

Is AI the Answer? Envisioning LLM Support for Responsible Computer Science Research

Zijing (Crystal) Zhao¹, Jessica Chen², Julie Walsh³, Eni Mustafaraj⁴

¹ Wellesley College

² Wellesley College

³ Department of Philosophy, Wellesley College

⁴ Department of Computer Science, Wellesley College

jc123wellesley.edu, zz102@wellesley.edu, julie.walsh@wellesley.edu, eni.mustafaraj@wellesley.edu

Abstract

CS researchers are grappling with emerging technologies changing their field at an increasing pace, yet lack support from their institutions for upholding research integrity. To address this lack of support in conducting ethical research, we investigate the potential of deploying an LLM-powered tool to provide easily accessible guidance on research ethics practices. We present findings from interviews with 16 computer science researchers at various stages in their careers, exploring their visions for how they would like to receive such guidance from an open-source LLM-based tool. Results from qualitative coding of the interviews reveal a wide range of possibilities: researchers imagined specific forms of support for ideation, development, revision, and final feedback. These findings highlight the potential benefits and challenges of using LLMs to support research ethics in CS.

1 Introduction

Computer science (CS) research can have profound societal implications, yet traditional oversight mechanisms like Institutional Review Boards (IRBs) often fail to address the unique ethical challenges that stem from the collection and analysis of large-scale user data (Metcalf and Crawford 2016; Vitak et al. 2017). Although researchers indicate that they have a strong interest in ethics support (Grosz et al. 2018; Singer 2018), existing design and reflection toolkits (e.g., Envisioning Cards, Value Cards) have low adoption due to friction and mismatch with researchers’ actual workflows (Friedman and Hendry 2019; Shen et al. 2021; Collins 2024; Do et al. 2023).

Large Language Models (LLMs) present an opportunity to offer researchers the ethical guidance they desire in a way that is conducive to their workflows. Recent systems specialized tools like Farsight and BLIP show that AI can effectively surface real-world ethical risks and brainstorm unintended consequences (Wang et al. 2024; Pang et al. 2024).

Previously, we conducted two studies on the topic of using generic off-the-shelf LLMs for ethical analysis. In the first study, we explored the concept of creating a tool for providing real-time ethics consults for researchers, using off-the-shelf, open-source LLMs (Hey et al. 2025). We noted that relying on generative Artificial Intelligence (GenAI) for ethical reasoning has its own ethical concerns, including privacy, risk of researcher deskilling, and the conditions of the

creation of the LLM model and its ongoing impact on environmental resources (Alicia Cappello 2024; Ranaldi and Pucci 2025; Selim et al. 2024). To further explore the potential efficacy of LLMs for ethics consults, in a second study, we compared an ethics consult produced by an off-the-shelf LLM to those produced by two human ethicists. We found that the LLMs produce comparable results to the ethicists and thus can already offer reliable support without fine tuning (Hey, Walsh, and Mustafaraj 2025).

While LLMs are effective at producing research ethics consults, it remains unclear whether researchers actively desire LLM-powered ethics assistance, and if so, in what specific form. Thus, we aim to answer this question:

RQ Do CS researchers want LLM support for research ethics and, if so, in what form?

To our best knowledge, this study is novel in providing and analyzing empirical results about social computing researchers’ perspectives on an LLM-powered tool for ethical research practices. By highlighting a diversity of researchers, ranging from experienced senior researchers to undergraduates, we enrich the ongoing conversation about developing community norms around how LLMs may be used to support research ethics.

2 Data & Methods

We recruited 16 computer science researchers with varying levels of research experience (2 to 20 years) and conducted interviews about their views on using an LLM-based tool to support their processes.

To prototype the ethics support tool, we selected Qwen3-14B (Qwen 2025), an open-source model on HuggingFace, following pilot evaluations comparing it against Deepseek-R1, Gemma-3-12b, and Mistra-small-3.1. Two ethics experts on our team preferred Qwen3-14B for its superior instruction-following capabilities and human preference alignment. Selecting an open-source model was important for research privacy and intellectual property, avoiding the risk of sharing ongoing research with commercial tools like Claude.

For each participant, we locally prompted the model with two inputs: a brief excerpt of their work (a paper *introduction* or ongoing research *protocol*, depending on status of research) to simulate limited early-stage information, and a

Table 1: Participant Demographics. The field “Input” refers to what was inputted to the LLM tool for the ethical analysis. Some researchers provided their introduction to a recently published paper and others a preliminary research protocol, depending on the status of their research.

