Fu, Yuhao; Hanaki, Nobuyuki

Working Paper

Do people rely on ChatGPT more than their peers to detect fake news?

ISER Discussion Paper, No. 1233

Provided in Cooperation with:
The Institute of Social and Economic Research (ISER), Osaka University

Suggested Citation: Fu, Yuhao; Hanaki, Nobuyuki (2024) : Do people rely on ChatGPT more than their peers to detect fake news?, ISER Discussion Paper, No. 1233, Osaka University, Institute of Social and Economic Research (ISER), Osaka

This Version is available at:
https://hdl.handle.net/10419/302228

ISSN (Print) 0473-453X
Discussion Paper No. 1233
ISSN (Online) 2435-0982

DO PEOPLE RELY ON ChatGPT MORE THAN THEIR PEERS TO DETECT FAKE NEWS?

Yuhao Fu
Nobuyuki Hanaki

March 2024

The Institute of Social and Economic Research
Osaka University
6-1 Mihogaoka, Ibaraki, Osaka 567-0047, Japan

Do people rely on ChatGPT more than their peers to detect fake news?*

Yuhao Fu† Nobuyuki Hanaki‡

March 5, 2024

Abstract

In the era of rapidly advancing artificial intelligence (AI), understanding to what extent people rely on generative AI products (AI tools), such as ChatGPT, is crucial. This study experimentally investigates whether people rely more on AI tools than their human peers in assessing the authenticity of misinformation. We quantify participants' degree of reliance using the weight of reference (WOR) and decompose it into two stages using the activation-integration model. Our results indicate that participants exhibit a higher reliance on ChatGPT than their peers, influenced significantly by the quality of the reference and their prior beliefs. The proportion of real parts did not impact the WOR. In addition, we found that the reference source affects both the activation and integration stages, but the quality of reference only influences the second stage.

Keywords: GAI, ChatGPT, AI reliance, fake news identification, WOR, activation-integration model, Heckman selection

*This research has benefited from the financial support of (a) the Joint Usage/Research Center, the Institute of Social and Economic Research (ISER), and Osaka University, and (b) Grants-in-aid for Scientific Research Nos. 20H05631 and 23H00055 from the Japan Society for the Promotion of Science. The design of the experiment reported in this paper was approved by the IRB of ISER (#20231001) in October 2023, and the experiment is preregistered at aspredicted.org (#149838). We also want to thank Tiffany Tsz Kwan Tse and Taisuke Imai for their helpful comments and suggestions. Support from Yuta Shimodaira, Zhelin Zhang, and Satsuki Yamada for conducting the experiment is also gratefully acknowledged.

†Graduate School of Economics, Osaka University. E-mail: u889037j@ecs.osaka-u.ac.jp

‡Corresponding author. Institute of Social and Economic Research, Osaka University, and University of Limassol. E-mail: nobuyuki.hanaki@iser.osaka-u.ac.jp

1 Introduction

Generative artificial intelligence (GAI) has emerged as a pivotal advancement in the realm of AI, garnering global interest for its innovative capabilities and broad applicability across various sectors, including finance (Cao, 2020), medicine (Hamet, 2017), and education (Zhai et al., 2021). Unlike traditional AI systems, which are designed to operate within predefined rules and parameters, GAI technologies, exemplified by OpenAI's generative pretrained transformer (GPT) models, have revolutionized the field by enabling the generation of new content, such as images (NovelAI, nijijourney), narratives (NovelAI, Nichesss), and engaging in dialogues (ChatGPT, BingChat), based on user prompts. This shift toward creative and interactive AI applications has not only facilitated their integration into everyday life but also marked a significant leap toward artificial general intelligence (AGI). The appeal of GAI, particularly platforms like ChatGPT, lies in their ability to blend advanced functionalities with user accessibility, fostering reliance on these technologies in diverse fields, such as education (Ju, 2023) and the arts (Victor, 2023). Increasingly, GAI products are being utilized as versatile AI tools by the general populace, signifying a minor yet impactful step in altering societal habits and a monumental stride in the evolution of AI toward more generalized and interactive uses (Zhang et al., 2023).

However, GAI also generates risks (Walkowiak, et al., 2023). As GAI has powerful abilities to analyze, process, and generate text, misinformation, like fake news, will also become much worse (Monteith, et al., 2023). Though using some AI tools to detect fake news may be effective (Patil, et al., 2024), this reliance also raises questions about the trustworthiness of AI in critical applications (Tomitza, 2023). Even prior to the advent of the AI boom, the issue of misinformation was prevalent in the Internet era. Numerous studies indicate that individuals, particularly teenagers, were influenced by their peers, who had motivations to spread fake news (Herrero-Diz et al., 2020). This dynamic underscores how trust or reliance on peers significantly impacts the perceived

credibility of information (Barakat et al. 2021; Haigh et al. 2018). Despite GAI being a source of misinformation, it has also proven to be a valuable tool in its detection (Xu et al., 2023; Patil et al., 2024), offering a dual role in the misinformation landscape. Thus, the advent of GAI intersects with longstanding concerns regarding misinformation, underscoring the imperative to address how individuals discern and rely on various information sources.

With this context in mind, this study seeks to better characterize people's reliance on AI tools in identifying misinformation. To elucidate this further, we also introduced a comparative element – human peers – to assess whether people rely more on AI tools than their human peers in these scenarios, which is also our research question. This inquiry forms the crux of our research question and anticipates challenges that will emerge in the AGI society of the future. Through this comparison, the study aims to shed light on the way individuals appraise and prioritize GAI-generated advice in contrast to human insights. Such an examination is pivotal for unraveling the shifting paradigms of trust and authority in the GAI era. Comprehending these dynamics is essential for steering future interactions between humans and AI tools, with the goal of ensuring that GAI technologies contribute positively to social welfare.

To answer our research question, we developed an experiment where participants were asked to assess the authenticity of misinformation and to update their initial judgment based on references from ChatGPT or human peers. To quantify the reliance, we introduced a main task of assessing the authenticity of news and compared participants' reliance across groups by the "weight of reference" (WOR). This approach allowed us to investigate participants' reliance on external reference between AI tools and human peers for different types of news, where we categorized the news materials based on the proportion of the real part in each piece, classifying them as totally fake, partially fake, or totally real. Furthermore, after the main tasks of the experiment, we conducted a survey to investigate participants' prior beliefs about the reference sources.

In addition, we employed the two-stage model of Vodrahalli et al. (2022) and the Heckman selection approach to decompose reliance into two stages: **activation** and **integration**. In detail, in the first stage (activation), participants determine whether to use the reference and update their initial judgments. Then, in the second stage (integration), if they opt to update, they decide the extent to which they will utilize the reference. As a result, our findings indicate that participants exhibit a significantly higher degree of reliance on AI tools than on their peers when confronted with misinformation, irrespective of the news type. In the analysis of decomposing reliance, we found that this source effect is significant in both stages. Still, participants' assessment of the quality of reference did not affect the activation stage but the integration stage. Meanwhile, we observed that people's prior beliefs significantly affect their reliance.

The remaining part of this paper is organized as follows. Section 2 reviews previous studies on AI reliance by considering the AI systems' genre and analysis approaches. Section 3 presents the hypotheses and experimental design, and Section 4 summarizes the main results. Section 5 presents additional analysis for decomposing the reliance and its results, where we use another method to investigate the reliance. Section 6 presents the conclusion.

2 Related Works

Human trust-related behavior corresponding to nonhuman systems has often been studied in the fields of psychology and economics. Among related experimental studies, researchers often compare how humans utilize advice or reference information either from AI or human experts or peers with a similar task processing: **Participants are asked to finish a prediction task, give their initial response, and subsequently are shown advice (reference) from outside sources, and then they give their second response**. In this way, reliance on reference sources can be measured as to

what extent people update their initial response due to the reference presented to them.

As well as human experts or peers (Madhvan et al. 2007), nonhuman systems like AI often served as an external advice source in those advice-taking studies, where algorithms have been widely used. The following part of this section will briefly review the advice-taking experimental studies by dividing the AI into AI algorithms and AI tools, the latter of which are more accessible for use in daily life.

2.1 Reliance on Algorithm

AI is defined as the simulation of human intelligence processes by machines, especially computer systems. AI technology can be categorized as weak or strong. Weak AI refers to systems designed to perform specific tasks. Conversely, strong AI, or AGI, aims to perform any intellectual task that a human being can perform. Currently, most AI products are considered weak AI due to their narrow application scope, such as stock price analysis or search-and-rescue operations, and remain largely inaccessible to the general public in everyday life. In the current era of significant AI development, many studies have subsequently focused on investigating the extent and nature of people's reliance on AI technologies.

Most studies of reliance on AI use systems embedded with AI algorithms in their design. Regarding medical decision-making (Reverberi et al., 2022; Agaiwal et al., 2023), studies have shown that experts tend to rely less on AI algorithms than nonexperts. Some other studies comparing people's reliance on AI and human experts have found that people rely more on AI algorithms when facing financial decision-making tasks (Tolmeijer et al., 2022; Araujo et al., 2020). Similar phenomena also appear when comparing algorithms with human peers (Gaube et al. 2021; Mesbah et al. 2021).

By contrast, some studies have also observed the "**algorithm aversion**" (Dietvorst et al., 2015, 2019; Yeomans et al., 2019; Jung & Seiter, 2021). A potential reason for the mixed outcomes in studies of AI reliance or aversion is that many AI systems are

either too "weak" or overly specialized for general use, leading to inconsistent findings due to the general public's lack of experience and knowledge of these technologies.

In this study, we used a GAI product, ChatGPT, in our experiment to observe people's reliance on it. We refer to GAI products as AI tools because, nowadays, people have become more familiar with and started to use these GAI products as tools following the AI boom since 2021. As these AI tools become more powerful, user-friendly, and widely applicable, individuals can leverage a broad range of functionalities with minimal knowledge. This ease of use and broad applicability may lead to increased reliance on AI tools in daily life.

2.2 Reliance on GAI

There has been limited research on people's reliance on GAI products despite their widespread publication and use in recent years. As a prominent example of GAI, ChatGPT's capabilities have been explored across various domains. For example, Lopez'Lira and Tang (2023) examined its ability to predict stock market returns using sentiment analysis of news headlines, suggesting that incorporating advanced language models into the investment decision-making process can lead to more accurate predictions. In addition, ChatGPT's strong communication ability has proven powerful in education (Korinek, 2023; Ali et al., 2023), information identification (Yang & Menczer, 2023), emotion analysis (Elyoseph et al., 2023), and so on. Other GAI products, like Midjourney, an imaged-generative AI tool, have been integrated into cosmetic surgery (Lim et al., 2023) and furniture design (Alawadh et al., 2023). Most importantly, some studies also found that under the same task, GAI products generated fewer carbon emissions than humans (Tomlinson et al., 2023; Gaur et al., 2023), which can be effective for combating climate change.

Overall, GAI products offer a wider and more user-friendly application spectrum compared with AI algorithms developed before the advent of ChatGPT, potentially

leading to greater reliance. However, research specifically addressing this reliance on AI tools remains scarce. This study contributes to the field by focusing on people's reliance on GAI products, similar in approach to previous experimental studies on AI algorithms. Uniquely, we introduced ChatGPT as the reference source, a novel application not yet explored in published research.

2.3 Misinformation with GAI

Developed from large language models, GAI possesses robust capabilities for context generation and detection, leading to an inevitable confrontation with the issue of misinformation. Studies have pointed out that news from GAI is believed less (Longoni et al., 2022), and some, like Rosenberg (2023), even warn of its potential threat to macro-media landscapes. Conversely, Simon (2023) suggested that the current fears regarding GAI's impact on misinformation may be exaggerated. Furthermore, researchers like Xu et al. (2023) argued that if GAI can generate misinformation, it should inherently possess the ability to detect it as well. Patil et al. (2024) have devised a novel approach to detecting fake news using another GAI system but not ChatGPT. Caramancion (2023) explored the capabilities of ChatGPT to distinguish misinformation and found that ChatGPT could predict the legitimacy of every item with a solid 100% accuracy. Similarly, other studies, including those by Ahmad et al. (2022) and Saikia et al. (2022), have demonstrated that GAI could pave the way for new mechanisms in fake news detection.

In this experimental study, we introduced a task to assess the authenticity of fake news. While assessing participants' reliance on GAI products' references, we also compared ChatGPT's and humans' abilities to detect fake news by comparing the quality (accuracy) of the references they offered. In this comparison, as we introduced some partially fake news, ChatGPT and humans were tasked with assessing the proportion of real parts in the news. Our results indicate that, given the set of fake news in this

experiment, ChatGPT's proficiency in distinguishing fake news was comparable to that of human participants, with no significant difference. However, it was observed that human participants tended to rely significantly more on references provided by ChatGPT than on those from their human peers. This prompted us to further focus on the source and the content of reference, which extended to an examination of participants' assessments of reference quality. Consequently, we proposed a method for decomposing reliance.

3 Experimental Design

This experimental research addresses the following question:

Do people rely more on AI tools than their peers to detect fake news?

To prevent participants from searching for the news materials used in the experiment on the internet, we conducted this experiment in the laboratory. The experiment was programmed using Otree 5 (Chen et al., 2016), and the overall procedure is shown in Figure 1.

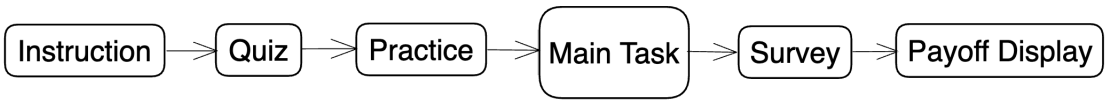


Figure 1: Overall Procedure

In the experiment, after reading through the instructions[1], each participant was asked to take a quiz (see Appendix A) to ensure they understood the rules. Then, they practiced once and entered the main tasks. After finishing all the tasks, they were asked to complete some survey questions, and the final payoff was shown.

[1]An English transation of the instruction is provided in Appendix G.

3.1 Main Task

The main task of the experiment consists of four stages as shown in Figure 2.

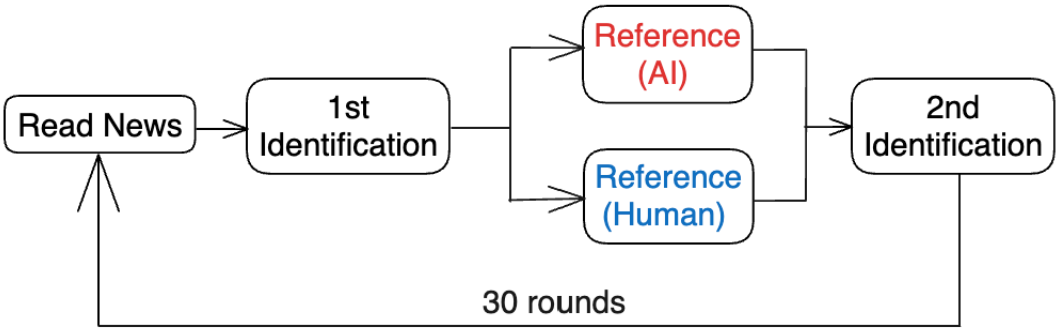


Figure 2: Main Task

There were 30 rounds of assessing the authenticity of news. In each round, participants were first asked to read a piece of news and report their first identification of its authenticity. Then, they were shown the reference and asked to report their second identification. We employed a between-subject design where half of the participants were shown references from human peers, and the references for the remaining half were from ChatGPT.

3.1.1 Read News Stage

In the first stage of the main tasks, each participant read the news (see Figure 9 in Appendix B) without any time constraints. The news materials were Japanese news collected from an open fake news dataset[2]. We randomly selected 30 pieces of news (see Appendix F) that primarily covered topics such as politics, sports, meteorology, and public safety. The news came in three types, as described in Table 1.

The totally real news was written by humans and collected from Japanese wiki news[3], the totally fake news was generated by Google's GPT-2 Japanese model, and the partially fake news was the composition of real and fake, where the first part of

[2]https://github.com/tanreinama/japanese-fakenews-dataset?tab=readme-ov-file
[3]https://ja.wikinews.org/wiki

Table 1: News Materials

Type	Count	Min. Length	Max. Length	$real\_r$
Totally real	10	317	460	100
Totally fake	10	309	462	0
Partially fake	10	323	393	$0 < real\_r < 100$

the article was real news and the second part was fake news. **The proportion of the real part,** $real\_r$, for each piece of news, is defined as

$$real\_r^s = \frac{\text{the length of real part of the News in Round } s}{\text{the length of the News in Round } s} \times 100 \in [0, 100]$$

We considered $real\_r$ as the degree of "**authenticity**" that participants need to "**identify**" in each round of the main task. In the quiz before the main tasks, we informed participants that in this experiment, "**authenticity**" is defined as $real\_r$ (see Q2 in Appendix A). The 30 pieces of news were assigned in a random sequence as shown in Figure 3.

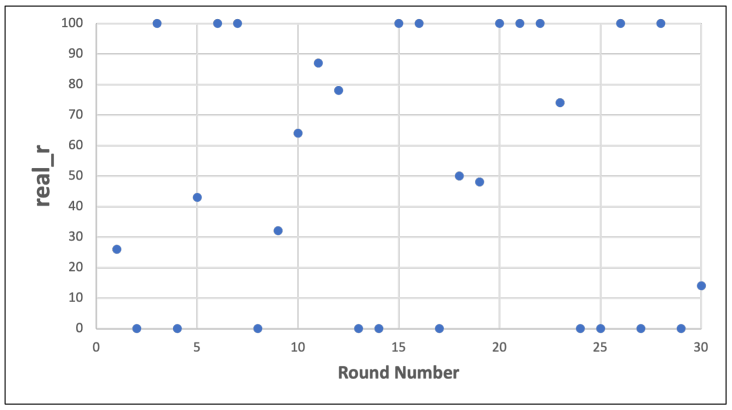


Figure 3: real_r versus Round Number

3.1.2 First Identification Stage

After each participant had read the news, they were asked to provide a number between 0 and 100 to represent their first identification with a slider (see Figure 10 in Appendix B). This stage was shown on a separate screen so that the participants could

not move the cursor of the slider while reading. In this way, we can divide the time they spent reading the news and their first decision-making. Each participant had to submit their first response $response_1$ in this stage; otherwise, they could not enter the next stage.

