Assessing Authorship Influence: The Effect on Persuasiveness and Engagement with Government Foresight Reports

Lukas Zumbrunn[1]* and Ali A. Guenduez[1]

1University of St. Gallen, Dufourstrasse 40a, 9000 St. Gallen, Switzerland

* Corresponding author: Lukas.zumbrunn@unisg.ch

Version for EGOS 2024 Conference. Please do not cite or distribute.

Abstract:

This study investigates the impact of authorship on the convincingness and willingness to read government foresight reports. Using an experimental design, this research examines how perceived and actual authorship (human experts vs. generative AI) influence these factors. Thereby, three questions are being addressed: (1) How does the perception of authorship affect convincingness? (2) How does actual authorship influence perceived convincingness? (3) How does convincingness impact the willingness to read the full report and its perceived relevance?

Our findings show a significant interaction between perceived and actual authorship. Reports perceived to be written by AI but actually authored by experts were less convincing, while reports perceived to be written by experts were more convincing. Higher perceived convinc-ingness correlated with a greater willingness to read the full text. Hence, the present research highlights biases and trust issues in AI-generated content, emphasizing the importance of align-ing perceived and actual authorship to enhance government communications.

Keywords:

Generative AI, Government Report, Government Communication, Foresight, Foresight Report, Public Foresight

1 Introduction

Governments often must communicate in highly uncertain times and on critical topics. These include, for example, the consequences of nuclear waste management (Andreiadis et al., 2019) or measures during a global health crisis (Hyland-Wood et al., 2021). Such topics must be communicated to citizens effectively so that core messages are well received. This also includes that citizens are convinced by the communication despite high uncertainty (Van Der Bles et al., 2020).

In recent years, Generative Artificial Intelligence (GAI) is increasingly being used to create and edit content, as it offers many advantages for example in journalism and science. It even produces material that can be crucial for electoral decisions (Bowers & Glenday, 2021; Fuerth & Faber, 2012; OPSI, 2022). GAI has the ability to create headlines, text, images, and spot mistakes in manuscripts (Dwivedi et al., 2023). Given this recent rise of interest in the field, this phenomenon was paired with research on its impacts and challenges in governance (see, for example, Margetts & Dorobantu, 2019; Selten et al., 2023; Straub et al., 2023). Present studies show, for example, that headlines labeled as AI-generated are less likely to be believed and shared, even when they are true or generated by humans (Altay & Gilardi, 2023).

This leads to the raising of the question of how the use of GAI impacts the convincingness and relevance of government reports. All across the world, public sectors already benefit from implementing GAI in their work, and potential impacts are already visible (Mökander & Schroeder, 2024; Nalbandian, 2022; Sun & Medaglia, 2019). However, research on this topic is highly new and sparse.

Building on this emerging line of research, this research aims to foster the understanding of this challenge for government reports. Using an experimental research design, we investigate the extent to which the convincingness of governmental foresight reports depends on whether they are written by humans or GAI. With this backdrop, we aim to answer the following research questions:

RQ_1: How does the convincingness of government foresight reports differ between government experts and GAI as perceived authors?

RQ_2: How does the convincingness of government foresight reports differ between government experts and GAI as actual authors?

RQ_3: How does higher convincingness affect the assumed relevance of the report?

Present study makes the following contributions. First, it investigates the perception of readers towards headlines and content (assumably) written by GAI. Second, it further develops the idea of convincingness in the realm of future-oriented texts with the example of governmental foresight reports. Third, it emphasizes the importance of recognizing the potential and drawbacks of the use of GAI in future applications of foresight when communicating to citizens.

2 Theory

2.1 Generative Artificial Intelligence

GAI works with data on the world population. Combined with multi-agent simulations, Artificial Intelligence (AI) is assumed to be able to make similar predictions compared to experts (Rao & Firth-Butterfield, 2020). On the other hand, recent studies have shown that news generated by AI are not perceived as convincing as news generated by humans (Longoni et al., 2022). It does not achieve to convince humans to the same extent as other texts. This human skepticism remains even if only the headlines are being labeled as generated by AI, despite being true and written by humans (Altay & Gilardi, 2023). This derives our first hypothesis:

H_1: Participants who are informed that a government report was written by experts will tend to attribute higher convincingness to the report than those informed that it was generated by Artificial Intelligence.

The human input is crucial for current-day government reports. Frequently, a collaborative approach is chosen in the development of relevant content for the report, combining the public with experts on a specific subject matter (Amanatidou & Guy, 2008, p. 544). The final write-up and implementation of reports on public projects remain with the public administration (Paliokaitė & Sadauskaitė, 2023). Therefore, government reports are a suitable research subject. It is a domain that is currently dominated by human authority. Nevertheless, the ability to work with a large amount of data and recognize patterns not visible to humans makes Artificial Intelligence a rapidly improving competitor for humans involved in the creation of government reports, especially due to its capability to interpret and find patterns in large data sets. Also, as research shows, when the content produced by GAI aligns with the individual judgment of public servants, they see an increase in convincingness in the artificial recommendations (Selten et al., 2023). Following this, we arrive at the second hypothesis:

H_2: H_1 is also true independent from the actual authorship of the report the participants are asked to read.

Additionally, two covariates ought to be considered. First, trust in government significantly influences how citizens perceive the convincingness of government communications. Higher trust in government institutions leads to a greater acceptance of information disseminated by these bodies (Chohan et al., 2021). As both reports, independent from the author, replicate or aim to replicate a governmental communication, this control variable influences H_1 and H_2.