ID	Field of Research	Current Role	Gender	Input
P1	Cybersecurity	Associate Professor	M	Intro
P2	Preference Learning	5th year PhD student	F	Intro
P3	Model Fairness	5th year PhD student	F	Intro
P4	Cybersecurity	Recent PhD graduate	M	Intro
P5	Machine Learning	Assistant Professor	M	Protocol
P6	Human-Computer Interaction	Assistant Professor	F	Protocol
P7	Statistical modeling	Postdoctoral Researcher	M	Intro
P8	Machine Learning	2nd year PhD student	F	Intro
P9	Computational Social Science	Postdoctoral Researcher	M	Intro
P10	Data Science	Undergraduate researcher	F	Protocol
P11	NLP and CSS	Postdoctoral Researcher	F	Intro
P12	Computational Social Science	5th year PhD Student	M	Intro
P13	Data Science	Undergraduate researcher	F	Protocol
P14	Machine Learning and NLP	Recent PhD graduate	M	Intro
P15	Multimodal models and biases	5th year PhD student	M	Intro
P16	LLMs in survey research	2nd year PhD student	M	Intro

step-by-step evaluation protocol to generate a structured ethical analysis report.

Procedure

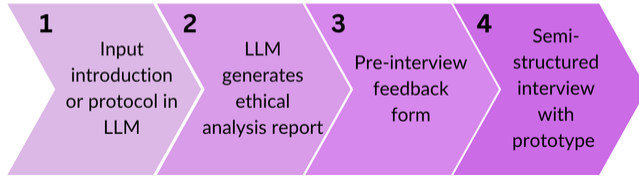


Figure 1: For each participant, we followed a four-step procedure to generate an ethics consult tailored to the participant’s research.

To evaluate LLM-based ethics support, participants reviewed a tailored ethical analysis report and interacted with a functional prototype. The report was structured across four categories: *Research Design*, *Data Collection and Analysis*, *Potential Impact of Study*, and *Risk-Benefit Analysis*. Following a pre-interview feedback form, participants completed semi-structured interviews paired with prototype testing to share their perceptions and envisioned uses of LLM tools in their research workflows.

3 Results

Participants provided a wide range of responses throughout our interviews. We group our findings according to two central insights, revealing the potential benefits and challenges of employing LLM-generated support for research ethics.

3.1 Insight 1: Potential Benefits of LLM Support

Ethics from Multiple Perspectives Participants (P1, P3, P5, P12, P15) imagined that LLM-support could offer dis-

tinct expert perspectives. P15 appreciated how the support would “provide another perspective into a potential ethical consideration or challenge” that they may not have considered or have been able to access. P12 imagined interacting with the tool in a “back and forth” manner so that it could “deliberately try and poke holes.” P3 and P5 liked to receive ethics support that represented “different stakeholders,” like policymakers and the “social science perspective” that is often lacking from a CS research project.

Suggestions for Specific Revisions Participants (P3, P9, P11) imagined that LLMs could be used as a different approach to revising and submitting research at CS conferences. To supplement current ethics checklists, which only include general considerations, P9 envisioned a tool with a “more interactive and personalized approach.” Furthermore, P3 and P11 wanted the tool to refine their paper after the writing process and before submitting to a conference.

Guidance for Young Researchers Participants (P3, P14) also expressed hope for using LLM-support to fill in gaps in education. P14 said that as a young researcher, “you don’t really know some of the ethics to consider.” P14 added that if inputting a research idea into the platform can generate a list of tailored, specific ethical concerns, then “it will be very beneficial to someone that is just starting out as a researcher.” Similarly, P3 also liked the idea of a “roleplay agent” that can “support brainstorming.”

Concrete and Structured Recommendations Participants (P2, P3, P4, P7, P9, P10, P12, P15) appreciated receiving concrete, structured outputs. To clarify dense or technical reports, P2, P3, and P4 emphasized the value of interactive follow-ups, step-by-step instructions, and grounded examples. P3 emphasized the importance of “ask[ing] follow-up questions” for particular issues so they could receive a “follow up response.” P4 also expressed a desire for the

option to ask for “concrete [instructions] and to have each step be accompanied with very clear, very detailed examples” when it comes to LLM-raised suggestions. Additionally, participants valued the report’s structured format for efficiency; P7 and P10 noted that a comprehensive layout allowed them to quickly prioritize key sections and engage the chatbot only when seeking deeper clarification on specific issues.

3.2 Insight 2: Potential Challenges of LLM Support

Reaching Beyond Research Scope Many participants (P1, P5, P6, P7, P10, P13) observed that the LLM-tool struggled delineating between what ethical concerns could or could not be addressed. P1 emphasized that there would always be limitations or ethical issues attached to a project, and it would be unfair to expect the researcher to fully resolve everything. P9 agreed that there was a gap between idealism and practicality, and said that they did not know “how to reach out to stakeholders, or which stakeholders to reach out to.” P3 further noted that interdisciplinary collaboration was not very useful for their own paper, since their research was “more technical.”

