-$value$<$ 0.001), it is only of 1.1 percentage points lower in magnitude. As Panel A of Fig. 2 shows, the distribution of accuracy based on human responses, in purple, is similar to the distribution based on chance.

Figure 2: Human Accuracy in Evaluating LLM Output



Notes: Panel A shows the distribution of human accuracy (in purple). It also shows the distribution of accuracy if from 58,640 random draws which evaluate an answer as correct with a 70% chance, and incorrect with a 30% chance (white). Panel B shows average accuracy of humans, by country, (1) if the LLM is correct (blue squares), (2) if the LLM is incorrect (green triangles), and (3) it shows average accuracy by country (orange circles). The bars underlying each symbol are 95% confidence intervals, from covariate-adjusted means that included socio-demographic characteristics and fixed effects for each question-answer and for whether the LLM was provided a source. See Online Appendix Tables B.2 and B.3 for detailed results.

When examining participants' ability to evaluate answers provided by the LLM, it is important to distinguish between their ability to identify correct answers and their ability to identify mistakes by the LLM. The data show that individuals make mistakes both when confronted with correct and incorrect LLM-generated responses. The rate of mistakes, in fact, parallels their beliefs about the LLM's accuracy (i.e., their trust in the LLM on average).

Conditional on the LLM being correct (blue squares), humans correctly rate the LLM 75.1% of the time, which is close to their belief about how often the LLM is accurate in general. Consistent with this finding, when the LLM is incorrect (green triangles), which occurs in a small fraction of cases (26%), humans correctly rate the LLM in 26.7% of the cases.

The differences across countries mirror differences observed in trust of the LLM. Participants in Kenya who overly trust the LLM exhibit a significantly lower ability to rate the LLM accurately when it is incorrect, and participants in the US who distrust the LLM exhibit a lower ability to rate the LLM accurately when it is correct (as shown in detail in Online Appendix Table B.2). These findings lead to Result 2.

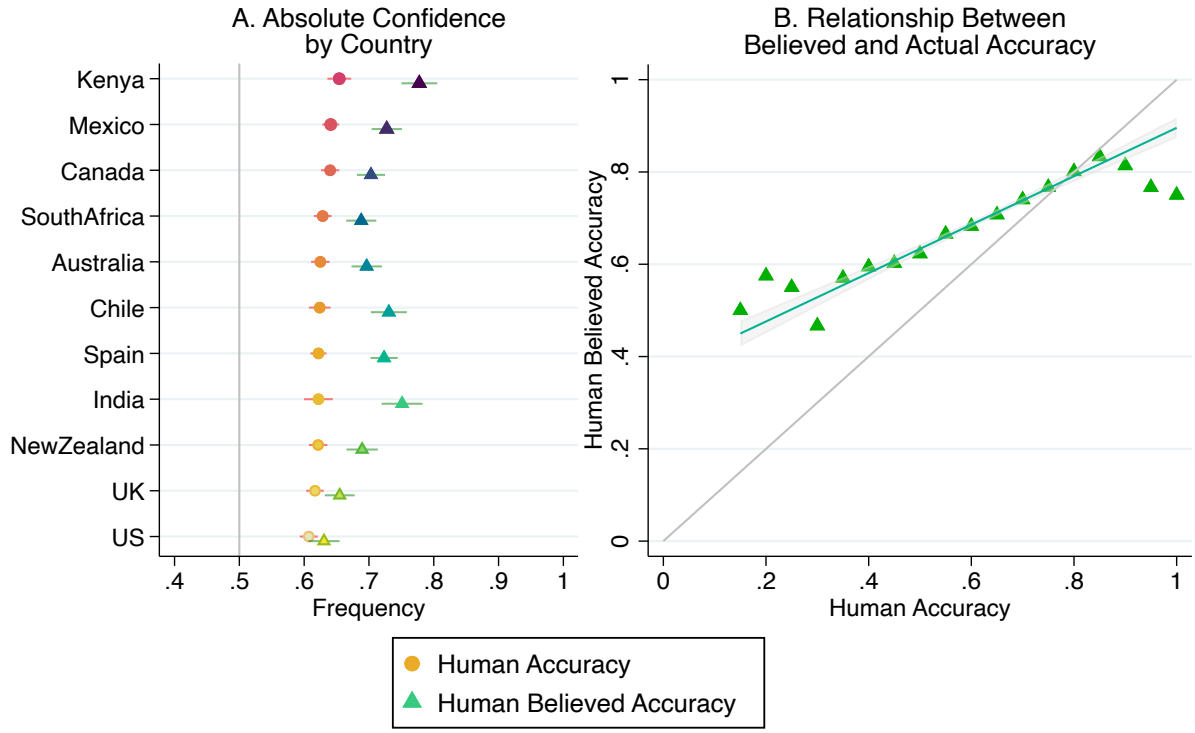
Result 2. *Individuals' ability to accurately assess the LLM is limited, only marginally better than chance, both if the LLM is correct and if it is incorrect.*

An important question that follows from these findings is whether individuals are aware of their limited ability to assess LLM-generated answers. In a majority of the cases, individuals are not. We first consider individuals' absolute confidence, based on how often they believe they correctly rated the LLM's answers. Individuals are overly confident: They believe they can accurately rate 70% of the LLM-generated answers, which is significantly larger than their actual average ability of 62.7% ($t-$test, $p-$value< 0.001), as shown in Panel A of Fig. 3. The difference between actual and believed ability varies by country, with Kenya and India showing the highest confidence (78% and 75%, respectively), and the US and the UK showing the lowest confidence (63% and 66%, respectively).

Although individuals overestimate their ability, their beliefs are not simply noisy. Panel B of Figure 3 shows that there is a significant correlation between participants' believed and actual performance – the Spearman rank correlation coefficient is 0.32 ($p-$value< 0.001). This finding shows that their confidence provides some information about their actual ability, although this information is biased (a finding that is consistent in all countries, as shown in Online Appendix Figure B.1).

Individuals' perceived ability relative to others (relative overconfidence) is also biased (as shown in Online Appendix Figure B.2). Their average reported quartile in the distribution of ability is 1.91, which is lower than the expected average of 2.5 ($t-$test, $p-$value< 0.001).

Figure 3: Human Confidence in their Ability to Assess LLM Accuracy



Notes: Panel A shows average beliefs about accuracy when evaluating the LLM responses (green triangles), comparing them to actual accuracy (orange circles), by country. Panel B plots the relationship between believed accuracy and actual accuracy, across all countries. In Panel B each triangle indicates a demi-decile of the distribution, and the fitted line shows is the estimated linear relationship, with a 95% confidence interval shown in the gray are. The solid gray line indicates the 45 degree line. See Online Appendix Table B.4 for regression results, and Figure B.1 for a figure of Panel B by country.

Comparing participants' beliefs to their actual placements, only 19 percent of participants in the third quartile of ability place themselves in that quartile. Similarly, only 3.3 percent of participants in the bottom quartile of ability place themselves in the correct quartile. These beliefs violate the condition within Bayesian updating that the largest group of subjects placing themselves in each quartile must belong to that quartile, consistent with overconfidence (Moore and Healy, 2008; Burks et al., 2013; for detailed results by country see Online Appendix Table B.4, column (3)).

Result 3. *Individuals overestimate their ability to accurately assess the LLM, in absolute and relative terms.*

Gender differences in confidence are also present in the data, consistent with previous

evidence (Coffman, 2014; Bordalo et al., 2019; Carvajal et al., 2024). Women are significantly less confident in both absolute and relative terms (as shown in Online Appendix Table B.4). In terms of absolute confidence, women believe their accuracy is 5.5 percentage points lower than men – out of 20 questions, they believe they correctly evaluate one LLM answer less than men, on average. In relative terms, they also display significantly less confidence: women's belief about the quartile in which they performed in the distribution of ability is 0.21 higher (2.01 on average, compared to 1.80 for men).

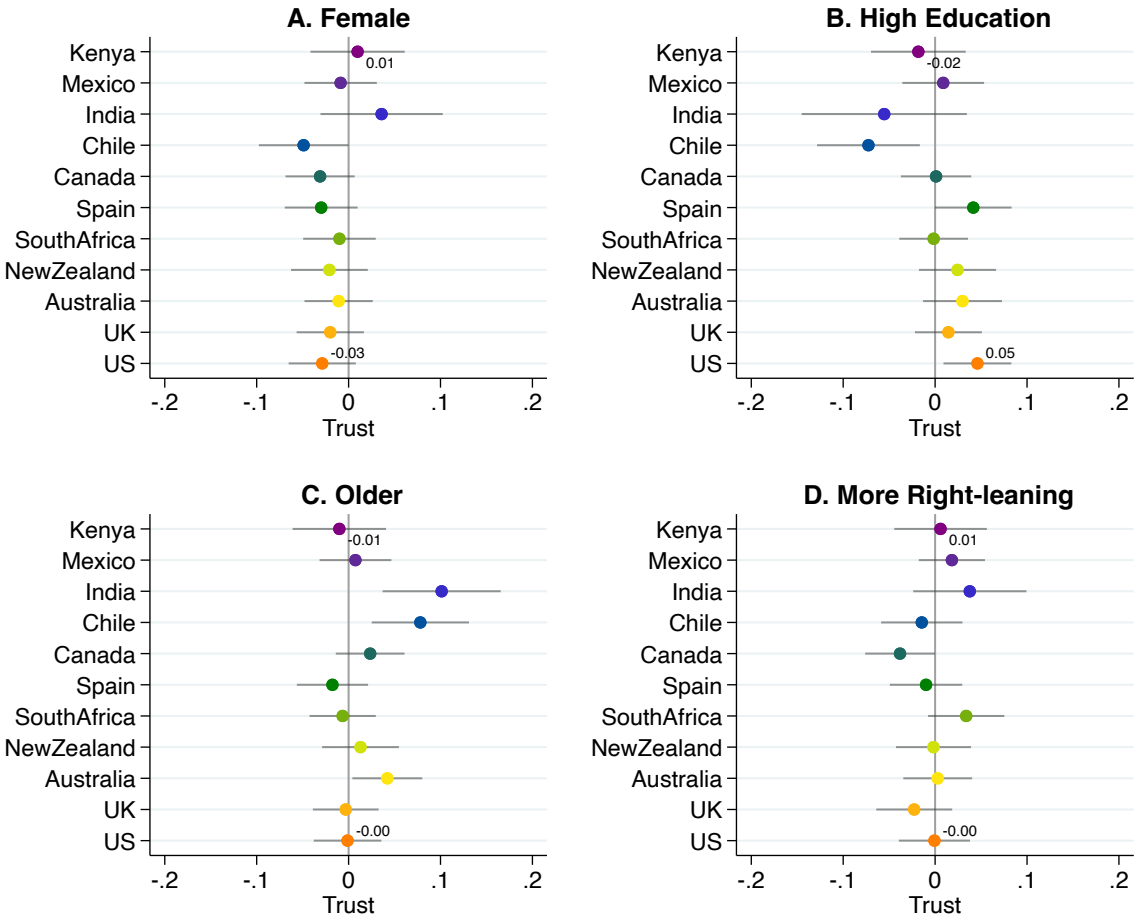
3.3 Determinants of Trust in the LLM

Within-country heterogeneity. We investigate whether individuals of different gender, age, educational level and politics exhibit systematically different patterns of trust in the LLM to better understand what the key determinants of trust may be.