Second, previous research shows that people tend to rate content as being more convincing when the findings conform with the already-held beliefs of the individual (Jurma, 1981). Also in recent studies, especially given the rise of GAI, the question of agreement with generated content and its connection to trust has received attention (see, e.g., Longoni et al., 2022; Selten et al., 2023). Thus, present study also measures the participants' levels of agreement as a control variable for both H_1 and H_2.

2.2 Relevance of the report

There is a further element that must be taken into consideration. Governments have to consider expectations of citizens towards how the respective state has to work and what should be provided via public resources substitute for private sector goals (Batalli, 2011). This could either be seen as an advantage or a challenge for both foresight projects and the implementation of GAI in them.

An important measure for the perception of a government report is relevance. Relevance includes that the report fosters a personal and meaningful connection with the individual reading the text (Priniski et al., 2018). This means that the report is actually of interest for the people who read it, thereby enhancing the impact the results have. Similarly, the results of Selten et al. (2023) highlight that the inclusion of relevant information is a factor influencing the decision to follow recommendations. Hence, the third hypothesis is:

H_3: Higher convincingness – as per H_1 and H_2 – leads to a higher willingness to read the full text, indicating a higher assumed relevance for citizens.

Based on the literature review, the main research model we analyzed included the following propositions:

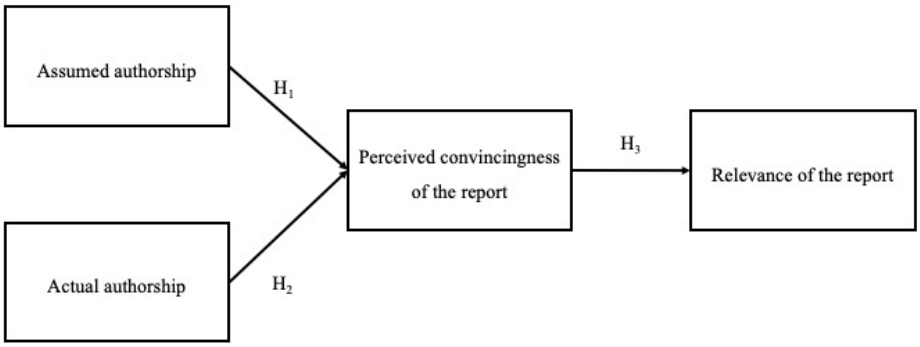

Figure 1: Research Model

2.3 Foresight

Governments publish a plethora of reports. One concrete example is the publication of results of foresight projects. Such, so called foresight reports are the exemplary items with which we assess the three hypotheses. They are the main way in which governments communicate the results of their foresight projects (see, for example, "We the UAE 2031", Singapore 2019 Report, UK Government Office for Science reports). These reports are predominantly written, and the projects are conducted by experts within the public administration.

In recent years, the field of foresight in the public sector has received increased attention in both research and practice (Bowers & Glenday, 2021; Fuerth & Faber, 2012; OPSI, 2022). To conduct foresight, two key sources must be given (Bolger & Wright, 2017). The first is data, which can be in a qualitative or a quantitative form, or in a mixed form. The second is the access to experts, who know how to work with and interpret the given data.

Foresight reports are suitable for investigating the underlying questions of this paper for mainly three reasons. Firstly, the rise of GAI has had a profound impact on the interpretation of the second key source of foresight, which, until recently, was a purely human domain. With the development of GAI, however, there is a possibility to exchange human experts with the capabilities of Artificial Intelligence (Noy & Zhang, 2023; West et al., 2023). The capabilities raise novel questions and challenges to governments publishing such reports.

Secondly, they tackle topics of high uncertainty and with potentially high impact on the country and its citizens. Thus, also the concept of convincingness is also not new in the field of foresight report. Haasnoot et al. (2018, p. 277) conceptualize "convincibility" of critical values in the evaluation of foresight reports as the consistency "with perceived change or other information already available, when they trust the source as authoritative, and/or when the assessment process has been according to scientific standards (trustworthiness of source and procedures)".

Lastly, compared to other reports such as evaluations of previously implemented policies, they actively communicate the perception of the government on issues related to the future. Again,

this is important for citizens, as political decisions predominantly affect what is yet to come (Birch, 2023). In this context, foresight and its reports are crucial tools helping individuals and organizations making sense of uncertain futures and taking concrete measures in preparing despite uncertainty (Piirainen & Gonzalez, 2015).

3 Experimental methods and measurements

The two independent variables are *IV_1: Information on authorship* (corresponding to H_1); and *IV_2: Concrete content of the written abstract* (corresponding to *H_2*). In other words, this means, that the stimuli for the experiment participants will be the claimed authorship of the report as well as the content that is written in the abstract. This builds on the previous research by Altay and Gilardi (2023). For the dependent variable, *DV: Perceived convincingness of Foresight report* is chosen. It is the measurement of the answers by the participants who have completed the experiment, as they will be asked to rate how convincing they think the foresight predictions would be.

The experimental design follows a two-by-two matrix approach, including both independent variables in the experiments. This helps to integrate the possibility of an interaction between the independent variables, and by measuring groups 1-4, this can be further analyzed in the design as presented in Table 1.de

Table 1: Groups in Experimental Design

	GAI written Foresight Report	**Expert written Foresight Report**
Participants told it is an Expert Report	Group 1	Group 2
Participants told it is a GAI Report	Group 3	Group 4

3.1 Procedure of the experiment

By highlighting in red the authorship of the reports, the emphasis is clearly put on the authorship of the respective report. To underscore that the extract of the report was written by GAI, the introduction text will include a note that provides the participants with the concrete prompt that the AI was working with. Thereby, the participants know the origin of the reported results and how they were produced. Of course, an exact reproduction is not possible, given the nature of GAI outputs. Nevertheless, given that GAIs such as ChatGPT have now passed the Turing Test (Biever, 2023), this information will probably not influence the perception of the quality of the text itself. It will, importantly, still underscore the potential and capabilities of the GAI at use.

