sociological science

Supplement to:

Alvero, AJ, Dustin S. Stoltz, Oscar Stuhler, and Marshall A. Taylor. 2025. “Generative AI in Sociological Research: State of the Discipline” Sociological Science 13: 45-62.

Appendix A. Example Recruitment Materials

Appendix A.1. Invitation Email

Dear [REDACTED],

I hope this email finds you well. Together with my colleagues [REDACTED], I am conducting a survey on how sociologists and their collaborators use and think about generative AI.

We are reaching out to you specifically because you recently published an article in a sociology journal. We would like to invite you to participate in our short online survey. We are interested in your thoughts, regardless of whether you have experience using generative AI. By participating, you will shape discussions on generative AI in sociology and help us understand its role, if any, in the sociological toolkit. Most people take between 5 and 10 minutes to complete the survey. All responses will be kept confidential.

To take this survey, please follow this personalized link: [REDACTED]

Or copy and paste the URL below into your internet browser: [REDACTED]

The above link is unique to your email address. If you prefer to take part anonymously, please use this link: [REDACTED]

Appendix A.2. Generative AI Explainer

For the sake of this survey, we will use GenAI to refer to generative models that produce content (e.g., GPT-4), as well as specific applications or platforms built on top of them (e.g., ChatGPT).

There is wide variation in how GenAI might be used in social scientific research. Its potential applications range from writing/outlining, transcription, translation, internet search, summarization and literature review, to data annotation/labeling, coding/debugging, and simulating human behavior and interactions.

Some specific GenAI tools include:

- OpenAI's ChatGPT browser-based chat interface
- OpenAI's Whisper transcription service
- Google's Gemini, providing, for example, summaries in search results
- Perplexity's conversational academic search engine
- GitHub's Copilot, providing automated code writing
- Microsoft's Copilot, generating PowerPoint slides
- Meta's Llama models, which can run local chat interfaces
- Mistral's models, which can run local chat interfaces
- Stability's Stable Diffusion, which generates images from text

Some models (GPT-4) run on remote servers and require an internet connection to access (e.g., through ChatGPT or OpenAI's API), however other models (e.g., Llama) can be operated "locally" on a desktop or laptop without the need for an internet connection.

Considering all these applications and models, as well as others we may not have listed, how often do you use GenAI in your research practices?

Appendix B. Sampling Methods

Appendix B.1. Design

A scholar's expertise (contributory or interactional à la Collins & Evans, 2019) in relation to GenAI is likely to shape both usage and attitudes. Specifically, we expected that *computational* scholars would bring a potentially unique perspective on GenAI as it has emerged out of the same computational techniques that many such scholars use (i.e., machine learning and NLP). However, comparatively, this group is a very small subset of the broader sociological community. We therefore make an effort to both identify these scholars from our larger sampling frame and oversample them. In addition to having a better understanding of how these technologies operate, we also expected that computational scholars will have a higher response rate due to intrinsic interest and will be more likely to use GenAI in some capacity. Furthermore, based on prior surveys of (generative and non-generative) AI usage in academia, we expected that men will also be more likely to respond (possibly, in part, because they are overrepresented among computational scholars). We therefore design the survey to allow us to adjust for some parameters of our population. Although we attempt to control for such bias, inferences to the entire field of sociology must be made with caution. See the schematic representation of our sampling strategy in the main text(Figure 1).

We build our sampling frame using a bibliometric approach with the Clarivate Web of Science (WoS) database. We collected all articles published in 50 sociology journals (see Table B.1) in the last 5 years (2020-2025).[6] This yielded a total of roughly 18,600 authors. Note that we include both sociologists (in title) and their collaborators.

Next, we classified an author in our sampling frame as "computational" if they authored an article that references specific terms (see Table B.2) in their title, abstract, or keywords. Our intent was to be broad in our boundary drawing, and thus capture the diversity of perspectives within the field of computational sociology. For additional precision, we also manually read a selection of titles and abstracts to ensure that we were capturing papers that engage with computational ideas and methods. In total, this yielded 985 authors who have written 405 computational articles in the last five years in sociology journals. To contact the authors, we combined the email addresses from the WoS records with manually searching for addresses posted publicly on faculty pages. Where an email "bounced" (i.e., rejected by the receiving server), one attempt was made to find an accurate address and resend the email. This resulted in a total of 879 authors in our computational sample with valid email addresses.