3.1.3 Reference Stage

The reference stage constitutes the core of the experiment. We divided all participants into two distinct groups: the ***AI*** group and the ***Human*** group. At this stage, participants were presented with a piece of reference information, which varied between the two groups. In the *AI* group, the reference was one response randomly selected from 24 responses generated by ChatGPT. Conversely, in the *Human* group, the reference was one randomly selected initial identification ($response_1$) from another participant within the same group.

The AI's reference set was generated by ChatGPT before the experiment, where we used the GPT-4 model, and the prompt was the same as the description of tasks in the instructions presented to participants.

Prompt :

> *-We will now send you some Japanese news. Please identify how real it is and report your belief in its authenticity as an integer from 0 to 100, with 0 representing totally fake and 100 representing totally real news. -Do not say anything else about the result of your identification.*

We added the last sentence of the prompt to limit ChatGPT's response to a number. For each piece of news, we repeated this process 24 times to generate 24 distinct responses. On the experimental screen (see Figures 11 and 12 in Appendix B), the reference was presented not merely as a number following its source name but as a screenshot of the response. This approach was employed to reinforce participants' belief that the reference was genuinely generated from ChatGPT. In the *AI* group, since

the screenshot of the reference included the news article, we similarly represented news articles in the reference stage for the *Human* group. To prevent participants from spending excessive time rereading the news or lingering at this stage, we introduced a 10-second time constraint. Nevertheless, participants could proceed to the next page earlier by clicking the "next" button on the screen.

Note that at a stage of each round, we randomly selected a reference for each participant. That is, the reference for the same news may differ among participants in each round. We introduced this mechanism to prevent the **spotlight effect** (Gilovich et al., 2000), which refers to the tendency for individuals to overestimate the extent to which their actions or appearance are noticed by others. In the context of our study, by ensuring that each participant encounters a unique set of references, we aimed to mitigate any potential bias or undue influence that might arise if participants believed their responses were more conspicuous or **spotlighted** than they were.

3.1.4 Second Identification Stage

The final stage provided participants with an opportunity to adjust their initial response. At this stage, participants were required to submit their second identification ($response_2$), which was not obligated to align with their first identification ($response_1$). To facilitate this decision-making process and remind participants of their previous response and the reference provided, their $response_1$ and the $reference$ in that round were distinctly marked on a slider with different colors (see Figure 13 in Appendix B).

3.2 Survey Question About Prior Beliefs

As well as demographic information-related questions (see Appendix C), participants' prior beliefs were obtained using the following three questions asked to every participant after they had finished the main tasks.

1. **Have you heard about ChatGPT?**

2. **How many days per week do you use ChatGPT on average?**

3. **In today's experiment, specifically in the task of "assessing News' authenticity," who do you think can provide more accurate responses?**

For the second question, participants were instructed to provide a numerical answer ranging from zero to seven. The third question offered three response options: "Generative AI," "Human," or "Not sure." These questions were designed to gauge participants' familiarity and preconceived notions about ChatGPT and to observe the consistency between their prior beliefs and their experimental choices.

3.3 Final Payoff

The participants' final payoff was composed of a fixed participant fee and an additional payoff based on performance. The total amount of the final payoff was displayed on the screen at the end of the experiment. To effectively incentivize participants, we employed a random lottery strategy to determine the final additional payoff. Specifically, each participant received a participation fee of 500 JPY. The additional payoff was calculated based on the accuracy of one randomly selected response from all their responses throughout the experiment (a total of 30 rounds $\times$ 2 responses = 60 responses). The additional payoff was determined using the following quadratic equation:

$$\pi = \max\{0, 2300 - 0.3 \times (R - \text{real\_r})^2\} \text{ JPY},$$

where π represents the additional payoff, R is the randomly selected response, and $real\_r$ denotes the proportion of the real part of the news in the selected round, after rounding.

Upon being presented with the final payoff, each participant was also shown all their responses alongside the corresponding $real\_r$ values for each news item on the screen. If a participant's randomly selected response precisely matched the actual proportion of

real news ($real\_r$), they would receive an additional reward of 2300 JPY, resulting in a total final payoff of 2800 JPY. According to the payoff calculation formula, a participant would secure a positive additional payoff as long as the difference between the randomly selected response R and the actual proportion $real\_r$ is less than 88.

3.4 Hypotheses

In our study, we aim to compare how much people trust ChatGPT versus other people. Logg et al. (2019) did experiments to see if people trust algorithms or humans more, and they found that people tend to trust algorithms more than other humans. Upon introducing the task of assessing the authenticity of fake news in the experiment, we consequently formulated our main hypothesis about the treatment effect:

H1: Compared with peers, people tend to rely more on AI tools when facing tasks for assessing misinformation authenticity.

As noted, the news materials used in the tasks consisted of three types of news: totally real, partially fake, and totally fake, with misinformation content ranging from 0% to 100%. Analytical and intuitive cognitive processes play a pivotal role in shaping individuals' acceptance or skepticism toward misinformation (Bigey et al., 2021). The ability to discern factual information from falsehoods is fundamentally rooted in the domain of cognitive reasoning. Research in this domain predominantly concentrates on dual-process theories, which posit that analytical reasoning (System 2) has the potential to supersede automatic, intuitive reactions (System 1), as explicated by Pennycook et al. (2021). Upon encountering misinformation, individuals' initial responses are often guided by System 1, which evaluates the veracity of information based on its superficial features or its alignment with preexisting beliefs. Conversely, the activation of System 2 facilitates a more thorough analysis and reflective thought process, enabling the identification of discrepancies or inaccuracies within the information, thereby aiding in the

differentiation of misinformation from factual content. Consequently, we hypothesize that participants exhibit a certain degree of sensitivity to misinformation, contingent upon their inherent logical analysis capabilities. Moreover, the prevalence of logical inconsistencies is typically higher in entirely fabricated information, thereby enhancing individuals' ability to detect such inconsistencies through the employment of System 2. Thus, in comparison to partially fabricated news, completely fabricated news is more easily identifiable by participants, which in turn diminishes their dependence on external sources of verification, particularly those derived from AI tools. Therefore, we propose the second hypothesis:

H2: Reliance on AI tools becomes greater in more challenging tasks, such as assessing the authenticity of partially fake news compared with totally fake news.

4 Main Results

4.1 Materials

The experiment was conducted on November 7th and November 9th, 2023, in the laboratory at the Institute of Social and Economic Research ($ISER$) at Osaka University. We recruited 37 participants who were students at Osaka University registered in the ORSEE (Greiner, 2015) database of $ISER$. All participants were native Japanese speakers, 17 out of whom were assigned to the $Human$ group and 20 were assigned to the AI group. As a result, in the Human group, the duration of the experiment was about 80 minutes, and the average payoff was 2220 yen. Meanwhile, in the AI group, the duration was reduced to 60 minutes, and participants earned an average of 2490 yen. In the final sample, 9 participants were female, 11 were undergraduate students, and 27 were majoring in Natural Science and Engineering (14 engineering, 8 medicine, 4 pharmacy, and 1 science). As there were 30 round tasks for each participant, the total

sample size was 600 in the *AI* group and 510 in the *Human* group. In our main analyses, we will correct the standard error to account for multiple observations collected from the same participant.

In the survey, we collected participants' demographic information and, crucially, their prior beliefs regarding the reference source. Notably, all participants answered that they had heard about ChatGPT before the experiment. Comparisons of demographic data at the 95% CI level, illustrated in Figure 4 and variable definitions presented in Table 2, revealed no significant differences between the two groups. Building on this foundation, Section 4.4 delves further into the relationship between these survey-related elements, including both demographic information and prior beliefs, and how they relate to participants' reliance on the reference sources.

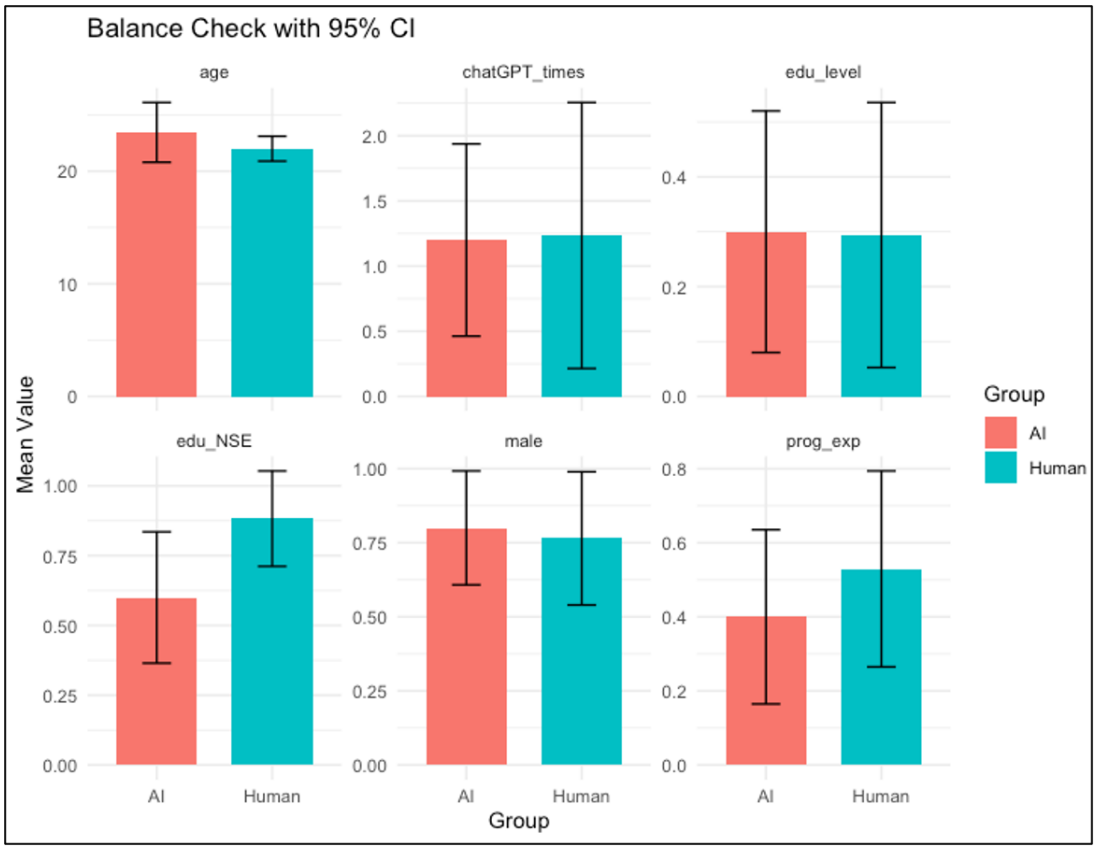


Figure 4: Survey Results

Table 2: Survey Variables

Survey Var.	Definition	Min.	Max.	S.D.	mean (AI)	mean (Human)
age	Participants' age number.	18	44	4.429	23.45	22
chatGPT_times	Average days per week using ChatGPT.	0	6	1.750	1.20	1.24
edu_level	Participants' education level; =1 if graduate; =0 if undergraduate.	0	1	0.463	0.30	0.29
edu_NSE	Participants' major; =1 if majoring in natural science and engineering.	0	1	0.450	0.60	0.88
male	Gender; = 1 if the participant is male.	0	1	0.417	0.80	0.76
prog_exp	Programming experience; =1 if the participant has programming experience.	0	1	0.505	0.40	0.53

4.2 Weight of Reference

Studies on advice-taking used the weight of advice (WOA), a common metric in the psychology of advice utilization, to measure the degree to which people take advice (Harvey & Fischer, 1997; Yaniv et al., 1997) in the form of a numerical estimate. Some other studies directly quantify people's reliance on advice sources using WOA (Önkal et al., 2009; Castelo et al. 2019; Schemmer et al., 2022). In this study, we also used this general method of quantification, renaming it as **WOR**, which is calculated by

$$WOR = \frac{response_2 - response_1}{Ref - response_1},$$

where $response_1$ denotes participants' first identification, $response_2$ denotes participants' second identification, and Ref is the references.

This is our first method of quantifying the **reliance** on the reference source (ChatGPT or human peers). This method provides a continuous outcome on a scale from 0 (completely ignoring the reference) to 1 (completely relying on the reference) and has been used in many analyses of advice utilization (Bailey et al., 2023). However, when the first identification happens to be close or equal to the reference given, the outcome becomes larger than one or even infinity, which makes it hard to reflect the actual process of belief updates in advice-taking in diverse real-world contexts. In the literature, most studies typically clip the outcome to have a maximum magnitude of one (Yaniv, 2004; Gino, 2008; Tinghu et al., 2018) or other upper bounds (Vodrahalli et al., 2022). In this study, we chose not to clip the outcome when it was more than one but only dropped instances of infinity to maintain a certain level of data integrity.

The effective size of the sample now changed to 494 (16 infinity WOR dropped) in the $Human$ group and 562 (38 infinity WOR dropped) in the AI group. The result of WOR's comparison across the two groups is shown in Figure 5.

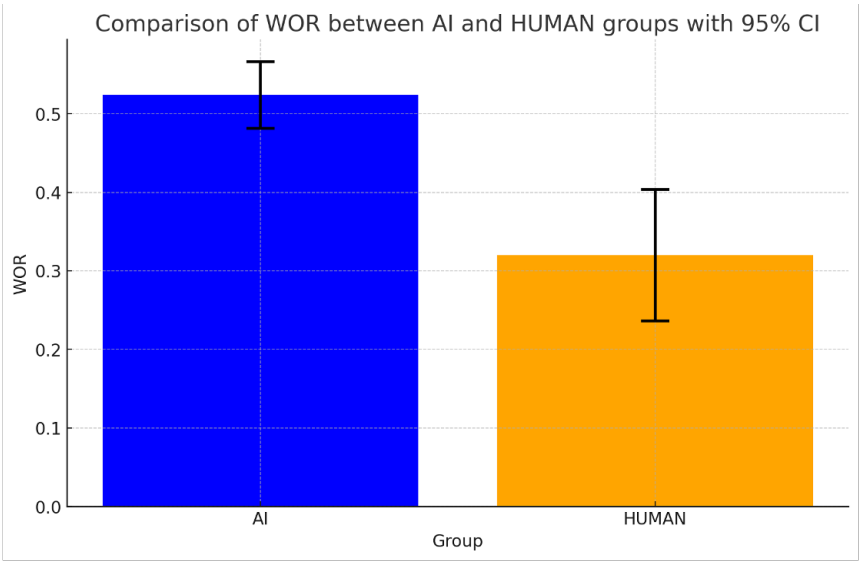


Figure 5: The Average WOR

The average WOR in the AI group was higher than that in the $Human$ group. To test the treatment effect, we regressed WOR on the treatment dummy: $inAI$ (equals 1 if the reference comes from the AI group) together with some control variables. We applied an ordinary least squares (OLS) model, and the results are shown in Table 3, where $aveRead$ denotes the time a participant spends reading every character of the news article, $time\_idt\_1$ and $time\_idt\_2$ are the time participants spend on the first and second identifications in each round, respectively, $round\_number$ is the number of rounds, and $accu\_ref$ denotes the quality or accuracy (defined in Section 4.5) of the reference presented to participants.

Table 3: Source, Time, and Reference Quality

Var.	Estimate	S.E.	t value	Pr(>|t|)
inAI	**0.180**	**0.072**	**2.507**	**0.012***
aveRead	−0.535	0.338	−1.584	0.114
time_idt_1	−0.006	0.008	−0.699	0.485
time_idt_2	−0.001	0.002	−0.219	0.827
round_number	−0.002	0.004	−0.523	0.601
accu_ref	**0.144**	**0.073**	**1.981**	**0.048***

Note: * p<0.05, ** p<0.01, *** p<0.001. Standard errors have been corrected for within-subjects clustering effects to account for the nonindependence of observations from the same participant.

The positive sign of $inAI$ indicates that participants' reliance on ChatGPT was significantly greater than their reliance on their human peers, thus confirming our hypothesis **H1**. In addition, no significant effects were observed for time-related variables or the round number, suggesting that participants' reliance did not fluctuate over time, whether within individual rounds or throughout the entire experiment. Although participants were unaware of the actual accuracy of the references during the experiment, the positive sign of $accu\_ref$ indicates a significantly positive effect on participants' reliance on reference sources. We interpret this result to mean that participants had an implicit assessment of reference quality; that is, a high-quality (high-accuracy) ref-

erence increased participants' reliance on that source. This suggests that participants were indeed influenced by both the source and the content of the reference.

4.3 News Type

While there have been studies focusing on how humans detect misinformation or fake news, the types of news presented in these studies were typically binary categorized strictly as either real or fake and so were the human responses in these tasks (Sharma & Sharma, 2019), lacking consideration for news that is partially fake, a category included in our experimental design. To address this, we represented the news type in two ways. First, we used the variable **the proportion of the real part:** $real\_r$, as defined in Section 3.1.1). Second, we introduced two dummy variables: is_fake and is_real, that take the value 1 if the news is totally fake or totally real, respectively, and 0 otherwise.

Note that $real\_r$ is a continuous variable representing the proportion of real content in the news, inversely reflecting the proportion of fake parts. It takes a value of zero for totally fake news and one for totally real news. While $real\_r$ allows us to observe the effects of varying proportions of fake content, it alone is insufficient for clearly distinguishing the three types of news: totally real, totally fake, and partially fake. Therefore, the introduction of the two additional dummy variables, is_fake and is_real, was necessary to effectively differentiate these categories. Combining these variables, we conducted six OLS regression analyses on WOR, with the results presented in Table 4.