3.2 Data collection

Data collection was done via an online experiment conducted on Amazon MTurk on 10 April 11, 2024. The recruiting intended to have 450 people conduct the survey. The criterion for selection via MTurk was that the worker has a HIT Approval Rate[1] of above 90%. This is amongst the most common qualifications used in research (Karpinska et al., 2021).

In fact, a total of 479 people clicked on the survey, of which 443 have finalized the experiment. Demographic-wise, only 204 participants declared their gender, of which 79 identified themselves as female and 125 as male. The age was collected in seven age groups, (<25, 25-34, 35-44, 45-54, 55-64, 65-74,75+). Only people over 25 and under 75 years old have indicated their age. The distribution can be found in the following table.

Table 2: Distribution of Age Groups of Participants

Age group	**Frequency**	**Percent**
25-34	70	22.8
35-44	34	11.1
45-54	16	5.2
55-64	5	1.6
65-74	1	.3

Also, participants had to give initial consent concerning data protection at the start of the experiment and informed consent after the experiment. 13 participants decided to withdraw their consent, resulting in the N=430.

3.3 Ensuring validity in online surveys

Online surveys and experiments have been noted as having a low quality of their data (Douglas et al., 2023). One explanation that has been frequently noted is a possibly higher distractions of online participants (Necka et al., 2016). Nevertheless, despite such challenges, MTurk is still

[1] A human intelligence task (HIT) refers to a question the participants have to answer. The approval rate is the ratio of completion to the standards of the researchers by the participants, and therefore considered an index of quality.

considered to be a viable source for online experiments, even in comparison to other tools and even in-person experiments (Kees et al., 2017). The same can be noted in the specific case of research in public administration (Stritch et al., 2024).

When using such tools, there are several methods that enable researchers to ensure the attentiveness of participants in an experiment, one of which is the task to recall essential information (Abbey & Meloy, 2017). This type of "memory recall" is frequently used to improve the validity of the data, with the ability to ensure retention and awareness.

Thus, the participants all had to pass an attention check in order to be viable for analysis. The attention check ensured that the participants understood the core elements of the task. In present experiment, participants had to correctly indicate the author that was displayed on the text they read right before the attention check. This further reduced the N to 307 with participants not passing the attention check. This corresponds to 29.61% of participants not being able to pass the attention check.

Previous research indicates that the attention of MTurk participants can go in both directions, either outperforming or underperforming compared to other control groups (Goodman et al., 2013; Hauser & Schwarz, 2016). Therefore, it is not possible to say whether the present results are indicative of systemic issues of the empirical approach. What they do underscore, however, is the importance of ensuring the quality of the results via the use of attention checks.

3.4 Operationalization and data analysis

In the following, a short operationalization of the variables for the experiment will be presented. In this study, the operationalization of our *IV*s involves distinctly categorizing the authorship of reports into two groups: human experts and GAI. This categorization ensures a clear differentiation between reports authored by experts and those generated by AI systems.

Meanwhile, the operationalization of the *DV* centers on measuring the level of convincingness of the abstract of the reports, quantified on a scale from 1 (representing full convincingness) to 5 (indicating a lack of convincingness). This scale allows for a nuanced assessment of how the perceived authorship influences the perception of the convincingness among readers or participants.

To construct the study stimuli, we use abstracts of foresight reports, focusing on the same topic. Abstracts often are a decisive factor whether people read the full version of the text or not, hence implying the relevance of the text for the reader. These abstracts form the core stimuli of the analysis– one authored by human experts and the other by GAI.

As for the data analysis, one- and two-way ANOVAs and a correlation analysis are used. For H_1 and H_2 individually, they are analyzed using a one-way ANOVA. However, it is still assumed that there could be an interaction effect between the *IV*s in H_1 and H_2. Hence, we see the two-way ANOVA as the most appropriate methodological tool, which is most used to measure in two variables between-subject designs. For H_3, a correlation analysis and, to increase reliability, a (one-way) Welch ANOVA were used to analyze the results.

4 Results

In the following section, we will discuss the results of our analysis. The table below provides an overview of the number of participants in our study. Overall, the results are based on N=307 participants. According to our experimental setup, the distribution is as follows:

The row labeled "Perceived Author" in Table 3 indicates the number of participants who, at the beginning of the experiment, were informed that the author was either "Government Experts" or "Generative AI." Specifically, 141 participants were told that the author was "Government Experts", and 166 participants were told that the author was "Generative AI."

Table 3: Descriptive Statistics of Perceived Authorship on Convincingness

Perceived Author	**Mean**	**Std. Deviation**	**N**
Government Experts	2.04	1.061	141
Generative AI	2.12	1.043	166

The results indicate that the perceived authorship alone does not significantly influence the convincingness of the results presented in a foresight report (p = 0.5, see Appendix Table 1). This is true for both texts perceived to be written by "Generative AI" as well as for "Government Experts." However, all defined covariates have a significant impact on the model. Additionally, it must be noted that experts still score slightly higher (mean 2.04 and 2.12, respectively).

Continuing with the assessment of H_2: The row labeled "Actual Author" in Table 4 indicates who actually wrote the text. For 157 participants, the text was authored by "Government Experts", while for 150 participants, the text was authored by "Generative AI."

