



We <3 Skynet: Or, How We Learned to Stop Worrying and Love the Bot Responses in Our Survey Data

Lynda Olman¹, Benjamin J. Hatchett², Meghan Collins³, Anne Heggli³, Kathryn Lambrecht⁴, Zach Tolby⁵, and Dominic Charles⁶

¹University of Nevada, Reno, Reno, Nevada, USA

²Colorado State University, Cooperative Institute for Research in the Atmosphere, Fort Collins, CO, USA

³Desert Research Institute, Reno, NV, USA

⁴Arizona State University, Mesa, AZ, USA

⁵National Oceanic and Atmospheric Administration, Global Systems Laboratory, Boulder, CO, USA

⁶Cognizant Technology Solutions, Teaneck, NJ, USA

Correspondence: Benjamin J. Hatchett (benjamin.hatchett@gmail.com)

Abstract. This case study in methods reveals a way that false responses to online surveys can be leveraged to add rhetorical value to technical communication research. Serendipitously, we discovered during our study of community commonplaces of weather-forecast discussions that open-ended bot responses contained phrases very close to rhetorical commonplaces generated by humans. Using freely available text-mining tools, technical communication researchers can easily mine false responses for indicators of community beliefs, norms, and values. This LESSONS Report (Limitations, Errors, Surprises, Shortcomings, Opportunities for New Science) exemplifies a surprising situation that could introduce limitations or errors (if unaddressed) but instead is converted to an opportunity to increase the robustness of survey data.

1 Introduction

“Victim”: Our team burst out laughing when we read this response to a survey question asking for respondents’ “role with respect to decision-making about weather hazards”. We were analyzing survey returns from a project funded by the Weather Program Office at the National Oceanic and Atmospheric Administration (NOAA) that aimed to improve the visualization of probabilistic forecast information for end users (Lambrecht et al., 2025; Olman et al., 2026; Lambrecht et al., 2026). But our laughter quickly died into silence as we read down the next 15 rows of our dataset: “victim, victim, victim, victim...”. By the time we had culled false responses generated by survey bots (programs written to find and fill out surveys) and farmers (humans who take as many surveys as possible, often with the assistance of bots, in order to “farm” the offered incentives), the robust return we had patted ourselves on the back for had dwindled to 20% of its original size (from $\approx$ 2500 total responses to 500), leaving us barely enough responses in cities like Pocatello ($n=46$) for our planned A/B testing. We had been bottled. In retrospect we realized we could have avoided false responses by following recommendations from the literature: including Captcha-type questions, distributing the survey directly by email, holding raffles for incentives rather than delivering incentives on completion, and eliminating example text from questions that can be scraped and entered (Griffin et al., 2022; Loebenberg et al., 2023; Storozuk et al., 2020). While we had adhered to many of these guidelines, our survey had no Captcha question and



contained example text that the bots and farmers—which we collectively referred to as Skynet after the artificial intelligence (AI) in the Terminator movies—could mimic. Readers may have experienced similar frustrations in their own research, as researchers are increasingly turning to online survey instruments to quickly harvest a broad range of user perceptions (see Luo et al. (2022) and Carradini (2020) for reviews of this trend and Fuller and La Sala (2021) and Dong and Gao (2023) for examples). Accordingly, we’re sharing the silver lining we found in the cloud of our Skynet responses, namely a reflection of user community norms, beliefs, and values—or commonplaces.

