Science Written by Generative AI is Perceived as Less Intelligent, but More Credible and Trustworthy than Science Written by Humans

David M. Markowitz[1]

Affiliations
[1] Department of Communication, Michigan State University, East Lansing, MI 48824

Abstract

This paper evaluated the effectiveness of using generative AI to simplify science communication and enhance public trust in science. By comparing lay summaries of journal articles from PNAS, yoked to those generated by AI, this work assessed linguistic simplicity across such summaries and public perceptions. Study 1a analyzed simplicity features of PNAS abstracts (scientific summaries) and significance statements (lay summaries), observing that lay summaries were indeed linguistically simpler, but effect size differences were small. Study 1b used GPT-4 to create significance statements based on paper abstracts and this more than doubled the average effect size without fine-tuning. Finally, Study 2 experimentally demonstrated that simply-written GPT summaries facilitated more favorable public perceptions of scientists (their credibility, trustworthiness) than more complexly-written human PNAS summaries. AI has the potential to engage scientific communities and the public via a simple language heuristic, advocating for its integration into scientific dissemination for a more informed society.

Significance Statement

Across several studies, this paper revealed that generative AI can simplify science communication, making complex concepts feel more accessible and enhancing public trust in scientists. By comparing traditional scientific summaries from the journal PNAS to AI-generated summaries of the same work, this research demonstrated that AI can produce even simpler and clearer explanations of scientific information that are easier for the general public to understand. Importantly, these simplified summaries can improve perceptions of scientists' credibility and trustworthiness as experimentally demonstrated in this work. With small, language-level changes, AI has the potential to be effective science communicators and its possible deployment at scale makes it an appealing technology for clearer science communication.

Keywords

science communication, generative AI, large language models, trust, credibility

Science Written by Generative AI is Perceived as Less Intelligent, but More Credible and Trustworthy than Science Written by Humans

Scientific information is essential for everyday decision-making. People often use science, or information communicated by scientists, to make decisions in medical settings (1), environmental settings (2), and many others (3). For people to use such information effectively, however, they must have some amount of scientific literacy (4) or at least trust those who communicate scientific information to them (5). Overwhelming evidence suggests these ideals are not being met, as trust in scientists and scientific evidence have decreased over time for nontrivial reasons (e.g., distrust in institutions, political polarization, among many others) (6–8). The public's decreasing trust in scientists and scientific information is unrelenting, which requires more thoughtful research into countermeasures and possible remedies that can be scaled across people and populations.

Several remedies have been proposed to make science more approachable, and to improve the perception of scientists. For example, some propose that being transparent about how research was conducted and disclosing possible conflicts of interest (9, 10), having scientists engage with the public about their work (11), or improving scientists' ability to tell a compelling story (12) can increase public trust. While there is no panacea for dwindling public trust in science and scientists, extant evidence suggests this is an issue worth taking seriously, and it is imperative that scientists discover ways to best communicate their work with the hope of improving how people perceive them and their research.

Against this backdrop, the current work argues that *how* one's science is communicated matters, and that language-level changes to scientific summaries can significantly improve perceptions of a scientist. Critically, the evidence in this paper suggests scientists may not be the best messengers to communicate their work if one goal is to communicate science simply. In other words, it may be difficult for experts to write for non-experts. Instead, as the current research demonstrates, generative AI can effectively summarize scientific writing in ways that are more approachable for lay readers, and such tools can be scaled to improve science communication efforts at a system level.

The Benefits of Simple Writing

The idea that simple language patterns can improve perceptions of scientists is supported by decades of processing fluency research (13–15). This literature suggests people tend to use their feelings when consuming information (16, 17), and simple (fluent) information feels better to most people than complex (disfluent) information. Support for this contention suggests people engage with, approach, and prefer content that is written in simple versus complex terms (e.g., simple synonyms of the same concept compared to complex synonyms) (18–20). Indeed, much of this research supports the *simpler-is-better* hypothesis, which claims that people will engage with content that is communicated in simple versus complex terms, absent some instrumental goal being activated (18).