From the remaining authors (who had not authored a computational paper as defined above), we randomly selected 2,200 individuals. We again used the email addresses from the WoS records and manually searched for those with missing addresses, and attempted to find more up-to-date email addresses in the case of bounces. This

[6] As of 10 February 2025.

Table B.1: 50 Journals Used to Select Sociology Authors

Acta Sociologica	*Social Currents*
American Journal of Cultural Sociology	*Social Forces*
American Journal of Sociology	*Social Networks*
American Sociological Review	*Social Problems*
Annual Review of Sociology	*Social Psychology Quarterly*
British Journal of Sociology	*Social Science Research*
City & Community	*Society and Mental Health*
Criminology	*Socio-Economic Review*
Cultural Sociology	*Sociological Forum*
Current Sociology	*Sociological Inquiry*
Demography	*Sociological Methodology*
Environmental Sociology	*Sociological Methods and Research*
Ethnic and Racial Studies	*Sociological Perspectives*
European Sociological Review	*Sociological Science*
Gender & Society	*Sociological Theory*
International Journal of Sociology	*Sociology*
Journal for the Scientific Study of Religion	*Sociology Compass*
Journal of Health and Social Behavior	*Sociology of Education*
Journal of Marriage and Family	*Sociology of Health & Illness*
Journal of Mathematical Sociology	*Sociology of Race and Ethnicity*
Poetics	*Sociology of Religion*
Rationality and Society	*Socius*
Rural Sociology	*Symbolic Interaction*
Sexualities	*Work and Occupations*
Social Compass	*Work Employment and Society*

resulted in 1,943 invitations sent to individuals in our non-computational sample with valid email addresses.

We began fielding the survey at the end of January 2025, and continued until the end of June 2025. Of the roughly 3,000 individuals invited to complete the survey, 219 completed the survey from the computational sample (~24.1% response rate) and 214 completed the survey from the general sample (~11% response rate), for a total of 433 respondents.

Table B.2: Terms Used to Select "Computational" Articles

Category	Term
Generative AI	generative AI, genAI, chatGPT, OpenAI, GPT-2, GPT-3, GPT-4, DALL-E, Llama, MIXTRAL, MISTRAL, Stable Diffusion, prompt engineering, large language models, LLMs
Machine Learning	machine learning, artificial intelligence, deep learning, reinforcement learning, neural network
Natural Language Processing	natural language processing, NLP, text analysis, text mining, text as data
General	computational, data science, big data, algorithm, agent based modeling

Appendix B.2. Weighting

By inferring the gender of authors using their names and/or pronouns (in authors' biographies), we estimate that roughly 39% of the total sampling frame are men. However, this jumps to 64% for our computational subsample. Furthermore, using the base domains[7] of email addresses, we inferred authors' locations. While roughly 24% of authors in the total sampling frame were in the United States, this jumps to about 42% in the computational subsample. Thus, oversampling on computational authors, as well as likely self-selection, shifted our sample toward men in the United States (see Table B.3). We therefore use weights to correct for these discrepancies.

Specifically, we use raking (DeBell & Krosnick, 2009). This procedure forces univariate distributions of the *self-reported* variables gender and location to equal the *inferred* population parameters in Table B.3 (in the "Total" column). We also condition on the size of our computational subsample (i.e., those publishing a computational article in a sociology journal in the last 5 years), given that we oversampled computational scholars and also found that this subsample had a higher response rate (24%) compared to the general subsample (10%). All together, then, our raking weights force the marginal distributions of gender, US-based, and subsample type to equal the first column of Table B.3. For the sake of raking, we use a binarized version of the respective variables (cis-man or not, US-based or not, and computational sample or not).

[7]In the few cases where the email address used a generic domain like gmail.com, we confirmed the location of the primary affiliation using their faculty biographies.