2 AI And Commonplaces

Our survey intended to harvest community commonplaces from responses to open-ended questions such as “What aspects of the forecast graphic were the most helpful”? Commonplaces are compact and stable rhetorical expressions of a particular community’s beliefs, values, and norms around managing a particular risk (Lambrecht et al., 2019). We were studying weather risk—specifically, visualizations of weather hazards featured in forecasts disseminated online. The pitch that won us the NOAA grant was as follows: if National Weather Service (NWS) weather forecast offices (WFOs) could know their individual forecast community’s commonplaces for discussing and managing weather hazards—which we could gather using an online survey—each WFO could better locally-tailor its forecasts by including those commonplaces in its forecast graphics (Lambrecht et al., 2019, 2021). This rhetorical strategy would promote more engagement with the WFOs and better risk decisions in the forecast community. We defined “forecast community” as the people who engaged with a WFO’s forecasts via their social media (non-experts) or core-partner email lists (experts), with each other in those forums, and with our survey when it was posted to those forums; in doing so, we were following Warner (2002)’s rhetorical definition of a community or “public” as a group that forms and interacts around a matter of common concern. Sure enough, our survey of expert and non-expert weather-forecast communities in five Western urban areas (Reno, Nevada; Portland, Oregon; Pocatello, Idaho; Phoenix, Arizona; and San Diego, California) surfaced multiple relevant commonplaces for each community: for example, Reno respondents shared commonplaces like, “I don’t worry about anything under 1 inch (2.54 cm) of snow”, and Pocatello respondents, “If it’s not windy as well, I don’t care how cold it is”. We also collected less-technical commonplaces bearing on personal or social norms, such as, “I’m retired, so I can just stay home and wait it out”, “I wonder if people with visual disabilities would be able to read this graph”, and “I’m fine, but my grandmother isn’t going to be able to shovel her walk if we get that much snow”. We identified these commonplaces with two coders, calculating interrater reliability using Cohen’s Kappa. We then incorporated the commonplaces into revised forecast graphics and compared reactions to the originals using A/B testing in a follow-up survey. In short, we found incorporating community commonplaces, both visually and verbally, into weather hazard forecasts generally produced faster and more confident risk decisions in the community. But our survey results are not actually the point of this article: we’re here to talk about Skynet. After our lead author had spent hours coding open-ended survey data for commonplaces, she happened to glance back over the false-response dataset and noticed something interesting: a lot of the commonplaces she had just coded were popping up in the Skynet set as well. How was this possible? She did a bit of research into how AI text generators worked and learned that they form sentences using a stochastic algorithm, adding

We tested this intuition by running our Skynet data through an open-source text mining suite called Orange: in particular, their word cloud and trigram analyses. Readers interested in using Orange can find a complete suite of YouTube tutorials and example workflows here: <https://orangedatamining.com/getting-started/>.) Figure 2.1 provides an example of partial word-cloud results.





Readers are likely familiar with word clouds, which express the frequency of key words in a text corpus (i.e., nouns, verbs, and modifiers) by font size and distance to the center of the cloud; in general, larger and closer means more frequent. However, readers may not be familiar with “n-grams” such as bigrams and trigrams. These collocations of two and three key words appear in sequence (ignoring determiners and prepositions) more often than they appear separately in a corpus; thus, they indicate common phrases or themes (see Table 1 for examples of trigrams). When we reviewed the results of the word cloud, we found that the most frequent words tended to reference the weather hazard in question (in this case, a forecast for freezing temperatures). However, the second tier or “halo” of words echoed language in our hand-coded commonplaces. These correspondences got stronger when we looked at the most frequent trigrams returned by Orange (Table 1).

Commonplace Family (with absolute frequency in 145 coded segments)	Examples of Coded Commonplaces within Family	Related Trigrams from Orange with frequency scores of 3 or above (with PMI scores)
Quantitative (n=30)	A good forecast graphic should help me think clearly about the degree or amount of impact for the hazard being forecast. It's hard to assess the risk of a hazard without sufficient information about interacting/intersecting hazards.	-different rainfall ranges (11.2) -specific wind forecasts (15) -associated potential risks (15.4) -rainfall intensity amount (9)
Personal/Community (n=14)	Graphics should help me pursue or protect my personal interests. Personal health and/or disability are important factors in assessing risks indicated by hazard forecast.	-consider business affect (15.3) -destinations types activities (19.3) -could affect travel (12.6) -goals priorities guide (21.5) -exigencies individual needs (20)
Cause/Effect (n=6)	Graphics should prepare viewers to take a specific action. The consequences of the hazard should be clear in the graphic.	-emergency response capabilities (19) -water seep building (19.9) -flooding soil waterlogging (17.9) -lead localized flooding (15.4)
Orientation (n=4)	Physical proximity to the hazard impacts my decisionmaking. It's important to know where high/low risk areas are to make the best risk decisions.	-location prevention contingency (15.6) -height basin location (15.7)
Timeline (n=1)	When impacts will occur is important to represent in a graphic.	-forecast early warning (16.9) -Reno Thursday night (16.4)

Table 1. Hand-coded commonplaces compared with trigram results from Orange.

It's important to note that Skynet couldn't “see” our forecast graphics as survey bots are generally unsophisticated AIs that can only search text that appears in question prompts (Storozuk et al., 2020); so, it produced nonsensical answers to open-



ended questions about the graphics themselves. Similarly, survey farmers won't spend the time required to analyze complex technical graphics so they can answer detailed questions about them; in our survey they took less than a minute on average (mean) to complete a survey that took valid human respondents about seven minutes to complete. However, when it came to echoing the other commonplaces in our dataset concerning the type of weather hazard, its impacts, and preventive actions, Skynet performed more closely to our valid human respondents (Table 1). As an example, the trigram "height basin location" appeared in the following response from Reno Skynet to the question, "What factors influenced your decision?": "The predicted rainfall and the height of the basin in the location, prevention and contingency". Meanwhile, a related commonplace from a human respondent in Reno read, "I consider the impact of geography on rainfall and factor that into my decision making". While the first is a somewhat garbled machine sentence and the second a coherent human argument, the two expressions nonetheless share the theme of considering geographic exposure when calculating weather risks.