The most common linguistic fluency dimension is lexical fluency, which considers the degree to which people use common and everyday terms in the communication. People perceive scientists to be more intelligent if their work is written with simple words (e.g., the word *job*) compared to complex words (e.g., the word *occupation*) (15). In most cases, people prefer simple synonyms for a concept compared to complex synonyms of the same concept because it is more of a challenge to interpret and comprehend complexity, and people are economical with their effort and attention (21, 22). Another fluency dimension is analytic writing fluency (23). This dimension considers one's communication style and *how* people communicate, instead of what they are communicating about (24, 25). According to prior work, a simple communication style is informal and reflects a story (e.g., it contains more pronouns, adverbs) compared to a complex communication style, which is formal and contains high rates of articles and prepositions (26–

28). Finally, another relevant fluency concept is structural fluency, which considers the length of words and sentences. Longer words (e.g., *occupation* vs. *job*) and sentences with more words tend to require more effort to process (29). Therefore, the final marker of fluency relevant to the current work is operationalized by readability, which considers verbal simplicity/complexity in terms of word and sentence length.

The Current Work

The current empirical package evaluates fluency effects in the context of science writing and has several aims. The first aim is to evaluate how lay summaries of scientific articles (called significance statements in many journals) are indeed linguistically simpler compared to scientific summaries of the same articles (abstracts). It is unclear if scientists are aware of how to effectively summarize their work for non-experts (30), making this effort worthwhile to empirically test if ideals of a journal like simple and approachable writing are being realized (Study 1a). The second aim is to evaluate if such lay summaries can be made simpler. Study 1b had generative Artificial Intelligence (AI) and a popular large language model (GPT-4) create lay summaries based on paper abstracts and compared the linguistic properties of such texts.

Finally, building on this progression of studies, Study 2 tested the causal impact of reading scientific writing generated by AI (versus reading scientific writing generated by humans) on perceptions of scientists. Participants were randomly assigned to AI or human versions of a scientific summary, and they made judgments about the credibility, trustworthiness, and intelligence of the authors. To foreshadow the results: people preferred the simple (AI) version of each summary compared to the complex (human) version, yet ironically, people believed that the complex version was more likely to be AI than human.

Study 1a: Method

Data Collection

To first evaluate if lay summaries had a simpler linguistic style than scientific summaries, significance statements and academic abstracts were respectively extracted from the journal *Proceedings of the National Academy of Sciences* (PNAS). This journal was selected because it is a widely read, high-impact general science journal that was one of the first outlets to require authors to provide traditional scientific summaries (e.g., abstracts) and lay summaries that appeal to average readers. PNAS also has topical breadth, scale, and longevity relative to other journals that may require lay summaries in that significance statements began in 2012 (31).

A total of 42,022 publications were extracted from PNAS between January 2010 and March 2024 to capture possible papers that included both academic abstracts and significance statements. Only those with both summary types were included in this paper to create a yoked comparison within the same article. The final dataset included 34,584 papers (34,584 significance statements and 34,584 abstracts), totaling 10,799,256 words.

Automated Text Analysis

All texts were evaluated with Linguistic Inquiry and Word Count (LIWC), an automated text analysis tool that counts words as a percentage of the total word count per text (32). LIWC contains a validated internal dictionary of social (e.g., words related to family), psychological (e.g., words related to cognition, emotion), and part of speech dimensions (e.g., pronouns, articles, prepositions), and the tool measures the degree to which each text contains words from its respective dictionary categories. For example, the phrase “This science aims to improve society” contains 6 words and counts the following LIWC categories, including but not limited to: impersonal pronouns (*this*; 16.67% of the total word count) and positive tone words (*improve*; 16.67% of the total word count). All texts were run through LIWC-22 unless otherwise stated.

Measures

To evaluate how lay versus scientific summaries compared in terms of verbal simplicity, three measures were used from prior work to approximate simple language patterns (23):

common words (e.g., the degree to which people use common and simple terms like *job* instead of uncommon and more complex terms like *occupation*), one's analytic writing style (e.g., the degree to which people are formal and complex in their writing style compared to informal and their writing reflects a story), and readability (e.g., the number of words per sentence and big words in a person's communication output).

Consistent with prior work (23, 33–35), common words were operationalized with the LIWC dictionary category. LIWC's dictionary represents a collection everyday words in English (36, 37). Therefore, texts that use more words from this dictionary are simpler than texts that use fewer words from this dictionary. One's analytic writing style was operationalized with the LIWC analytic thinking index, which is a composite variable of seven style word categories. Style words represent *how* one is communicating rather than what they are communicating about (24, 38). This index contains high rates of articles and prepositions, but low rates of conjunctions, adverbs, auxiliary verbs, negations, and pronouns (26, 39, 40).[1] Finally, readability was operationalized with the Flesch Reading Ease metric (29) and calculated using the *quanteda.textstats* package in R (41). High scores on the Flesch Reading Ease metric suggest more readable and simpler writing (e.g., texts with smaller words and shorter sentences) compared to low scores. These language dimensions were evaluated as an index by first standardizing (z-scoring) each variable and then applying the following formula: Common Words + Readability – Analytic Writing. High scores are linguistically simpler than low scores.