We explore this within-country heterogeneity in Figure 4, which shows differences in trust, depending on whether the participants are female (relative to male), have higher educational attainment, are older (above median in age) and more right-leaning in politics (above median in their political views) for each country. Panel A shows that trust in the LLM is directionally lower among women in most countries. Trust in the LLM varies by educational level. Individuals with higher education levels are significantly more likely to trust the LLM, especially in high-income countries like the US and Canada. By contrast, they are less trusting in low-income countries.

The differences in trust by individuals' demographic characteristics, highlight that there are significant differences in trust across low- and middle-income countries, relative to high-income countries. We show this difference in Table 2. As column (2) shows, individuals in high-income countries who have a higher education exhibit more trust, while such a relationship does not emerge for low- and middle-countries. In the latter group of countries, individuals with more right-leaning political views show directionally higher trust in the LLM, a relationship that reverses for high-income countries. Given these findings, we examine differences in trust in the LLM, accuracy, and confidence, focusing on two groups of countries: high-income compared to low- and middle-income countries.

Figure 4: Differences in Trust in LLM by Individual's Demographic Characteristics



Notes: Results shown the estimated differences for four demographic characteristics by country. Panel A shows the average difference in trust in the LLM of female, relative to male participants. Panel B shows the average difference in trust in the LLM of individuals with higher education (completed 4-year college education), compared to low and mid education levels. Panel C shows the average difference for older individuals (above median age) relative to younger individuals. Panel D shows the average difference for individuals who indicate their political position is more right-leaning (above median position). Medians are calculated at the country level. Confidence intervals are shown as dark gray lines, and are calculated using the Delta method, for standard errors clustered at the participant level. See Table B.6 for details.

Country income level. Figure 5 shows trust, accuracy and confidence by country income group. Individuals in low and middle-income countries trust the LLM more, as described earlier. They are not better able to identify the accuracy of the LLM, but they believe they are. Specifically their ability does not differ by a large magnitude: 63.3% accuracy in low- and middle-income countries, compared to 62.3% in high-income countries, $t-$test, $p-$value$=0.044$). By contrast, in low- and middle-income countries, individuals believe their

Table 2: Heterogeneity in Trust by Demographics, Question-Level

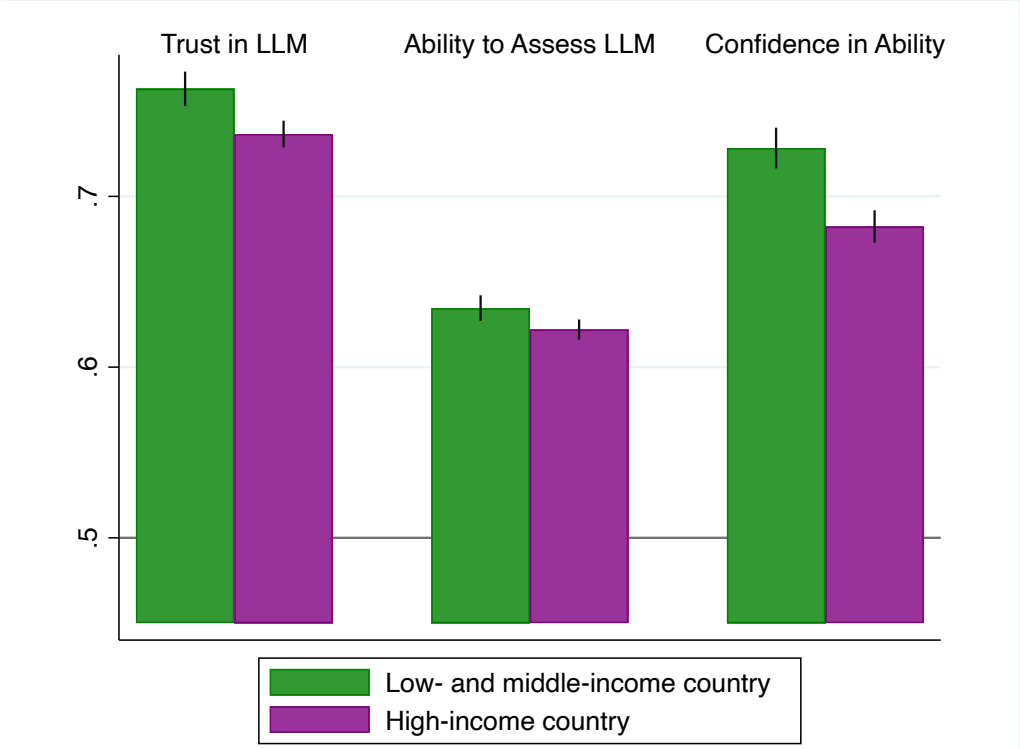
	(1)	(2)
	Trust the LLM	
Country Income Group		
High-Income Country=1	-0.024***	-0.025*
	(0.006)	(0.014)
Demographic characteristics		
Female=1	-0.019***	-0.010
	(0.006)	(0.010)
High Education=1	0.017***	-0.009
	(0.007)	(0.011)
Older age (above median)=1	0.012*	0.020*
	(0.006)	(0.010)
Right-leaning politics (above median)=1	-0.000	0.021**
	(0.006)	(0.010)
Female=1 × High-Income Country=0		0.000
		(.)
Interaction Terms		
Female=1 × High-Income Country=1		-0.015
		(0.013)
High Education=1 × High-Income Country=0		0.000
		(.)
High Education=1 × High-Income Country=1		0.039***
		(0.014)
Older age (above median)=1 × High-Income Country=0		0.000
		(.)
Older age (above median)=1 × High-Income Country=1		-0.011
		(0.013)
Right-leaning politics (above median)=1 × High-Income Country=0		0.000
		(.)
Right-leaning politics (above median)=1 × High-Income Country=1		-0.032**
		(0.013)
Constant	0.723***	0.726***
	(0.020)	(0.021)
Observations	58,560	58,560
Clusters	2,928	2,928

Notes: This table displays coefficients from linear regressions on an indicator variable for whether the individual trusted the LLM, at the question level. The regressions include fixed effects for question (100 questions in total), and presence of a source in it. Robust standard errors, clustered at the participant level, in parentheses. * p<.10; ** p<.05; *** p<.01

answers are correct 72.4% of the time, while individuals in high-income countries believe their evaluation of the LLM is correct 68.5% of the time ($t-$test, $p-$value< 0.001).

Result 4. *In low- and middle-income countries individuals trust the LLM more, especially if they have more right-leaning political views. Their ability to assess the LLM's accuracy is not different from that of individuals in high-income countries, but they are more confident in their ability.*

Figure 5: Trust in LLM and Individual Accuracy, by Country Income



Notes: Results are shown for low and middle income countries (Chile, Kenya, India, Mexico, and South Africa) and for high income countries (Australia, Canada, New Zealand, Spain, UK and US). The first two columns show individual trust in the LLM, measured as the individual trusting the LLM to be accurate more than the median, by income level. The second pair of columns shows average ability to accurately evaluate the LLM answers. The third pair of columns shows individuals' confidence in their ability to accurately evaluate the LLM. Confidence intervals are shown as dark gray lines, and are calculated using the Delta method, for robust standard errors (HC3).

Question-level characteristics. We also explore whether the characteristics of the questions asked to the LLM could affect individuals' trust and in turn suggest types of questions about which individuals show particularly high levels of trust (shown in Appendix Figure B.3). We examine whether trust in the LLM depends on whether it is asked about a matter concerning the US or an international topic, and whether the LLM is given a source to use to answer the question or not. Although the accuracy of the LLM does not vary significantly, the participants trust the LLM marginally more when it is asked about an international matter (76.2%) than a US matter (73.0%, $t-$test, $p-$value$<$ 0.001). Their trust is equal when it provided a source for their answer (74.3%) and when it did not (75.1%, $t-$test, $p-$value$=$ 0.213), and it does not vary significantly by the topic of the question (climate, crime, economics, education, health, or politics).

3.4 Broader Measures of Trust and Trust in the LLM: Using WVS to Understand Trust

The findings thus far raise the question: Does trust in the LLM reflect individuals' generalized trust in others, their trust in information sources, or is it a distinct construct altogether? We address this question using data from the World Values Survey (WVS).

We begin by comparing measures of trust in the WVS to trust in the LLM, as measured in the experiment. Figure 6 presents trust in the LLM in Panel A, generalized trust from the WVS in Panel B, trust in two traditional information sources—the press and television—in Panels C and D, and trust in government and political parties in Panels E and F. To ensure comparability, all trust measures are coded as binary variables.[4]

The main result is that trust in the LLM differs from generalized trust. At the country level, the relationship between these two measures of trust is negative. The Spearman rank correlation coefficient between generalized trust and trust in the LLM across the 11 countries is -0.5909 (p-val$= 0.0556$). By contrast, trust in the LLM is strongly correlated with trust in traditional information sources. The Spearman rank correlation coefficient for the correlation with the press is 0.75 (p-val$= 0.0085$) and with TV is 0.74 (p-val$= 0.0098$). Trust in the government and in political parties shows a weaker relationship to trust in the LLM: the Spearman correlation coefficient for government is 0.55 (p-val$= 0.083$) and with TV is 0.52 (p-val$= 0.1025$).

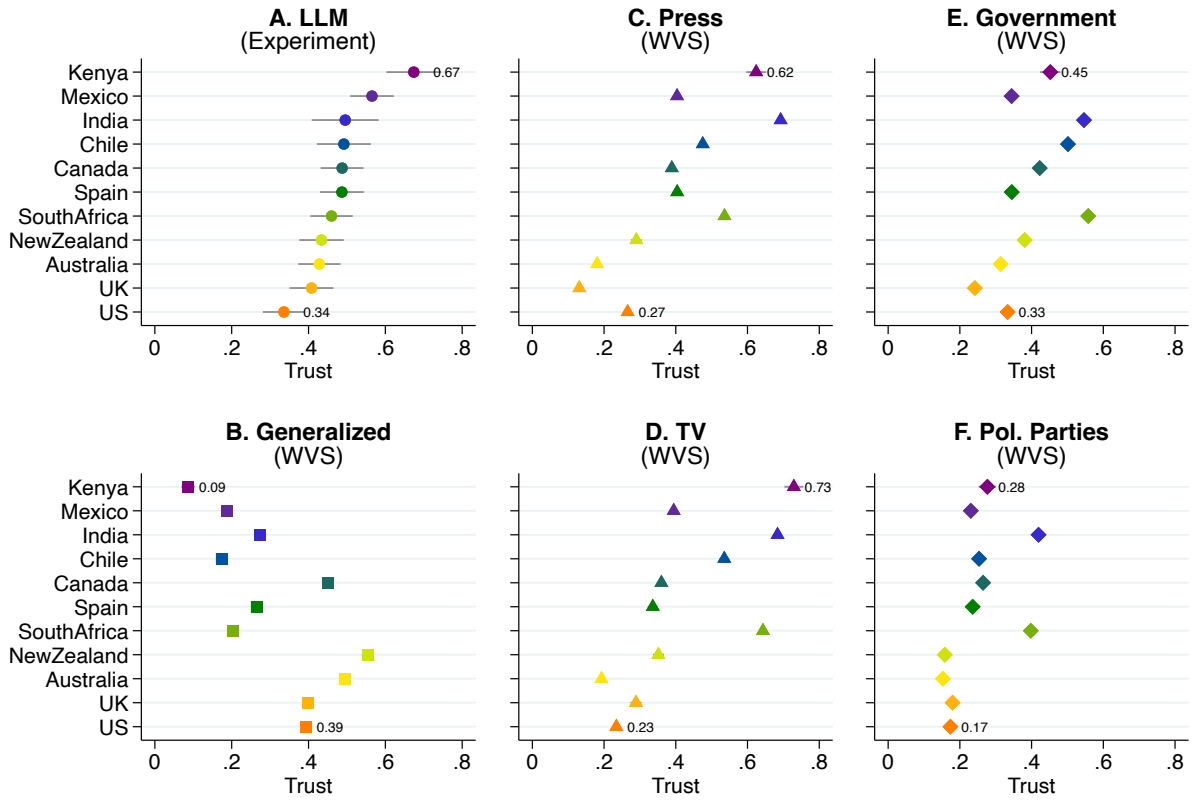
Taken together, these results suggest that trust in the LLM reflects a broader form of trust—specifically, trust in traditional information sources. In countries in which the press and TV is considered trustworthy, individuals also show greater trust in the LLM.