Table B.3: Population, Sample, and Respondent Demographics

	Total (Inferred)	General (Inferred)	Computational (Inferred)	Respondents (Self-Reported)
Gender (% Men)	39.40%	38.1%	64.3%	54.3 %
Location (% US)	24.40%	23.4%	42.5%	46.7 %
Computational (Self-Reported)	4.95%	-	-	50.6%

The sum of the rake weights equals the total number of respondents in the total sample ($n = 433$). Each respondent-specific weight quantifies "how much" of the sample they represent as a function of their gender identity, location, and subsample. The maximum weight is 3.28 for non cis-men not based in the U.S. and who *have not* published a computational article in the last 5 years; the minimum weight is 0.04 for cis-men based in the U.S. who *have* published a computational article in the last 5 years.

In Table B.4, we present the raw proportion for a range of descriptive variables, and their rake-weighted counterparts (note that each variable is presented as a binary response). Importantly, we are inferring to the population defined by the 50 sociology journals, which may not represent the field of sociology by all definitions. After weighting across these three population parameters, we can infer that roughly 30% of the population contributing to the sociological literature considers themselves a quantitative (as opposed to qualitative or mixed-methods) scholar; 39% identify as cis-men; 36% are native English speakers; 34% are tenured professors; and 60% consider themselves racially white.[8]

Table B.4: Raw Proportions and Rake-Weighted Proportions on Select Variables

	Proportion (Raw)	Proportion (Rake)
Gender (% Cis-Men)	52.9%	39.40%
Location (% US)	45.7%	24.40%
Computational	54.7%	34.90%
Sociologist	71.8%	63.00%
Quantitative	40.0%	30.40%
Language (% English)	47.1%	36.70%
Race/Ethnicity (% White)	62.8%	60.90%
Position (% Tenured Prof.)	34.4%	34.07%

[8]Although our sample is quite international, the gender and race proportions are nevertheless closely aligned with the demographic breakdowns of the 2024 membership of the American Sociological Association: 39.2% cis-men and 54.4% white (https://www.asanet.org/diversity-equity-inclusion/dei-at-asa/asa-membership/current-membership-2024/).

Appendix C. Descriptive Bar Charts

The following Figure C.5 breaks down the research stages by individual tasks. For each panel the percentages sum to 100% per specific use case with the never-users omitted from the visualization.

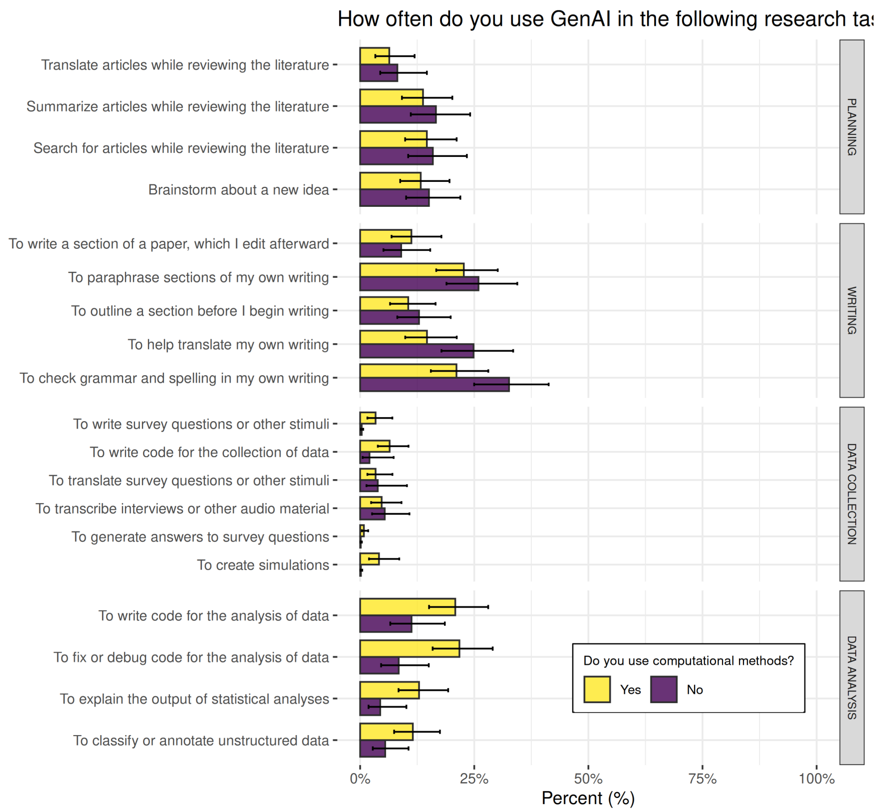

Figure C.5: Scholars GenAI Use Frequency by Specific Research Tasks

Note: $n = 334$, 216, 244, 143, and 200 for overall, planning, writing, data collection, and analysis, respectively, after removing respondents who responded having never used GenAI, and listwise deletion. Error bars are 95% confidence intervals.