Analytic Plan

Since each article contained one lay summary and one scientific summary from the same article, independent samples t-tests were conducted for the simplicity index and each individual dimension of the index. All data across studies are located on the Open Science Framework: https://osf.io/64am3/?view_only=883926733e6e494fa2f2011334b24796.

Study 1a: Results

Descriptive statistics for each language dimension and intercorrelations are in Table 1. As expected, lay summaries were linguistically simpler than scientific summaries of the same article, Welch's $t(65793) = 40.62$, $p < .001$, Cohen's $d = 0.31$, 95% CI [0.29, 0.32].[2] At the item level of the simplicity index, lay summaries ($M = 69.77\%$, $SD = 7.14\%$) contained more common words than scientific summaries ($M = 67.79\%$, $SD = 6.60\%$), Welch's $t(68741) = 37.79$, $p < .001$, Cohen's $d = 0.29$, 95% CI [0.27, 0.30]. Lay summaries ($M = 92.34$, $SD = 7.95$) also had a simpler linguistic style than scientific summaries ($M = 94.31$, $SD = 5.19$), Welch's $t(59561) = -38.52$, $p < .001$, Cohen's $d = 0.29$, 95% CI [0.28, 0.31]. Finally, lay summaries ($M = 12.96$, $SD = 13.93$) were more readable than scientific summaries as well ($M = 12.49$, $SD = 12.46$), Welch's $t(68320) = 4.67$, $p < .001$, Cohen's $d = 0.036$, 95% CI [0.02, 0.05].

Together, while lay summaries were indeed linguistically simpler than scientific summaries at PNAS, the effect sizes between such groups were quite small and it is therefore unclear if general readers would be able to recognize or appreciate such differences. Can lay summaries be written even simpler, using generative AI tools, to produce more substantive effect sizes while maintaining the core content of each text? In the next study, a random selection of abstracts was submitted to a popular large language model, GPT-4, and were given the same instructions as PNAS authors on how to construct a significance statement.

Study 1b: Method

An *a priori* power analysis using a small effect size (Cohen's $d = 0.20$) powered at 80% suggested 788 cases were needed to detect a difference between GPT significance statements

---

[1] Analytic writing = [articles + prepositions - pronouns - auxiliary verbs - adverb - conjunctions - negations] from LIWC scores (40).

[2] 95% Confidence Intervals were bootstrapped with 5,000 replicates.

and PNAS significance statements. Therefore, a random selection of 800 abstracts from Study 1a was used in this study. Using the OpenAI API, the large language model GPT-4 was fed each abstract individually and given the following prompt, which was drawn from descriptions of what PNAS authors should communicate in their significance statements (31):

> *The following text is an academic abstract from the journal Proceedings of the National Academy of Sciences. Based on this abstract, create a significance statement. This statement should provide enough context for the paper's implications to be clear to readers. The statement should not contain references and should avoid numbers, measurements, and acronyms unless necessary. It should explain the significance of the research at a level understandable to an undergraduate-educated scientist outside their field of specialty. Finally, it should include no more than 120 words. Write the significance statement here*:

The same text analytic process was performed on these data as Study 1a. Each GPT significance statement received scores based on common words (LIWC dictionary category), analytic writing (LIWC analytic writing category), and readability (Flesch Reading Ease).

Study 1b: Results

Distributions of the comparisons in this study are reflected in Figure 1. Indeed, GPT significance statements were written in a simpler manner than PNAS significance statements for the simplicity index, Welch's $t(1492.1) = 11.55$, $p < .001$, Cohen's $d = 0.58$, 95% CI [0.47, 0.69]. Specifically, GPT significance statements ($M = 75.53\%$, $SD = 5.57\%$) contained more common words than PNAS significance statements ($M = 69.84\%$, $SD = 7.45\%$), Welch's $t(1478.7) = 17.31$, $p < .001$, Cohen's $d = 0.87$, 95% CI [0.76, 0.97]. GPT significance statements ($M = 17.59$, $SD = 11.15$) were also more readable than PNAS significance statements ($M = 12.86$, $SD = 14.27$), Welch's $t(1510) = 7.39$, $p < .001$, Cohen's $d = 0.37$, 95% CI [0.27, 0.47]. However, GPT significance statements ($M = 92.73$, $SD = 6.89$) had a statistically equivalent analytic style as PNAS significance statements ($M = 92.32$, $SD = 7.48$), Welch's $t(1587.7) = 1.16$, $p = .246$, Cohen's $d = 0.06$, 95% CI [-0.04, 0.16]. All results were maintained when comparing GPT significance statements to PNAS abstracts and PNAS significance statements as well.