To further explore these findings, we examine whether demographic differences in trust in information sources (press and television) mirror the patterns observed in trust in the LLM. To simplify the comparison, we create two groups of countries by income (low and middle-income countries and high-income countries), as above.

Figure 7 shows the similarities between trust in the LLM and trust in other information

[4]To construct an individual-level binary measure of trust in the LLM, we use an indicator variable equal to one if the participant's trust in the LLM is above the sample median. The cross-country ranking of trust remains qualitatively similar using this definition.

Figure 6: Measures of Trust: Experimental and World Values Survey

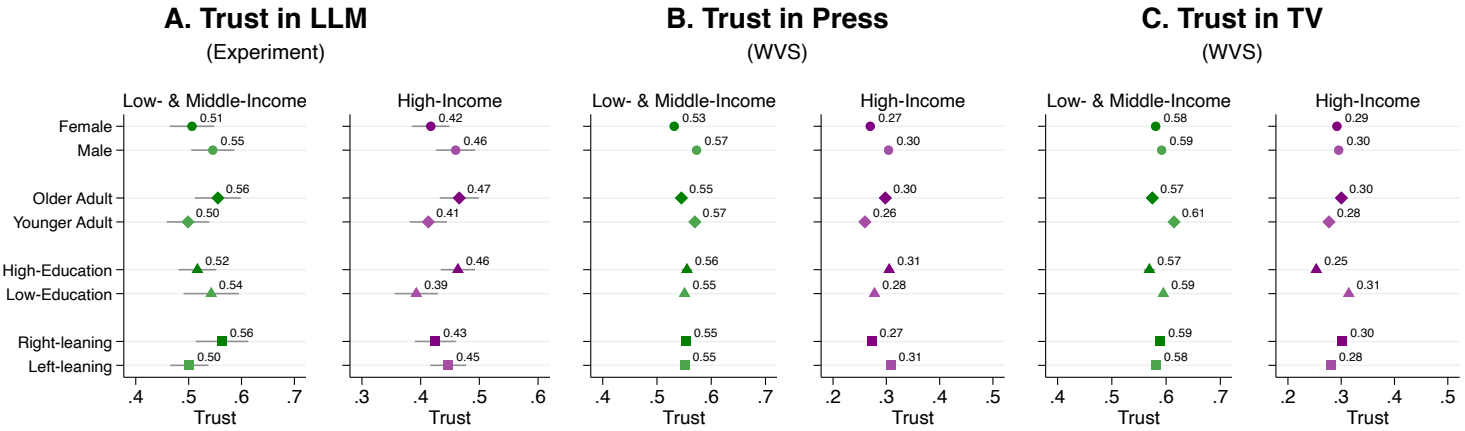


Notes: Panel A shows trust in the LLM measured in the experiment. Panel B shows generalized trust in others in society, measured in the WVS. This variable takes value one if the individual indicates that most others can be trusted, and 0 otherwise. Panels C through D show trust in the press, TV, government, and political parties. These variables take value one if the individual indicates that they have a "great deal" or "quite a lot" of confidence in each entity.

sources, measured by the WVS. Full regression estimates are provided in Table 3. Focusing on gender, we find that females are less trusting of all information sources, with significant differences in the World Values Survey. Consistent with the experimental results on trust in the LLM, this gender difference holds across both high-income and low- and middle-income countries.

Turning to political views, we find that in low- and middle-income countries, right-leaning individuals exhibit higher trust in the LLM and directionally greater trust in the press and television. In contrast, in high-income countries, the relationship reverses: right-leaning individuals show significantly lower trust. In these countries, individuals with right-leaning political views show, on average, lower trust in LLMs and the press. These findings

Figure 7: Determinants of Trust Across Information Sources



Notes: This figure shows the estimated trust in LLM (Panel A), in the press (Panel B) and in TV (Panel C) by demographic group (gender, age, education and political views) and country income level. Older and younger adults are defined relative to the median age in the experiment. Similarly, right-leaning compared to left-leaning individuals are defined relative to the median position in the experiment. Estimated coefficients are show in Table 3.

are consistent with evidence that right-leaning individuals trust mainstream media sources less (e.g., Tsfati and Ariely, 2014). Regarding education, the data shows that individuals with higher education levels in high-income countries trust the LLM and the press more. However, trust in TV follows somewhat different patterns in both cases, suggesting that trust in LLMs may be more closely aligned with trust in the press.

Finally, with regards to age, older individuals in high-income countries tend to trust traditional information sources more (consistent with Sutter and Kocher, 2007) and their patterns are similar for trust in AI. By contrast, in low-income countries, older adults trust traditional information sources less, while they trust AI more. This finding suggests that there may be a different age gradient for AI than traditional sources, depending on the media landscape in different countries.

4 Discussion

LLM prevalence. A concern regarding the experimental results is that they may reflect cross country differences in familiarity with the LLM. To address this concern, we examine whether the heterogeneity in trust in LLM that we document is related to experience with

Table 3: Heterogeneity in Different Types of Trust

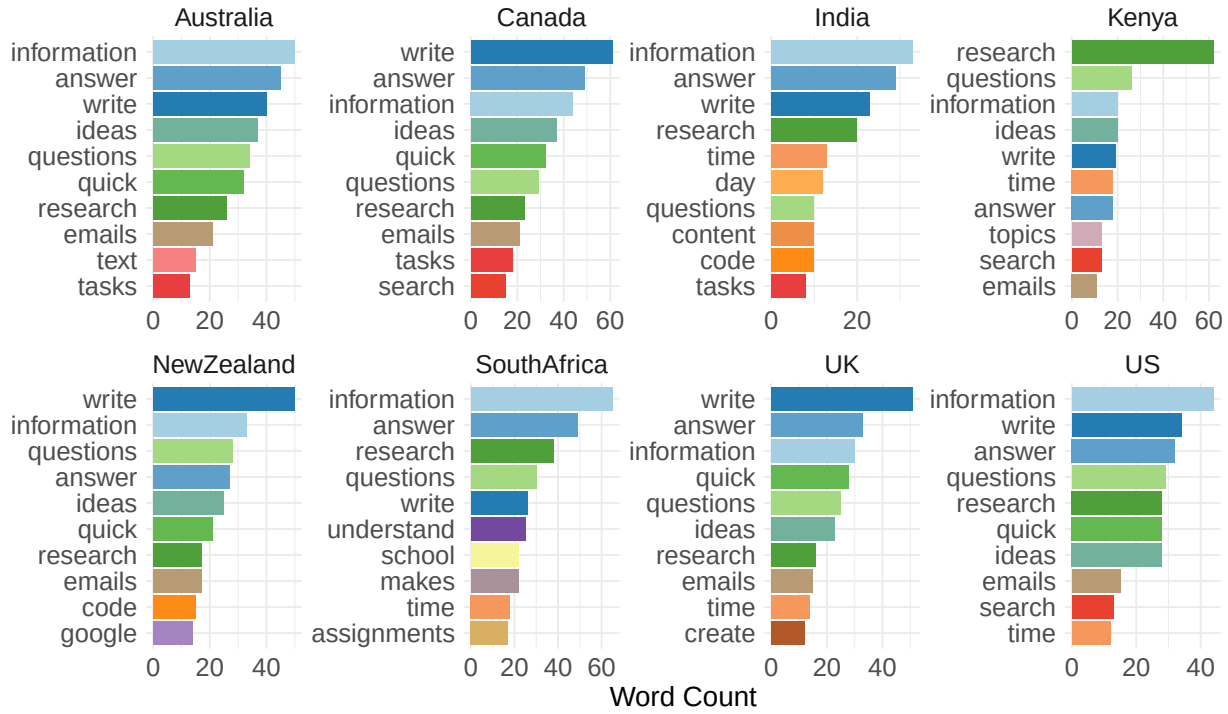
	(1)	(2)	(3)
		Trust	
	in LLM	in Press	in TV
Income group			
High-income Country=1	-0.114***	-0.299***	-0.338***
	(0.043)	(0.009)	(0.009)
Demographic characteristics			
Female=1	-0.039	-0.041***	-0.011**
	(0.030)	(0.005)	(0.005)
High Education=1	-0.026	0.004	-0.026***
	(0.032)	(0.006)	(0.006)
Older age (above median)=1	0.057*	-0.025***	-0.040***
	(0.030)	(0.005)	(0.005)
Right-leaning politics (above median)=1	0.062**	0.003	0.008
	(0.031)	(0.005)	(0.005)
Interaction Terms			
Female=1 × High-income Country=1	-0.003	0.007	0.008
	(0.038)	(0.007)	(0.007)
High Education=1 × High-income Country=1	0.097**	0.023***	-0.035***
	(0.040)	(0.008)	(0.008)
Older age (above median)=1 × High-income Country=1	-0.004	0.063***	0.063***
	(0.038)	(0.008)	(0.008)
Right-leaning politics (above median)=1 × High-income Country=1	-0.084**	-0.038***	0.013*
	(0.039)	(0.007)	(0.007)
Constant	0.511***	0.611***	0.363***
	(0.033)	(0.009)	(0.013)
Observations	2928	76219	72840

Notes: This table displays coefficients from linear regressions on an indicator variable for whether the individual showed trust in the LLM above the median (in the experiment) for column (1), and whether the individual indicated high or quite a lot of confidence in the press and in TV within the World Values Survey, columns (2) and (3) respectively. Wave fixed effects are included in the latter two specifications. Robust standard errors in parentheses.

its use. We find no significant association between trust in the LLM and regular use of it. As shown in Online Appendix Table B.10, the relationship between trust and experience shows a coefficient of -0.018 ($p-$value= 0.374), in a regression analysis that controls for country and individual characteristics. The analysis also shows that there is heterogeneity in use of the LLM: In low- and middle-income countries, almost 50% of participants report to use GPT regularly (49.7%), while in high-income countries only 36.7% of individuals report to use it with the same frequency. In the data, however, patterns of trust in LLM as an information sources are not explained by these differences in experience with it.