The next figures visualize the distributions for all questions across each Likert-type category. The percentages sum to 100% per reason—meaning that, for example in C.6, about 17% of scholars who answer the question about GenAI affording them the opportunity to do more meaningful research are non-computational scholars who neither disagree nor agree with this reason.

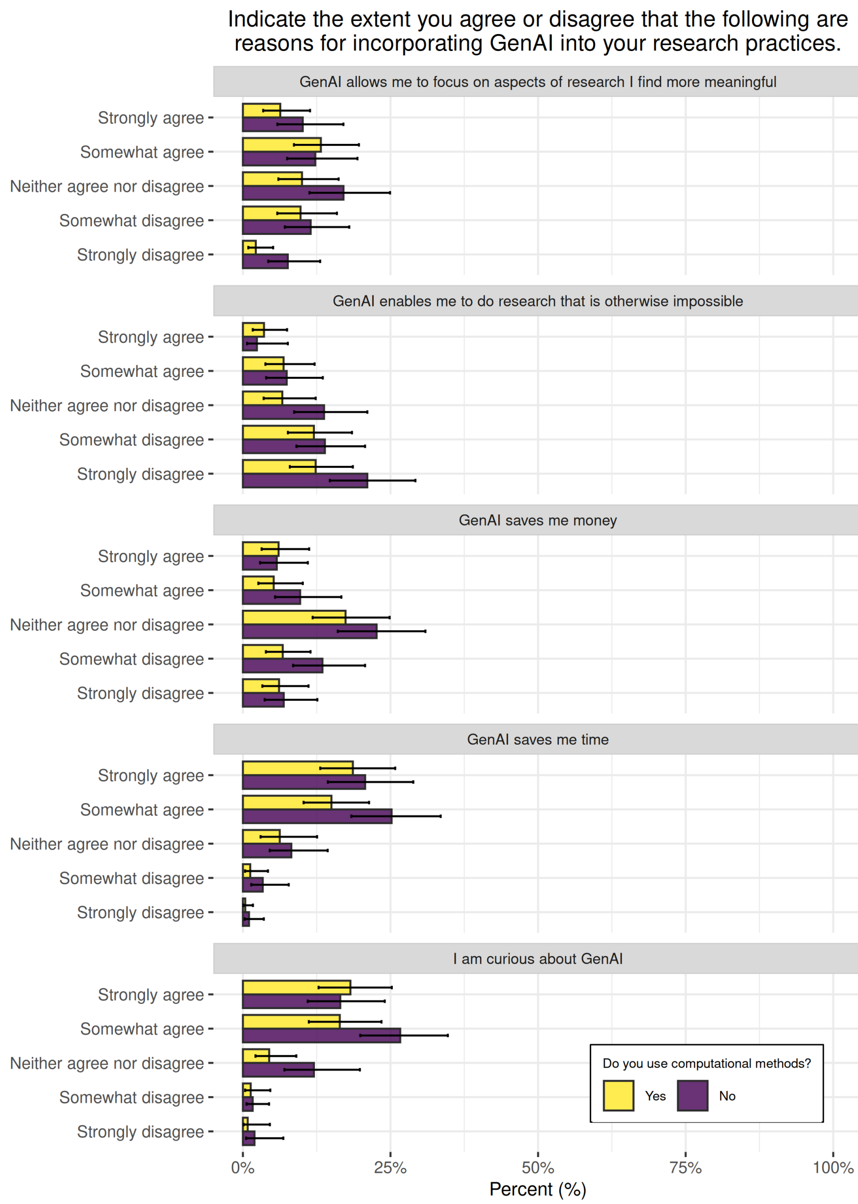

Figure C.6: Reasons for Using GenAI in Research

Note: $n = 310$ after listwise deletion. Error bars are 95% confidence intervals.