Table 4: Descriptive Statistics of Actual Authorship on Convincingness

Actual Author	**Mean**	**Std. Deviation**	**N**
Government Experts	2.06	1.066	157

Generative AI	2.11	1.037	150

Like the results for the perceived author, the results indicate that the actual author alone also does not significantly influence the convincingness of the results presented in a foresight report. Furthermore, the results seem to be similar also in the closeness of perceived convincingness. However, the results show a higher level of significance, being significant at the 0.1 threshold ($p = .064$, see Appendix Table 2).

We further investigate the assumption of an interaction effect between the independent variables perceived and actual authorship. Therefore, in the following, the combined effect will be further examined using a two-way ANOVA analysis. The results of this analysis presented in Table 5 provide insights into the effects of the independent variables and the covariates and their interactions on the perceived convincingness of the findings. The results are in favor of the assumption that the independent variables of perceived and actual author have an interacting effect on the results of the dependent variables.

Overall, the ratings of the convincingness of the Foresight Reports were similar close to 2, regardless of the perceived author, indicating that the texts were generally perceived as convincing. The descriptive statistics, which can be found in Appendix Table 3, reveal that when participants believed that the text was authored by "Government Experts", they found the results less convincing if the text was in fact written by "Generative AI" (Mean = 2.22) compared to "Government Experts" (Mean = 1.86). Conversely, when participants believed that the text was authored by "Generative AI," they found the results more convincing if the text was actually written by "Generative AI" (Mean = 2.00) compared to "Government Experts" (Mean = 2.23). In summary, the findings suggest that the perceived convincingness of the reports varies depending on the perceived and actual author. Participants tend to find texts written by "Generative AI" more convincing if they believed they were authored by "Generative AI," and texts written by "Government Experts" more convincing if they thought they were authored by "Government Experts."

Table 5: Results of Two-Way ANOVA

	df	**Mean Sq**	**F**	**Sig.**
Corrected Model	5	27.875	42.286	<.001

Intercept	1	13.637	20.687	<.001
Trust in national government	1	44.850	68.036	<.001
Agree with findings	1	14.918	22.631	<.001
Perceived Author	1	.198	.301	.584
Actual Author	1	1.777	2.696	.102
Perceived Author * Actual Author	1	2.820	4.279	.039

a. R Squared = .413 (Adjusted R Squared = .403)

The two-way ANOVA results show that the adjusted model explains a substantial portion of the variance in the perceived convincingness of the findings (Adjusted R Squared = 0.403). Both covariates, trust in the national government and agreement with the findings, are significant predictors, while the perceived and actual authorship individually do not have significant effects. However, a significant interaction exists between perceived and actual author, indicating an interplay between these factors in influencing perceived convincingness. Figure 2 further visualizes this interaction.

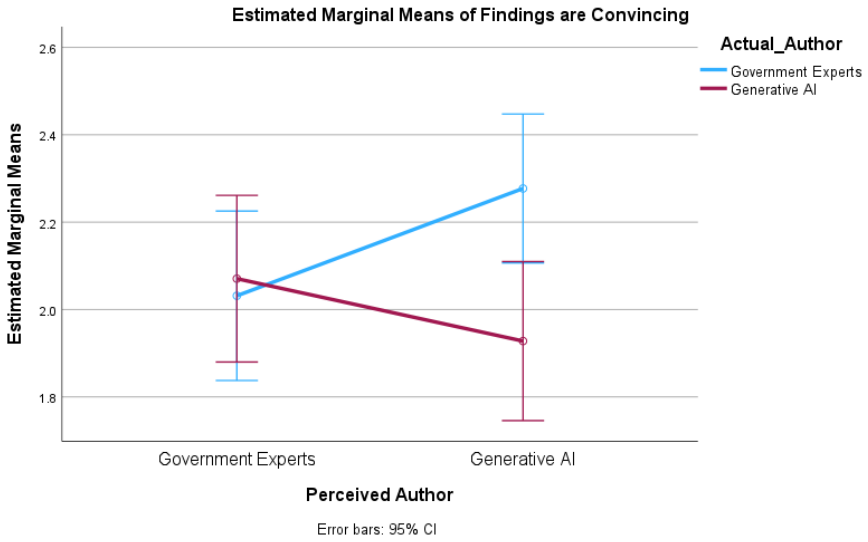

Figure 2: Interaction Diagram of Perceived and Actual Authorship

The plot highlights that the convincingness of the findings is influenced by the interaction between perceived and actual authorship. Specifically, texts perceived to be written by "Generative AI" but actually authored by "Government Experts" tend to be rated as less convincing.

Overall, these results underscore the complex interplay between perception and actual authorship in shaping participants' evaluations of the text's convincingness. The interaction plot illustrates that the alignment between perceived and actual authorship directly impacts the perceived

convincingness of the report. When the actual author aligns with the perceived author, participants rate the findings as more convincing.

In the next analysis step, we examined the link between the report's perceived convincingness and the results' relevance, as described in H_3. We employed the Welch ANOVA to ensure robust analysis.

Our findings indicate a clear positive relationship between the perceived convincingness of the text and the desire to read the full text. Specifically, the mean interest in reading the full text decreases together with the perceived convincingness (1 being the highest, 5 the lowest). These results suggest that participants are more likely to engage further with the material when they perceive the initial findings as highly convincing.