Uses of LLMs. We can also explore whether there are significant differences in how individuals use the LLM across countries. We use the responses provided via an open-ended

Figure 8: Reported Use of LLM, by country



Notes: This figure shows the 10 most frequently used words by participants when responding to the question "In what ways, if any, is GPT useful for you?". Each word is consistently colored across panels, for comparability. Stop words are filtered out, as well as words referring to GPT and help are excluded as they are the most common way answers start.

question at the end of the experiment, to examine whether the most frequently used words are similar across participants in different countries. We focus on English-speaking countries and participants' responses to the question "In what ways, if any, is GPT useful to you?" Figure 8 shows the top 10 most frequent words by country. Five of the most common words appear consistently across countries: "information," "answer," "write," "research," and "ideas." These descriptive findings provide suggestive evidence that the way individuals use the LLM across countries does not vary significantly.

5 Conclusion

LLMs are increasingly used as information sources, providing answers to a wide range of user-generated questions. Over time, LLMs may replace traditional search engines—reducing users' search costs but potentially compromising the accuracy of their beliefs. As LLMs

become more integrated into users' daily information-seeking behaviors, it becomes essential to examine the extent of public trust in these tools and the patterns that trust follows.

This paper documents significant cross-country differences in beliefs about the LLM's reliability as an information source, and in individuals' perceptions of their ability to assess the LLM's information. The data reveal that trust in LLMs is distinct from generalized trust but closely mirrors trust in traditional information sources, such as the press. These findings highlight that trust in the LLM can reflect trust in institutions of society that perform similar roles. These may slowly be substituted by LLMs. As LLMs evolve into primary information sources, patterns of use and trust are likely to vary substantially across countries. Effective regulation of LLMs will require an understanding of the sources of user trust and a careful accounting of cross-country differences.

Differences in trust in LLMs raise concerns about potential inequities emerging across countries over time—particularly in time-sensitive domains such as public health (Nan et al., 2022; De Angelis et al., 2023; Meyrowitsch et al., 2023), voting (Schneid, 2024), and national security (Horowitz and Kahn, 2024). In countries where individuals overly trust LLMs, educational interventions may be necessary to improve users' critical engagement. Conversely, in contexts where trust is unduly low, trust-building initiatives could help ensure the effective use of LLMs. As new regulations are introduced (Gibney, 2024; Warren, 2024), the interaction between societal norms and regulations (e.g., Lane et al., 2023) should be carefully considered, as cross-country differences in trust could significantly alter the impacts of different policies.

References

1. Bao, L., Huang, D., & Lin, C. (2024). Can artificial intelligence improve gender equality? Evidence from a natural experiment. *Management Science*.

2. Belot, M., & van de Ven, J. (2017). How Private Is Private Information? The Ability to Spot Deception in an Economic Game. *Experimental Economics* 20 (1): 19–43.

3. Bond, C. F., & B. M. DePaulo. 2006. Accuracy of Deception Judgments. *Personality and Social Psychology Review* 10 (3): 214–34.

4. Bordalo, P., Coffman, K.B., Gennaioli, N. & Shleifer, A. (2019). Beliefs about Gender. *American Economic Review* 109 (3): 739–73.

5. Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y.T., Li, Y., Lundberg, S. & Nori, H., (2023). Sparks of artificial general intelligence: Early experiments with gpt-4. arXiv preprint arXiv:2303.12712.

6. Burks, S. V., Carpenter, J.P., Goette, L. , & Rustichini, A. (2013). Overconfidence and Social Signalling. *Review of Economic Studies* 80 (3): 949–83.

7. Caplin, A., Deming, D. J., Li, S., Martin, D. J., Marx, P., Weidmann, B., & Ye, K. J. (2025). The ABC's of Who Benefits from Working with AI: Ability, Beliefs, and Calibration. *Management Science*, forthcoming.

8. Carvajal, D., Franco, C., & Isaksson, S. (2024). Will Artificial Intelligence Get in the Way of Achieving Gender Equality?. NHH Dept. of Economics Discussion Paper, (03)

9. Chen, C., & Shu, K. (2023). Can llm-generated misinformation be detected?. arXiv preprint arXiv:2309.13788.

10. Coffman, K. B. (2014.) Evidence on Self-Stereotyping and the Contribution of Ideas. *Quarterly Journal of Economics* 129 (4): 1625–60.

11. Dargnies, M. P., Hakimov, R., & Kübler, D. (2024). Aversion to hiring algorithms: Transparency, gender profiling, and self-confidence. *Management Science*.

12. De Angelis, L., Baglivo, F., Arzilli, G., Privitera, G. P., Ferragina, P., Tozzi, A. E., & Rizzo, C. (2023). ChatGPT and the rise of large language models: the new AI-driven infodemic threat in public health. *Frontiers in public health*, 11, 1166120.

13. Del Rio-Chanona, M., Laurentsyeva, N., & Wachs, J. (2023). Are large language models a threat to digital public goods? evidence from activity on stack overflow. arXiv preprint arXiv:2307.07367.

14. Doshi, A. R., & Hauser, O. P. (2024). Generative AI enhances individual creativity but reduces the collective diversity of novel content. *Science Advances*, 10(28), eadn5290.

15. Farquhar, S., Kossen, J., Kuhn, L., & Gal, Y. (2024). Detecting hallucinations in large language models using semantic entropy. *Nature*, 630(8017), 625-630.

16. Feuerriegel, S., DiResta, R., Goldstein, J. A., Kumar, S., Lorenz-Spreen, P., Tomz, M., & Pröllochs, N. (2023). Research can help to tackle AI-generated disinformation. *Nature Human Behaviour*, 7(11), 1818-1821.

17. Gibney, E. (2024). What the EU's tough AI law means for research and ChatGPT. *Nature* 626, 938-939.

18. Goldstein, J. A., Chao, J., Grossman, S., Stamos, A., & Tomz, M. (2024). How persuasive is AI-generated propaganda? *PNAS nexus*, 3(2), pgae034.

19. Gorodnichenko, Y., & Roland, G. (2011). Which dimensions of culture matter for long-run growth? *American Economic Review*, 101(3), 492-498.

20. Greevink, I., Offerman, T., & Romagnoli, G. (2023). AI-Powered Promises: The Influence of ChatGPT on Trust and Trustworthiness. Technical report.

21. Guiso, L., Sapienza, P., & Zingales, L. (2004). The role of social capital in financial development. *American Economic Review*, 94(3), 526-556.

22. Horowitz, M. C., & Kahn, L. (2024). Bending the Automation Bias Curve: A Study of Human and AI-Based Decision Making in National Security Contexts. *International Studies Quarterly*, 68(2), sqae020.

23. Inglehart, R., C. Haerpfer, A. Moreno, C. Welzel, K. Kizilova, J. Diez-Medrano, M. Lagos, P. Norris, E. Ponarin & B. Puranen (eds.). 2022. World Values Survey: All Rounds - Country-Pooled Datafile. Madrid, Spain & Vienna, Austria: JD Systems Institute & WVSA Secretariat. Dataset Version 3.0.0.

24. Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y., Chen, D., Dai, W., Madotto, A., & Fung, P. (2023). Survey of Hallucination in Natural Language Generation. ACM Computing Surveys, 55, 1 - 38.

25. Ji, Z., Chen, D., Ishii, E., Cahyawijaya, S., Bang, Y., Wilie, B., & Fung, P. (2024). LLM Internal States Reveal Hallucination Risk Faced With a Query. ArXiv, abs/2407.03282.

26. Jones, C. I. (2023). The ai dilemma: Growth versus existential risk (No. w31837). National Bureau of Economic Research.

27. Karlan, D., Mobius, M., Rosenblat, T., & Szeidl, A. (2009). Trust and social collateral. *The Quarterly Journal of Economics*, 124(3), 1307-1361.

28. Knack, S., & Keefer, P. (1997). Does social capital have an economic payoff? A cross-country investigation. *The Quarterly Journal of Economics*, 112(4), 1251-1288.

29. Kreps, S., McCain, R. M., & Brundage, M. (2022). All the news that's fit to fabricate: AI-generated text as a tool of media misinformation. *Journal of experimental political science*, 9(1), 104-117.

30. Lane, T., Nosenzo, D., & Sonderegger, S. (2023). Law and norms: Empirical evidence. *American Economic Review*, 113(5), 1255-1293.

31. Lee, M. D., & Lee, M. N. (2017). The Relationship between Crowd Majority and Accuracy for Binary Decisions. *Judgment and Decision Making* 12 (4): 328–43.

32. Lin, S.C., Hilton, J., & Evans, O. (2022). Teaching Models to Express Their Uncertainty in Words. Trans. Mach. Learn. Res., 2022.

33. Meyrowitsch, D. W., Jensen, A. K., Sørensen, J. B., & Varga, T. V. (2023). AI chatbots and (mis) information in public health: impact on vulnerable communities. Frontiers in Public Health, 11, 1226776.

34. Moore, D. A., & Healy, P.J., (2008). The Trouble with Overconfidence. *Psychological Review* 115 (2): 502–17.

35. Nan, X., Wang, Y., & Thier, K. (2022). Why do people believe health misinformation and who is at risk? A systematic review of individual differences in susceptibility to health misinformation. Social Science & Medicine, 314, 115398.

36. Narayanan, A., & Kapoor, S. (2024). AI snake oil: What artificial intelligence can do, what it can't, and how to tell the difference. *Princeton University Press*.

37. Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, 381(6654), 187-192.

38. Peng, S., Kalliamvakou, E., Cihon, P., & Demirer, M. (2023). The impact of ai on developer productivity: Evidence from github copilot. arXiv preprint arXiv:2302.06590.

39. Reddit (2024a). "Anyone else basically done with Google search in favor of ChatGPT?". https://www.reddit.com/r/ChatGPT/comments/13ik8wh/ anyone_else_basically_done _with_google_search_in/, consulted October 18, 2024.

40. Reddit (2024b). 'When you need a question answered or problems solved, which do you use first: an LLM (ChatGPT or not) or a search engine?" consulted October 18, 2024.

41. Rowlands, C. (2025). "Goodbye Google? People are increasingly switching to the likes of ChatGPT, according to major survey – here's why." Techradar, https://www.techradar.com/tech/people-are-increasingly-swapping-google-for-the-likes-of-chatgpt-according-to-a-major-survey-heres-why.

42. Schneid, R. (2024, September 18). How to Protect Yourself from AI Election Misinformation. Times. https://time.com/7022120/ai-election-misinformatioen-2024/

43. Serra-Garcia, M. & Gneezy, U. (2021). Mistakes, overconfidence, and the effect of sharing on detecting lies. *American Economic Review,* 111(10), 3160-3183.