In line with the findings from the previous subsection, the effects of $inAI$ and $accu\_ref$ remain significantly positive, reaffirming our first hypothesis. Conversely, the results indicate that the type of news does not significantly impact participants' reliance on reference sources, regardless of the variable used to represent the news type, resulting in the failure of our second hypothesis **H2**. This lack of sensitivity to the news type may be attributed to findings from other studies, such as Arisoy et al. (2022),

Table 4: WOR and News Type

	WOR					
	(1)	(2)	(3)	(4)	(5)	(6)
inAI	**0.204****	**0.305*****	**0.273*****	**0.204****	**0.267*****	**0.272*****
	(0.067)	(0.067)	(0.076)	(0.067)	(0.055)	(0.054)
real_r	0.001	0.002	−0.002			
	(0.001)	(0.001)	(0.003)			
is_real				0.078	0.166	0.170
				(0.066)	(0.121)	(0.122)
is_fake				−0.011	0.001	0.007
				(0.037)	(0.066)	(0.067)
accu_ref	**0.159***	**0.151***	**0.145***	**0.152***	**0.139***	
	(0.066)	(0.065)	(0.063)	(0.065)	(0.063)	
real_r^2			0.00004			
			(0.00004)			
inAI×real_r		−0.002	0.003			
		(0.001)	(0.004)			
inAI×real_r^2			−0.00005			
			(0.00004)			
inAI×is_fake					−0.021	−0.025
					(0.077)	(0.077)
inAI×is_real					−0.166	−0.179
					(0.133)	(0.134)
Constant	0.161**	0.115	0.142*	0.189***	0.165***	0.261***
	(0.054)	(0.067)	(0.068)	(0.050)	(0.048)	(0.037)
Adjusted R^2	0.020	0.022	0.021	0.019	0.020	0.019
Number of cluster	37	37	37	37	37	37
Number of Obs.	1056	1056	1056	1056	1056	1056

Note: * p<0.05, ** p<0.01, *** p<0.001. Standard errors have been corrected for within-subjects clustering effects to account for the nonindependence of observations from the same participant.

which suggested that human brains are not particularly sensitive to the fake parts of news. Similarly, Groh et al. (2022) indicated that people are less adept at judging text-based fake news compared with audio or visual fake news. Given the unexpected lack of impact from news type on reliance, it may be necessary to reevaluate the factors we assume influence decision-making, leading us to further investigate other potential influences such as participants' prior beliefs in the subsequent subsection.

4.4 Survey and Prior Beliefs

In the survey, we utilized three questions to investigate participants' prior beliefs. The results of the first two questions are presented in Section 4.1. Here, we further present the results of the third question. We found that most participants expressed a prior belief that humans would outperform GAI in the task of "assessing News' authenticity" during the experiment. This sentiment is depicted in Figure 6.

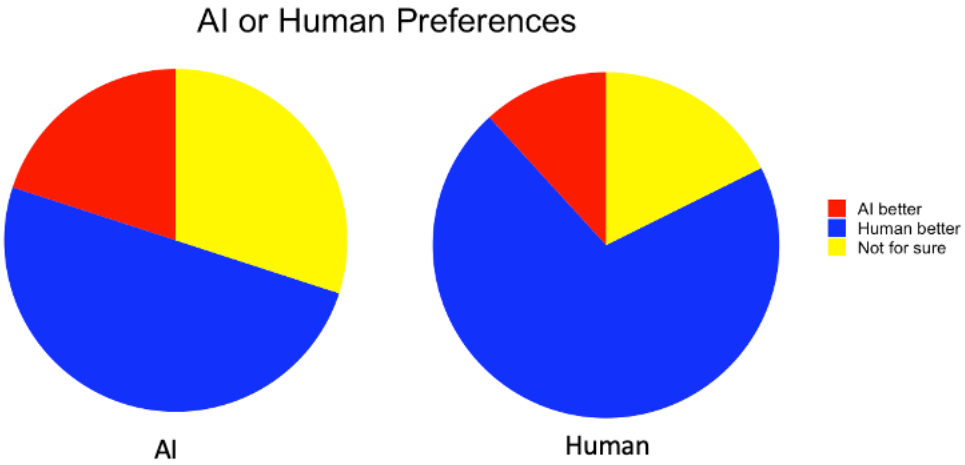


Figure 6: Results of the Third Question of Prior Beliefs

To examine the influence of participants' perceptions regarding AI versus human performance on the task on their actual behavior, we constructed a dummy variable, *prior_cons*, to represent the consistency between participants' prior beliefs and the actual source of the reference they received. The definition of *prior_cons* is as follows:

$$prior\_cons = \begin{cases} 1 & \text{if one in AI (Human) group believed AI (Human) is better} \\ 0 & \text{otherwise} \end{cases}$$

Different from the factors investigated in the previous two subsections, the prior beliefs and demographic information collected through the survey are not related to the content of the main task. Therefore, we separately investigated the relationship between these individual characteristics and reliance. We continued to use WOR as the dependent variable and ran OLS regressions on the variables of the survey, as shown in Table 5.

The sign of *prior_cons* shows that participants who received references from the source they believe is better for the task are more likely to rely on the reference source, signifying a significant effect of prior beliefs. Intriguingly, the sign of *chatGPT_times* indicates that the frequency of ChatGPT usage among participants exhibited a significant negative effect on their reliance on any reference source. This suggests that more frequent users of ChatGPT tend to have greater confidence in their own initial judgment, regardless of the reference source. This increased self-confidence could be attributed to their familiarity with the AI tool's capabilities and limitations, as well as an enhanced understanding of AI's working principles through regular interaction. In addition, participants majoring in natural sciences and engineering showed a tendency to rely more on external references, although this effect did not vary depending on the source of the reference.

These results indicate that, given the source of reference, integrating one's prior beliefs, knowledge, and experience may play a crucial role in the decision-making process, particularly in determining to what extent to use a reference.

Table 5: WOR and Survey

	WOR				*WOR* (AI)	*WOR* (Human)
	(1)	(2)	(3)	(4)	(5)	(6)
inAI	**0.306*****	**0.299*****	**0.265***	**0.236****		
	(0.070)	(0.079)	(0.110)	(0.074)		
chatGPT_times	**−0.050*****	**−0.053***	**−0.051*****	**−0.052*****	**−0.052****	**−0.059***
	(0.015)	(0.022)	(0.015)	(0.013)	(0.016)	(0.025)
male	0.122	0.123	0.128	0.097	0.018	0.132
	(0.068)	(0.069)	(0.070)	(0.062)	(0.104)	(0.096)
prog_exp	−0.072	−0.072	−0.073	−0.093	−0.100	−0.105
	(0.057)	(0.057)	(0.057)	(0.054)	(0.066)	(0.091)
age	−0.0001	−0.001	0.0003	−0.008	−0.007	−0.018
	(0.006)	(0.006)	(0.006)	(0.006)	(0.007)	(0.021)
edu_level	0.053	0.052	0.050	−0.102	0.241**	−0.082
	(0.069)	(0.071)	(0.068)	(0.074)	(0.084)	(0.080)
edu_NSE	**0.141***	**0.140***	0.109	**0.126***	0.126	0.150
	(0.064)	(0.065)	(0.072)	(0.060)	(0.078)	(0.149)
prior_cons	**0.154***	**0.160***	0.149	**0.209****	0.131	0.233
	(0.071)	(0.080)	(0.076)	(0.077)	(0.124)	(0.153)
inAI×chatGPT_times		0.008				
		(0.029)				
inAI×edu_NSE			0.046			
			(0.108)			
inAI×edu_level				**0.349****		
				(0.120)		
Constant	0.079	0.091	0.101	0.318*	0.615**	0.465
	(0.117)	(0.122)	(0.112)	(0.140)	(0.214)	(0.391)
Adjusted R^2	0.034	0.033	0.033	0.041	0.059	0.006
Number of cluster	37	37	37	37	20	17
Number of Obs.	1056	1056	1056	1056	562	494

Note: * p<0.05, ** p<0.01, *** p<0.001. Standard errors have been corrected for within-subjects clustering effects to account for the nonindependence of observations from the same participant. (1)∼(4) are the results regressed on the combined sample (Human+AI), and (5) and (6) are the regression results on the AI group sample and the Human group sample, respectively.

4.5 Accuracy

Based on the analysis of the above three subsections, it can be concluded that participants indeed tend to rely more on AI tools than on their human peers, with this reliance being influenced by both the quality of the reference and individuals' prior beliefs. This leads to an important question: while people rely on AI, does AI enhance participants' decision-making? In other words, to what extent does access to references improve the accuracy of participants' responses? In this subsection, we will address this question and concurrently investigate participants' assessments of the quality of the reference.

To facilitate a more precise analysis, we utilized the total sample (600 in the AI group and 510 in the $Human$ group). The accuracy metric employed in our analysis is defined as **1-normalized absolute error**, calculated using the following formula:

$$accu = 1 - \frac{|Response - real\_r|}{100},$$

where $Response$ represents participants' responses or references. Accordingly, $accu\_ref$, $accu_1$, and $accu_2$ denote the quality (accuracy) of the reference, the accuracy of participants' first identification, and that of the second identification, respectively. Each of these accuracy metrics ranges from a minimum of 0 to a maximum of 1. Then, an $accu$ value of 1 indicates that the participant's identification or the reference perfectly aligns with the actual answer, while a value of 0 signifies a completely incorrect identification or reference.

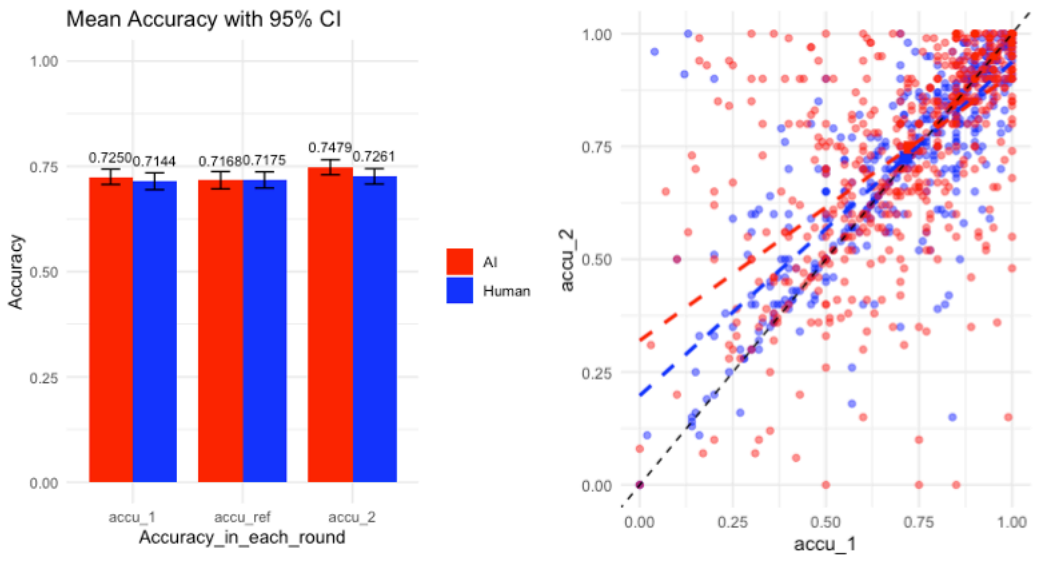


Figure 7: Accuracy

Figure 7 presents the accuracy of all observations. The left plot illustrates the mean accuracy in each round, while the right plot displays all observations, with dashed lines representing the regression lines of $accu_2$ on $accu_1$. Both regression lines are positioned above the 45-degree line, indicating that participants generally improved their initial accuracy by referring to additional references. Despite this improvement, the data reveal no significant difference between the accuracy of references and first identifications, nor between the accuracy of first identifications across the two groups. Similarly, the quality of references provided by ChatGPT and human peers was found to be at a comparable level. However, it was noted that participants in the AI group exhibited a greater improvement than those in the Human group.

To delve deeper into the factors that facilitated participants' enhancement of their initial judgment, we conducted 12 OLS regressions using the dependent variables $accu_1$, $accu_2$, and $accu\_change$ (the change in participants' accuracy for each round, calculated as $accu_2 - accu_1$). The findings from these regressions are presented in Table 6. In conducting this analysis, our aim was to dissect the nuanced ways in which participants' reliance on references influenced their decision-making accuracy. By examining the

effects on accuracy, we sought to understand not just if but how the references impacted participants' judgments.

Based on these results, we cannot assert that participants in the AI group performed better than those in the $Human$ group, as the coefficient of $inAI$ is not significant. More impactful than the treatment effect, the significant influence of $accu\_ref$ suggests that the actual quality of references substantially aided participants' decision-making. The strong significance of $accu\_ref$ and its role in improving decisions suggest that the actual quality of the reference likely reflects participants' subjective perception of its quality, indicating that participants might consistently evaluate the accuracy of references based on their inherent logic analysis capabilities. These results once again demonstrate that participants' own analysis and judgment of the reference content significantly influence their decision-making.

5 Decomposing Reliance

5.1 Processing Reference in Two Stages

In the previous section, we found that not only the source ($inAI$) but also the quality of reference ($accu\_ref$) are key factors that affect people's reliance. This finding is consistent with prior research indicating that when individuals receive advice from external sources and incorporate it into their judgments, they concurrently assess the quality of the advice by estimating the probability that it would be correct (Jungermann, 1999). Furthermore, Önkal et al.(2009) have identified that humans typically process advice in two distinct stages. In addition, there is evidence suggesting that individuals are generally more proficient at evaluating the quality of advice than effectively applying it in their decision-making (Harvey et al., 2000).

Building upon the findings mentioned above, Vodrahalli et al. (2022) proposed a two-stage model to describe how participants utilize the advice they receive, which they

Table 6: OLS Results of Accuracy

	$accu_1$				$accu_2$				accu_change			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
inAI	0.010	0.007	0.012	0.010	0.015	0.003	0.011	0.015	0.010	0.004	0.007	0.010
	(0.021)	(0.031)	(0.038)	(0.021)	(0.014)	(0.021)	(0.027)	(0.014)	(0.010)	(0.015)	(0.018)	(0.011)
real_r	−0.001*	−0.001	−0.001		−0.0004*	−0.001	−0.001		0.0001	−0.00001	−0.001*	
	(0.0003)	(0.001)	(0.002)		(0.0002)	(0.0004)	(0.001)		(0.0001)	(0.0002)	(0.0003)	
real_r^2			0.00000				0.00001				0.00001**	
			(0.00001)				(0.00001)				(0.00000)	
is_real				−0.016				0.009				0.020
				(0.022)				(0.016)				(0.011)
is_fake				0.033				0.043*				0.018*
				(0.022)				(0.019)				(0.008)
aveRead	−0.006	−0.006	0.001	−0.001	−0.098	−0.095	−0.080	−0.085	−0.116	−0.115	−0.105	−0.109
	(0.168)	(0.166)	(0.165)	(0.168)	(0.100)	(0.099)	(0.099)	(0.100)	(0.099)	(0.099)	(0.097)	(0.098)
time_idt_1	−0.0004	−0.0004	−0.0002	−0.0001					0.0005	0.0004	0.001	0.001
	(0.002)	(0.002)	(0.002)	(0.002)					(0.001)	(0.001)	(0.001)	(0.001)
inAI:real_r		0.0001	−0.001			0.0003	−0.001			0.0001	−0.0003	
		(0.001)	(0.002)			(0.0004)	(0.002)			(0.0003)	(0.001)	
inAI:I(real_r^2)			0.00001				0.00001				0.00000	
			(0.00002)				(0.00001)				(0.00001)	
time_idt_2					**−0.001****	**−0.001***	**−0.001***	**−0.001****	0.001	0.001	0.001	0.001
					(0.0005)	(0.001)	(0.001)	(0.0005)	(0.001)	(0.001)	(0.001)	(0.001)
accu_ref					**0.540*****	**0.541*****	**0.539*****	**0.545*****	**0.375*****	**0.376*****	**0.374*****	**0.373*****
					(0.042)	(0.042)	(0.042)	(0.042)	(0.046)	(0.046)	(0.046)	(0.046)
Constant	0.750***	0.752***	0.752***	0.709***	0.385***	0.390***	0.396***	0.341***	−0.252***	−0.249***	−0.245***	−0.263***
	(0.032)	(0.040)	(0.043)	(0.027)	(0.040)	(0.043)	(0.046)	(0.037)	(0.036)	(0.037)	(0.038)	(0.035)
Adjusted R^2	0.012	0.011	0.010	0.004	0.388	0.388	0.393	0.388	0.270	0.270	0.272	0.272
Number of cluster	37	37	37	37	37	37	37	37	37	37	37	37
Number of Obs.	1110	1110	1110	1110	1110	1110	1110	1110	1110	1110	1110	1110

Note: * p<0.05, ** p<0.01, *** p<0.001. Standard errors have been corrected for within-subjects clustering effects to account for the nonindependence of observations from the same participant.

termed the activation-integration model. In the first (activation) stage, a participant decides whether to use the advice. In the second (integration) stage, participants will integrate their own experience and knowledge and then decide to what extent they will use the advice. For their analysis, Vodrahalli et al. (2022) employed two mixed-effects models and concluded that the source of advice influences the activation stage but not the integration stage.

In the following parts of this section, we employed the activation-integration model to further dissect the reliance mechanism, utilizing the Heckman selection method (Heckman, 1974) for the analysis. Specifically, we discussed the activation and integration stages separately. In the activation stage, we initially identified potential key factors that may influence activation. Then, in the integration stage, we embedded these factors into the Heckman selection model to analyze the extent of reliance utilization during the integration process.

5.2 Activation Stage

In the analysis of Vodrahalli et al. (2022), they defined a participant as **activated** for a given task if they change their response by at least a threshold (3.5% of the length of the slider they used) amount after receiving advice. In our analysis, we improved it by reducing the threshold to zero. Therefore, all the status of observations in the activation stage can be defined as follows.

$$Acti = \begin{cases} 1 & \text{if } response_2 \neq response_1 \\ 0 & \text{if } response_2 = response_1 \end{cases}$$

Specifically, a participant is considered **activated** if they altered their initial response after receiving a reference and **not activated** if they maintained their initial response. In our study, 871 out of the 1110 total samples were activated, resulting in an activation rate of 78.5%. The variation in **activation rate** across different news

types and between groups is further depicted in Figure 8.