Table 6: Impact of Convincingness on Willingness to Read the Full Text

Findings are Convincing	Mean	Std. Deviation	N
1	1.59	.733	103
2	2.13	.804	117
3	2.74	.955	57
4	3.39	1.145	18
5	4.42	.515	12
Total	2.22	1.063	307

The Welch ANOVA (Appendix Table 6) confirmed significant differences between groups, stressing the importance of perceived convincingness in motivating engagement with the material. Higher convincingness leads to greater desire to read the full report, supporting H_3.

Additionally, the results of the between-subjects effects test provide further validation. The corrected model is highly significant ($F(4, 302) = 51.003$, $p < .001$), explaining a substantial portion of the variance in the willingness to read the full text (Adjusted R Sq = .395). The significant effect of the convincingness variable ($F(4, 302) = 51.003$, $p < .001$, Partial Eta Sq = .403) highlights that the perceived convincingness of the findings has a large and significant impact on participants' willingness to read the full text. Thus, this indicates the high perceived relevance the participants give to the report.

5 Discussion and Conclusion

Our analysis highlights an interesting aspect of human perception and bias. The findings indicate that participants rated the Foresight Report as more convincing when the perceived and actual authorship aligned, particularly favoring texts actually authored by “Government Experts” when they were perceived to be written by "Government Experts." Conversely, when the text was perceived to be authored by “Generative AI,” it was rated as less convincing if the actual author was “Government Experts.” This suggests that participants may have higher expectations or a more favorable bias towards texts perceived to be generated by AI when, in reality, these texts are authored by established experts.

Additionally, the presence of overlapping confidence intervals when the perceived author is “Government Experts” and slightly overlapping intervals when the perceived author is “Generative AI” further supports the significance of this interaction effect. The significant interaction term in the ANOVA results ($p = 0.039$) highlights the necessity of considering both the perceived and actual authorship when evaluating the convincingness of government reports. This interaction indicates that the effect of the actual author on the perceived convincingness of the text depends on who the participants believe the author is. Specifically, the data show that when there is a mismatch between the perceived and actual authors, such as believing the text was written by “Generative AI” when it was actually authored by “Government Experts,” the perceived convincingness can vary significantly. This finding emphasizes the complex interplay between perception and reality in shaping opinions on the convincingness and persuasiveness of government information.

In summary, while individual effects of perceived and actual authorship are not statistically significant, their interaction reveals critical insights. The interplay between who the participants believe wrote the text and who actually wrote it impacts their perception of the text's convincingness. These findings suggest that convincingness and expectations play a vital role in how information is received and evaluated, particularly in contexts involving GAI and expert authorship.

Furthermore, when considering the impact, the convincingness has on the perceived relevance for the citizens, results indicate that this is a highly significant indicator. Of course, it must be considered whether this is a case of self-selection. However, this is unlikely in the present case. Of all the people who clicked on the experiment, 95.82% decided to finish the study. This means that the participants do not show signs of a selection bias.

Overall, our analysis underscores the critical role of perceived convincingness and relevance in government reports. Ensuring that the perceived and actual authorship align can enhance the perceived convincingness and, consequently, the willingness of citizens to read the full text. This has important implications for how information is presented in various contexts, particularly those involving GAI and expert authorship.

First, our results highlight the critical role of both perception and reality in determining how information is received and evaluated. Understanding this interaction can provide valuable insights into how to present information more convincingly in various contexts, particularly those involving GAI and expert authorship. Furthermore, the strong correlation between the convincingness and the perceived relevance of the text further adds to the importance of understanding.

Second, our findings suggest that enhancing the convincingness of findings could be a key strategy to increase audience engagement. Ensuring that the findings are perceived as convincing and convincing may lead to a greater interest in reading of and engagement with the full report. This insight can be particularly valuable for researchers and practitioners aiming to disseminate their work more effectively.

Third, our study emphasizes that "staying true" is still true despite very convincing results from GAI. Correctly reporting the authorship behind a text enhances the convincingness of results. This is particularly interesting, especially in light of recent studies presenting evidence that humans seem to struggle to tell apart artificially generated texts from human-generated ones when not informed accordingly (Fui-Hoon Nah et al., 2023; Nov et al., 2023).

For theory, the present study adds to previous research on the importance of authorship credibility, especially the effect of discrepancies between human and artificial intelligence (Andrews & Gutkin, 1991). Additionally, the experiment forwards the scholarly understanding of questions concerning the relevance of human expertise in the sensemaking of citizens of uncertain futures (Bolger & Wright, 2017). This further adds to studies of the relevance of the disseminating body of the government report (Felli & Castree, 2012).

The present study contributes to our understanding of how people view the utility of AI – and the domains in which it potentially challenges human authority. With a continuously growing field of possible applications, challenging even typical domains of human authority, understanding how citizens see the convincingness of its current outputs in the public sector. The results highlight the critical role of both perception and reality in determining how information is received and evaluated. Hence, it considers possibilities for further advancements of AI into what is still considered to be a majority human domain.

6 Sources

Abbey, J. D., & Meloy, M. G. (2017). Attention by design: Using attention checks to detect inattentive respondents and improve data quality. *Journal of Operations Management*, *53-56*(1), 63-70. https://doi.org/10.1016/j.jom.2017.06.001

Altay, S., & Gilardi, F. (2023). *Headlines Labeled as AI-Generated Are Less Likely to Be Believed and Shared, Even When True or Human-Generated*. Center for Open Science. https://dx.doi.org/10.31234/osf.io/83k9r

Amanatidou, E., & Guy, K. (2008). Interpreting foresight process impacts: Steps towards the development of a framework conceptualising the dynamics of 'foresight systems'. *Technological Forecasting and Social Change*, *75*(4), 539-557. https://doi.org/10.1016/j.techfore.2008.02.003

Andreiadis, E., Pereira, G. D., Cha, J., Kiegiel, K., Maffia, P., Pattupara, R., Yao, S., Kim, T., Xiong, T., & Meng, W. (2019). CHANGE. CONNECT. CONVINCE. *Networks for Nuclear Innovation*, 49-57.