Alternative Explanations

One possible explanation for the Study 1b results is that there are content differences across the PNAS and GPT texts explaining or impacting such differences across groups. This concern was addressed in two ways. First, PNAS has various sections that authors submit to, and LIWC has categories to approximate words associated with such sections. For example, the LIWC category for political speech would approximate papers submitted the Social Science section, specifically Political Sciences. Several linguistic covariates were therefore examined to account for content-related differences across GPT and PNAS texts. After including overall affect/emotion and cognition (to control for topics within the Psychological Sciences section of PNAS), political speech (to control for topics within the Political Science section of PNAS), and physical references to the multivariate models (to control for topics within the Biological Sciences section of PNAS), all results were maintained except for Analytic writing, where GPT texts were more analytic than PNAS texts, which is also consistent with prior work (42). Please see the online supplement for additional LIWC differences across these text types.

Content effects were also evaluated in a bottom-up manner using the Meaning Extraction Method to measure dominant themes across the GPT and PNAS texts (43, 44). The evidence in the online supplement states there were 8 themes reliably extracted from the data, ranging from basic methodological and research information to gene expression and cancer science. Controlling for these themes, including the prior LIWC content dimensions, revealed consistent results as well (see supplement). Therefore, Study 1b evidence is robust to content.

Altogether, human authors write simpler for lay audiences than for scientific audiences (Study 1a), but Study 1b demonstrated artificial intelligence and large language models can do so more effectively (e.g., the effect size differences between GPT significance statements and PNAS significance statements was larger than human in Study 1a). The findings thus far are correlational and therefore need causal evidence to demonstrate the impact of these effects on human perceptions. In Study 2, participants were randomly assigned to read a GPT significance statement or PNAS significance statement from pairs of texts that appeared in the previous studies. Participants made perceptions about the author (e.g., intelligence, credibility, trustworthiness), judged the complexity of each text, and they rated how much they believed the author of each text was human or artificial intelligence. Only perceptions of the author were made because prior work suggests people generally report consistent ratings when asked about both scientists and their science in similar studies (9).

Study 2: Method

Participants in the US were recruited from Prolific and paid $4.00 for their time in a short study (median completion time < 7 minutes). People were told that they would read scientific summaries and make judgments about the authors of such texts.

Participants and Power

Based on this study's preregistration (https://aspredicted.org/C3K_T31), a minimum of 164 participants were required to detect a small effect powered at 80% in a within-subjects study (f = 0.10, α = two-tailed, three measurements). A total of 274 participants were recruited to ensure enough participants were in the study. Most participants self-identified as men (n = 139; 50.7%; women n = 127, other n = 7), they were 36.74 years old on average (SD = 12.47 years), and were mostly White (n = 190; 69.3%). On a 7-point political ideology scale (1 = extremely liberal, 7 = extremely conservative), participants leaned liberal (M = 2.97, SD = 1.63).

Procedure

Five pairs of stimuli from Study 1b were selected for the experiment, having had the greatest difference in common words scores between the PNAS and GPT texts (Pair 1 GPT = 79.31%, Pair 1 PNAS = 48.65%; Pair 2 GPT = 76.64%, Pair 2 PNAS = 46.32%; Pair 3 GPT = 79.07%, Pair 3 PNAS = 52.14%; Pair 4 GPT = 85.00%, Pair 4 PNAS = 59.66%; Pair 5 GPT = 87.10%, Pair 5 PNAS = 62.81%). Participants were randomly assigned to read stimuli from three out of a possible five pairs (see the online supplement for the stimuli texts), and within these randomly selected pairs, participants were randomly assigned to the GPT (simple) or PNAS (complex) version of each pair. Participants were told to read each summary of a scientific paper and then answer questions below each summary. They were specifically told "we are not expecting you to be an expert in the topic discussed below. Instead, make your judgments based on how the summary is written."