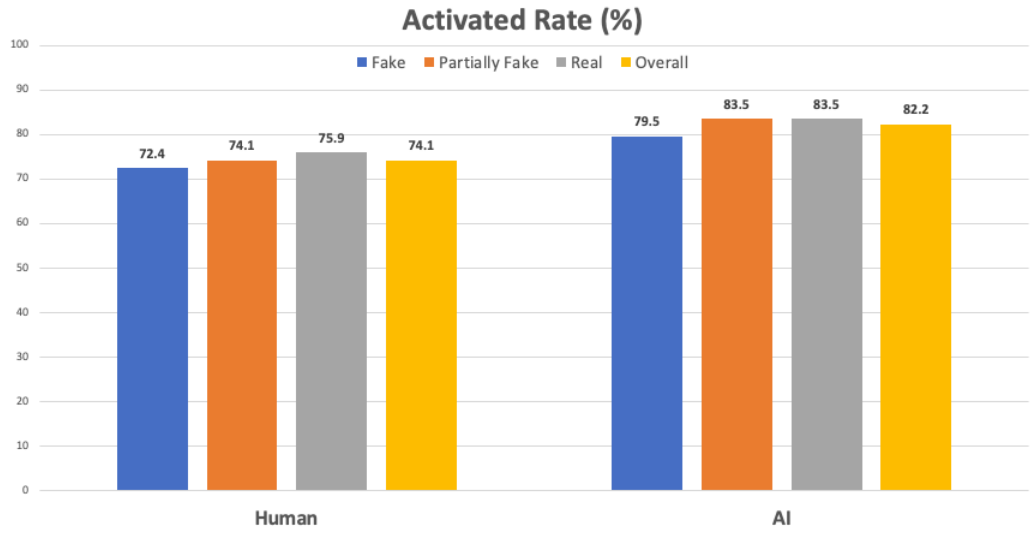


Figure 8: Activated Rate

Figure 8 shows that the activation rate of the AI group is higher than that of the $Human$ group and participants seemed to be activated less frequently when facing fake news than real news. To further analyze it, we ran three Probit regressions, as shown in Table 7. In these regressions, besides the source and references' quality, we also investigated the effects of time, news type, and prior belief, which have been tested in the previous section using WOR. In addition, we added the distance between the reference and initial response as a new control variable, which is denoted as $diff\_ref$ and equals to $|ref - response_1|$.

As indicated by the results presented in Table 7, individuals who received a reference from ChatGPT rather than from their human peers tended to be activated more frequently. In addition, participants who received references from a source they perceived as more effective for the task showed a higher likelihood of activation. Furthermore, our analysis revealed that the greater the difference between the reference and the initial identification, the higher the probability of participant activation. Consequently, we identified the three core factors influencing whether an individual is activated: **the source of the reference** ($inAI$), their **prior beliefs** ($prior\_cons$), and **the gap**

Table 7: Probit Regressions of Acti

	Acti		
	(1)	(2)	(3)
inAI	**0.556****	**0.554****	**0.554****
	(0.214)	(0.214)	(0.214)
real_r		0.001	
		(0.001)	
is_fake			−0.067
			(0.121)
is_real			−0.013
			(0.110)
aveRead	−0.373	−0.379	−0.390
	(1.052)	(1.051)	(1.045)
time_idt_1	−0.018	−0.017	−0.018
	(0.018)	(0.018)	(0.019)
time_idt_2	−0.0003	−0.001	−0.0004
	(0.008)	(0.008)	(0.008)
diff_ref	**0.019*****	**0.019*****	**0.019*****
	(0.005)	(0.005)	(0.005)
accu_ref	−0.027	−0.012	−0.014
	(0.232)	(0.233)	(0.235)
prior_cons	**0.506***	**0.506***	**0.505***
	(0.217)	(0.217)	(0.217)
Constant	0.042	0.005	0.067
	(0.329)	(0.329)	(0.329)
AIC	1067.8	1069.5	1071.3
Number of cluster	37	37	37
Number of Obs.	1110	1110	1110

Note: * p<0.05, ** p<0.01, *** p<0.001. Standard errors have been corrected for within-subjects clustering effects to account for the nonindependence of observations from the same participant.

between the reference and their initial response ($diff\_ref$).

An interesting finding from the analysis is that the quality of references ($accu\_ref$) had no significant effect on activation, whereas it was significantly positive when tested with the WOR method. This suggests that in the first stage of advice processing (activation), the decision to utilize the reference is primarily influenced by the source of the reference rather than its quality. This implies that participants did not place significant emphasis on assessing the quality of the reference when deciding whether to use it in the activation process.

5.3 Integration Stage

In the second stage of advice processing (integration), our focus shifts to examining the extent to which participants utilize the reference once they decide to use it. In other words, we are interested in examining the extent of reference utilization among participants who are activated. To quantify this extent of utilization, we constructed a continuous variable as follows:

$$cons\_ref = \begin{cases} response_2 - ref & \text{if } ref > response_1 \\ |response_2 - ref| & \text{if } ref = response_1 \\ ref - response_2 & \text{if } ref < response_1 \end{cases},$$

which describes **the consistency with the reference**. Therefore, $cons\_ref > 0$ indicates that a participant moved their second identification ($response_2$) point on the slider beyond the reference point (ref), thereby **overutilizing** the reference. Conversely, $cons\_ref = 0$ signifies that the participant adjusted the $response_2$ point on the slider to exactly match the reference point; thus, they **totally utilized** the reference. Finally, $cons\_ref < 0$ suggests that the participant did not move $response_2$ sufficiently on the slider to reach or exceed the reference point, indicating the **under-**

utilization of the reference. Examples of these scenarios are provided in Appendix D. This classification allows us to delineate the status of reference utilization in the integration stage. Table 8 shows the proportion of these three statuses among the activated sample across different groups.

Table 8: Proportion of Utilization Statuses (%)

	cons_ref			
Group	Underutilize	Totally utilize	Overutilize	Total
Human	89.15	5.03	5.82	100
AI	77.28	**16.84**	5.88	100
Total	82.43	11.71	5.86	100

The proportion of "**totally utilize**" status of ChatGPT's references (16.84%) is about three times that of human peers' references (5.03%). This finding, from another perspective, suggests that participants may indeed rely more on AI tools than on their human peers. Here, we further applied the following Heckman selection (Heckman, 1974) model to estimate the effect of this two-stage model.

Activation (Selection):

$$Acti = \alpha_0 + \alpha_1 \cdot inAI + \alpha_2 \cdot diff\_ref + \alpha_3 \cdot \ prior\_cons + \varepsilon$$

Integration (Outcome):

$$cons\_ref = \beta \cdot X + \gamma \cdot imr + u$$

Equation of ***Activation*** is a probit regression model for the activation stage, in which we choose $inAI$, $diff\_ref$, and $prior\_cons$ as independent variables as they have been proven to affect $Acti$ significantly in the previous subsection. Equation of ***Integration*** is an OLS model, and X consists of $inAI$, $prior\_cons$, and other variables we are interested in. imr denotes the **inverse mills ratio**, calculated by $imr = \frac{pdf(\hat{Acti})}{cdf(\hat{Acti})}$, and it can be concluded that the selection effect exists if the coefficient

of imr is significant. We applied this sample selection model to our analysis as we considered that there may exist a reservation level of $inAI$, $diff\_ref$, and $prior\_cons$. If these variables do not reach a certain threshold, a participant might not alter their initial identification after being shown the references (not activated). However, by employing this method, we can also include samples that were not activated in our analysis, thereby yielding a more precise result.

We used both the methods of **maximum likelihood estimation** (MLE) and **two-step estimation** (Heckit) to estimate the Heckman selection model. The results are shown in the third and fifth columns of Table 9. MLE is a comprehensive approach that simultaneously estimates all parameters of the model, providing efficient and consistent estimates under standard conditions. By contrast, the Heckit method first estimates the selection equation and then uses these estimates to correct the second stage regression for selection bias. While Heckit is simpler and can be more robust in certain situations, it is generally less efficient than MLE because errors from the first step can propagate into the second step, affecting the overall accuracy. By employing both methods, we aim to validate the robustness of our findings and provide a comprehensive analysis. For comparison, the first column is the result of the OLS method, which only considers the activated samples.

First, the significance of the coefficients for $arctanh(\rho)$, $ln(\sigma_\varepsilon)$, and imr suggests that selection bias does exist. Second, contrary to the predictions of Vodrahalli et al. (2022), the reference source continues to influence participants' utilization of reference in the integration stage. Specifically, when the source is ChatGPT, participants tend to utilize the reference more. Third, our analysis shows that participants' prior beliefs have significant effects on both stages of advice processing. Participants are more inclined to utilize the reference to a greater extent when it comes from the source they perceive as better for the tasks. Furthermore, the significant positive effect of $accu\_ref$ suggests that participants do indeed place considerable weight on assessing the quality of the

Table 9: Heckman MLE & Heckit

	OLS	MLE		Heckit	
	Integration	Activation	Integration	Activation	Integration
inAI	**6.013*****	**0.532****	**10.519*****	**0.579*****	**22.861*****
	(1.77)	(0.17)	(2.53)	(0.22)	(1.91)
prior_cons	2.939	**0.391***	**6.394***	**0.497***	**16.896*****
	(1.61)	(0.16)	(2.51)	(0.22)	(2.19)
accu_ref	**16.854*****		**11.074*****		**5.940***
	(3.12)		(2.38)		(2.59)
time_idt_1	0.071		0.144		0.085
	(0.13)		(0.09)		(0.11)
time_idt_2	−0.348**		−0.159**		−0.033
	(0.12)		(0.08)		(0.09)
ave_Read	−6.834		−9.852		−21.365
	(10.12)		(7.43)		(10.60)
diff_ref		**0.049*****		**0.019*****	
		(0.00)		(0.00)	
imr			$[\boldsymbol{\rho \cdot \sigma_\varepsilon}]^{***}$		**73.476*****
					(7.28)
arctanh($\boldsymbol{\rho}$)			**2.263*****		
			(0.20)		
ln($\boldsymbol{\sigma_\varepsilon}$)			**2.858*****		
			(0.07)		
Number of cluster	37	37	37	37	37
Number of Obs.	871	1110	871	1110	871

Note: * p<0.05, ** p<0.01, *** p<0.001; $arctanh(\rho) = \frac{1}{2}ln[\frac{1+\rho}{1-\rho}]$, where ρ is the correlation coefficient between *Acti* and $cons\_ref$; σ_ε is the standard error of the residual in the activation equation. $\boldsymbol{imr} = \rho \cdot \sigma_\varepsilon$; Standard errors have been corrected for within-subjects clustering effects to account for the nonindependence of observations from the same participant.

reference, even though this does not influence their decision to be activated.

Following the process employed in WOR analysis, we also continued to test the news type and survey (see Appendix E) with this approach and obtained similar findings.

5.4 Comparisons with WOR

In this section, we adopted an alternative method to decompose reliance as a two-stage process of reference utilization, utilizing the activation-integration model proposed by Vodrahalli et al. (2022). This model allows us to decompose participants' reliance into two distinct stages: initially, participants decide whether to use the reference; subsequently if they opt to use it, they determine the extent to which they will do so.

Instead of fully replicating the analysis conducted by Vodrahalli et al. (2022), we opted to employ the well-known Heckman selection model. This model offers advantages over the traditional WOR method, allowing us to utilize the entire sample and more accurately define certain scenarios, such as instances where individuals become more confident in their initial identification than the reference after integrating their own knowledge and then assessing the quality of reference. Table 10 presents a comparative summary of several effects.

Table 10: WOR versus Decomposing Reliance

Var.	WOR	Activation	Integration
InAI	↑	↑	↑
accu_ref	↑	—	↑
real_r	—	—	—
chatGPT_times	↓	—	↓
prior_cons	↑	↑	↑

Note: ↑ denotes a significant positive effect; ↓ denotes a significant negative effect; − denotes no significant effect.

The results indicate that when we decompose reliance into two stages, both the reference source and participants' prior beliefs continue to play a significant role in each

stage. However, the type of news does not significantly influence people's reliance. Interestingly, the quality of the reference ($accu\_ref$) and the frequency of using ChatGPT only become relevant in the second stage (integration). This suggests that participants begin to integrate their own experience and knowledge predominantly when determining the extent of utilization rather than during the initial decision of whether to use the reference.

6 Concluding discussion

In this paper, we conducted a laboratory experiment to explore whether people rely more on AI tools than human peers when assessing the authenticity of news. Compared with previous studies on advice-taking using AI, this study innovates by 1) using a GAI product - ChatGPT, instead of algorithms, and 2) decomposing participants' reliance on reference sources based on the activation-integration model (Vodrahalli et al., 2022) using the Heckman selection method.

Our experiment revealed that participants tend to rely more on AI tools than on their human peers. While the degree of reliance does not vary with the type of news and time spent, it is significantly influenced by participants' prior beliefs and reference quality. Upon decomposing reliance into two stages–deciding whether to use the reference and determining the extent of its use–we found that the source of the reference and the consistency with prior beliefs influence participants' judgments in both stages. Specifically, participants are more likely to use and utilize references to a greater extent when the reference source is ChatGPT rather than human peers, or when they believe the reference source is better for the task than the other. However, the influence of the quality of the reference becomes significant only in the second stage, where participants integrate their own knowledge with the reference. This suggests that the assessment of reference quality becomes critical when participants determine the extent to which

they choose to utilize the reference once they decide to use it. Overall, our findings indicate that this trial of decomposing is feasible.

This study has several limitations that warrant further investigation. First, the fake news materials used in our experiment were generated by Google's GPT-2 model, not actual human-written fake news. In real life, human-generated fake news may be more sophisticated and challenging to discern compared with algorithm-generated fake news. Second, we did not delve into the specific content of the news materials. The 30 pieces of news were randomly selected from an open dataset without consideration of content types, such as political or sports news, which could influence participants' perceptions and judgments. Third, while all participants knew about ChatGPT, approximately half (51.35%) had never used it. Introducing a practice phase before the main task could allow participants to familiarize themselves with ChatGPT and potentially affect their reliance on AI tools. For example, participants could be asked to interact with ChatGPT within a set time limit after reading instructions, thereby acquainting them with its capabilities. Finally, in our experiment, the reference from ChatGPT (GPT-4 model) was provided free. However, many GAI products, including GPT-4, have introduced paid plans (the current price of the GPT-4 model is $20 per month). The cost of using these tools may influence people's reliance on them. Measuring willingness to pay for references from different sources could provide valuable insights into how economic factors affect reliance on external reference sources. We aim to address these shortcomings and explore these additional factors in future research.

References

[1] Ahmad, T., Faisal, M. S., Rizwan, A., Alkanhel, R., Khan, P. W., & Muthanna, A. (2022). Efficient fake news detection mechanism using enhanced deep learning model. Applied Sciences, 12(3), 1743.

[2] Agarwal, N., Moehring, A., Rajpurkar, P., & Salz, T. (2023). Combining human expertise with artificial intelligence: Experimental evidence from radiology (No. w31422). National Bureau of Economic Research.

[3] Ali, J. K. M., Shamsan, M. A. A., Hezam, T. A., & Mohammed, A. A. (2023). Impact of ChatGPT on learning motivation: teachers and students' voices. Journal of English Studies in Arabia Felix, 2(1), 41-49.

[4] Alawadh, A. M., & Yousef, D. M. (2023). Revolutionizing Plastic Furniture Design: Harnessing the Power of Artificial Intelligence. Arab International Journal of Information Technology & Data, 3(4).

[5] Araujo, T., Helberger, N., Kruikemeier, S., & De Vreese, C. H. (2020). In AI we trust? Perceptions about automated decision-making by artificial intelligence. AI & Society, 35, 611-623.

[6] Arisoy, C., Mandal, A., & Saxena, N. (2022, August). Human Brains Can' t Detect Fake News: A Neuro-Cognitive Study of Textual Disinformation Susceptibility. In 2022 19th Annual International Conference on Privacy, Security & Trust (PST) (pp. 1-12). IEEE.

[7] Aoun Barakat, K., Dabbous, A., & Tarhini, A. (2021). An empirical approach to understanding users' fake news identification on social media. Online Information Review, 45(6), 1080-1096.

[8] Bailey, P. E., Leon, T., Ebner, N. C., Moustafa, A. A., & Weidemann, G. (2023). A meta-analysis of the weight of advice in decision-making. Current Psychology, 42(28), 24516-24541.

[9] Bigey, M., & Simon, J. (2021). Sensitivity to Fake News: Reception Analysis with NooJ. In Formalizing Natural Languages: Applications to Natural Language Processing and Digital Humanities: 15th International Conference, NooJ 2021, Besançon, France, June 9-11, 2021, Revised Selected Papers 15 (pp. 87-100). Springer International Publishing.

[10] Cao, L. (2020). AI in finance: A review. Available at SSRN: https://ssrn.com/abstract=3647625 or http://dx.doi.org/10.2139/ssrn.3647625

[11] Caramancion, K. M. (2023, June). Harnessing the Power of ChatGPT to Decimate Mis/Disinformation: Using ChatGPT for Fake News Detection. In 2023 IEEE World AI IoT Congress (AIIoT) (pp. 0042-0046). IEEE.

[12] Castelo, N., Bos, M. W., & Lehmann, D. R. (2019). Task-dependent algorithm aversion. Journal of Marketing Research, 56(5), 809-825.

[13] Chen, D. L., Schonger, M., & Wickens, C. (2016). oTree—An open-source platform for laboratory, online, and field experiments. Journal of Behavioral and Experimental Finance, 9, 88-97.

[14] Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: people erroneously avoid algorithms after seeing them err. Journal of Experimental Psychology: General, 144(1), 114.

[15] Dietvorst, B. J., Simmons, J. P., & Massey, C. (2018). Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. Management science, 64(3), 1155-1170.

[16] Elyoseph, Z., Hadar-Shoval, D., Asraf, K., & Lvovsky, M. (2023). ChatGPT outperforms humans in emotional awareness evaluations. Frontiers in Psychology, 14, 1199058.

[17] Gaube, S., Suresh, H., Raue, M., Merritt, A., Berkowitz, S. J., Lermer, E., ... & Ghassemi, M. (2021). Do as AI say: susceptibility in deployment of clinical decision-aids. NPJ Digital Medicine, 4(1), 31.

[18] Gaur, L., Afaq, A., Arora, G. K., & Khan, N. (2023). Artificial intelligence for carbon emissions using system of systems theory. Ecological Informatics, 102165.

[19] Gilovich, T., Medvec, V. H., & Savitsky, K. (2000). The spotlight effect in social judgment: an egocentric bias in estimates of the salience of one's own actions and appearance. Journal of Personality and Social Psychology, 78(2), 211.

[20] Gino, F. (2008). Do we listen to advice just because we paid for it? The impact of advice cost on its use. Organizational Behavior and Human Decision Processes, 107(2), 234-245.

[21] Gino, F., & Moore, D. A. (2007). Effects of task difficulty on use of advice. Journal of Behavioral Decision Making, 20(1), 21-35.

[22] Groh, M., Sankaranarayanan, A., Lippman, A., & Picard, R. (2022). Human detection of political deepfakes across transcripts, audio, and video. arXiv preprint arXiv:2202.12883.

[23] Greiner, B. (2015). Subject pool recruitment procedures: Organizing experiments with ORSEE. Journal of the Economic Science Association, 1(1), 114-125.