Andrews, L. W., & Gutkin, T. B. (1991). The effects of human versus computer authorship on consumers' perceptions of psychological reports. *Computers in Human Behavior*, *7*(4), 311-317. https://doi.org/10.1016/0747-5632(91)90018-V

Batalli, M. (2011). Impact of Public Administration Innovations onEnhancing the Citizens' Expectations. *International Journal of e-Education, e-Business, e-Management and e-Learning*. https://doi.org/10.7763/ijeeee.2011.V1.25

Biever, C. (2023). ChatGPT broke the Turing test — the race is on for new ways to assess AI. *Nature*, *619*(7971), 686-689. https://doi.org/10.1038/d41586-023-02361-7

Birch, S. (2023). Voting for the Future: Electoral Institutions and the Time Horizons of Democracy. *Political Studies Review*. https://doi.org/10.1177/14789299231204550

Bolger, F., & Wright, G. (2017). Use of expert knowledge to anticipate the future: Issues, analysis and directions. *International Journal of Forecasting*, *33*(1), 230-243. https://doi.org/10.1016/j.ijforecast.2016.11.001

Bowers, A. P., & Glenday, P. (2021). *Effective foresight by governments: an international view*. Retrieved 2023/07/08 from https://oecd-opsi.org/blog/effective-foresight-by-governments-an-international-view/

Chohan, S. R., Hu, G., Khan, A. U., Pasha, A. T., & Sheikh, M. A. (2021). Design and behavior science in government-to-citizens cognitive-communication: a study towards an inclusive framework. *Transforming Government: People, Process and Policy*, *15*(4), 532-549. https://doi.org/10.1108/tg-05-2020-0079

Douglas, B. D., Ewell, P. J., & Brauer, M. (2023). Data quality in online human-subjects research: Comparisons between MTurk, Prolific, CloudResearch, Qualtrics, and SONA. *PLOS ONE*, *18*(3). https://doi.org/10.1371/journal.pone.0279720

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., . . . Wright, R. (2023). Opinion Paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, *71*. https://doi.org/https://doi.org/10.1016/j.ijinfomgt.2023.102642

Felli, R., & Castree, N. (2012). Neoliberalising Adaptation to Environmental Change: Foresight or Foreclosure? *Environment and Planning A: Economy and Space*, *44*(1), 1-4. https://doi.org/10.1068/a44680

Fuerth, L. S., & Faber, E. M. H. (2012). *Anticipatory Governance Practical Upgrades: Equipping the Executive Branch to Cope with Increasing Speed and Complexity of Major Challenges*. National Defense University, Institute for National Strategic Studies. https://apps.dtic.mil/sti/pdfs/ADA585519.pdf

Fui-Hoon Nah, F., Zheng, R., Cai, J., Siau, K., & Chen, L. (2023). Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration. *Journal of Information Technology Case and Application Research*, *25*(3), 277-304. https://doi.org/10.1080/15228053.2023.2233814

Goodman, J. K., Cryder, C. E., & Cheema, A. (2013). Data Collection in a Flat World: The Strengths and Weaknesses of Mechanical Turk Samples. *Journal of Behavioral Decision Making*, *26*(3), 213-224. https://doi.org/10.1002/bdm.1753

Haasnoot, M., van 't Klooster, S., & van Alphen, J. (2018). Designing a monitoring system to detect signals to adapt to uncertain climate change. *Global Environmental Change*, *52*, 273-285. https://doi.org/10.1016/j.gloenvcha.2018.08.003

Hauser, D. J., & Schwarz, N. (2016). Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behav Res Methods*, *48*(1), 400-407. https://doi.org/10.3758/s13428-015-0578-z

Hyland-Wood, B., Gardner, J., Leask, J., & Ecker, U. K. H. (2021). Toward effective government communication strategies in the era of COVID-19. *Humanities and Social Sciences Communications*, *8*(1). https://doi.org/10.1057/s41599-020-00701-w

Jurma, W. E. (1981). Evaluations of Credibility of the Source of a Message. *Psychological Reports*, *49*(3), 778-778. https://doi.org/10.2466/pr0.1981.49.3.778

Karpinska, M., Akoury, N., & Iyyer, M. (2021). The perils of using Mechanical Turk to evaluate open-ended text generation. *arXiv preprint arXiv:2109.06835*.

Kees, J., Berry, C., Burton, S., & Sheehan, K. (2017). An Analysis of Data Quality: Professional Panels, Student Subject Pools, and Amazon's Mechanical Turk. *Journal of Advertising*, *46*(1), 141-155. https://doi.org/10.1080/00913367.2016.1269304

Longoni, C., Fradkin, A., Cian, L., & Pennycook, G. (2022). *News from Generative Artificial Intelligence Is Believed Less* Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency, Seoul, Republic of Korea.