44. Serra-Garcia, M. & Gneezy, U. (2025). Improving Human Deception Detection with Algorithmic Feedback. *Management Science*, forthcoming.

45. Similarweb (2024). Chat.openai.com Website traffic checker. https://www.similarweb.com/website/chat.openai.com/#demographics, retrieved October 16, 2024.

46. Spitale, G., Biller-Andorno, N., & Germani, F. (2023). AI model GPT-3 (dis) informs us better than humans. *Science Advances*, 9(26), eadh1850.

47. Sutter, M., & Kocher, M. G. (2007). Trust and trustworthiness across different age groups. *Games and Economic Behavior*, 59(2), 364-382.

48. Surowiecki, J. (2005). *The Wisdom of the Crowds*. New York: Anchor Books.

49. Tabellini, G. (2010). Culture and institutions: economic development in the regions of Europe. *Journal of the European Economic Association*, 8(4), 677-716.

50. Tsfati, Y., & Ariely, G. (2014). Individual and contextual correlates of trust in media across 44 countries. Communication research, 41(6), 760-782.

51. Wang, W., Gao, G., & Agarwal, R. (2024). Friend or foe? Teaming between artificial intelligence and workers with variation in experience. *Management Science*, 70(9), 5753-5775.

52. Warren, Z. (2024). Legalweek 2024: Current US AI regulation means adopting a strategic — and communicative — approach. Thomson Reuters, https://www.thomsonreuters.com/en-us/posts/legal/legalweek-2024-ai-regulation.

53. Xu, Y., Dai, H., & Yan, W. (2024). Identity disclosure and anthropomorphism in voice chatbot design: a field experiment. *Management Science*.

Online Appendix

Understanding Trust in AI as an Information Source: Evidence from 11 Countries

by Sanchaita Hazra (University of Utah) and
Marta Serra-Garcia (UC San Diego)

Table of Contents

A LLM Accuracy: Additional Results

Table A.1 shows the fraction of questions that were asked with and without source, the fraction that concerned the US, relative to other geographical locations, and the distribution of topics across the questions. By design, half of the questions contained a source, and half of the questions concerned the US. Across questions the topics varied, with Economics (40%) and Politics (36%) being the most frequent.

Table A.1: LLM Questions Asked: Descriptive Statistics

	(1) Fraction of Questions
Question included source	0.50
Question concerns US	0.50
Topic=Climate	0.05
Topic=Crime	0.03
Topic=Economics	0.40
Topic=Education	0.03
Topic=Health	0.13
Topic=Politics	0.36
Observations	200

Table A.2 examines the determinants of accuracy for the LLM. None of the covariates significantly affect the accuracy of the LLM. The LLM was provided the source to consult (78% accuracy) and when it was not provided a specific source (70% accuracy, t-test, p-value= 0.202). The geographical region to which the question refers, or the topic of the question do not correlate with accuracy of the LLM either.

Exploring the LLM's responses, we find that in most cases (78%) the LLM is rated the same, as correct or incorrect, irrespective of the presence of a source (78 cases out of 100, of which 63 cases are correct and 15 cases incorrect). In the remaining 22 cases, there are 15 cases in which LLM with source is correct, while the LLM without source is incorrect. There are only 7 cases in which the LLM without source is correct, but the LLM with source is incorrect.

The LLM listed the source or sources it used to generate the response (under searched sites). When given a source, it used the source given in 84% of the cases, and did not mention an explicit source in the remaining cases. When not given a source, the LLM provided on average 3.41 sources and provided at least 1 source in 99% of the cases. When not given a source link, in 38% of the cases, one of the sources it provided by itself coincided with the source given to the LLM (in the source cases).

Several mistakes from the LLM occurred when the answer to the question was provided in a figure, instead of text. In 25% of the cases, the factual response was presented in a figure in the source link (or in at least one of the sources if two or more sources were used by the LLM to find the answer). The average accuracy when the factual answer is in a figure is 62%, compared to 78% when the answer is presented in the text or a table.

Table A.2: Determinants of LLM Accuracy

	(1)	(2)
	LLM provides accurate response= 1	
Question included source	0.080	0.080
	(0.062)	(0.062)
Question concerns US		0.005
		(0.066)
Crime		-0.033
		(0.228)
Economics		0.059
		(0.153)
Education		-0.202
		(0.229)
Health		-0.049
		(0.168)
Politics		0.075
		(0.153)
Constant	0.700***	0.660***
	(0.044)	(0.143)
Observations	200	200

Notes: This table displays the estimated coefficients from linear regressions on likelihood that the LLM answers the question it was asked correctly, according to three independent raters. The variable Question included source is an indicator for whether the LLM was given a link to the source to consult for the question asked, and Question concerns US is an indicator for whether the question asked concerned the US. Crime, Economics, Education, Health, and Politics are indicator variables for topics of the questions asked, where Climate is the omitted variable. Robust standard errors, clustered at the participant level, in parentheses. * p<.10; ** p<.05; *** p<.01

B Experimental Data: Additional Results

B.1 Demographics

Table B.1 shows the average values of each socio-demographic characteristic across countries.

Table B.1: Demographics by Country

	Australia	Canada	Chile	Kenya	India	Mexico
Female	0.49	0.49	0.34	0.30	0.39	0.49
Age	34.98	35.05	29.05	32.15	27.73	29.59
High Education	0.67	0.65	0.65	0.85	0.67	0.74
Political leaning (towards right)	4.18	4.45	4.20	6.05	6.26	4.33
Observations	305	307	200	127	175	301

	NewZealand	Spain	SouthAfrica	UK	US	Total
Female	0.56	0.71	0.37	0.62	0.56	0.50
Age	35.92	29.47	32.53	40.35	39.02	33.67
High Education	0.60	0.59	0.66	0.59	0.56	0.65
Political leaning (towards right)	4.31	5.97	4.15	4.39	5.00	4.75
Observations	284	324	301	300	304	2928

B.2 Human Accuracy

Throughout the analyses in the paper, to measure accuracy we focus on how often participants correct evaluate an LLM answer, conditional on the answer of the LLM being correct or incorrect. These measures are closely related to Type I and II errors. Although we expected to use them initially, the fact that LLM accuracy is significantly better than 50% (74%), means that the frequency of Type I and II errors is highly dependent on the LLM accuracy rate. In the extreme case that the LLM were always correct, the frequency of Type I errors, participants believing it is correct when it is not, would always be 0%. Hence, we measure accuracy conditional on whether the LLM is correct or incorrect.

Table B.2: Human Accuracy

	(1)	(2)	(3)
		Human Accuracy	
	Human Accurate if LLM is correct = 1	Human Accurate if LLM is incorrect = 1	Human Accurate All
Canada	0.015	0.011	0.015
	(0.014)	(0.019)	(0.010)
Chile	0.007	-0.010	0.000
	(0.016)	(0.020)	(0.011)
India	0.000	-0.001	-0.001
	(0.020)	(0.025)	(0.013)
Kenya	0.064***	-0.055**	0.031***
	(0.017)	(0.023)	(0.012)
Mexico	0.036***	-0.015	0.018*
	(0.014)	(0.020)	(0.010)
NewZealand	-0.013	0.030	-0.003
	(0.015)	(0.020)	(0.010)
SouthAfrica	0.008	-0.003	0.005
	(0.014)	(0.019)	(0.010)
Spain	0.003	-0.009	-0.001
	(0.014)	(0.019)	(0.009)
UK	-0.008	-0.009	-0.008
	(0.014)	(0.019)	(0.010)
US	-0.035**	0.035*	-0.019*
	(0.014)	(0.020)	(0.010)
Female	-0.015**	0.022**	-0.005
	(0.007)	(0.009)	(0.005)
Age	0.001*	-0.001***	0.000
	(0.000)	(0.000)	(0.000)
High Education	0.012*	-0.024***	0.003
	(0.007)	(0.009)	(0.005)
Politically leaning right (standardized)	-0.001	-0.002	-0.002
	(0.003)	(0.004)	(0.002)
Constant	0.689***	0.334***	0.688***
	(0.024)	(0.028)	(0.021)
Observations	43,504	15,056	58,560
Clusters	2,928	2,924	2,928

Notes: This table displays the estimated coefficients from linear regressions on likelihood that the human correctly rates the LLM if the LLM answered correctly (column (1)), the human correctly rates the LLM if the LLM answered incorrectly (column(2)) and the human correctly rates the LLM in all cases (column (3)). The regressions include fixed effects for question (100 questions in total) and presence of a source in it. Each country name represents an indicator variable for that country, where the omitted country is Australia. Robust standard errors, clustered at the participant level, in parentheses. * p<.10; ** p<.05; *** p<.01

Table B.3: Human Accuracy: Country Level Averages

	(1)	(2)	(3)
	Human Accuracy		
	Human Accurate if LLM is correct = 1	Human Accurate if LLM is incorrect = 1	Human Accurate All
Australia	0.746	0.267	0.624
	(0.010)	(0.014)	(0.007)
Canada	0.761	0.278	0.639
	(0.010)	(0.013)	(0.007)
Chile	0.753	0.257	0.624
	(0.013)	(0.015)	(0.009)
India	0.747	0.266	0.623
	(0.017)	(0.022)	(0.011)
Kenya	0.810	0.212	0.655
	(0.014)	(0.018)	(0.009)
Mexico	0.783	0.252	0.642
	(0.010)	(0.014)	(0.006)
NewZealand	0.734	0.297	0.621
	(0.011)	(0.014)	(0.007)
SouthAfrica	0.754	0.264	0.630
	(0.010)	(0.013)	(0.007)
Spain	0.749	0.258	0.623
	(0.010)	(0.013)	(0.006)
UK	0.739	0.258	0.616
	(0.010)	(0.013)	(0.007)
US	0.712	0.302	0.606
	(0.010)	(0.014)	(0.007)
Observations	43504	15056	58560

Notes: This table displays regression-adjusted averages of human accuracy if the LLM is correct (column (1)), if the LLM is incorrect (column (2)), and for all answers (column(3)), for each country, based on the estimation in Table B.2. Robust standard errors, using the Delta-method, in parentheses.