Only this sort of holistic comparison is possible between the trigrams and the commonplaces because they are different things and are recognized differently. Orange returns by default a pointwise mutual information (PMI) score for each trigram (though you may select other statistics, such as chi-square, etc.). A PMI score is a log ratio (base 2) that compares the sequential frequency of the three words in the trigram to their independent frequency in the corpus; for example, a PMI score of 5 means the trigram in question showed up 32 times more often as a sequence than it did as 3 non-sequential words (for more about PMI scores in corpus analysis, see Recchia and Jones (2009)). When we manually coded commonplaces, by contrast, we holistically identified them and normed that identification between raters to an 80% or higher Kappa score (percent agreement adjusted for random agreement), a standard method in discourse analysis (Craggs and Wood, 2005; Walsh, 2018).

So then, why perform a collocation analysis of false open-ended responses if it doesn't yield all the results, or the rigor, of hand-coding? We argue, having performed both analyses, that the most frequent Skynet trigrams were a *reasonable* index to commonplaces in our human forecast-user communities—and the trigram analysis took a half hour to run while the hand-coding (and norming between raters) took from 60 to 120 hours for each dataset (the expert set was about half the size of the non-expert). As a result, on our next survey with open-ended responses, we plan to begin with a trigram analysis to generate a starter codebook; this new approach should speed up coding and norming significantly.

3 Conclusions

After comparing collocation analyses of false responses to our survey of community attitudes toward weather forecast graphics to our hand-coded analysis of the valid human responses, we believe that the former adds value to survey research in technical communication. Obviously, this "hack" won't apply to surveys without open-ended questions. But if researchers want to accelerate their hand-coding in a commonplace analysis, frame analysis (Lindekilde, 2014), or critical discourse analysis (Fairclough, 2013) of open-ended responses, our suggested hack should quick-start their codebook. Even if researchers are not planning a discourse analysis, but they find themselves with a batch of open-ended false responses (or intentionally false responses, e.g., Argyle et al. 2023), we suggest running them through a text miner like Orange to see what patterns appear.



Increasingly sophisticated AI-based bots will likely require additional survey safeguards and techniques to detect bot responses
110 (Höhne et al., 2025).

The resulting glimpse into community commonplaces can provide a starting point for a follow-up discourse analysis, or a focus-group exploration of community attitudes. If Skynet is going to meddle in our surveys, we should make it work for us, to push our technical-communication research in interesting new directions.

4 Ethical Statement

115 After review by the Arizona State University Institutional Research Board (IRB), our research was determined to be exempt (IRB #00012947). We also received NOAA Office of Management and Budget (OMB) approval (OMB #06900030).

Code availability. No code is available.

Data availability. No datasets are available.

Author contributions. LO: Data analysis, Methodology, Formal analysis, Visualization, Writing - Original draft. LO, BJH, MC, KL: Funding
120 acquisition, Supervision. All authors: Conceptualization, Writing - review and editing

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. This project was funded by the National Oceanic and Atmospheric Administration's Weather Program Office Social Sciences Program under Grant NA21OAR4590208. The statements, findings, conclusions, and recommendations are those of the authors and do not reflect the views of NOAA or the U.S. Department of Commerce.