Margetts, H., & Dorobantu, C. (2019). Rethink government with AI. *Nature*, *568*(7751), 163-165. https://doi.org/10.1038/d41586-019-01099-5

Mökander, J., & Schroeder, R. (2024). Artificial Intelligence, Rationalization, and the Limits of Control in the Public Sector: The Case of Tax Policy Optimization. *Social Science Computer Review*. https://doi.org/10.1177/08944393241235175

Nalbandian, L. (2022). An eye for an 'I:' a critical assessment of artificial intelligence tools in migration and asylum management. *Comparative Migration Studies*, *10*(1). https://doi.org/10.1186/s40878-022-00305-0

Necka, E. A., Cacioppo, S., Norman, G. J., & Cacioppo, J. T. (2016). Measuring the Prevalence of Problematic Respondent Behaviors among MTurk, Campus, and Community Participants. *PLOS ONE*, *11*(6). https://doi.org/10.1371/journal.pone.0157732

Nov, O., Singh, N., & Mann, D. (2023). Putting ChatGPT's Medical Advice to the (Turing) Test: Survey Study. *JMIR Med Educ*, *9*, e46939. https://doi.org/10.2196/46939

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, *381*(6654), 187-192. https://doi.org/10.1126/science.adh2586

OPSI. (2022). *Anticipatory innovation governance : towards a new way of governing in Finland*. https://oecd-opsi.org/wp-content/uploads/2022/06/OECD-Finland-Anticipatory-Report-FINAL.pdf

Paliokaitė, A., & Sadauskaitė, A. (2023). Institutionalisation of participative and collaborative governance: Case studies of Lithuania 2030 and Finland 2030. *Futures*, *150*, 103174. https://doi.org/10.1016/j.futures.2023.103174

Piirainen, K. A., & Gonzalez, R. A. (2015). Theory of and within foresight — "What does a theory of foresight even mean?". *Technological Forecasting and Social Change*, *96*, 191-201. https://doi.org/10.1016/j.techfore.2015.03.003

Priniski, S. J., Hecht, C. A., & Harackiewicz, J. M. (2018). Making Learning Personally Meaningful: A New Framework for Relevance Research. *J Exp Educ*, *86*(1), 11-29. https://doi.org/10.1080/00220973.2017.1380589

Rao, A., & Firth-Butterfield, K. (2020). *3 ways COVID-19 is transforming advanced analytics and AI*. World Economic Forum. Retrieved 2023/11/17 from https://www.weforum.org/agenda/2020/07/3-ways-covid-19-is-transforming-advanced-analytics-and-ai/

Selten, F., Robeer, M., & Grimmelikhuijsen, S. (2023). 'Just like I thought': Street-level bureaucrats trust AI recommendations if they confirm their professional judgment. *Public Administration Review*, *83*(2), 263-278. https://doi.org/10.1111/puar.13602

Straub, V. J., Morgan, D., Bright, J., & Margetts, H. (2023). Artificial intelligence in government: Concepts, standards, and a unified framework. *Government Information Quarterly*, *40*(4), 101881. https://doi.org/10.1016/j.giq.2023.101881

Stritch, J. M., Pedersen, M. J., & Pezo, I. (2024). Crowdsourced data in public administration research: A review and look to the future. *Public Administration Review*. https://doi.org/10.1111/puar.13823

Sun, T. Q., & Medaglia, R. (2019). Mapping the challenges of Artificial Intelligence in the public sector: Evidence from public healthcare. *Government Information Quarterly*, *36*(2), 368-383. https://doi.org/10.1016/j.giq.2018.09.008

Van Der Bles, A. M., Van Der Linden, S., Freeman, A. L. J., & Spiegelhalter, D. J. (2020). The effects of communicating uncertainty on public trust in facts and numbers. *Proceedings of the National Academy of Sciences*, *117*(14), 7672-7683. https://doi.org/10.1073/pnas.1913678117

West, P., Lu, X., Dziri, N., Brahman, F., Li, L., Hwang, J. D., Jiang, L., Fisher, J., Ravichander, A., & Chandu, K. (2023). THE GENERATIVE AI PARADOX:"What It Can Create, It May Not Understand". The Twelfth International Conference on Learning Representations,

7 Appendix

7.1 Visual Approach of experiment and content

The procedure, simplified, looks like the following Figure 3.