B.3 Confidence

Table B.4: Human Confidence

	(1)	(2)	(3)
		Human Accuracy and Confidence	
	Human Accurate	Human Absolute Confidence (% out of 20)	Human Relative Confidence (1-4 Quartile)
Canada	0.015	0.008	-0.106*
	(0.010)	(0.016)	(0.058)
Chile	-0.005	0.039**	-0.222***
	(0.012)	(0.018)	(0.065)
India	-0.003	0.059***	-0.444***
	(0.013)	(0.020)	(0.078)
Kenya	0.030**	0.089***	-0.483***
	(0.012)	(0.018)	(0.062)
Mexico	0.016	0.040**	-0.235***
	(0.010)	(0.017)	(0.057)
NewZealand	-0.003	-0.007	-0.028
	(0.010)	(0.017)	(0.062)
SouthAfrica	0.008	-0.003	-0.202***
	(0.010)	(0.017)	(0.060)
Spain	-0.007	0.031*	-0.100*
	(0.010)	(0.016)	(0.059)
UK	-0.005	-0.039**	0.151**
	(0.010)	(0.017)	(0.060)
US	-0.019*	-0.067***	0.139**
	(0.010)	(0.017)	(0.060)
Female	-0.004	-0.053***	0.203***
	(0.005)	(0.008)	(0.027)
Age	0.000	0.001***	-0.003**
	(0.000)	(0.000)	(0.001)
High Education	0.001	0.039***	-0.138***
	(0.005)	(0.008)	(0.028)
Politically leaning right (standardized)	-0.002	0.011***	-0.049***
	(0.002)	(0.004)	(0.013)
Constant	0.625***	0.651***	2.092***
	(0.011)	(0.018)	(0.063)
Observations	58,560	2,928	2,928
Clusters	2,928		

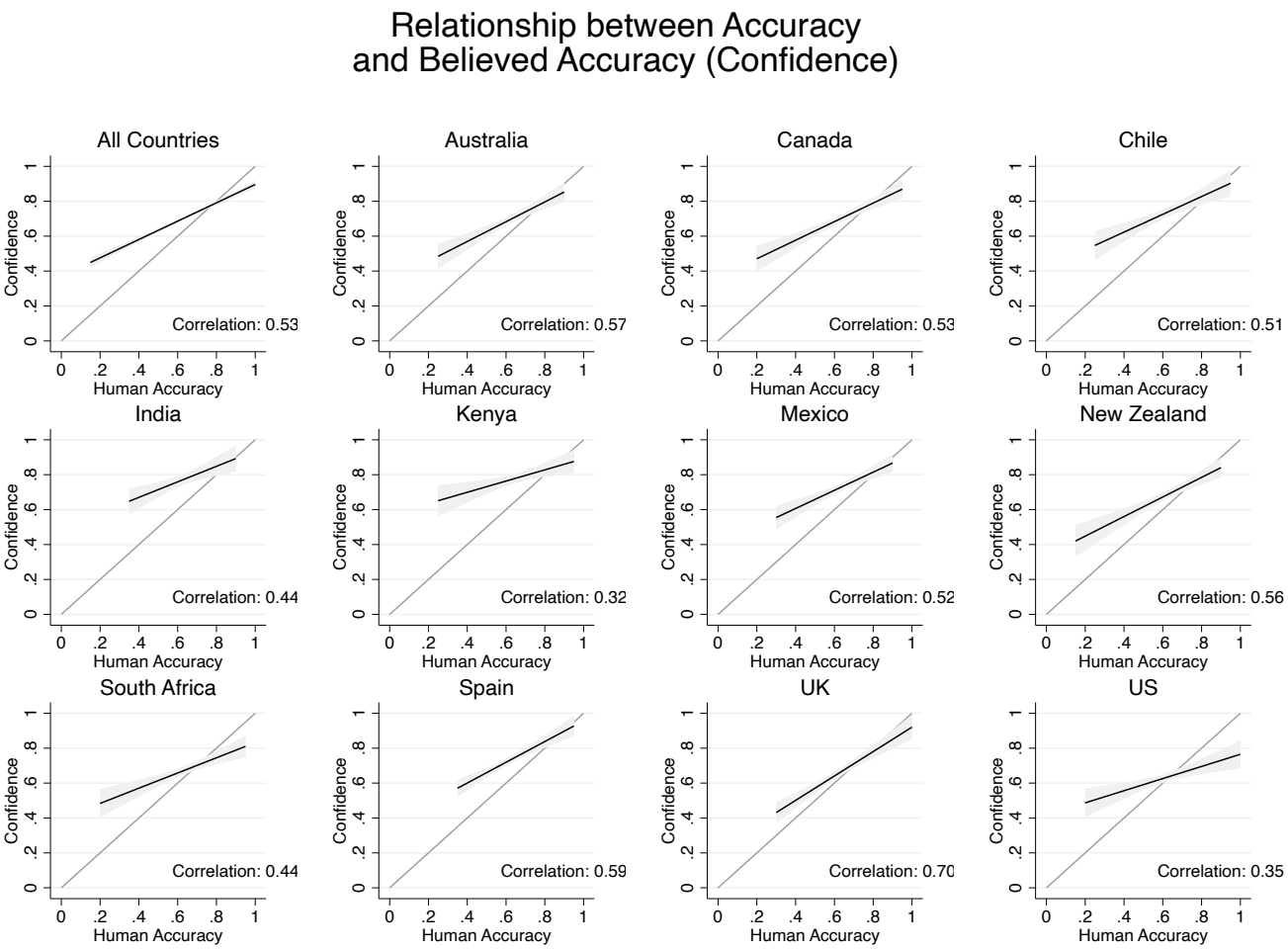
Notes: This table displays the estimated coefficients from linear regressions on likelihood that the human correctly rates the LLM (column (1)), their belief in their absolute ability to correctly rate the LLM, measured as the fraction of answers they believe to have correctly rated out of 20 (column(2)) and their belief in their relative ability to rate the LLM, measured as their believed quartile of ability relative to others (column (3)). Each country name represents an indicator variable for that country, where the omitted country is Australia. Robust standard errors, clustered at the participant level, in parentheses. * p<.10; ** p<.05; *** p<.01

Table B.5: Human Confidence: Country Level Averages

	(1)	(2)	(3)
	Human Accuracy and Confidence		
	Human Accurate	Human Absolute Confidence (% out of 20)	Human Relative Confidence (1-4 Quartile)
Australia	0.625	0.693	2.013
	(0.007)	(0.012)	(0.042)
Canada	0.640	0.701	1.908
	(0.007)	(0.011)	(0.040)
Chile	0.620	0.732	1.791
	(0.009)	(0.014)	(0.050)
India	0.622	0.752	1.570
	(0.011)	(0.016)	(0.066)
Kenya	0.654	0.782	1.530
	(0.010)	(0.014)	(0.045)
Mexico	0.641	0.733	1.779
	(0.007)	(0.012)	(0.038)
NewZealand	0.622	0.686	1.985
	(0.007)	(0.012)	(0.045)
SouthAfrica	0.633	0.691	1.812
	(0.007)	(0.012)	(0.042)
Spain	0.618	0.724	1.914
	(0.007)	(0.011)	(0.041)
UK	0.620	0.654	2.164
	(0.007)	(0.012)	(0.043)
US	0.605	0.627	2.153
	(0.007)	(0.012)	(0.043)
Observations	58560	2928	2928

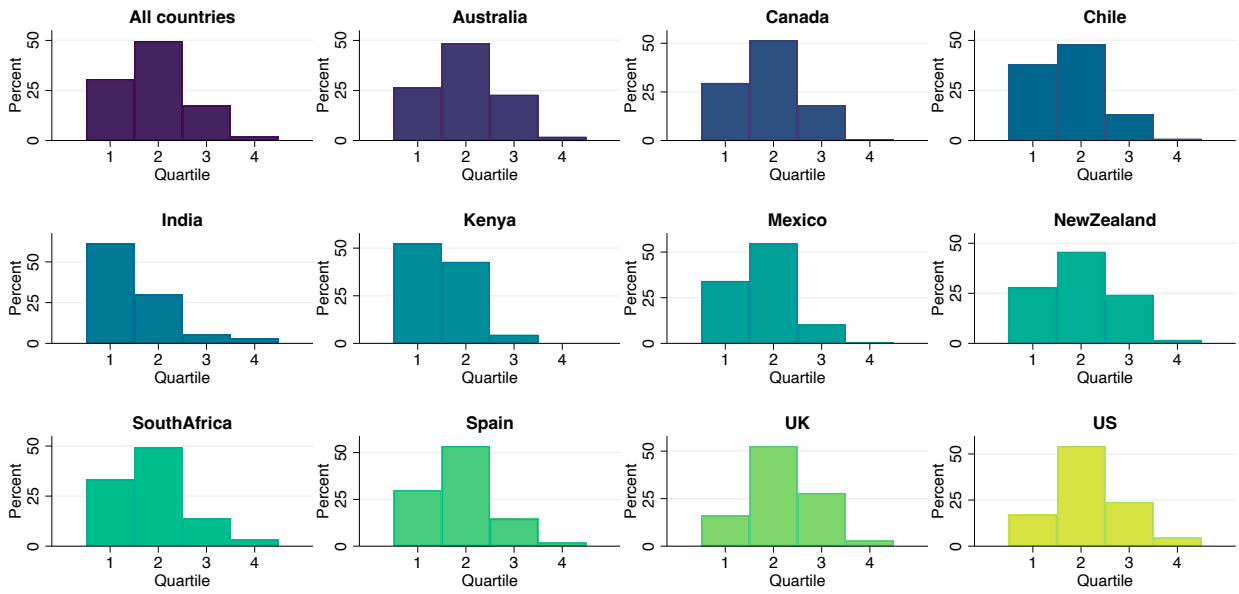
Notes: This table displays regression-adjusted averages of human accuracy (column (1)), absolute confidence (column (2)), and relative confidence (column(3)), for each country, based on the estimation in Table B.4. Robust standard errors, estimated using the Delta-method, in parentheses.

Figure B.1: Correlation between Accuracy and Believed Accuracy (Absolute Confidence)



Notes: Linear relationship between human accuracy and their believed accuracy in evaluating LLM output. Confidence intervals are shown as gray areas underlying each black line. The 45-degree line is shown in light gray in each plot. Each plot shows the linear correlation coefficient, corresponding to the slope shown in the figure.

Figure B.2: Relative Confidence by Country



*Notes:*This figure shows the distribution of relative confidence, plotting how frequently participants in each country reported to believe their quartile of ability was the 1st, 2nd, 3rd or 4th.