Discouraging Research Participants (P1, P6, P12) emphasized the need for a constructive tone that avoids discouraging researchers from pursuing an idea. P1 commented on this balance between encouragement and critique as they wanted the tool to “be friendly, but not completely barricade every research there is.” P6 added that any potential worry should be phrased “more as warnings rather than concrete issue[s],” especially if the tool is considering an early-stage idea. P12 stated that “softening the tone a little bit” is appropriate as researchers will continue to think about “what steps they might want to take” after receiving feedback.

Potential for Deskilling A few researchers with 7+ years in the field (P1, P5, and P6) warned that inadequately governed LLM support would replace critical thinking for earlier-career researchers. P1 stressed the importance of inconvenience in ethical thinking. P1 said, “Ethical deliberation requires writing. I ask my students to write ethical sections by themselves. [...] I don’t want them to slip into convenience.” P5 echoed this concern: “the point [of ethical analysis] not the product, it’s the process that led to it.” P6 worried that because early researchers may lack ethical reasoning skills, they are at risk of overly trusting LLM-output as they are unable to think about its suggestions critically and would just “assume the tool knew way more than I did” and “take all the considerations seriously.”

Misalignment with Research Process Two participants (P15, P16) worried about aligning LLM-support with their respective research processes. The LLM-based tool we provided required the researcher to input a piece of writing, like an abstract, a protocol, or a pre-registration. Participants reflected that it was difficult to present these writings at the outset of their study. A challenge is to delineate during which stage of research the LLM-tool should be embedded.

4 Discussion

Our goal in this paper was to evaluate if CS researchers imagined receiving LLM-generated ethical support, and how that support would look like. This resulted in the question:

RQ Do CS researchers want LLM support for research ethics and, if so, in what form?

First, we found that some participants (N=12) do want to receive LLM-support to fill current gaps in ethical research support. Participants envision that LLM support could help bridge gaps in the lack of accessibility to multiple perspectives on ethics, guidance on reacting to ethics-related conference reviews, specificity of ethics checklists, and overall ethics guidance for young researchers. These are aligned with current gaps in supporting responsible research ethics in CS, where there is a lack of access to cross-disciplinary collaborators (Moats et al. 2025), conference guidance (Frachtenberg and Koster 2020), and ethics education (Brown et al. 2024). This suggests that these gaps could be filled by employing tailored LLM-based tools.

Second, our results highlighted the challenges of integrating LLM-support for ethical research. This supports previous findings where people do not immediately trust LLMs assisting them in certain tasks, as they fear potential cognitive offloading and inaccuracies. Particularly, in our sample, researchers with more experience had more substantial concerns with LLM-ethics support.

Moreover, a number of participants (N=10) wanted LLM-support in some cases but not others, pointing to a tension within how each researcher wants support. Indeed, our findings reveal that researchers are typically cautious about the use of LLMs, especially in ethics, despite appreciating the unique benefits of such a tool. Still, as shown by previous studies, the starting baseline for ethics support is currently so low. As such, taking into account researchers’ cautions, we can use LLMs to create a higher baseline for ethics support to begin with.

4.1 Envisioning an ideal tool

An iterative, open-source support tool can democratize ethical reflection and raise current support baselines, which could be an improvement from the minimal human support that currently exists. However, guardrails are essential.

First, a tool must avoid one-size-fits-all approaches. Researchers of different backgrounds approach both research and ethics differently and so require different kinds of support. Constrained customizability of support is thus important for ensuring responsible research at all levels. For example, a tool could consist of multiple fine-tuned models adjusted to various modes, like providing concrete solutions for mid-career researchers and surfacing general potential issues for early-career researchers.

Second, the tool should integrate mechanisms, to reduce skill atrophy and the friction necessary to train researchers in ethical reasoning, such as the implementation of Socratic Dialogue.

Third, there is a unique risk of relying on LLMs: it can further reduce interdisciplinary dialogue at a moment when

it is more needed than ever. Therefore, there must still be simultaneous structural support for ethics in CS education and research, along with consideration for how a tool may be integrated to promote such support.

5 Limitations and Future Work

This study has two primary limitations. First, our small sample size (N=16) restricts generalizability across the CS community, though these qualitative findings establish the foundation for future large-scale surveys mapping broader community ideas on how LLMs could be integrated to support research ethics. Such future studies would provide more data on their needs and corresponding worries would inform a more thoughtful design of a tool (or set of tools). Second, using a comprehensive ethical report prototype as a baseline may have anchored participants toward techno-solutionism or constrained alternative visions of LLM support, though open-ended questioning helped mitigate this artifact bias.