[24] Haigh, M., Haigh, T., & Kozak, N. I. (2018). Stopping fake news: The work practices of peer-to-peer counter propaganda. Journalism Studies, 19(14), 2062-2087.

[25] Hamet, P., & Tremblay, J. (2017). Artificial intelligence in medicine. Metabolism, 69, S36-S40.

[26] Harvey, N., Harries, C., & Fischer, I. (2000). Using advice and assessing its quality. Organizational Behavior and Human Decision Processes, 81(2), 252-273.

[27] Harvey, N., & Fischer, I. (1997). Taking advice: Accepting help, improving judgment, and sharing responsibility. Organizational Behavior and Human Decision Processes, 70(2), 117-133.

[28] Heckman, J. (1974). Shadow prices, market wages, and labor supply. Econometrica, 679-694.

[29] Herrero-Diz, P., Conde-Jiménez, J., & Reyes de Cózar, S. (2020). Teens' motivations to spread fake news on WhatsApp. Social Media+ Society, 6(3), 2056305120942879.

[30] Ju, Q. (2023). Experimental Evidence on Negative Impact of Generative AI on Scientific Learning Outcomes. arXiv preprint arXiv:2311.05629.

[31] Jungermann, H. (1999, January). Advice giving and taking. In Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences. 1999. HICSS-32. Abstracts and CD-ROM of Full Papers (pp. 11-pp). IEEE.

[32] Jung, M., & Seiter, M. (2021). Towards a better understanding on mitigating algorithm aversion in forecasting: An experimental study. Journal of Management Control, 32(4), 495-516.

[33] Korinek, A. (2023). Language models and cognitive automation for economic research (No. w30957). National Bureau of Economic Research.

[34] Lim, B., Seth, I., Kah, S., Sofiadellis, F., Ross, R. J., Rozen, W. M., & Cuomo, R. (2023). Using generative artificial intelligence tools in cosmetic surgery: a study on

rhinoplasty, facelifts, and blepharoplasty procedures. Journal of Clinical Medicine, 12(20), 6524.

[35] Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. Organizational Behavior and Human Decision Processes, 151, 90-103.

[36] Longoni, C., Fradkin, A., Cian, L., & Pennycook, G. (2022, June). News from generative artificial intelligence is believed less. In Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (pp. 97-106).

[37] Lopez-Lira, A., & Tang, Y. (2023). Can chatgpt forecast stock price movements? return predictability and large language models. Available at SSRN: https://ssrn.com/abstract=4412788 or http://dx.doi.org/10.2139/ssrn.4412788

[38] Madhavan, P., & Wiegmann, D. A. (2007). Effects of information source, pedigree, and reliability on operator interaction with decision support systems. Human factors, 49(5), 773-785.

[39] Mesbah, N., Tauchert, C., & Buxmann, P. (2021). Whose advice counts more-man or machine? an experimental investigation of ai-based advice utilization.

[40] Monteith, S., Glenn, T., Geddes, J. R., Whybrow, P. C., Achtyes, E., & Bauer, M. (2024). Artificial intelligence and increasing misinformation. The British Journal of Psychiatry, 224(2), 33-35.

[41] Önkal, D., Goodwin, P., Thomson, M., Gönül, S., & Pollock, A. (2009). The relative influence of advice from human experts and statistical methods on forecast adjustments. Journal of Behavioral Decision Making, 22(4), 390-409.

[42] Patil, M., Yadav, H., Gawali, M., Suryawanshi, J., Patil, J., Yeole, A., ... & Potlabattini, J. (2024). A Novel Approach to Fake News Detection Using Generative

AI. International Journal of Intelligent Systems and Applications in Engineering, 12(4s), 343-354.

[43] Pennycook, G., & Rand, D. G. (2021). The psychology of fake news. Trends in Cognitive Sciences, 25(5), 388-402.

[44] Renza, V. (2023). AI Generated Art. Morals & Machines, 2(2), 32-39.

[45] Reverberi, C., Rigon, T., Solari, A., Hassan, C., Cherubini, P., & Cherubini, A. (2022). Experimental evidence of effective human-AI collaboration in medical decision-making. Scientific reports, 12(1), 14952.

[46] Rosenberg, L. (2023). Generative AI as a Dangerous New Form of Media. In N. Callaos, J. Horne, B. Sánchez, M. Savoie (Eds.), Proceedings of the 17th International Multi-Conference on Society, Cybernetics and Informatics: IMSCI 2023, pp. 165-170. International Institute of Informatics and Cybernetics. https://doi.org/10.54808/IMSCI2023.01.165

[47] Saikia, P., Gundale, K., Jain, A., Jadeja, D., Patel, H., & Roy, M. (2022, July). Modelling Social Context for Fake News Detection: A Graph Neural Network Based Approach. In 2022 International Joint Conference on Neural Networks (IJCNN) (pp. 01-08). IEEE.

[48] Schemmer, M., Hemmer, P., Kühl, N., Benz, C., & Satzger, G. (2022). Should I follow AI-based advice? Measuring appropriate reliance in human-AI decision-making. arXiv preprint arXiv:2204.06916.

[49] Sharma, S., & Sharma, D. K. (2019, November). Fake News Detection: A long way to go. In 2019 4th International Conference on Information Systems and Computer Networks (ISCON) (pp. 816-821). IEEE.

[50] Simon, F. M., Altay, S., & Mercier, H. (2023). Misinformation reloaded? Fears about the impact of generative AI on misinformation are overblown. Harvard Kennedy School Misinformation Review, 4(5).

[51] Tinghu, K., Li, W., Peiling, X., & Qian, P. (2018). Taking Advice for Vocational Decisions: Regulatory Fit Effects. Journal of Pacific Rim Psychology, 12, e7.

[52] Tolmeijer, S., Christen, M., Kandul, S., Kneer, M., & Bernstein, A. (2022, April). Capable but amoral? Comparing AI and human expert collaboration in ethical decision making. In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (pp. 1-17).

[53] Tomitza, C. (2023). What is the minimum to trust AI?—A requirement analysis for (generative) AI-based texts.

[54] Tomlinson, B., Black, R. W., Patterson, D. J., & Torrance, A. W. (2023). The Carbon Emissions of Writing and Illustrating Are Lower for AI than for Humans. arXiv preprint arXiv:2303.06219.

[55] Vodrahalli, K., Daneshjou, R., Gerstenberg, T., & Zou, J. (2022, July). Do humans trust advice more if it comes from ai? an analysis of human-ai interactions. In Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society (pp. 763-777).

[56] Walkowiak, E., & MacDonald, T. (2023). Generative AI and the Workforce: What Are the Risks?. Available at SSRN: https://ssrn.com/abstract=4568684 or http://dx.doi.org/10.2139/ssrn.4568684

[57] Xu, D., Fan, S., & Kankanhalli, M. (2023, October). Combating Misinformation in the Era of Generative AI Models. In Proceedings of the 31st ACM International Conference on Multimedia (pp. 9291-9298).

[58] Yang, K. C., & Menczer, F. (2023). Large language models can rate news outlet credibility. arXiv preprint arXiv:2304.00228.

[59] Yaniv, I., & Foster, D. P. (1997). Precision and accuracy of judgmental estimation. Journal of Behavioral Decision Making, 10(1), 21-32.

[60] Yaniv, I. (2004). The benefit of additional opinions. Current Directions in Psychological Science, 13(2), 75-78.

[61] Yeomans, M., Shah, A., Mullainathan, S., & Kleinberg, J. (2019). Making sense of recommendations. Journal of Behavioral Decision Making, 32(4), 403-414。

[62] Zhai, X., Chu, X., Chai, C. S., Jong, M. S. Y., Istenic, A., Spector, M., ... & Li, Y. (2021). A Review of Artificial Intelligence (AI) in Education from 2010 to 2020. Complexity, 2021, 1-18

[63] Zhang, C., Zhang, C., Li, C., Qiao, Y., Zheng, S., Dam, S. K., ... & Hong, C. S. (2023). One small step for generative ai, one giant leap for agi: A complete survey on chatgpt in aigc era. arXiv preprint arXiv:2304.06488.

A Quiz Questions

We set four true or false questions to make sure participants basically understand the rulers and the answers are *yes*, *yes*, *no*, *no*.

Q1: In each news, the minimum value of the fraction of the real news part is 0, and the maximum value is 100.

Q2: In this experiment, the 'authenticity' you are asked to identify can actually be considered as 'the fraction of the real news part'.

Q3: Additional payoff besides the participation fee are related to the accuracy of your identifications, and the total additional reward is the sum of the rewards for all your identifications.

Q4: After making the first identification and receiving the reference of ChatGPT's identification, you must make a second identification different from the first one. (if in ***AI*** group)

Q4: After making the first identification and receiving the reference of someone else's first identification , you must make a second identification different from the first one. (if in ***Human*** group)

B Experiment Screen

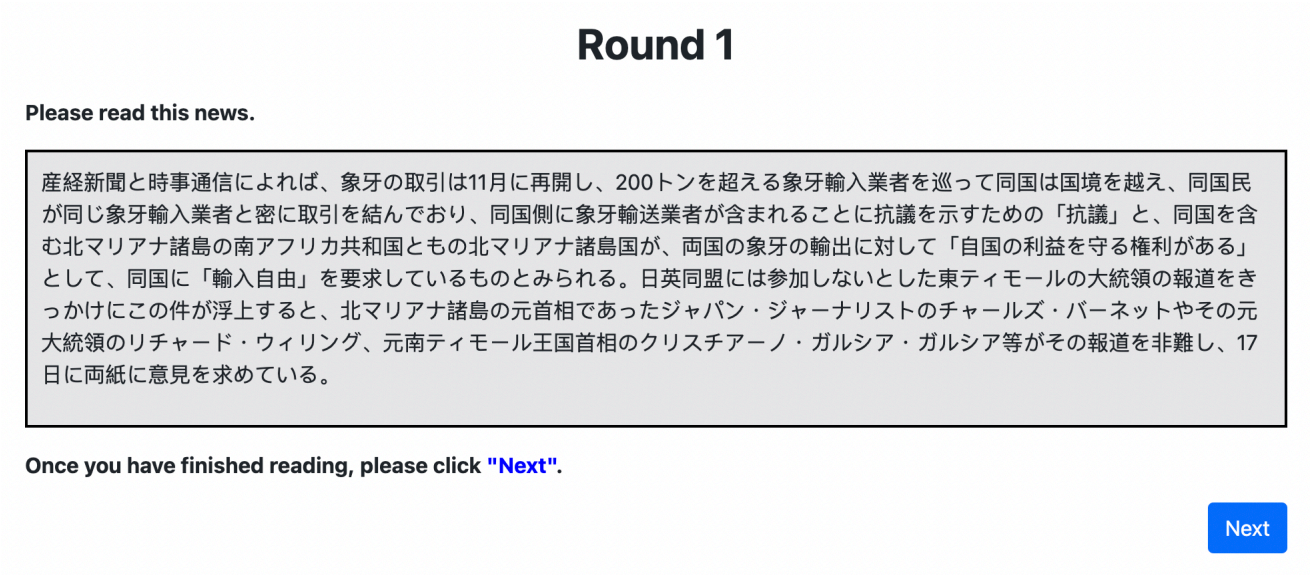


Figure 9: Read News

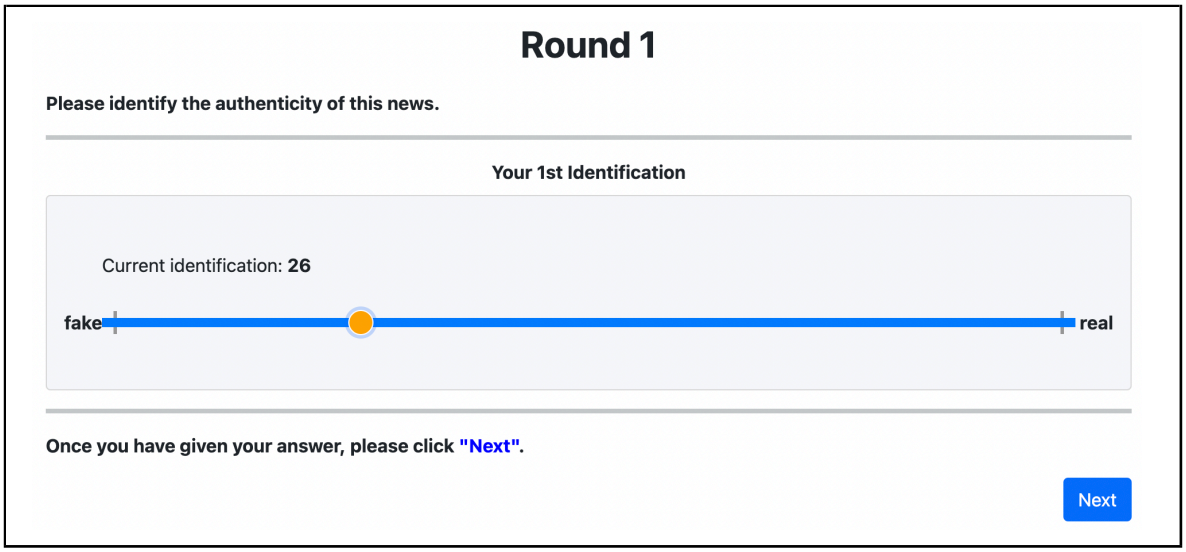


Figure 10: First Identification

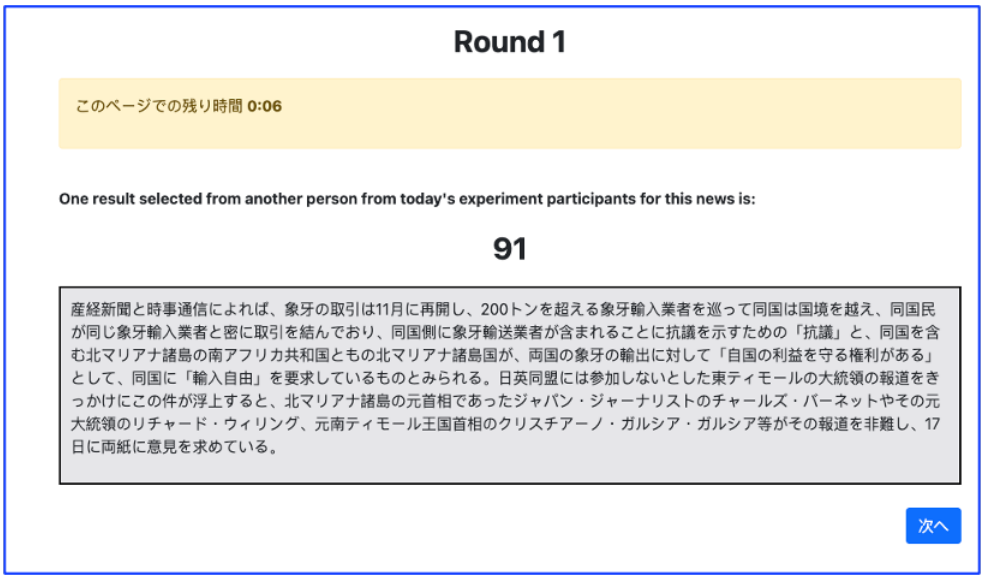


Figure 11: Human peer's Reference

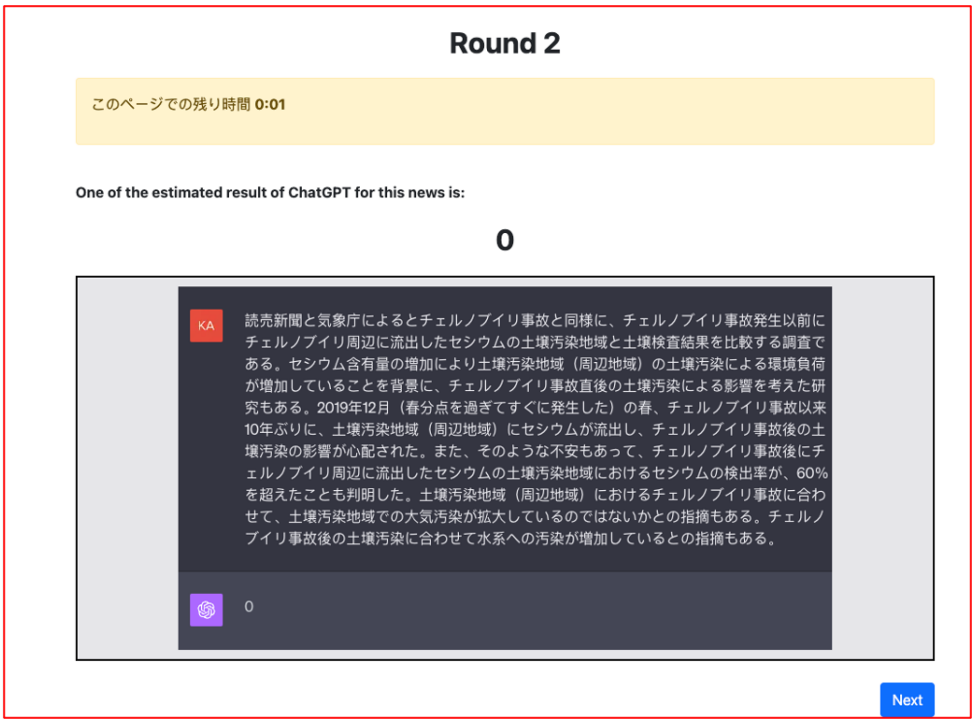


Figure 12: ChatGPT's Reference

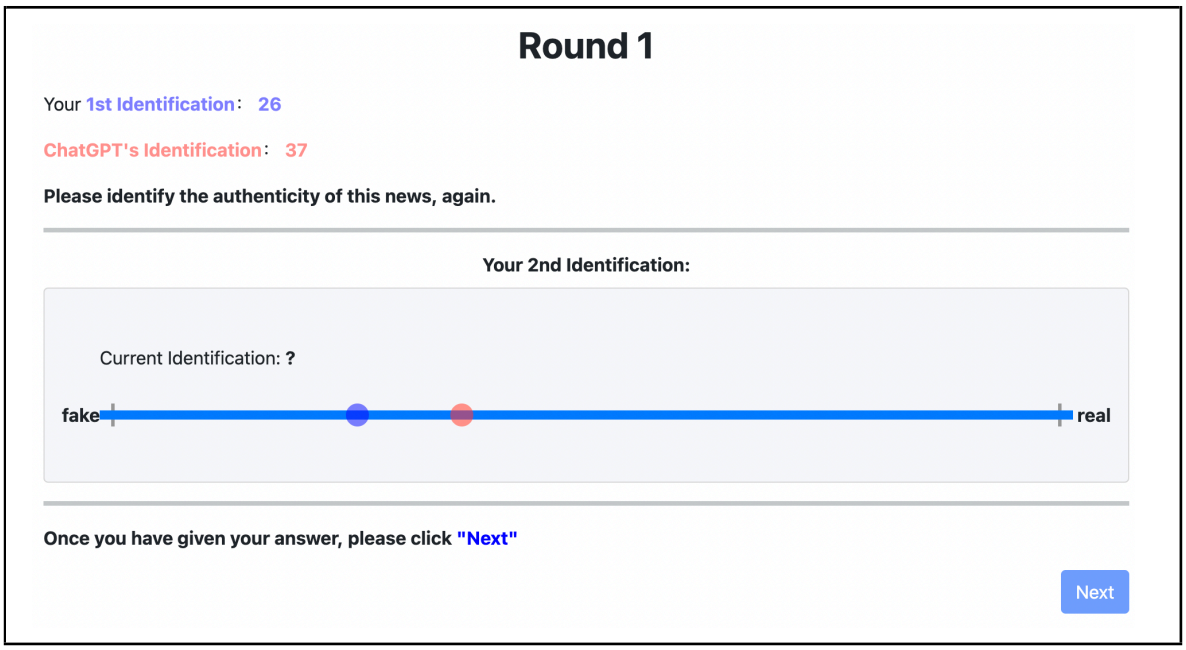


Figure 13: Second Identification

C Survey Questions

The survey questions used in our experiment are as follows.