125 References

- Argyle, L. P., Busby, E. C., Fulda, N., Gubler, J. R., Rytting, C., and Wingate, D.: Out of One, Many: Using Language Models to Simulate Human Samples, *Political Analysis*, 31, 337–351, <https://doi.org/10.1017/pan.2023.2>, 2023.
- Carradini, S.: A comparison of research topics associated with technical communication, business communication, and professional communication, 1963–2017, *IEEE Transactions on Professional Communication*, 63, 118–138, 2020.
- 130 Craggs, R. and Wood, M. M.: Evaluating discourse and dialogue coding schemes, *Computational Linguistics*, 31, 289–296, 2005.
- Dong, L. and Gao, Z.: Who are China’s Technical Communicators? A survey on the state of diversity, equity, and inclusion of the profession, *Technical Communication*, 70, 42–62, 2023.
- Fairclough, N.: Critical discourse analysis, in: *The Routledge Handbook of Discourse Analysis*, pp. 9–20, Routledge, New York, 2013.
- Fuller, R. P. and La Sala, A.: Crisis communication preparedness practices among US charitable organizations: Results from a national
 135 survey, *SAGE Open*, 11, 21582440211014 516, 2021.
- Griffin, M., Martino, R. J., LoSchiavo, C., Comer-Carruthers, C., Krause, K. D., Stults, C. B., and Halkitis, P. N.: Ensuring survey research data integrity in the era of internet bots, *Quality & Quantity*, 56, 2841–2852, <https://doi.org/10.1007/s11135-021-01252-1>, 2022.
- Höhne, J. K., Claassen, J., Shahania, S., and Broneske, D.: Bots in web survey interviews: A showcase, *International Journal of Market Research*, 67, 3–12, <https://doi.org/10.1177/14707853241297009>, 2025.
- 140 Lambrecht, K., Hatchett, B. J., VanderMolen, K., and Feldkircher, B.: Identifying community values related to heat: recommendations for forecast and health risk communication, *Geoscience Communication*, 4, 517–525, <https://doi.org/10.5194/gc-4-517-2021>, 2021.
- Lambrecht, K., Olman, L., Collins, M., Hatchett, B., Heggli, A., and Tolby, Z.: Tailoring probabilistic information to community needs: A rhetorical CRIB sheet, NOAA Technical Memorandum, OAR GSL 72, <https://doi.org/https://doi.org/10.25923/k9kh-y971>, 2025.
- Lambrecht, K., Olman, L. C., Collins, M., Heggli, A., Hatchett, B., Tolby, Z., and Wadhwa, I.: Bridging the Gap Between Numbers and Expe-
 145 rience: Identifying Community Risk Thresholds in Extreme Weather Communication, *Journal of Business and Technical Communication*, 0, 10506519261455 873, <https://doi.org/10.1177/10506519261455873>, 2026.
- Lambrecht, K. M., Hatchett, B. J., Walsh, L. C., Collins, M., and Tolby, Z.: Improving visual communication of weather forecasts with rhetoric, *Bulletin of the American Meteorological Society*, 100, 557–563, <https://doi.org/10.1175/BAMS-D-18-0186.1>, 2019.
- Lindekilde, L.: Discourse and frame analysis, in: *Methodological Practices in Social Movement Research*, edited by Della Porta, D., pp.
 150 195–227, Oxford, London, 2014.
- Loebenberg, G., Oldham, M., Brown, J., Dinu, L., Michie, S., Field, M., Greaves, F., and Garnett, C.: Bot or not? Detecting and managing participant deception when conducting digital research remotely: Case study of a randomized controlled trial, *Journal of Medical Internet Research*, 25, e46 523, <https://doi.org/10.2196/46523>, 2023.
- Luo, M., DeWitt, D., and Alias, N.: Mapping the evolutionary characteristics of global research related to technical communication: A
 155 scientometric review, *Technical Communication*, 69, 73–87, 2022.
- Olman, L. C., Lambrecht, K., Heggli, A., Collins, M., Tolby, Z., Hatchett, B., and Dannies, Q.: Visualizing Probabilistic Forecast Information Around Community Risk Thresholds: An Action-Research Collaboration Between Environmental Rhetoricians and the National Weather Service, *Commun. Des. Q. Rev.*, 14, 22–37, <https://doi.org/10.1145/3794916.3794919>, 2026.
- Perez-Castro, A., Martínez-Torres, M. D. R., and Toral, S. L.: Efficiency of automatic text generators for online review content generation,
 160 *Technological Forecasting and Social Change*, 189, e122 380, <https://doi.org/10.1016/j.techfore.2023.122380>, 2023.



- Recchia, G. and Jones, M. N.: More data trumps smarter algorithms: Comparing pointwise mutual information with latent semantic analysis, *Behavior Research Methods*, 41, 647–656, 2009.
- Storozuk, A., Ashley, M., Delage, V., and Maloney, E. A.: Got Bots? Practical Recommendations to Protect Online Survey Data from Bot Attacks, *The Quantitative Methods for Psychology*, 16, 472–481, <https://doi.org/10.20982/tqmp.16.5.p472>, 2020.
- 165 Walsh, L.: Visual invention and the composition of scientific research graphics: A topological approach, *Written Communication*, 35, 3–31, 2018.
- Warner, M.: Publics and counterpublics (abbreviated version), *Quarterly Journal of speech*, 88, 413–425, 2002.