B.4 Determinants of Trust in the LLM: Demographic Characteristics

Table B.6: Heterogeneity in Trust by Demographics, Question-Level

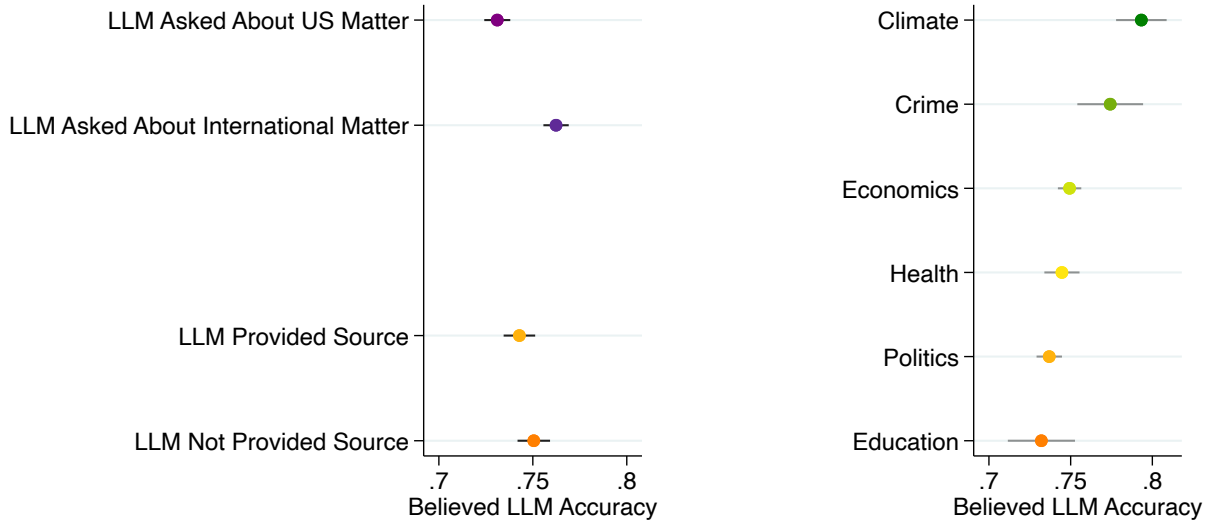
	(1)		(2)	
	Trust the LLM			
Countries				
Canada	0.01	(0.013)	0.07*	(0.035)
Chile	0.00	(0.015)	0.07**	(0.033)
India	-0.00	(0.019)	0.01	(0.051)
Kenya	0.06***	(0.016)	0.10***	(0.037)
Mexico	0.03**	(0.013)	0.05	(0.033)
NewZealand	-0.02	(0.014)	0.01	(0.035)
SouthAfrica	0.00	(0.013)	0.03	(0.031)
Spain	0.00	(0.014)	0.04	(0.032)
UK	-0.00	(0.013)	0.04	(0.032)
US	-0.03**	(0.013)	-0.01	(0.032)
Demographic characteristics				
Female=1	-0.02***	(0.006)	-0.01	(0.019)
High Education=1	0.02**	(0.007)	0.03	(0.022)
Older age (above median)=1	0.01*	(0.006)	0.04**	(0.019)
Right-leaning politics (above median)=1	-0.00	(0.006)	0.00	(0.019)
Interaction Terms				
Female=1 × Canada			-0.02	(0.027)
Female=1 × Chile			-0.04	(0.031)
Female=1 × India			0.05	(0.039)
Female=1 × Kenya			0.02	(0.032)
Female=1 × Mexico			0.00	(0.028)
Female=1 × NewZealand			-0.01	(0.029)
Female=1 × SouthAfrica			0.00	(0.028)
Female=1 × Spain			-0.02	(0.028)
Female=1 × UK			-0.01	(0.027)
Female=1 × US			-0.02	(0.027)
High Education=1 × Canada			-0.03	(0.029)
High Education=1 × Chile			-0.10***	(0.036)
High Education=1 × India			-0.09*	(0.051)
High Education=1 × Kenya			-0.05	(0.034)
High Education=1 × Mexico			-0.02	(0.032)
High Education=1 × NewZealand			-0.01	(0.031)
High Education=1 × SouthAfrica			-0.03	(0.029)
High Education=1 × Spain			0.01	(0.031)
High Education=1 × UK			-0.02	(0.029)
High Education=1 × US			0.02	(0.029)
Older age (above median)=1 × Canada			-0.02	(0.027)
Older age (above median)=1 × Chile			0.04	(0.033)
Older age (above median)=1 × India			0.06	(0.038)
Older age (above median)=1 × Kenya			-0.05	(0.032)
Older age (above median)=1 × Mexico			-0.03	(0.028)
Older age (above median)=1 × NewZealand			-0.03	(0.029)
Older age (above median)=1 × SouthAfrica			-0.05*	(0.027)
Older age (above median)=1 × Spain			-0.06**	(0.028)
Older age (above median)=1 × UK			-0.05*	(0.027)
Older age (above median)=1 × US			-0.04	(0.027)
Right-leaning politics (above median)=1 × Canada			-0.04	(0.027)
Right-leaning politics (above median)=1 × Chile			-0.02	(0.030)
Right-leaning politics (above median)=1 × India			0.03	(0.037)
Right-leaning politics (above median)=1 × Kenya			0.00	(0.032)
Right-leaning politics (above median)=1 × Mexico			0.02	(0.027)
Right-leaning politics (above median)=1 × NewZealand			-0.00	(0.028)
Right-leaning politics (above median)=1 × SouthAfrica			0.03	(0.029)
Right-leaning politics (above median)=1 × Spain			-0.01	(0.028)
Right-leaning politics (above median)=1 × UK			-0.03	(0.028)
Right-leaning politics (above median)=1 × US			-0.00	(0.027)
Constant	0.70***	(0.022)	0.68***	(0.030)
Observations	58,560		58,560	
Clusters	2,928		2,928	

Notes: This table displays coefficients from linear regressions on an indicator variable for whether the individual showed trust in the LLM above the median (in the experiment). The omitted country (reference category) is Australia. Robust standard errors in parentheses.

B.5 Determinants of Trust in the LLM: Question-Level Characteristics

We examine whether beliefs about the accuracy of the LLM depend on whether it is asked about a matter concern the US or an international topic, and whether it is given a source to use to answer the question. Although the accuracy of the LLM does not vary significantly, the left panel of Figure B.3 shows that participants believe that the LLM is marginally more accurate when asked about an international matter (76.2%) than a US matter (73.0%, $t-$test,$p-$value$<$ 0.001), but equally likely to be accurate when it is provided a source (74.3%) and when it is not (75.1%, $t-$test,$p-$value$=$ 0.213).[5]

Figure B.3: Trust and Question-Level Characteristics



Notes: Average believed LLM accuracy by whether the topic was related to the US or International, by whether the LLM was provided with a source, and by topic. Average are based on regression-adjusted estimates, including country fixed effects and socio-demographic characteristics (gender, age, education, urban location, social media use, and politics). Confidence intervals are shown as gray horizontal lines, and are calculated using the Delta method, for standard errors clustered at the participant level. See Table B.7 for details.

To explore other determinants of accuracy, we test whether beliefs differ depending on the topic of the question and find significant differences between questions about climate, where believed accuracy is high (79.3%), while for other topics, excluding crime, it is significantly lower. For example, education and politics are topics that individuals believe the LLM provides accurate responses in 73.2% and 73.7% of the cases.

[5]The country of the participant and the country to which the question asked to the LLM relates is the same in 6.9% of the cases. Comparing the accuracy of participants in these cases to the accuracy in other questions we do not find significant differences ($t-$test, $p-$value$=$ 0.842).

Table B.7: Determinants of Trust in the LLM's Answer

	(1)	(2)	(3)	(4)
	Human Believes LLM is Accurate = 1			
LLM provided source	-0.008			-0.008
	(0.006)			(0.006)
LLM asked about US matter		-0.032***		-0.026***
		(0.003)		(0.004)
Crime			-0.019	-0.019
			(0.012)	(0.012)
Economics			-0.059***	-0.043***
			(0.008)	(0.008)
Education			-0.069***	-0.061***
			(0.013)	(0.013)
Health			-0.063***	-0.048***
			(0.009)	(0.009)
Politics			-0.069***	-0.056***
			(0.008)	(0.008)
Female	-0.017***	-0.017***	-0.017***	-0.017***
	(0.006)	(0.006)	(0.006)	(0.006)
Age	0.001**	0.001**	0.001**	0.001**
	(0.000)	(0.000)	(0.000)	(0.000)
High Education	0.015**	0.015**	0.015**	0.015**
	(0.007)	(0.007)	(0.007)	(0.007)
Politically leaning right (standardized)	-0.000	-0.000	-0.000	-0.000
	(0.003)	(0.003)	(0.003)	(0.003)
Constant	0.730***	0.735***	0.778***	0.789***
	(0.017)	(0.015)	(0.016)	(0.019)
Observations	58,560	58,560	58,560	58,560
Clusters	2,928	2,928	2,928	2,928

Notes: This table displays the estimated coefficients from linear regressions on likelihood that the human believes the LLM answered correctly. The regressions include country fixed effects. The variables Crime, Economics, Education, Health and Politics are indicators for the topic of the question asked to the LLM where Climate is the omitted category. The regressions include sociodemographic characteristics as controls. Robust standard errors, clustered at the participant level, in parentheses. * p<.10; ** p<.05; *** p<.01

B.5.1 Checking Outside Sources

Table B.8: Outside Consult: Country Level Averages

	(1)	(2)	(3)
	Individual Behavior & Beliefs		
	Checked Online LLM Output	Human Accuracy if did NOT check	Human Accuracy if checked
Australia	0.291	0.604	0.639
	(0.026)	(0.033)	(0.052)
Canada	0.296	0.687	0.611
	(0.026)	(0.032)	(0.052)
Chile	0.322	0.571	0.579
	(0.034)	(0.043)	(0.059)
India	0.531	0.667	0.631
	(0.044)	(0.062)	(0.058)
Kenya	0.374	0.562	0.645
	(0.037)	(0.049)	(0.059)
Mexico	0.333	0.609	0.668
	(0.028)	(0.035)	(0.046)
NewZealand	0.315	0.578	0.661
	(0.027)	(0.036)	(0.051)
SouthAfrica	0.317	0.653	0.508
	(0.026)	(0.033)	(0.051)
Spain	0.295	0.642	0.563
	(0.026)	(0.033)	(0.053)
UK	0.145	0.615	0.707
	(0.020)	(0.031)	(0.076)
US	0.195	0.578	0.576
	(0.023)	(0.032)	(0.070)
Observations	2928	2067	861

Notes: This table displays regression-adjusted averages of how frequently participants report to have checked at least one answer provided by the LLM online (column (1)), their accuracy if they report not to have checked (column (2)), and their accuracy if the report to have checked (column(3)), for each country. The regressions include individual socio-demographic characteristics. Robust standard errors, estimated using the Delta-method, in parentheses.

B.6 Effect of Source in LLM Answers

Table B.9: Human Beliefs and Accuracy

	(1)	(2)	(3)
	Human Beliefs and Accuracy		
	Belief LLM is correct = 1	If LLM is correct, human is accurate= 1	If LLM is incorrect, human is accurate= 1
Question included source	-0.008	0.010	-0.009
	(0.006)	(0.009)	(0.006)
Constant	0.730***	0.296***	0.742***
	(0.017)	(0.025)	(0.018)
Observations	58,560	15,056	43,504
Individual participants	2,928	2,924	2,928

Notes: This table displays the estimated coefficients from linear regressions on likelihood that the human beliefs the LLM answered correctly (column (1)), on the likelihood that, if the LLM is correct, the human accurately detects it (column(2)), and on the likelihood that, if the LLM is incorrect, the human accurately detects it (column (3)). The regressions include country fixed effects and demographic characteristics. Robust standard errors, clustered at the participant level, in parentheses. * p<.10; ** p<.05; *** p<.01

Table B.10: Experience with and Trust in LLM

	(1)	(2)	(3)	(4)
	Use LLM regularly		**Trust in LLM**	
Female	-0.090***	-0.072***	-0.044**	-0.030
	(0.017)	(0.018)	(0.018)	(0.019)
Age	-0.006***	-0.006***	0.002**	0.002***
	(0.001)	(0.001)	(0.001)	(0.001)
High Education	0.111***	0.103***	0.041**	0.033*
	(0.018)	(0.018)	(0.019)	(0.019)
Politically leaning right (standardized)	0.031***	0.031***	0.001	0.001
	(0.009)	(0.009)	(0.009)	(0.009)
High-Income Country=1	-0.176***		-0.096***	
	(0.019)		(0.020)	
Canada		0.009		0.059
		(0.038)		(0.040)
Chile		0.069		0.065
		(0.045)		(0.046)
India		0.350***		0.073
		(0.047)		(0.053)
Kenya		0.370***		0.252***
		(0.041)		(0.047)
Mexico		0.036		0.138***
		(0.040)		(0.040)
NewZealand		-0.055		0.005
		(0.038)		(0.041)
SouthAfrica		0.169***		0.034
		(0.039)		(0.040)
Spain		0.082**		0.060
		(0.039)		(0.040)
UK		-0.089**		-0.022
		(0.036)		(0.040)
US		-0.038		-0.093**
		(0.038)		(0.040)
Use LLM regularly			-0.006	-0.017
			(0.020)	(0.020)
Constant	0.717***	0.527***	0.466***	0.345***
	(0.030)	(0.041)	(0.035)	(0.044)
Observations	2928	2928	2928	2928

Notes: This table displays coefficients from linear regressions on an indicator variable for whether the individual uses the LLM regularly (columns (1)-(2)) and whether the individual showed trust in the LLM above the median (in the experiment, columns (3)-(4)). Robust standard errors in parentheses.