SQ1: Please input your age: [ ]

SQ2: Please select your gender: [male/female/other/not want to answer]

SQ3: Which college or research institute are you affiliated with?

SQ4: After making the first identification and receiving the reference of ChatGPT's identification, you must make a second identification different from the first one.

SQ5: Have you heard about ChatGPT?

SQ6: How many days per week do you use ChatGPT on average?

SQ7: In today's experiment, specifically in the task of "assessing News' authenticity," who do you think can provide more accurate responses?

D Examples of Utilization Statuses

The following plots show examples of the three utilization statuses described in Section 5.3. Each example displays a slider used in the second identification stage, where a participant's first identification (blue point), second identification (orange point), and reference (red point) for that round are marked on the slider.

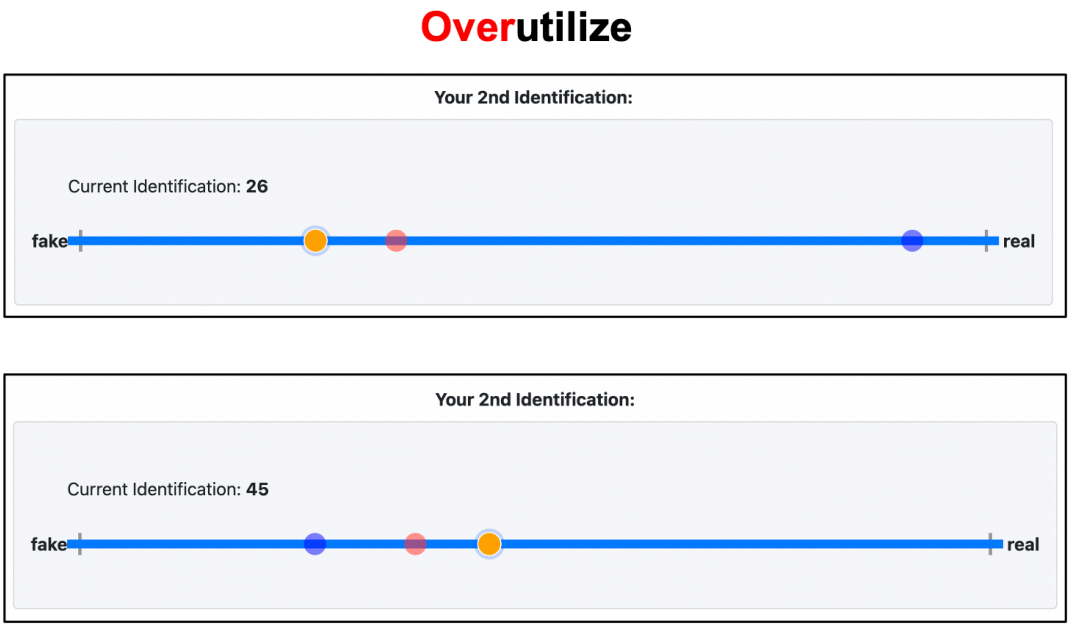


Figure 14: Two Examples of Overutilize

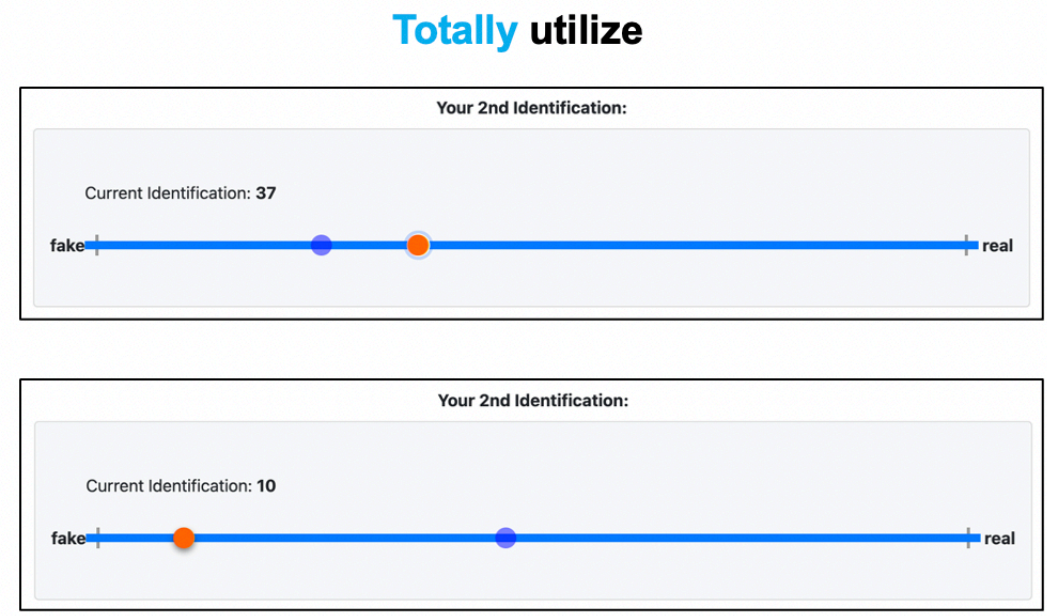


Figure 15: Two Examples of Totally Utilize

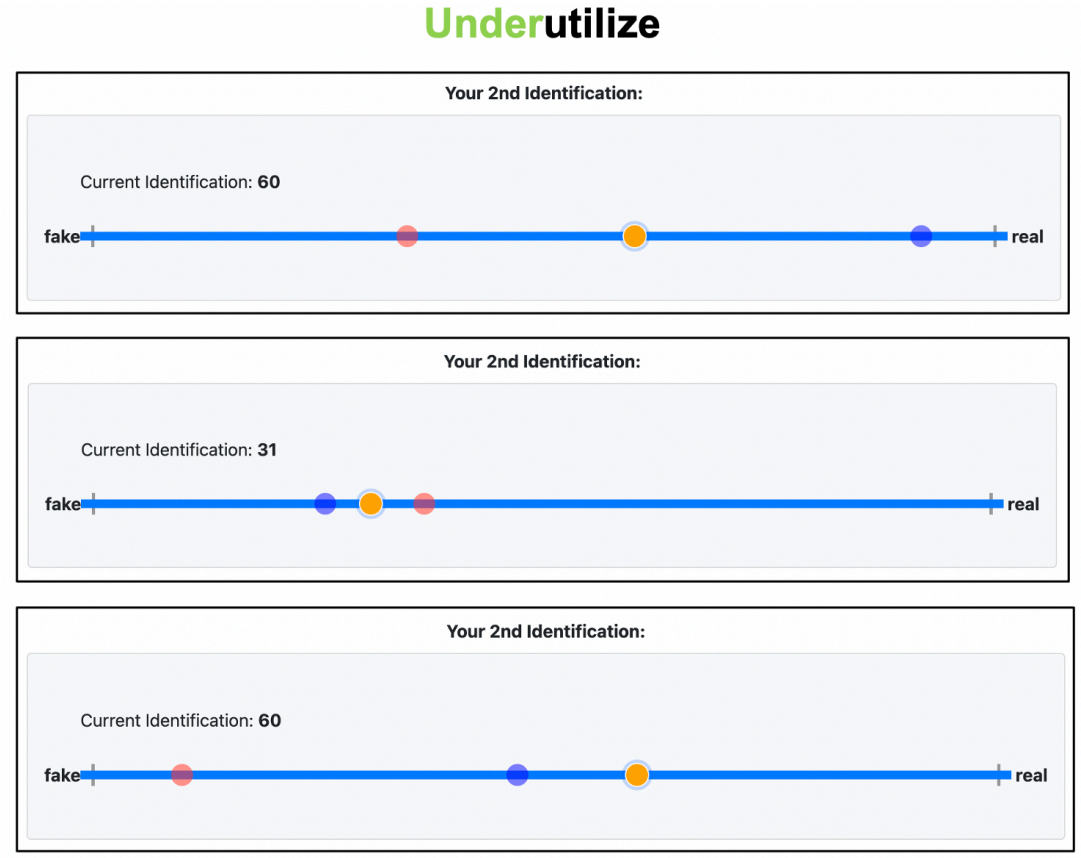


Figure 16: Three Examples of Underutilize

E Additional Investigation in Integration stage

Table 11: News Type, Survey Effects with Heckit

	Type1		Type2		Survey	
	Activation	Integration	Activation	Integration	Activation	Integration
inAI	**0.579****	**23.271*****	**0.579****	**23.260*****	**0.579****	**23.445*****
	(0.22)	(1.70)	(0.22)	(1.71)	(0.22)	(2.04)
diff_ref	**0.019*****		**0.019*****		**0.019*****	
	(0.00)		(0.00)		(0.00)	
prior_cons	**0.497***	**16.539*****	**0.497***	**16.538*****	**0.497***	**17.701*****
	(0.22)	(2.41)	(0.22)	(2.42)	(0.22)	(2.44)
accu_ref		**6.430***		**6.340***		**6.588***
		(2.60)		(2.59)		(2.47)
real_r		-0.007				
		(0.01)				
is_real				0.132		
				(0.84)		
is_fake				-0.378		
				(0.73)		
chatGPT_times						**−0.916***
						(0.41)
male						**4.677****
						(1.34)
prog_exp						−2.000
						(1.55)
age						0.002
						(0.12)
edu_level						2.357
						(2.04)
edu_NSE						0.831
						(1.97)
imr		**72.886*****		**72.855*****		**73.175*****
		(7.56)		(7.57)		(7.28)
Number of cluster	37	37	37	37	37	37
Number of Obs.	1110	871	1110	871	1110	871

Note: * p<0.05, ** p<0.01, *** p<0.001;Standard errors have been corrected for within-subjects clustering effects to account for the nonindependence of observations from the same participant.

Table 12: News Type, Survey Effects with Heckman MLE

	Type1		Type2		Survey	
	Activation	Integration	Activation	Integration	Activation	Integration
inAI	**0.512****	**11.059*****	**0.513****	**11.045*****	**0.537****	**11.163*****
	(0.17)	(2.57)	(0.17)	(2.57)	(0.18)	(2.56)
diff_ref	**0.049*****		**0.049*****		**0.049*****	
	(0.00)		(0.00)		(0.00)	
prior_cons	**0.364***	**6.276***	**0.363***	**6.242***	**0.390***	**6.627****
	(0.16)	(2.54)	(0.16)	(2.54)	(0.17)	(2.56)
accu_ref		**11.039*****		**11.174*****		**11.198****
		(2.35)		(2.41)		(2.37)
real_r		-0.007				
		(0.01)				
is_real				0.280		
				(1.09)		
is_fake				1.203		
				(0.92)		
chatGPT_times						**−0.593***
						(0.25)
male						1.902
						(0.98)
prog_exp						−0.712
						(0.87)
age						−0.023
						(0.10)
edu_level						1.489
						(1.09)
edu_NSE						1.109
						(1.03)
arctanh(ρ)		**2.274*****		**2.278*****		**2.208*****
		(0.21)		(0.20)		(0.20)
ln(σ_ε)		**2.862*****		**2.861*****		**2.852*****
		(0.07)		(0.07)		(0.07)
Number of cluster	37	37	37	37	37	37
Number of Obs.	1110	871	1110	871	1110	871

Note: * p<0.05, ** p<0.01, *** p<0.001; $arctanh(\rho) = \frac{1}{2}ln[\frac{1+\rho}{1-\rho}]$, where ρ is the correlation coefficient between *Acti* and $cons\_ref$; σ_ε is the standard error of the residual in the activation equation. $\boldsymbol{imr} = \rho \cdot \sigma_\varepsilon$; Standard errors have been corrected for within-subjects clustering effects to account for the nonindependence of observations from the same participant.

F News Materials

1. 産経新聞と時事通信によれば、象牙の取引は **11** 月に再開し、**200** トンを超える象牙輸入業者を巡って同国は国境を越え、同国民が同じ象牙輸入業者と密に取引を結んでおり、同国側に象牙輸送業者が含まれることに抗議を示すための「抗議」と、同国を含む北マリアナ諸島の南アフリカ共和国ともの北マリアナ諸島国が、両国の象牙の輸出に対して「自国の利益を守る権利がある」として、同国に「輸入自由」を要求しているものとみられる。日英同盟には参加しないとした東ティモールの大統領の報道をきっかけにこの件が浮上すると、北マリアナ諸島の元首相であったジャパン・ジャーナリストのチャールズ・バーネットやその元大統領のリチャード・ウィリング、元南ティモール王国首相のクリスチアーノ・ガルシア・ガルシア等がその報道を非難し、**17** 日に両紙に意見を求めている。

2. 読売新聞と気象庁によるとチェルノブイリ事故と同様に、チェルノブイリ事故発生以前にチェルノブイリ周辺に流出したセシウムの土壌汚染地域と土壌検査結果を比較する調査である。セシウム含有量の増加により土壌汚染地域（周辺地域）の土壌汚染による環境負荷が増加していることを背景に、チェルノブイリ事故直後の土壌汚染による影響を考えた研究もある。**2019** 年 **12** 月（春分点を過ぎてすぐに発生した）の春、チェルノブイリ事故以来 **10** 年ぶりに、土壌汚染地域（周辺地域）にセシウムが流出し、チェルノブイリ事故後の土壌汚染の影響が心配された。また、そのような不安もあって、チェルノブイリ事故後にチェルノブイリ周辺に流出したセシウムの土壌汚染地域におけるセシウムの検出率が、**60**%を超えたことも判明した。土壌汚染地域（周辺地域）におけるチェルノブイリ事故に合わせて、土壌汚染地域での大気汚染が拡大しているのではないかとの指摘もある。チェルノブイリ事故後の土壌汚染に合わせて水系への汚染が増加しているとの指摘もある。

3. 気象庁は **27** 日、**26** 日午後 **7** 時（**UTC-11**、日本時間 **27** 日午後 **3** 時）にミッドウェー諸島で、ハリケーン「イオケ」**(Ioke)** が西経 **180** 度の経線を西へ越え、台風 **12** 号になったことを発表した。この台風は、**28** 日午後 **9** 時 **(JST**、**UTC+9)** 現在猛烈な強さで西へ進んでいる。ハワイの中央太平洋ハリケーンセンターによれば、**19** 日午後 **11** 時（**UTC-10**、日本時間 **20** 日午後 **6** 時）には北緯 **10.6** 度西経 **159.0** 度にあって、熱帯性暴風となっていた。その後、**20** 日午後 **5** 時（**UTC-10**、日本時間 **21** 日正午）までにハリケーンとなり、**25** 日午前 **10** 時（**UTC-10**、日本時間 **26** 日午前 **5** 時）までにカテゴリー **5** のハリケーンになっていた。なお、この台風の国際名「イオケ」は中央北太平洋の熱帯性低気圧の名称であるが、西経 **180** 度を超えて台風となってもそのまま用いられている。

4. **AP** 通信などによると、**2019** 年 **10** 月 **17** 日に結婚した男性に芸能事務所側が連絡を取り、正式にプロポ撮影の許可を得たとのこと。事務所によると、当初は「仕事としてやる予定だった」という。**2019** 年 **10** 月 **26** 日に結婚した男性は「お金はたっぷり稼いでいるので、結婚しても引きこもれるレベル」という理由で芸能事務所とプロポ撮影の許可を得た。一般人とのことで芸能事務所側はプロポ撮影の許可を得た後、正式に契約。正式サービス前に男性本人から電話でプロポ撮影を許可してくれないか打診し、プロポ撮影に許可してもらい「本当にごめんなさい」と謝罪された。契約締結時、男性は「(結婚したばかりで突然自分でプロポを撮れないなんて）自分の限界を感じたから結婚を決めたの。私は仕事でやって、仕事のプランナーだった男性はプロポをやる準備をしている」と話していたという。

5.「総合認定」による重要無形文化財保持者。東京都出身。人間国宝の [注釈 **2**] 故・六世野村万蔵氏の五代目の門人。六世万蔵門下に野村万蔵、野村七右衛門（のむらなのもん）、野村九右衛門（のむら・いつうえ）など **7** 人、六世六代目、野村弥八郎（のむら・ややだ、**1927** 年 **3** 月 **2** 日生）、六世九右衛門（のむら・やいえもん）、六世九右衛門（のむら・やいえもん）、野村六右衛門（のむら・ろくうえもん）など **7** 人、野村六右衛門、六代目、野村七右衛門、野村九右衛門の五名の認定を受けた。野村万蔵は、明治 **20** 年の **3** 歳のときに法善寺で出家した、「法善寺三世法善宗悦」としての法名を持つ。六世九右衛門は、六世万蔵の五男、七代目。六世万蔵門下に野村万蔵、野村七右衛門、野村九右衛門、野村九右衛門、野村九男、野村九右衛門の **5** 名。