Future research should explore what kinds of ethics interventions CS researchers most desire, the concrete steps to implement those forms of LLM-support, and how these desires promote the practice of responsible research ethics. In conjunction, researchers should delve into generational and career-stage variations in perspectives on research ethics, contributing to findings about responsible research practices. Lastly, future studies should investigate how an LLM-supported ethical analysis tool can support institutional interdisciplinary collaboration and education efforts, promoting human-in-the-loop approaches to ethical analysis.

In summary, CS researchers show strong interest in exploring the use of LLMs for ethical reasoning, but they also remain aware of the risks inherent to such exploration. We thus suggest that researchers, developers, and institutions must reflect on the depth and breadth of the current gaps in ethical support, and what sorts of support could be facilitated by LLMs with relevant guardrails in place.

6 Acknowledgments

We are thankful to the 16 participants for engaging with us on this project. We also acknowledge funding from NSF ER2 2220772 grant.

References

- Alicia Cappello, V. G. R. K. C. S. H. S., Muhammed Dada. 2024. Large Language Models and the Disappearing Private Sphere.
- Brown, N.; Xie, B.; Sarder, E.; Fiesler, C.; and Wiese, E. S. 2024. Teaching Ethics in Computing: A Systematic Literature Review of ACM Computer Science Education Publications. *ACM Trans. Comput. Educ.*, 24(1).
- Collins, G. 2024. A Murky World: Navigating Ethical Considerations in Data-Driven Research - College of Information (INFO).
- Do, K.; Pang, R. Y.; Jiang, J.; and Reinecke, K. 2023. "That's important, but...": How Computer Science Researchers Anticipate Unintended Consequences of Their Research Innovations.
- Frachtenberg, E.; and Koster, N. 2020. A survey of accepted authors in computer systems conferences. *PeerJ Computer Science*, 6(e299): e299–e299.
- Friedman, B.; and Hendry, D. G. 2019. *Value sensitive design: Shaping technology with moral imagination*. MIT Press.
- Grosz, B. J.; Grant, D. G.; Vredenburg, K.; Behrends, J.; Hu, L.; Simmons, A.; and Waldo, J. 2018. Embedded EthiCS: Integrating Ethics Broadly Across Computer Science Education. arXiv:1808.05686.
- Hey, S. P.; Walsh, J.; and Mustafaraj, E. 2025. Comparing Human and LLM Ethical Analyses: A Case Study in Computational Social Science Research. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 8(2): 1245–1254.
- Hey, S. P.; Weijer, C.; Walsh, J.; and Mustafaraj, E. 2025. EthicsBot: Fine-Tuning Open-Source LLMs to Assist Scientific Investigators in Analyzing Ethical Issues in Research. In *2025 IEEE International Symposium on Ethics in Engineering, Science, and Technology (ETHICS)*, volume 1, 1–5. IEEE.
- Metcalf, J.; and Crawford, K. 2016. Where are human subjects in Big Data Research? the emerging ethics divide. *Big Data Society*, 3(1).
- Moats, D.; Holtrop, T.; Jan, Varga, J.; Waltman, L.; and Dechesne, F. 2025. Making problems: interdisciplinary collaboration and AI ethics. *Science as Culture*, 1–26.
- Pang, R. Y.; Santy, S.; Just, R.; and Reinecke, K. 2024. BLIP: Facilitating the Exploration of Undesirable Consequences of Digital Technologies. arXiv:2405.06783.
- Qwen. 2025. Qwen3 Technical Report. arXiv:2505.09388.
- Ranaldi, L.; and Pucci, G. 2025. When Large Language Models contradict humans? Large Language Models' Sycophantic Behaviour. arXiv:2311.09410.
- Selim, R.; Basu, A.; Anto, A.; Foscht, T.; and Eisingerich, A. B. 2024. Effects of large language model-based offerings on the well-being of students: qualitative study. *JMIR Formative Research*, 8: e64081.
- Shen, H.; Deng, W. H.; Chattopadhyay, A.; Wu, Z. S.; Wang, X.; and Zhu, H. 2021. Value Cards: An Educational Toolkit for Teaching Social Impacts of Machine Learning through Deliberation. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, FAccT '21*, 850–861. New York, NY, USA: Association for Computing Machinery. ISBN 9781450383097.
- Singer, N. 2018. Tech's Ethical "Dark Side": Harvard, Stanford and Others Want to Address It. *The New York Times*.
- Vitak, J.; Proferes, N.; Shilton, K.; and Ashktorab, Z. 2017. Ethics Regulation in Social Computing Research: Examining the Role of Institutional Review Boards. *Journal of Empirical Research on Human Research Ethics*, 12(5): 372–382.
- Wang, Z. J.; Kulkarni, C.; Wilcox, L.; Terry, M.; and Madaio, M. 2024. Farsight: Fostering Responsible AI Awareness During AI Application Prototyping. 1–40.