C World Values Survey Data: Additional Results

C.1 Demographics

Table C.1: Demographics by Country

	Australia	Canada	Chile	Kenya	India	Mexico
Female	0.55	0.52	0.51	0.40	0.48	0.48
Age	49.34	46.87	42.42	38.42	30.72	38.14
High Education	0.45	0.49	0.28	0.29	0.26	0.24
Political leaning (towards right)	5.34	5.34	5.23	5.70	5.50	6.08
Observations	5696	7196	3852	9640	1140	6964

	NewZealand	Spain	SouthAfrica	UK	US	Total
Female	0.52	0.49	0.49	0.54	0.50	0.49
Age	52.69	37.66	45.50	50.55	46.27	43.04
High Education	0.52	0.31	0.21	0.45	0.48	0.36
Political leaning (towards right)	5.66	5.65	4.73	5.05	5.61	5.52
Observations	2686	9825	3792	2975	8308	62074

C.2 Observations

Table C.2: Generalized Trust - Observations in WVS by country and wave

	1989-93	1994-98	1999-2004	2005-09	2010-14	2017-22	Total
Australia	0	1747	0	1304	1013	1632	5696
Canada	0	0	1639	1560	0	3997	7196
Chile	0	792	987	694	689	690	3852
India	1866	1175	1041	940	3428	1190	9640
Kenya	0	0	0	0	0	1140	1140
Mexico	0	1162	1043	1283	1901	1575	6964
NewZealand	0	775	0	593	560	758	2686
SouthAfrica	1834	2318	2607	0	3066	0	9825
Spain	0	829	942	1018	1003	0	3792
UK	0	0	0	856	0	2119	2975
US	0	1321	1119	1201	2152	2515	8308

Table C.3: Trust in the Press - Observations in WVS by country and wave

	1989-93	1994-98	1999-2004	2005-09	2010-14	2017-22	Total
Australia	0	1747	0	1304	1013	1632	5696
Canada	0	0	1639	1560	0	3997	7196
Chile	0	792	987	694	689	690	3852
India	1866	1175	1041	940	3428	1190	9640
Kenya	0	0	0	0	0	1140	1140
Mexico	0	1162	1043	1283	1901	1575	6964
NewZealand	0	775	0	593	560	758	2686
SouthAfrica	1834	2318	2607	0	3066	0	9825
Spain	0	829	942	1018	1003	0	3792
UK	0	0	0	856	0	2119	2975
US	0	1321	1119	1201	2152	2515	8308

D Participant Instructions

In countries where English is the main or one of the main languages (US, UK, Canada, Australia, New Zealand, South Africa, India and Kenya), all the instructions were presented in English. In Spanish-speaking countries (Chile, Mexico, and Spain), all the instructions were presented in Spanish. The questions to the LLM were asked in English and its responses were in English, and they were translated into Spanish and shown in that language. Participants knew about this translation.

Participants had to be located in the respective country, which was checked via their IP geolocation as well as through Prolific. They also had to have reported to Prolific that they hold that country's nationality.

The instructions included two control questions for the participants to answer before starting the experiment, as shown below. As pre-registered, we excluded subjects who failed to answer either one of the control questions correctly.

Welcome

This Study - You will make several decisions in this study.
Because any decision may affect your BONUS payment, please read closely all the information you are shown. Thank you!

Task Instructions

This Study
In this study, we will show you 20 questions we asked to ChatGPT (version 4).
We will also show you the answers that ChatGPT provided. If ChatGPT listed any sources, they will appear in the answer as sources or in brackets.

Your Task
Your task is to evaluate whether the answers provided by ChatGPT are correct. Each answer is either correct or incorrect.

Your Bonus
We will select one question at random.
If your evaluation is correct (your selected option matches whether GPT is correct or not), you will receive a bonus of $2.

Understanding Questions Below are two understanding questions that you are asked to answer before we proceed with the study. You will have two attempts to answer them correctly.
Understanding Question 1. What is your task?

- Answer questions that ChatGPT also answered.
- Evaluate the answers provided by ChatGPT as correct or incorrect.
- Evaluate the answers provided by ChatGPT on a scale between 0 and 10, where 0 is incorrect and 10 is correct.

Understanding Question 2. How is your bonus payment determined?

- I receive $1 no matter what my answers are.

- I receive $2 for completing the study and answering all the questions.
- I receive $2 if in a randomly selected question my evaluation for the GPT answer is correct.

Example from the Task

Question we asked ChatGPT (1 of 20)
Anti-Muslim bias in the United States has been rising, manifesting in increased discrimination, hate crimes, and negative stereotypes against Muslim individuals and communities. What is the difference in the number of Anti-muslim bias complaints received from the start to the end of Donald Trump's presidential rule in the United States 2017-2021? Please cite a source for your response.

Answer it provided:
Between the start and end of Donald Trump's presidency from 2017 to 2021, there was a significant increase in anti-Muslim bias complaints. In 2017, CAIR reported over 2,500 complaints of anti-Muslim bias incidents. By 2021, this number had increased to over 6,000 complaints. This indicates a notable rise in reported incidents of anti-Muslim discrimination and bias during this period (ArabAmericanNews).

What do you think?

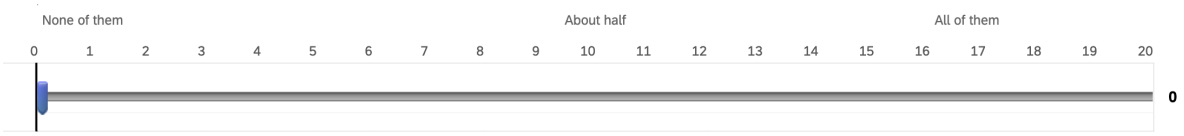
The answer is correct.

The answer is incorrect.

Additional Questions

Thank you for completing the task. In what follows, we will ask you a few more questions and then conclude the study.

How many of the 20 questions and answers by GPT that you evaluated do you think you evaluated correctly? If your answer is correct, you will receive an additional $1 bonus.



Compared with previous participants in this experiment, how well do you think you did? We ask you to choose a quartile. If your answer is correct, you will receive an additional $1 bonus. Relative to the other participants, my number of correct answers is in the following quartile:

- **Quartile 4:** 75th-100th percentile (better than at least 75% of participants).
- **Quartile 3:** 50th-75th percentile
- **Quartile 2:** 25th-50th percentile
- **Quartile 1:** 0th-25th percentile (worse than at least 75% of participants)

What is your experience with ChatGPT?

- Never used it.
- Used it only a few times
- Use it regularly

Do you have a paid subscription to ChatGPT?

- No
- Yes

In what ways, if any, is ChatGPT useful for you? Please explain briefly. [Text Box Entry]

To answer the questions on whether the GPT-provided answer was correct, did you check yourself what the correct answer was elsewhere (on ChatGPT, or online, or else)? Please respond truthfully, you will be paid your bonus regardless of your answer.

- I did not check elsewhere whether the GPT-provided answers were correct.
- I checked elsewhere whether (some of) the GPT-provided answers were correct.

About you

How old are you (in years)? [select number from list below]
What is your gender?

- Male
- Female
- Other [Text Box Entry]

In political matters, people talk of "the left" and "the right." How would you place your views on this scale, considering 1 is left and 10 is right, generally speaking? [scale presented]

Which of the following best describes the place where you currently live?

- A large city
- A suburb near a large city
- A small city or town
- A rural area

Please select the level of education that you have completed.

- Elementary School
- Middle School
- High School or equivalent
- Some college
- College Graduate with Associate's Degree (2 year)
- College Graduate with Bachelor's Degree (4 year)
- Master's Degree (MS)
- Doctoral Degree (PhD)
- Professional Degree (MD, JD, etc.)
- Other [Text Box Entry]

Which social media platforms do you regularly use? [multiple options could be selected]

- Facebook
- Instagram
- Tiktok
- Twitter
- Whatsapp
- Other, please list here [Text Box Entry]
- None, I do not use social media

E Coding Instructions: LLM Accuracy

Below we present the instructions used by coders to evaluate the accuracy of each LLM answer.

Each response given by GPT has to be coded into "Correct" or "Incorrect." Participants in an experiment saw "screenshots" of the question and answer, and provided their own rating (correct or incorrect). We will have 3 coders classify each statement from US and international, and 2 responses by GPT, depending on whether it was given the source or not. We will then meet and reach an agreement if there's a discrepancy between us.

What is CORRECT?

An answer is correct if it coincides with the "correct answer" column in the google sheet. [*Coders were provided with a sheet that contained all statements and correct answers according to the source consulted (and provided as a link). They also had access to the questions, answers and screenshots for the correct answers in separate documents.*]. Then it's from the source used in the source treatment. Or from a different source that republished/requoted the source from the source treatment.

If the answer does not "coincide" with the correct answer in the google sheet:

- It does not give ANY specific answer. This could be when it reports it's increased/decreased, but does not provide a %, although we asked for it.
  - Then, we classify it as CORRECT.
  - Because it's not explicitly incorrect, and from the perspective of the reader, they're not getting the information requested, but they're not getting incorrect information either.
- If it gives a "specific" or "numerical" answer, that differs from the source:
  - First, evaluate the sources (searched sites) for their reliability or the sources that they themselves use.
    - Official government websites
    - UN, WB, and other international organizations
    - Mainstream media
  - Case 1: The numbers reported by GPT in the answer correspond correctly to those listed in that source and the source is reliable, then the answer is CORRECT.
  - Case 2: The numbers reported by GPT in the answer do not correspond to those in the source. They could be on that source/site/page but apply to a different outcome/variable than the one we asked about. Regardless of whether the source is reliable or not, the answer is INCORRECT.
  - Case 3: If the numbers are similar in magnitude to those in the source listed (and ideally somewhat close to the original source), then the answer is CORRECT. Example would be: approximately 40% is the original source, GPT reports with a different source 36%, that's CORRECT.