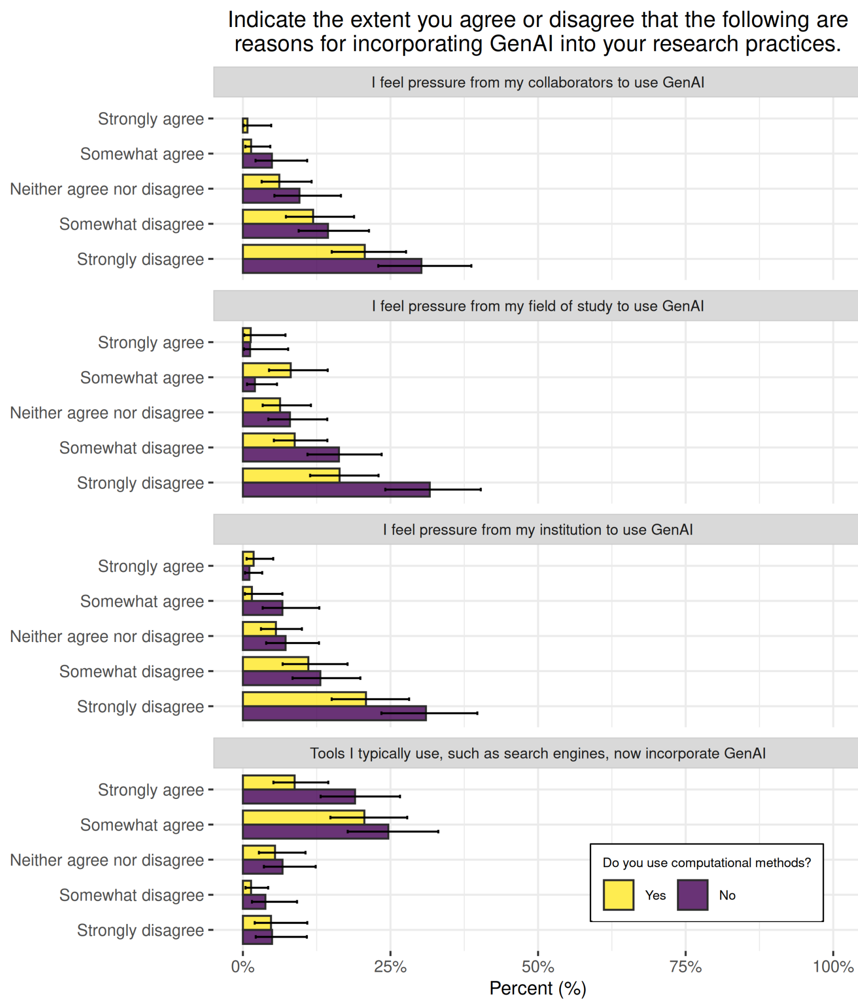

Figure C.7: Reasons for Using GenAI in Research: External Pressures

Note: $n = 308$ after removing respondents who responded having never used GenAI in their research, as well as listwise deletion.