Finally, participants made various perceptions of the author (e.g., intelligence, trustworthiness) based on prior work (14, 15, 34), judgments about the identity of who wrote the scientific summary (AI or human), and assessed the complexity in each text as a manipulation check. The order of these measures was randomized, and items within each block were randomized as well. This study was approved by the author's university research ethics board.

Measures

Manipulation Check

Based on prior work (15, 34), three questions asked participants to rate how clear ("How clear was the writing in the summary you just read?"), complex ("How complex was the writing in the summary you just read?"), and how well they understood each scientific summary ("How much of this writing did you understand?"). Ratings for the first two questions were made on 7-point Likert-type scales from 1 = Not at all to 7 = Extremely. The third question ranged from 1 = Not at all to 7 = An enormous amount.

Author Perceptions

Participants made three ratings about the author of each scientific summary: (1) "How intelligent is the scientist who wrote this summary?", (2) How credible is the scientist who wrote this summary?", and (3) "How trustworthy is the scientist who wrote this summary?" As a collection, these dimensions were highly reliable (Cronbach's α = 0.88) and therefore, they were averaged to create a general author perceptions index, while also being evaluated individually. All items were measured on 7-point Likert-type scales from 1 = Not at all to 7 = Extremely.

Author Identity Perceptions

Participants were asked for their agreement with two questions: (1) This summary was written by a human, and (2) This summary was written by Artificial Intelligence. All items were measured on 7-point Likert-type scales from 1 = Strongly disagree to 7 = Strongly agree.

Demographics

Basic demographic data were obtained from each participant, including their age, gender, ethnicity, and political ideology.

Analytic Plan

Since there were multiple observations per participant, linear mixed models with random intercepts for participant and stimulus were constructed (45, 46).

Study 2: Results

Manipulation checks were successful. Participants perceived the simpler GPT significance statements as clearer (B = 1.47, SE = 0.09, t = 16.70, p < .001, R^2m = .210, R^2c = .502)[3], less complex (B = -1.50, SE = 0.08, t = -19.28, p < .001, R^2m = .275, R^2c = .498), and they reported understanding more information in such summaries than the complex PNAS versions (B = 1.48, SE = 0.08, t = 18.74, p < .001, R^2m = .229, R^2c = .584).

Crucially, GPT significance statements were perceived more favorably than PNAS significance statements overall (B = 0.13, SE = 0.05, t = 2.44, p = .015, R^2m = .004, R^2c = .575). Analyses at the item level told a more nuanced story, however. GPT significance statements were perceived as more credible (B = 0.25, SE = 0.06, t = 3.95, p < .001, R^2m = .011, R^2c = .548) and more trustworthy than PNAS significance statements (B = 0.28, SE = 0.06, t = 4.63, p < .001, R^2m = .015, R^2c = .558), but they were also perceived as less intelligent (B = -0.15, SE = 0.06, t = -2.57, p = .010, R^2m = .005, R^2c = .500).

Ironically, participants agreed less with the idea that GPT significance statements were written by AI (B = -0.42, SE = 0.09, t = -4.29, p < .001, R^2m = .021, R^2c = .116), and more with the idea that GPT significance statements were written by humans (B = 0.51, SE = 0.09, t = 5.42, p < .001, R^2m = .033, R^2c = .165). In other words, people perceived complexity to be a trait of artificial intelligence more than a trait of humanness. All relationships were maintained after controlling for AI and human perceptions as fixed effects in the linear mixed models.

General Discussion

The current work explored the potential of generative AI to simplify scientific communication, enhance public trust in scientists, and increase engagement in the understanding of science. The evidence suggested that while lay summaries from a top general science journal, PNAS, were linguistically simpler than scientific summaries, the degree of difference between these texts could be enlarged and improved. Generative AI assisted in making scientific texts simpler and more approachable compared to the human-written versions of such summaries. Therefore, this paper is notable given current challenges of scientific literacy and the disconnect between scientific communities and the public — AI is indeed better at communicating like a human (or the intentions of writing simply) than humans (42, 48). As prior work suggests, decreasing trust in scientists and scientific institutions, exacerbated by complex communication barriers, call for inventive solutions that are scalable and relatively inexpensive. Those that are offered here, particularly through generative AI, represent one

---

[3] R^2m = variance explained by fixed effects alone; R^2c = variance explained by fixed and random effects. All values were calculated using the MuMIn package in R (47).

potential pathway toward simpler, more approachable, and improved science communication.