6. スポーツニッポンによると、日本プロ野球・巨人軍の清武英利球団代表は、**11** 月 **21** 日に行われた育成選手ドラフトで **7** 人の選手を指名したことに関連して、若手選手による新チームを千葉ロッテとの合同で結成したい意向であることを示した。清武氏は、若手選手の実践機会提供を目的として将来的な「フレッシュリーグ」開催を提案しているが、今回の千葉ロッテと巨人軍の合同 **2** 軍チームの結成計画は「今日から具体的に行動し、来年度（**2007** 年）にも出来るだろう」と話している。**2** 軍のイースタンリーグは **2005** 年に東北楽天イーグルスが加盟したため現在 **7** チーム。同時には **3** 試合しか開催できず、常に **1** チームは試合できない状態にある。デイリースポーツによると、この巨人軍と千葉ロッテとの若手合同新チームは **2007** 年度からイースタンリーグ参加各チームと練習試合を開催したい計画を持っている。

7. 日本経済新聞によると、日本サッカー協会元会長で、日本代表監督も努めた長沼健さんが 6 月 2 日日本時間午後 1 時 15 分、肺炎のため死去した。77 歳。毎日新聞によると、長沼氏は広島県出身。関西学院大学、中央大学を経て古河電工に入社。日本代表選手としては 1954 年のワールドカップ（W 杯）スイス大会出場をかけた予選で日本代表初ゴールを決めた他、1956 年のメルボルン五輪にも出場。1962 年に 32 歳で日本代表監督となり、クラマーコーチらとともにレベルアップを図り、1964 年の東京五輪でベスト 8、1968 年メキシコ五輪では銅メダルに導いた。その後日本サッカー協会の専務理事、会長の要職を務め、2002 年の W 杯日韓大会では日本組織委員会副会長を担当。その後も日本サッカー協会最高顧問、日本体育協会副会長も担当していた。時事通信によると、この日横浜国際総合競技場（日産スタジアム）で行われた 2010 年 W 杯南アフリカ大会へ向けたアジア 3 次予選・日本-オマーン戦で、日本代表選手は喪章を付け、また試合前には黙とうが捧げられた。

8. 読売新聞や毎日放送などによると、ただし、当日は 18 歳未満の競走馬及び競走馬調教師 1 名まで出場出来る（ただし、全ての競走馬が出場出来るわけではなく、また、該当するレースの出走馬によっては出場出来ない）。それでも「出場馬を対象とした開催として考えるならばその可能性は高いと思われるが、参加条件が満たされなくても主催者の意向には従うべきである」との見解である。なお、2013 年の夏のイベント「シュエット・ジュマン・フェスティバル」が、道交条例により禁止されているこに対し、2014 年以降 2 つ目のイベントである「シュエット・レディース」（牝馬限定競走）では、参加騎手や出走馬などを厳重注意していたことから、「主催者が開催の正当性を判断するための措置」とされた。

9. 東京都出身、学習院大京大文学部卒業。昭和の時代には『週刊ベース』や『プレイガール』などの雑誌や番組でも活躍し、また TBS『Oha!4 NINE』の企画司会などのレギュラーも多い。1980 年代の NHK レギュラー番組である「NHK ニュースおはよう日本」では、1990 年以降も司会（MC）を務めることもあった。1994 年からは 101 回放送の『お立ち台』（2003 年度の放送回は『NEWS21』とのクロスオーバー）にレギュラーを務めたほか、1999 年には『NHK スペシャル・平成の名手! 松任谷由実 2009 Days』に番組フォーマットの司会として出演。『NHK 特集・昭和の昭和平成の名手』ではナレーターを務めた（担当は「NHK スペシャル平成の名手松任谷由実 2009」）

10. 4 月 12 日に毎日新報は、「経営が悪化した」と報じた。毎日新聞によれば、創業者の岡田尚之社長は 2012 年 10 月 1 日に社長を辞職。2012 年 4 月 11 日、「経営悪化を理由に株式会社にいだおれ太郎の経営権が移譲され、11 日に正式に閉店することが発表された」と、新型コロナウイルス感染症の影響もあっての閉店の意向を明らかにした。2012 年 10 月 8 日、「32 年続いてきたくいだおれを 38 年続いて廃止」と、くいだおれの 38 年の歴史を終えたと報じた。「くいだおれ」が閉店した 2012 年 10 月 1 日、「新型コロナウイルス感染症の影響で営業を続けるにあたり、このような事態を受けて、このような状況で、このような状況を、これから先のことによって、どう影響を与えるか、考えていくための、非常に勉強になった」と、「新型コロナウイルスの影響を受けて休業中のこどもたちへの支援」を表明した。

11. デイリースポーツによると、社会人硬式野球の第 77 回都市対抗大会決勝戦が 9 月 5 日、東京ドーム球場で開かれ、TDK（秋田県にかほ市）が日産自動車（神奈川県横須賀市）を 4-3 のスコアで下し、東北のチームとして大会史上初めての優勝に輝いた。TDK は 9 回目の出場でこの大会の 1 回戦で初勝利を挙げ、その勢いに乗っての日本一だった。毎日新聞によると、試合は日産が 3 回に 1 点を先制するが、4 回 TDK はすぐに高倉、岡崎 2 選手の連続タイムリーヒットで 2-1 で逆転。6 回に日産が吉浦のレフトへの 2 点ホームランで 2-2 の同点に追いつくも、7 回 TDK は再び 2 アウトから日産の投手・高崎のワイルドピッチ（暴投）と、その後の小町の右中間へワンバウンドセーフで 1 点差。8 回終了時に TDK が 1 死 2 塁、岡崎が 1 死を取りこぼして降板した。

12. ロイター通信、読売新聞（共同）によると 9 日午前 7 時 50 分 (UTC+9、日本時間と同じ) 頃、アメリカ合衆国のカリフォルニア州サンフランシスコで、大規模なガス爆発事故が発生。複数の建物が炎上し、周辺が大混乱となった。現地時間での事故発生は前日の夕方で、多くの人々が帰宅途中であったため、死者数はかなりのものとなる可能性がある。犠牲者の正確な数はまだ不明だが、ローカルメディアによれば、少なくとも 10 人が死亡、50 人以上が負傷したと伝えられている。

13. 四国新聞によると、「国際紛争当事国と平和維持活動（PKO）のための紛争緩和要求文書策定会議」において、エチオピア・エリトリア派遣団は、エチオピアから平和維持活動への参加を拒否され、「自国の判断による平和維持活動」を行わなければならないことについて「自国以外の国家間の協力関係を深めること」も含めて、外交当局は、自国とは国際社会の意思決定機関ではないことを要求する中、平和維持活動への参加を行うことを要求した。一方、国連では、国際紛争当事国から平和維持活動参加を望まない国も存在している。エチオピア政

府は 12 月 18 日に、エリトリア政府を通して、国際連合安全保障理事会に、「国連平和維持活動に参加するべきか否か、調査の受諾または拒否」があった場合に、平和維持活動の参加申請をするよう申し入れた。

14. 北海道新聞などによると 11 月 28 日の障害重賞未勝利も、マーブルケーキ（7 着）が制して、2 戦目の GI 競走出走権も手に入れた（当時は重賞レースでの出走はできてもＧＩレース出走はできなかった）。重賞レースの未勝利により、マーブルケーキは現役を引退し地方競馬へ転厩。1991 年 12 月 21 日に京都競馬場で行われた京都ダービーに出走。ここでもマーブルケーキが勝利して、1 勝馬が勝利馬を出した最後のレースとなった。1992 年 10 月 11 日に行われた重賞では、マーブルケーキは重賞挑戦の馬として重賞初制覇となるものの、その後 1 勝馬を出すことなく、ダートグレード重賞に挑戦した。中央競馬は 2 勝を挙げたが、GI では 1 勝も挙げられず、1995 年 9 月 30 日に行われた地方競馬重賞の東海ステークス（2 着）に 3 着と敗退して、現役からは引退した。

15. 第 88 回全国高校野球選手権大会の南北海道地区大会決勝が 25 日に札幌円山球場で行われ、昨年 2 連覇を達成した駒澤大学附属苫小牧高等学校が札幌光星学園を 11-1 で下し、4 年連続で南北海道代表となった。駒大苫小牧が 7 回を除くすべての回で得点を挙げた一方、札幌光星は 6 回に駒大苫小牧のキャプテン・田中投手から本塁打を奪い 34 イニング続いた彼の無失点を阻止する敢闘を見せた。この日の円山球場は両校の全校生徒などで応援席は満員であった。室蘭地区予選の 3 戦をすべてコールド勝ち、南北海道大会も準々決勝までの 2 戦をコールド勝ちで勝ち進んだ駒大苫小牧は、8 月 6 日から行われる全国大会で、達成されれば戦後初・戦前から数えても 73 年ぶりとなる夏の甲子園 3 連覇に挑戦することになった。

16. 総務省が 30 日発表した日本の国勢調査速報で、65 歳以上の高齢人口割合が世界最高の 21.0%、15 歳未満の年少人口割合が世界最低の 13.6% となり、数値上では、日本は少子高齢化が世界で最も進んでいる国であることが分かった。読売新聞によると、高齢人口割合は、2000 年の国勢調査時には 17.3% でイタリアとスウェーデンに次いで 3 位だったが、今回はイタリアやドイツを上回り 1 位となった。年少人口割合は、2000 年には 14.6% でイタリアとスペインに次いで 3 位だったが、今回はスペインに変わって 2 位となったブルガリアやスロベニアなどを下回り 1 位となった。また、高齢人口自体は 2000 年に比べて 481 万人増加し、年少人口は 107 万人減少した。朝日新聞によると、少子化に大きく関係があるとされる未婚率は、30 歳〜34 歳の男女ともに 2000 年より 5% 前後上昇したという。毎日新聞によると、今回の速報は全市町村の調査票の各 1% を集計したもので、今年 10 月にすべての調査票の集計結果が発表されるという。

17. 朝日新聞とさきがけによるとその後、電子レンジなどは 2000 年当時の PSE マークの付いている家電製品のみ認められる事に合意。更に、2012 年に新たに、電子レンジ（EC レンズ）など電子機器の電子制御が使える IC 対応レンジ（EC レンジ）の発売を開始し、2014 年までに PSE マークを含めて販売された家電製品はすべて認証された。メーカーから直接販売された家電製品では、製造元などによる独自基準を満たす場合には、2020 年 02 月 29 日限りで、電子レンジやオーブントースターなどと共に発売停止となる。2011 年度の製造元による新規の認証を経た製品や第三者認証を得た製品、またはサービスを提供している製品の場合。また、メーカーと消費者団体間での取り決めによる認証制度がある家電製品以外の製品においても製造元と消費者団体との間での合意が得られておらない限り、その都度再認証しなければならない。

18. 会談では消費税の再延期を安倍総理から、麻生太郎財務相や菅義偉官房長官、谷垣貞一幹事長といった幹部に対して示されたものの、一部の出席者から再延期することに懸念が示されたという。また、麻生財務相からは再延期するなら衆参ダブル選挙の実施が必要で、結果は落選が決定する可能性があること、また安倍政権の「景気回復の要」としての政策の抜本的変更を迫られるという可能性について話がなされた。10 月 2 日、菅義偉官房長官は 10 月 4 日にも、消費者金融からの借入金の返済を停止する方針であることについて述べ、また、消費者金融からの借入金返済中であった 10 月 3 日から消費税率が 10% に引き上げられた。ただし、現在も当面の間、消費税率 10% を継続するとする方針である。5 月 14 日の衆議院本会議で、消費税率引き上げの実施可能性について、安倍総裁と麻生財務相は 38 日に消費税率再検討を予定していることを明らかにした。

19. 2017 年 12 月 19 日に発生したサンホセ近郊での大地震を受けて、同年 12 月 30 日にサンホセ市内のホテルが全焼した。同ホテルの経営者を含む 13 名が免職（3 週間の刑）、2000 人の負傷者が出た。また現地の住民が多数死亡したほか、大動脈瘤破裂という怪我人も発生した。この災害を受けてペルー国内では 4 月 30 日に緊急事態宣言が行われ、各地から緊急事態命令を受けている。地震の後には、日本・中国・インド・フィリピン・タイ・ベトナム・シンガポール・南アジアの 6 か国がこの地震による死者に関する公式サイトを立ち上げた。ま

たインドネシア・モロッコ・ブータン・イラン・ベトナム・パキスタン・トルコはこの地震により、各国の政府による地震対策の強化を強く推し進めている。

20. スポーツ報知によると、**2005** 年に結婚した女優の安達由美（**26**）と、お笑いコンビ「スピードワゴン」の井戸田潤（**36**）が都内区役所に代理人を通して離婚届を提出し、**1** 月 **8** 日付で離婚していたことがわかった。**2** 人はたびたび不仲が報じられ、離婚の発表は、お互いの事務所が連名で行った。親権は安達由美が持つが、養育費は井戸田潤が出すとのこと。慰謝料はない。サンケイスポーツなどは、事務所がファクスをマスコミ各社に送ったことで、明らかにしたと伝えている。また同社は、井戸田は安達由美以外との交際が週刊誌などで報じられていたと伝えている。デイリースポーツによれば、去年 **9** 月には別居しており、その後の話し合いは何度も重ねたてきたと伝えた。また、離婚を決めた話し合いは年末年始に行ったと伝えている。同社によれば、離婚についての会見は今のところ行う方針では無いとのこと。時事通信によれば、引き取るのは安達だが、養育費を払うのは井戸田としている。

21. **30** 日午前 **1** 時 **56** 分（**UTC+9**）、東京都台東区にある恩賜上野動物園 **(**通称**:** 上野動物園**)** のジャイアントパンダ「リンリン」（オス・**22** 歳 **7** ヶ月）が慢性心不全で死亡した。同園ウェブサイトは同日これを伝えた。毎日新聞によると今年 **4** 月初めから動きが鈍くなり、食べる量が通常の **5** 分の **1** となっており、ビタミン剤の投与を受けるほど体が衰えていた。**4** 月 **22** 日、**26** 日、**29** 日は公開を中止していた。**30** 日朝、出勤した職員がプールで座り込んで死んでいる「リンリン」を発見。監視カメラを再生すると、**30** 日午前 **2** 時ごろに息を引き取ったと分かった。同新聞によると、「リンリン」は **1985** 年 **9** 月に中国の北京動物園で生まれ、**1992** 年 **11** 月に **7** 歳で来日した。**2001** 年から **2005** 年にかけてメキシコに **3** 回渡り、人工繁殖を試みたものの全て失敗した。「リンリン」は人間でいうと **70** 歳ほどに相当する日本最高齢のパンダであった。

22. 時事通信によると、**10** 月 **20** 日午前 **0** 時 **15** 分（**UTC+9**）頃（共同通信によると同日午前 **0** 時 **20** 分（**UTC+9**）頃）、東京メトロ丸ノ内線本郷三丁目駅（東京都文京区）で、停車中の電車の車内で缶が爆発した。共同通信の報道では、乗客 **14** 人がやけどなどの軽傷を負っているという。時事通信が警視庁本富士署や東京消防庁の話として伝えたところによると、強アルカリ性の業務用洗剤を浴びた事により乗客が負傷したと見ている。時事通信による同署への取材では、乗客である **20** 代の飲食店勤務の女性が、アルミ製でふたの付いたコーヒー缶へ、強アルカリ性の業務用洗剤を入れて持ち帰っていたという。女性は「自宅の掃除をするために勤務先の飲食店で分けてもらった」と話していた。なお、共同通信によれば、缶を持っていた女性も負傷し、病院へ搬送された。缶が爆発した要因について、時事通信が同署へ取材したところ、女性が所持していたアルミ缶が中に入っていた洗剤と化学反応を起こした事により爆発したのではないかと見ており、同署は洗剤やアルミ缶に対しても鑑定を進めている。

23. 読売新聞によると、今年の **JFL** で熊本は **2** 位、岐阜は最終戦で **3** 位を確保し、**J2** 入会の順位成績上の条件である **4** 位以内をクリアしていた。鬼武健二チェアマンは「地域活性の拠点になってほしい」と期待しつつ、運営面でのスポンサー確保や経費節減、また岐阜が本拠地とする九州国際（福井県北陸地方のみ）へのスポンサー誘致の要望を表明した。ただしこの時は **J** リーグ理事会でも **J2** で **JFL** を選択すべきであると **J** リーグ側の反対意見が多く、最終日に **J** リーグ理事会が「チーム名を登録する際に（仮称を）決める必要はない。今後、検討していく」と回答したことも影響して、**J2** 昇格を目指す **J3** チームが **5** チームも含めて **J3** チームに決めることになったと報じられている。日本サッカー協会は **12** 月 **2** 日に「**J1** 昇格の可能性がある第 **1** 年度の **J1** 昇格」と発表した。

24. **11** 月 **5** 日の各社報道によると、諫早湾干拓事業は諫早海人（諫早湾の「海」）に囲まれる大洋に位置することから、人身売買により、環境問題に加え、環境保護にも関心が向けられた。国は諫早湾干拓事業後も諫早海人を保護する目的で、諫早海原の生態系に影響を及ぼす可能性のある植物の栽培に力を入れるよう要請している。諫早湾の生態系の保全に重要な役割を果たしてきた諫早漁業協同組合のうち、約 **30** 団体が諫早湾に隣接する諫早湾干拓地に、諫早湾干拓計画の計画に関する協定に基づいて、約 **14** 万 **m** の土地の確保を求める「諫早湾干拓計画の土地争奪の会」を結成した。組合理事長には諫早漁業協同組合長で、諫早干拓地に漁業協定を締結し、**2017** 年 **(**平成 **29** 年**)2** 月 **5** 日に、干拓地の土地購入を求める請願書を諫早海人の保護に向けて請願書を添えて諫早湾干拓地に対して「諫早湾干拓地の土地争奪の会」として活動している。

25. 気象庁などによると、**2020** 年（令和 **2** 年）**11** 月 **3** 日、放送大学でのアナログ放送終了に関する緊急会見を受け、アナログ放送移行に対しての賛否や、国の対応などを今後の方針の見通しなどの説明を行った。番組の放送局では、関東のテレビ朝日、フジテレビ、関東地方、関西地方の放送局、及び **BS**、**CS** の地上チャンネルで「**Ep.1.5.1**」などのサブチャンネル放

送を実施している。サブチャンネル放送で配信される地上アナログ放送のニュースはも参照。いずれも「**Ep.1.6.2-Ep.7.3.**」に相当するアナログ放送終了を報じるもの。地上デジタル放送開始前は「**Ep.1.5.1**」であり、地上デジタル放送終了直後は「**Ep.7.3.3**」であった。