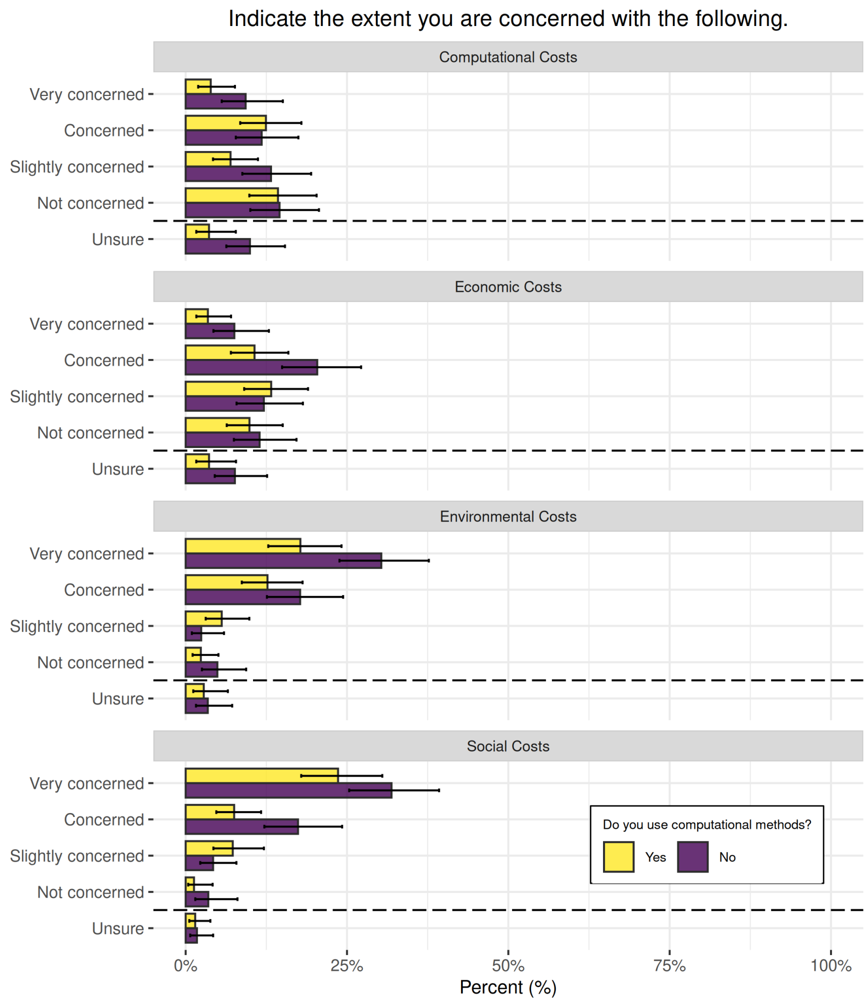

Figure C.8: Costs of Using GenAI in Research

Note: $n = 392$ after listwise deletion. Error bars are 95% confidence intervals.

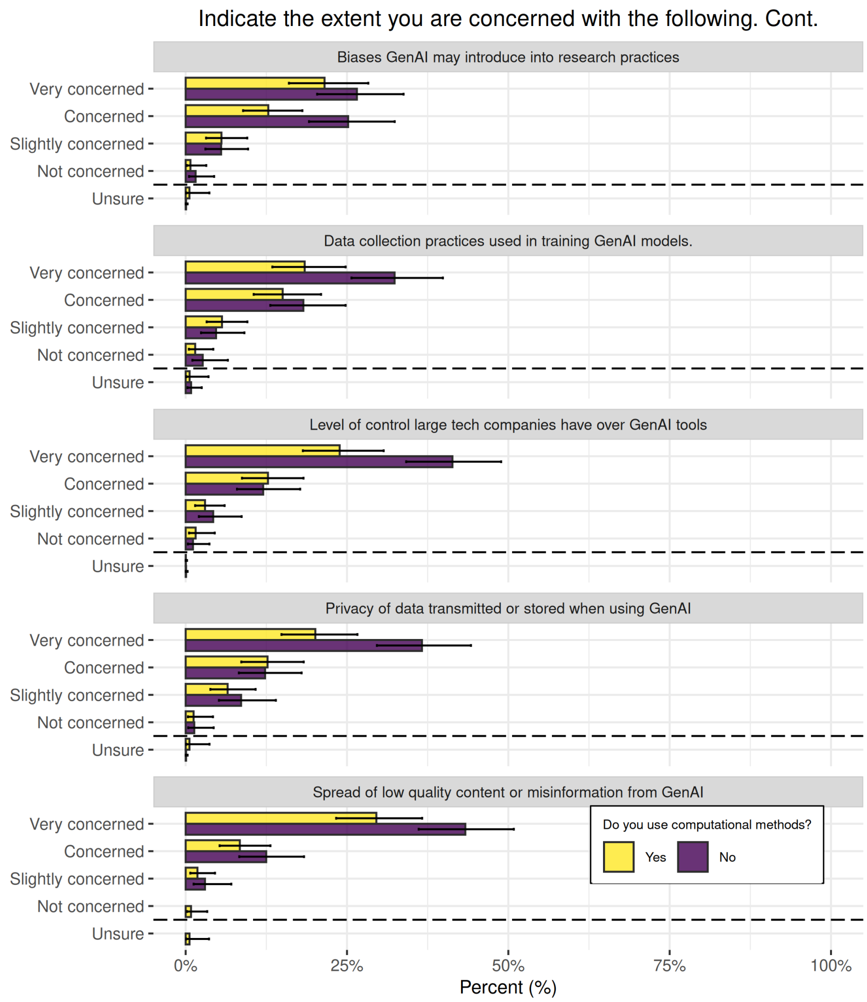

Figure C.9: Concerns about Using GenAI in Research

Note: $n = 393$ after listwise deletion. Error bars are 95% confidence intervals.