These data build on a body of existing fluency research and provide empirical support for the hypothesis that linguistic simplicity, facilitated by AI, can significantly influence public perceptions of scientists' credibility, trustworthiness, and intelligence. Generative AI, specifically large language models like GPT-4, can produce scientific summaries that are not only simpler, but also more accessible to lay audiences compared to those written by human experts. These results align with a broader scientific narrative (and interest) that advocates for clearer and more direct communication strategies in science dissemination (49).

The implications of this paper are twofold. First, the results suggest that leveraging AI in scientific communication can bridge scientific communities and the general public. This could be particularly beneficial in a time where science is increasingly central to everyday decision-making but is also viewed with skepticism or deemed inaccessible by non-experts. Second, the increased readability and approachability of AI-generated texts might contribute to a higher engagement with scientific content, thereby cultivating a more informed public.

Despite these positive outcomes and effects, it is important to acknowledge that the *simpler-is-better* hypothesis was not universally supported (18). While AI-generated summaries were rated higher in terms of credibility and trustworthiness, they were also perceived as less intelligent. This inconsistency underscores the complex interplay between content simplicity and perceived expertise, suggesting that while simpler language can enhance understanding and trust, it might simultaneously reduce perceived intelligence. In science, people may be perceived as smart but untrustworthy and not credible, which suggests a one-size-fits-all model of the relationship between complexity and person-perceptions is perhaps inaccurate.

Future research should aim to examine these dynamics further, potentially exploring how different domains of science (e.g., communicating about health, communicating about climate) might uniquely benefit from AI-mediated communication (50). Studies could investigate the long-term impact of AI-mediated communication strategies on public engagement with science and scientists. Finally, texts from only one journal were used in this paper across studies and therefore, texts from other journals should be used as well. As a general science journal that publishes high-impact research, however, using PNAS for this paper was purposeful and helped to ensure fluency effects were investigated across core domains of scientific inquiry.

Acknowledgements

The author thanks [Redacted] for their thoughts and comments on an earlier draft of this manuscript.

References

1. M. C. Silverstein, *et al.*, Operating on anxiety: Negative affect toward breast cancer and choosing contralateral prophylactic mastectomy. *Med Decis Making* **43**, 152–163 (2023).

2. C. J. Kirchhoff, M. C. Lemos, S. Dessai, Actionable knowledge for environmental decision making: Broadening the usability of climate science. *Annual Review of Environment and Resources* **38**, 393–414 (2013).

3. D. von Winterfeldt, Bridging the gap between science and decision making. *Proceedings of the National Academy of Sciences* **110**, 14055–14061 (2013).

4. R. C. Laugksch, Scientific literacy: A conceptual overview. *Science Education* **84**, 71–94 (2000).

5. Y. Algan, D. Cohen, E. Davoine, M. Foucault, S. Stantcheva, Trust in scientists in times of pandemic: Panel evidence from 12 countries. *Proceedings of the National Academy of Sciences* **118**, e2108576118 (2021).

6. B. Kennedy, A. Tyson, Americans' trust in scientists, positive views of science continue to decline. *Pew Research Center Science & Society* (2023). Available at: https://www.pewresearch.org/science/2023/11/14/americans-trust-in-scientists-positive-views-of-science-continue-to-decline/ [Accessed 25 February 2024].

7. A. Lupia, *et al.*, Trends in US public confidence in science and opportunities for progress. *Proceedings of the National Academy of Sciences* **121**, e2319488121 (2024).

8. S. Vazire, Quality uncertainty erodes trust in science. *Collabra: Psychology* **3**, 1 (2017).

9. H. Song, D. M. Markowitz, S. H. Taylor, Trusting on the shoulders of open giants? Open science increases trust in science for the public and academics. *Journal of Communication* **72**, 497–510 (2022).

10. T. Rosman, M. Bosnjak, H. Silber, J. Koßmann, T. Heycke, Open science and public trust in science: Results from two studies. *Public Underst Sci* **31**, 1046–1062 (2022).

11. J. Beard, M. Wang, Rebuilding public trust in science. *Boston University* (2023). Available at: https://www.bu.edu/articles/2023/rebuilding-public-trust-in-science/ [Accessed 18 April 2024].

12. S. Martinez-Conde, S. L. Macknik, Finding the plot in science storytelling in hopes of enhancing science communication. *Proceedings of the National Academy of Sciences* **114**, 8127–8129 (2017).