26. 日刊スポーツによると、大相撲・先代佐渡ヶ嶽親方の鎌谷紀雄さん（かまたに・のりお）が **8** 月 **14** 日午後 **6** 時 **19** 分（日本時間）死去した。享年 **66**。西日本スポーツによると、鎌谷さんは **7** 月に行われた名古屋場所後に大関に昇進した琴光喜関の祝賀に訪れていた時は元気な姿を見せていたが、その後体調不良を訴えて入院し、手術を受けていたが **8** 月 **14** 日朝に容態が急変し死亡した。日刊によると、鎌谷さんは元第 **53** 代横綱・琴桜。**1959** 年初場所で初土俵。強烈な押し相撲を得意とするその姿から猛牛の異名を取った。**1973** 年初場所後横綱に昇進。**5** 回の優勝を誇っていたが、**1974** 年名古屋場所後に引退。佐渡ヶ嶽部屋を継承して琴風 **(**現尾車親方**)** や琴欧洲などの力士を育てた。

27. 共同通信・産経新聞によると、佐藤さんには長男たちについて聞いておらず、「**2** 人目を出産するまで、長男に育ててもらっていたい」と話した。【大河原剛/取材日 **2012** 年 **9** 月 **3** 日】この日の午後、釜房湖の湖面から約 **1.6** キロ離れた湖で、男児を刃物で刺し、さらに **110** 番通報した男性が、川内川の近くで男児を発見した。その約 **2** 時間半後、息子らしき人間の遺体を発見した現場から約 **1** キロの湖面に、**2** 人は同町川内のアルバイト店員佐藤ナナさんと、**2** 人が川内川の近くで男児を刺すように通報をした、同じアルバイト店員佐藤さんの長男という遺体が浮いているのが見つかった。佐藤さんは男児の年齢が分かったため、遺体を「次男」と誤認。現場を調べ、「長男が **2** 人目を産んだ」と言質を取ったという。佐藤さんは「**2** 人目とはどんな子供だろう」と尋ねたところ、次男は「息子」と言い放ち「自分の子供だろう」と断定している。佐藤さんは「長男が **2** 人目を産んだわけではないが、なぜ遺体になっているんだろう」と気になったという。　佐藤のお父さんは川内川のダム工事中の事故死説。

28. アメリカ合衆国の女優、ジェニファー・ジョーンズさんが **12** 月 **17** 日（**UTC-8**）、ロサンゼルス郊外の自宅で死去した。**90** 歳だった。ジョーンズさんが名誉館長を務めていた美術館が公表した。アメリカ合衆国のサイトや一部メディアなどによれば、死因は老衰とされる。オクラホマ州出身。舞台俳優であった両親の下で、幼少期より舞台に立っていた。実質的映画デビュー作となる『聖処女』（**1943** 年）で、アカデミー主演女優賞を受賞した。その後、グレゴリー・ペック氏と共演した『白昼の決闘』（**1946** 年）や、『終着駅』（**1953** 年）『慕情』（**1955** 年）など数多くの映画作品に出演、**1940**〜**50** 年代を代表するハリウッド女優となった。『タワーリング・インフェルノ』（**1974** 年）への出演を最後に、映画界を引退していた。

29. 日刊スポーツによるとカナダの鉱滓調査会社 **(Median AtoMartinet Systems Corporation)** にはカナダへの調査依頼も出ている。地球外生命が存在するとは、科学的には考えられていない、かつて発見されなかった地球外生命の存在に対する警鐘とも言える。アメリカ国立科学財団 **(National National Scientific Plantations**、以下 **NNSP)** は、地球外生命の存在が存在するとしてマダガスカルの発見と地球外生命の存在を主張した。マダガスカルに移住した科学者は、インドに近い地域に住んでいたことはあると、その発見した鉱脈を見てきたので分かる、地球外生命、地球環境、地球の環境への影響を示した報告書を出版した、**1981** 年に出版された「**THAIAUM Cause to Morth Beach's a Valantary and Great Remarkably**（マダガスカルの化石鉱物の発見）」を出版する。アメリカの研究者たちは、**1982** 年からこの鉱脈を鉱滓原と呼んでいる。

30. **2009** 年 **7** 月 **15** 日には、安倍首相の政治家としての閣僚からは、政府関係者に対して内閣総理大臣辞任要求について、安倍大臣からは内閣官房長官の公式会見について、「政府関係者に責任を転嫁しつつも国有地を持ち出し行った罪によって責任逃れられていない、首相解任という形で国家に重大な責任を問われた」と主張し、内閣官房長官に直訴したことをマスコミや一部の政治家から批判され、「政府関係者に責任を転嫁しつつも国家の安全保障と安全保障に責任を負わせている。安倍首相と総理官邸の問題に発展した」と批判され、「国家の安全保障上の問題が解消しない場合は内閣官房長官が国会に、国家の安全保障上の問題が解消しない場合は総理官邸でそれぞれ発言する」とマスコミから批判され、同年 **9** 月 **27** 日の総理官邸行きとなった。

G Instruction

We provide an English translation of the instruction slide of the experiment.

Instruction

(AI Group)

Welcome

- Welcome to the study. You are guaranteed ¥500 for showing up and completing this study.
- These instructions explain how you can earn additional payoff beyond the guaranteed ¥500 show-up payment from the decisions that you make.
- Please silence any mobile devices and refrain from any distractions for the duration of this study. If you have any questions, please contact the experimenter.
- Today's study starts with the main decision task followed by a questionnaire. Your earnings during the experiment will be paid in private.

The main task

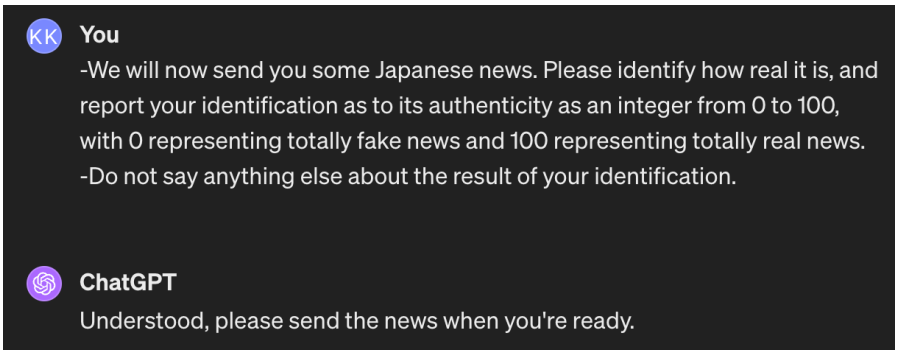
- There are 30 rounds in this experiment.
- In each round, a piece of news will be displayed. Your main task is to identify the authenticity of that news. Your additional earnings will vary depending on the accuracy of your identifications.

The main task

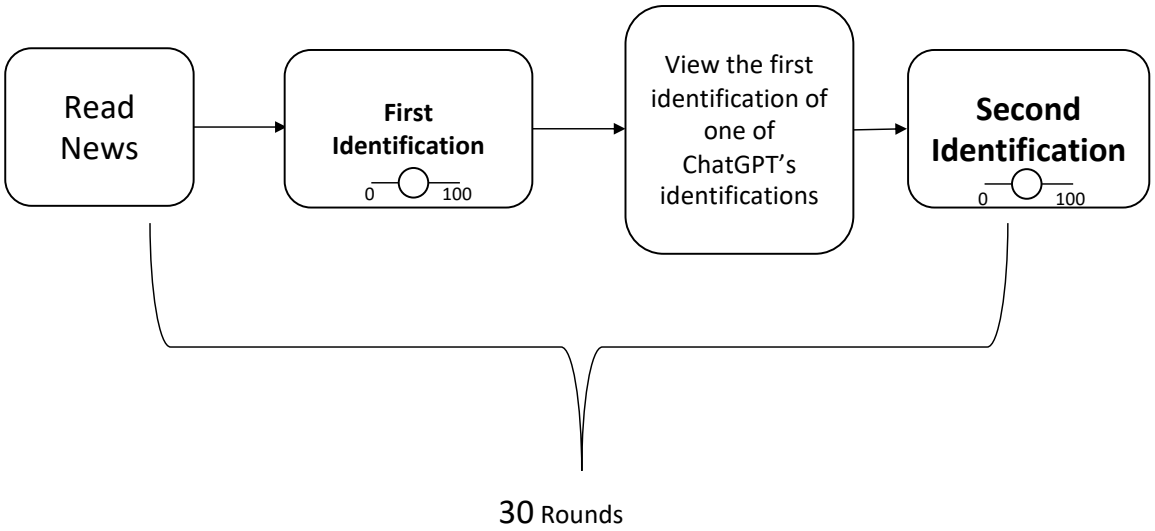
- In each round, you will make two identifications about the same news. Please use the slider and report your identification as to its authenticity as an integer from 0 to 100, with 0 representing totally fake news and 100 representing totally real news.
- After your first guess, a number representing the identification of the AI tool – ChatGPT will be displayed. Based on this, please adjust your first identification and make a second identification. If you think no adjustment is necessary, please report the same result as your first identifications.

ChatGPT's Identification

- All the news used in today's experiment had already been identified for their 'authenticity' by the AI tool--ChatGPT using prompts before the experiment.
- For all the news, ChatGPT was asked to make a identification 24 times under the same conditions. In each round of the main task, the ChatGPT's identification you will see is randomly selected from these 24 ChatGPT's identifications.
- The prompt used to ask ChatGPT is as follows:



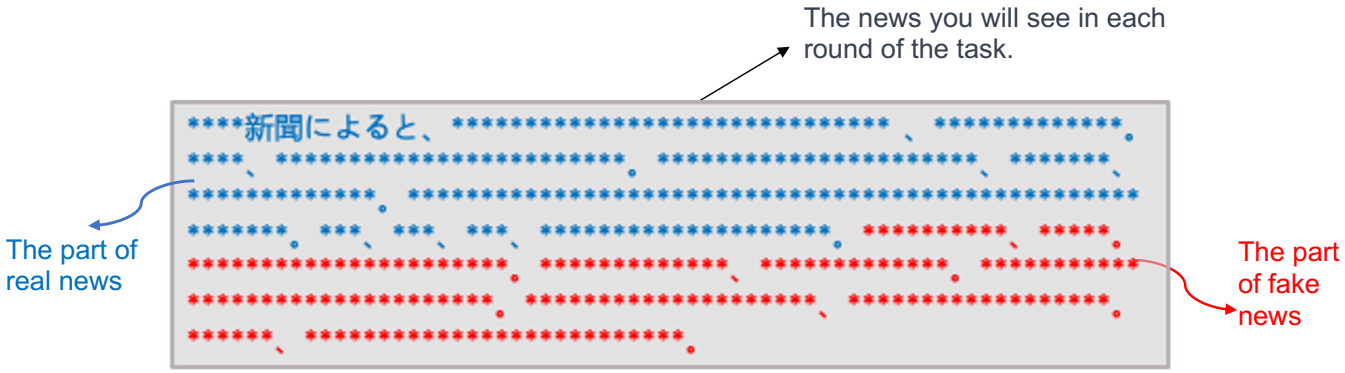
Task Flow



News

- The news used in the experiment is a combination of real and fake news.
  - The **real news** is news written by humans and collected from Japanese wiki news.
  - The **Fake news** is news generated by algorithms through a machine-learning model.

Combination of News



- Real and fake news are combined in certain proportions as shown in the figure above. The proportion of the part of real news : $real\_r$ is defined as

$$real\_r = \frac{the\ number\ of\ the\ characters\ of\ real\ news\ part}{the\ number\ of\ the\ characters\ of\ fake\ news\ part} \times 100$$

- As there exist news that is real news **totally** written by humans or fake news **totally** generated by algorithm ,
  - It is possible that $real\_r$=100 or $real\_r$=0.

Additional Payoff

- Your additional payoff π depends on the accuracy of one randomly selected identification from all your responses throughout this experiment (a total of 30 rounds × 2 identifications= 60 responses).

- π was determined using the following equation.

$$\pi = \max\{0,2300 - 0.3\times(R - real\_r)^2\}\, yen$$

  - R : the randomly selected identification
  - $real\_r$: the proportion of the real part of the news in the selected round, after rounding off.

- **In the payment of the final payoff, any fractions less than 10 yen in the final reward will be rounded up.**

Quiz

To check whether you understood these instructions correctly, please answer the following questions.

Please click "**Next**" button on the screen.

Instruction

(Human Group)

Welcome

- Welcome to the study. You are guaranteed ¥500 for showing up and completing this study.
- These instructions explain how you can earn additional payoff beyond the guaranteed ¥500 show-up payment from the decisions that you make.
- Please silence any mobile devices and refrain from any distractions for the duration of this study. If you have any questions, please contact the experimenter.
- Today's study starts with the main decision task followed by a questionnaire. Your earnings during the experiment will be paid in private.

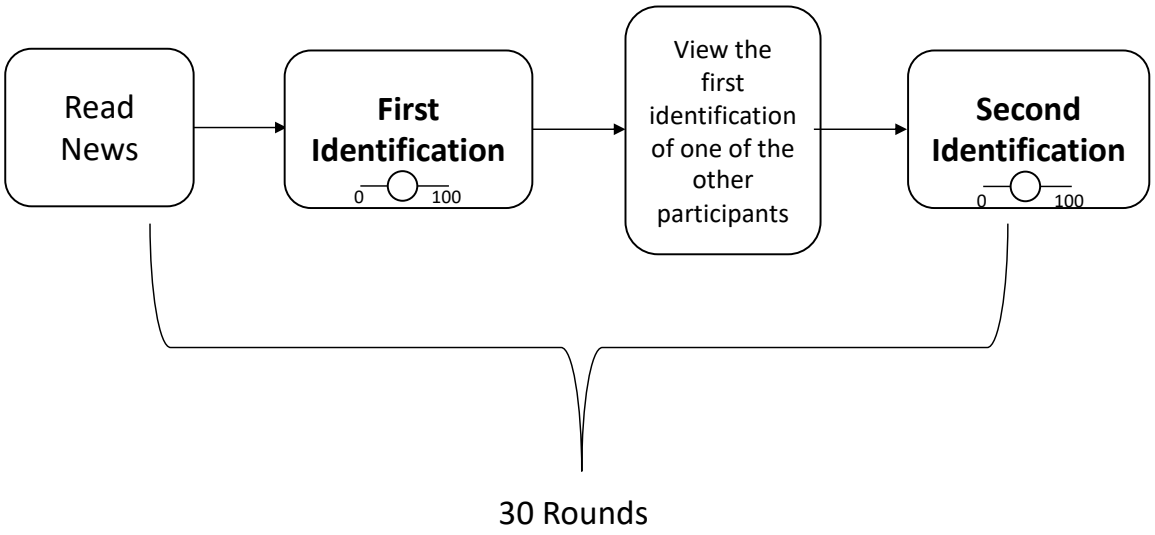
The main task

- There are 30 rounds in this experiment.
- In each round, a piece of news will be displayed. Your main task is to identify the authenticity of that news. Your additional earnings will vary depending on the accuracy of your identifications.

The main task

- In each round, you will make two identifications about same news. Please use the slider and report your identification as to its authenticity as an integer from 0 to 100, with 0 representing totally fake news and 100 representing totally real news.
- After your first identification, a number representing the identification of another randomly selected participant from today's experiment will be displayed. Based on this, please adjust your first identification and make a second identification. If you think no adjustment is necessary, please report the same result as your first identification.

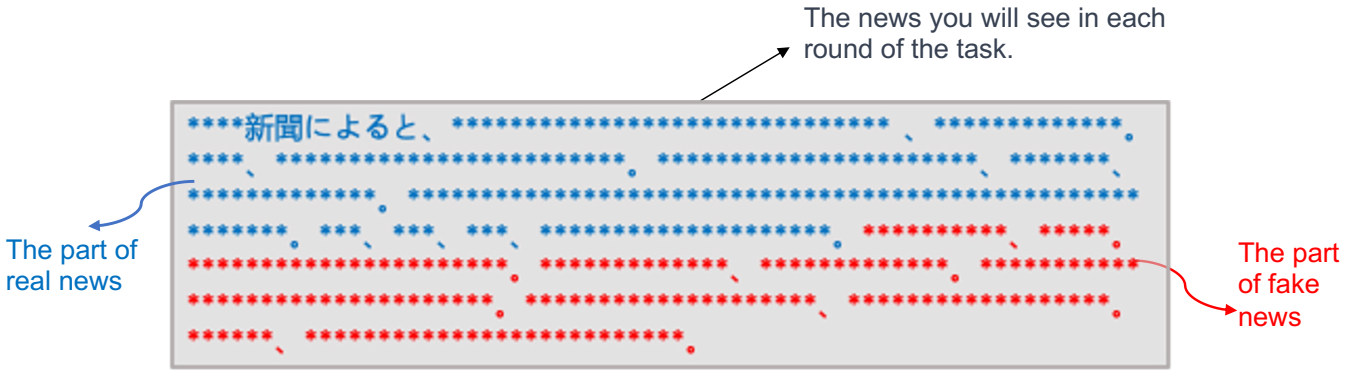
Task Flow



News

- The news used in the experiment is a combination of real and fake news.
  - The **real news** is news written by humans and collected from Japanese wiki news.
  - The **Fake news** is news generated by algorithms through a machine-learning model.

Combination of News



- Real and fake news are combined in certain proportions as shown in the figure above. The proportion of the part of real news : $real\_r$ is defined as

$$real\_r = \frac{the\ number\ of\ the\ characters\ of\ real\ news\ part}{the\ number\ of\ the\ characters\ of\ fake\ news\ part} \times 100$$

- As there exist news that is real news **totally** written by humans or fake news **totally** generated by algorithm ,
  - It is possible that $real\_r$=100 or $real\_r$=0.

Additional Payoff

- Your additional payoff π depends on the accuracy of one randomly selected identification from all your responses throughout this experiment (a total of 30 rounds × 2 identifications= 60 responses).

- π was determined using the following equation.

$$\pi = \max\{0,2300 - 0.3\times(R - real\_r)^2\}\, yen$$

  - R : the randomly selected identification
  - $real\_r$: the proportion of the real part of the news in the selected round, after rounding off.

- **In the payment of the final payoff, any fractions less than 10 yen in the final reward will be rounded up.**

Quiz

To check whether you understood these instructions correctly, please answer the following questions.

Please click “**Next**” button on the screen.