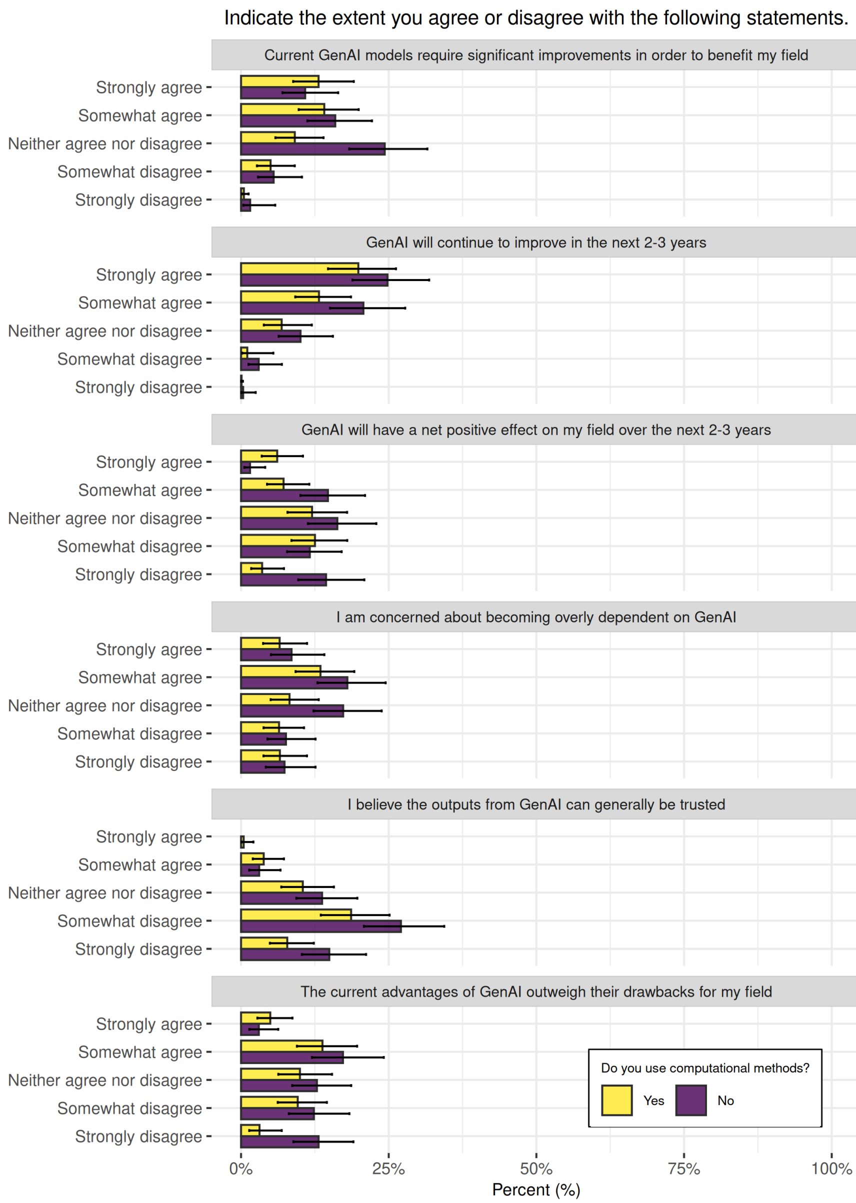

Figure C.10: Distrust and Optimism about Using GenAI in Research

Note: $n = 392$ after listwise deletion. Error bars are 95% confidence intervals.