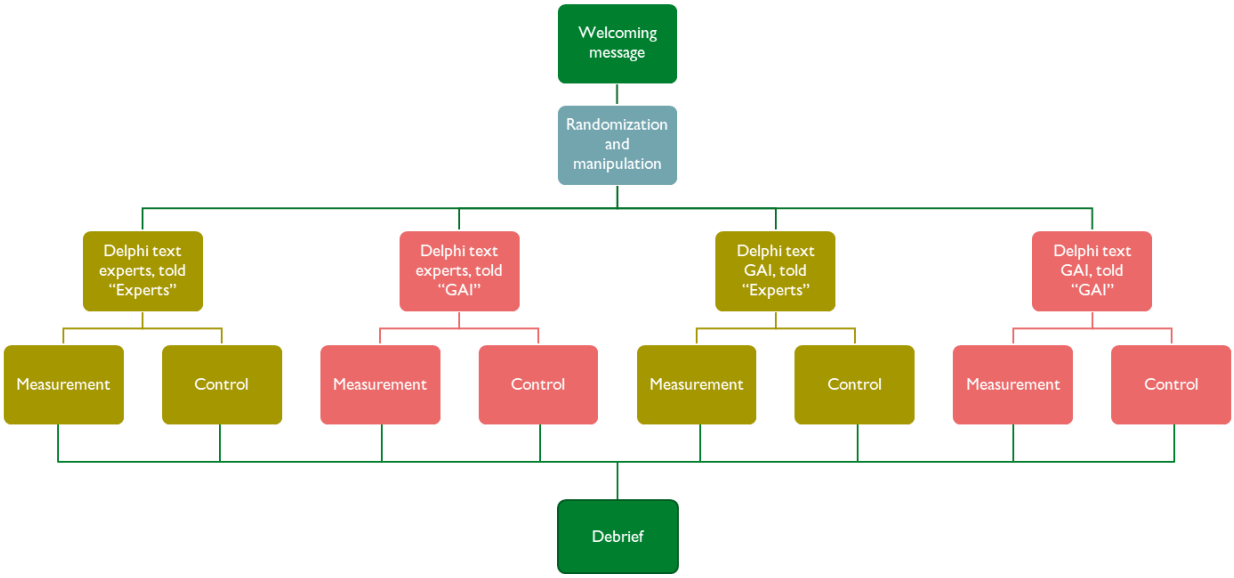

Figure 3: Procedure of the Experiment

Foresight Abstract on Digital Society 2040

The spread of digital technology into every area of life has caused the datafication of the economy and society, as the actions of people, companies, machines and even nature leave a lasting data footprint.

The datafication of everyday life increased noticeably as a consequence of the Covid-19 pandemic, as virtual school and work were added to online social communication and shopping. The next leap forward in data in everyday life will come as a wide range of domestic appliances are connected to the internet, which will happen together with the spread of data networks using 5G technology. New data are being added to the existing stockpiles at an ever increasing rate, as the volume of additional data created each year will double over the next three years.

Foresight Abstract on Digital Society 2040

By 2040, the digital society will be marked by an unparalleled integration of digital technologies in daily life, driven by advanced digital infrastructure and pervasive artificial intelligence (AI). This era will see seamless global connectivity facilitated by technologies like 6G, enabling real-time data exchange and interactions. Quantum computing will revolutionize data analytics, enhancing decision-making and innovation across sectors. AI will be deeply embedded in various aspects of life, from autonomous transportation and smart cities to personalized healthcare, significantly improving efficiency, convenience, and quality of life.

However, the promise of this digital utopia comes with significant challenges, particularly in ensuring digital inclusivity and safeguarding privacy and security. Bridging the digital divide will be critical to prevent widening socioeconomic disparities, requiring global efforts to enhance digital literacy and access. Moreover, the omnipresence of data collection underscores the need for robust privacy protections and ethical frameworks to govern AI and data use, ensuring a digital society that is equitable, secure, and respects individual rights.

Figure 4: Two visual approaches to the Foresight report and Abstract

7.2 *H1* & *H2* Insights

Source	df	Mean Square	F	Sig.
Corrected Model	3	44.773	66.672	<.001
Intercept	1	14.954	22.268	<.001
Agree with findings	1	16.184	24.099	<.001
Trust in national gov	1	42.594	63.427	<.001
Perceived Author	1	.306	.455	.500

a. R Squared = .398 (Adjusted R Squared = .392)

Appendix Table 1: Effect of Perceived Authorship on Convincingness

Source	df	Mean Square	F	Sig.
Corrected Model	3	45.439	68.333	<.001
Intercept	1	12.960	19.490	<.001
Agree with findings	1	16.632	25.012	<.001
Trust in national gov	1	45.061	67.765	<.001
Actual Author	1	2.302	3.462	.064

a. R Squared = .404 (Adjusted R Squared = .398)

Appendix Table 2: Effect of Actual Authorship on Convincingness

Perceived Author	Actual Author	Mean	Std. Deviation	N
Government Experts	Government Experts	1.86	1.047	69
	Generative AI	2.22	1.051	72
Generative AI	Government Experts	2.23	1.058	88
	Generative AI	2.00	1.019	78

Appendix Table 3: Between-Groups Analysis of Perceived Convincingness

7.3 *H3* Insights

		Levene Statistic	df1	df2	Sig.
Want to read full text	Based on Mean	2.534	4	302	.040
	Based on Median	1.275	4	302	.280
	Based on Median and with adjusted df	1.275	4	274.052	.280
	Based on trimmed mean	2.899	4	302	.022

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Dependent variable: Want to read full text

b. Design: Intercept + Findings are Convincing

Appendix Table 4: Levene's Test of Equality of Error Variances

		Levene Statistic	df1	df2	Sig.
Want to read full text	Based on Mean	2.534	4	302	.040
	Based on Median	1.275	4	302	.280
	Based on Median and with adjusted df	1.275	4	274.052	.280
	Based on trimmed mean	2.899	4	302	.022

Appendix Table 5: Tests of Homogeneity of Variances

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	139.294	4	34.824	51.003	<.001
Within Groups	206.198	302	.683		
Total	345.492	306			

Appendix Table 6: Welch ANOVA for “Want to read full text.”

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean	
					Lower Bound	Upper Bound
1	103	1.59	.733	.072	1.45	1.74
2	117	2.13	.804	.074	1.98	2.28
3	57	2.74	.955	.126	2.48	2.99
4	18	3.39	1.145	.270	2.82	3.96
5	12	4.42	.515	.149	4.09	4.74

Appendix Table 7: Descriptive Analysis for "Want to read full text."

Source	df	Mean Square	F	Sig.
Corrected Model	4	34.824	51.003	<.001
Intercept	1	1164.521	1705.573	<.001
Findings are Convincing	4	34.824	51.003	<.001

a. R Squared = .403 (Adjusted R Squared = .395)

Appendix Table 8: Tests of Between-Subjects Effects on "Want to read full text."

	Statistica	df1	df2	Sig.
Welch	79.632	4	52.853	<.001

a. Asymptotically F distributed.

Appendix Table 9: Robust Tests of Equality in Means on "Want to read full text."