13. A. L. Alter, D. M. Oppenheimer, Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review* **13**, 219–235 (2009).

14. D. M. Oppenheimer, The secret life of fluency. *Trends in Cognitive Sciences* **12**, 237–241 (2008).

15. D. M. Oppenheimer, Consequences of erudite vernacular utilized irrespective of necessity: problems with using long words needlessly. *Applied Cognitive Psychology* **20**, 139–156 (2006).

16. N. Schwarz, "Feelings-as-information theory" in *Handbook of Theories of Social Psychology: Volume 1*, P. A. M. Van Lange, A. W. Kruglanski, E. T. Higgins, Eds. (SAGE Publications Inc., 2012), pp. 289–308.

17. N. Schwarz, "Metacognition" in *APA Handbook of Personality and Social Psychology, Volume 1: Attitudes and Social Cognition*, APA handbooks in psychology®., M. Mikulincer, P. R. Shaver, E. Borgida, J. A. Bargh, Eds. (American Psychological Association, 2015), pp. 203–229.

18. D. M. Markowitz, H. C. Shulman, The predictive utility of word familiarity for online engagements and funding. *Proceedings of the National Academy of Sciences* **118**, e2026045118–e2026045118 (2021).

19. H. C. Shulman, M. D. Sweitzer, Advancing framing theory: Designing an equivalency frame to improve political information processing. *Human Communication Research* **44**, 155–175 (2018).

20. M. D. Sweitzer, H. C. Shulman, The effects of metacognition in survey research: Experimental, cross-sectional, and content-analytic evidence. *Public Opinion Quarterly* **82**, 745–768 (2018).

21. T. Rogers, J. Lasky-Fink, *Writing for busy readers* (Dutton, 2023).

22. H. Song, N. Schwarz, If it's hard to read, it's hard to do: Processing fluency affects effort prediction and motivation. *Psychol Sci* **19**, 986–988 (2008).

23. D. M. Markowitz, Instrumental goal activation increases online petition support across languages. *Journal of Personality and Social Psychology* **124**, 1133–1145 (2023).

24. C. K. Chung, J. W. Pennebaker, "The psychological functions of function words" in *Social Communication*, K. Fiedler, Ed. (Psychology Press, 2007), pp. 343–359.

25. J. W. Pennebaker, *The secret life of pronouns: What our words say about us* (Bloomsbury Press, 2011).

26. J. W. Pennebaker, C. K. Chung, J. Frazee, G. M. Lavergne, D. I. Beaver, When small words foretell academic success: The case of college admissions essays. *PLOS ONE* **9**, e115844–e115844 (2014).

27. S. Seraj, K. G. Blackburn, J. W. Pennebaker, Language left behind on social media exposes the emotional and cognitive costs of a romantic breakup. *Proceedings of the National Academy of Sciences* **118**, e2017154118–e2017154118 (2021).

28. D. M. Markowitz, Psychological trauma and emotional upheaval as revealed in academic writing: The case of COVID-19. *Cognition and Emotion* **36**, 9–22 (2022).

29. R. Flesch, A new readability yardstick. *Journal of Applied Psychology* **32**, 221–233 (1948).

30. J. Stricker, A. Chasiotis, M. Kerwer, A. Günther, Scientific abstracts and plain language summaries in psychology: A comparison based on readability indices. *PLOS ONE* **15**, e0231160 (2020).

31. I. M. Verma, PNAS Plus: Refining a successful experiment. *Proceedings of the National Academy of Sciences* **109**, 13469–13469 (2012).

32. J. W. Pennebaker, R. L. Boyd, R. J. Booth, A. Ashokkumar, M. E. Francis, Linguistic inquiry and word count: LIWC-22. (2022). Deposited 2022.

33. D. M. Markowitz, J. T. Hancock, Linguistic obfuscation in fraudulent science. *Journal of Language and Social Psychology* **35**, 435–445 (2016).

34. D. M. Markowitz, M. Kouchaki, J. T. Hancock, F. Gino, The deception spiral: Corporate obfuscation leads to perceptions of immorality and cheating behavior. *Journal of Language and Social Psychology* **40**, 277–296 (2021).

35. D. M. Markowitz, What words are worth: National Science Foundation grant abstracts indicate award funding. *Journal of Language and Social Psychology* **38**, 264–282 (2019).

36. Y. R. Tausczik, J. W. Pennebaker, The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology* **29**, 24–54 (2010).