Appendix D. Regression Model

Table D.5: Weighted Generalized Linear Estimates of GenAI Trust, Future Improvement, and Net Positive for Research

	Trust		Future Improvement		Net Positive	
	$\hat{\beta}$	se_{β}	$\hat{\beta}$	se_{β}	$\hat{\beta}$	se_{β}
Familiarity	0.044*	0.020	0.000	0.027	0.007	0.029
Use	0.138	0.727	−0.659	0.534	0.455	0.652
Familiarity × Use	0.035	0.064	0.098*	0.045	0.074	0.056
Not Cis-Man	−0.015	0.125	−0.144	0.115	−0.113	0.150
Intercept	1.566***	0.213	4.100***	0.292	2.325***	0.300
N	411		407		409	
Pseudo-R^2	0.144		0.115		0.300	
AIC	1025.377		982.277		1176.869	
Dev_{model}	286.636		264.318		426.790	
Dev_{null}	334.950		298.621		610.081	

Note: Coefficient estimates are unstandardized. Standard errors are design-adjusted (Lumley & Scott, 2017). $n = 411$, 407, and 409, respectively, after listwise deletion. Reference category for "Use" is "Not at least weekly." Reference category for "Not Cis-Man" is "Cis-Man." $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$ (two-tailed)

Appendix E. Alternative Model Specifications

Table E.6: Survey-Weighted Ordered Logistic Estimates of GenAI Trust, Future Improvement, and Net Positive for Research

	Trust		Future Improvement		Net Positive	
	$\hat{\beta}$	se_β	$\hat{\beta}$	se_β	$\hat{\beta}$	se_β
Familiarity	0.096^{*}	0.048	-0.006	0.062	0.000	0.052
Use	0.787	1.830	-2.260	1.420	0.461	1.190
Familiarity × Use	0.033	0.157	0.295^{*}	0.123	0.152	0.106
Not Cis-Man	-0.066	0.274	-0.345	0.285	-0.241	0.271
Intercepts						
$Y \leq$ Strongly Disagree	-0.039	0.529	-5.490^{***}	1.140	-1.180^{*}	0.564
$Y \leq$ Somewhat Disagree	2.110^{***}	0.544	-3.120^{***}	0.719	0.191	0.541
$Y \leq$ Neither Agree nor Disagree	3.960^{***}	0.576	-1.310	0.685	1.730^{***}	0.513
$Y \leq$ Somewhat Agree	6.120^{***}	0.849	0.405	0.677	3.390^{***}	0.580
N	411		407		409	
Dev_{model}	1005.255		888.167		1141.478	
Dev_{null}	1069.968		947.538		1286.297	
$\chi^2_{null-model}$	64.712^{***}		59.371^{***}		144.818^{***}	

Note: Standard errors are design-adjusted (Lumley & Scott, 2017). Estimates are expressed in log odds. $n = 411$, 407, and 409, respectively, after listwise deletion. Reference category for "Use" is "Not at least weekly." Reference category for "Not Cis-Man" is "Cis-Man." $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$ (two-tailed)

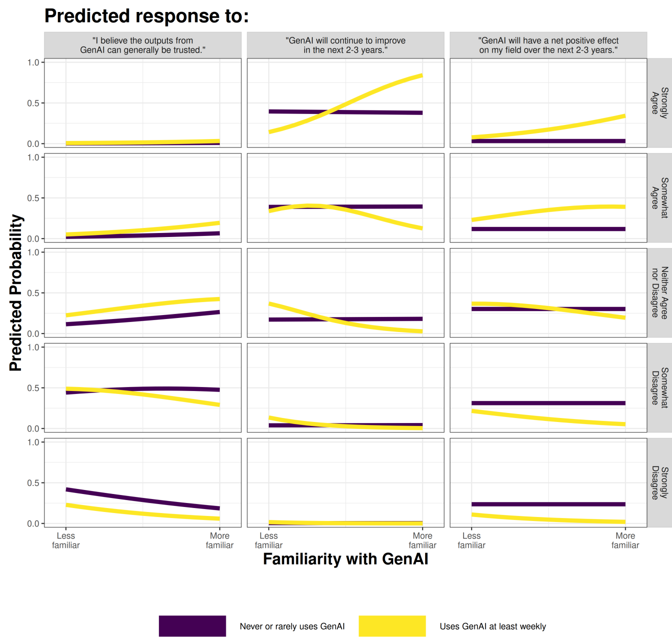

Figure E.11: Adjusted Predictions of GenAI Trust, Future Improvement, and Net Positive for Research from Survey-Weighted Ordered Logistic Regressions

Note: Coefficient estimates are unstandardized. Predictions based on n = 411, 407, and 409, respectively, after listwise deletion. Gender identity control is held at the modal value ("cis-men").