37. R. L. Boyd, A. Ashokkumar, S. Seraj, J. W. Pennebaker, "The development and psychometric properties of LIWC-22" (University of Texas at Austin, 2022).

38. M. E. Ireland, J. W. Pennebaker, Language style matching in writing: Synchrony in essays, correspondence, and poetry. *Journal of Personality and Social Psychology* **99**, 549–571 (2010).

39. K. N. Jordan, J. W. Pennebaker, C. Ehrig, The 2016 U.S. presidential candidates and how people tweeted about them. *SAGE Open* **8**, 215824401879121–215824401879121 (2018).

40. K. N. Jordan, J. Sterling, J. W. Pennebaker, R. L. Boyd, Examining long-term trends in politics and culture through language of political leaders and cultural institutions. *Proceedings of the National Academy of Sciences of the United States of America* **116**, 3476–3481 (2019).

41. K. Benoit, *et al.*, quanteda.textstats: Textual statistics for the quantitative analysis of textual data. (2021). Deposited 24 November 2021.

42. D. M. Markowitz, J. T. Hancock, J. N. Bailenson, Linguistic markers of inherently false AI communication and intentionally false human communication: Evidence from hotel reviews. *Journal of Language and Social Psychology* **43**, 63–82 (2024).

43. C. K. Chung, J. W. Pennebaker, Revealing dimensions of thinking in open-ended self-descriptions: An automated meaning extraction method for natural language. *Journal of Research in Personality* **42**, 96–132 (2008).

44. D. M. Markowitz, The meaning extraction method: An approach to evaluate content patterns from large-scale language data. *Frontiers in Communication* **6**, 13–13 (2021).

45. D. Bates, M. Mächler, B. M. Bolker, S. C. Walker, Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* **67** (2015).

46. A. Kuznetsova, P. B. Brockhoff, R. H. B. Christensen, S. P. Jensen, lmerTest: Tests in linear mixed effects models. (2020). Deposited 23 October 2020.

47. K. Bartoń, MuMIn. (2020). Deposited 2020.

48. M. Jakesch, J. T. Hancock, M. Naaman, Human heuristics for AI-generated language are flawed. *Proceedings of the National Academy of Sciences* **120**, e2208839120 (2023).

49. An act to enhance citizen access to Government information and services by establishing that Government documents issued to the public must be written clearly, and for other purposes. (2010).

50. J. T. Hancock, M. Naaman, K. Levy, AI-Mediated Communication: Definition, research agenda, and ethical considerations. *Journal of Computer-Mediated Communication* **25**, 89–100 (2020).

Table 1

Descriptive Statistics and Correlation Matrix for Study 1a

Variable	*M*	*SD*	1	2
1. Common words	68.78	6.94		
2. Analytic writing	93.33	6.79	-.18** [-.19, -.18]	
3. Readability	12.73	13.22	.24** [.24, .25]	-.07** [-.08, -.06]

Note. ** p < .01. Numbers in brackets are 95% Confidence Intervals

Table 2

Descriptive Statistics and Correlation Matrix for Study 2

Variable	*M*	*SD*	1	2	3	4	5	6	7
1. Intelligent	5.07	1.05	--						
2. Credible	4.60	1.16	.65** [.61, .69]	--					
3. Trustworthy	4.56	1.12	.65** [.60, .68]	.84** [.82, .86]	--				
4. AI	3.89	1.44	-.18** [-.24, -.11]	-.20** [-.27, -.14]	-.22** [-.29, -.16]	--			
5. Human	4.55	1.39	.24** [.18, .30]	.28** [.21, .34]	.33** [.27, .39]	-.79** [-.82, -.77]	--		
6. Clear	3.95	1.58	.16** [.10, .23]	.32** [.25, .38]	.33** [.27, .39]	-.26** [-.33, -.20]	.31** [.25, .37]	--	
7. Complex	4.74	1.41	.21** [.14, .27]	.03 [-.04, .10]	.01 [-.06, .08]	.14** [.07, .21]	-.13** [-.20, -.07]	-.55** [-.60, -.50]	--
8. Understand	3.46	1.55	.05 [-.02, .11]	.23** [.17, .30]	.25** [.19, .31]	-.20** [-.27, -.13]	.25** [.18, .31]	.72** [.69, .75]	-.56** [-.60, -.51]

Note. ** $p < .01$. Numbers in brackets are 95% Confidence Intervals.

Figure 1
Distributions of Key Comparisons in Study 1b