Check for updates

"I'm categorizing LLM as a productivity tool": Examining Ethics of LLM Use in HCI Research Practices

SHIVANI KAPANIA* and RUIYI WANG*, Carnegie Mellon University, USA
TOBY JIA-JUN LI, University of Notre Dame, USA
TIANSHI LI, Northeastern University, USA
HONG SHEN, Carnegie Mellon University, USA

Large language models are increasingly applied in real-world scenarios, including research and education. These models, however, come with well-known ethical issues, which may manifest in unexpected ways in human-computer interaction research due to the extensive engagement with human subjects. This paper reports on research practices related to LLM use, drawing on 16 semi-structured interviews and a survey with 50 HCI researchers. We discuss the ways in which LLMs are already being utilized throughout the entire HCI research pipeline, from ideation to system development and paper writing. While researchers described nuanced understandings of ethical issues, they were rarely or only partially able to identify and address those ethical concerns in their own projects. This lack of action and reliance on workarounds was explained through the perceived lack of control and distributed responsibility in the LLM supply chain, the conditional nature of engaging with ethics, and competing priorities. Finally, we reflect on the implications of our findings and present opportunities to shape emerging norms of engaging with large language models in HCI research.

CCS Concepts: • **Human-centered computing** → **Empirical studies in HCI**; **User studies**; • **Computing methodologies** → *Natural language generation.*

Additional Key Words and Phrases: research practices, research ethics, large language models, HCI research

ACM Reference Format:
Shivani Kapania, Ruiyi Wang, Toby Jia-Jun Li, Tianshi Li, and Hong Shen. 2025. "I'm categorizing LLM as a productivity tool": Examining Ethics of LLM Use in HCI Research Practices. *Proc. ACM Hum.-Comput. Interact.* 9, 2, Article CSCW102 (April 2025), 26 pages. https://doi.org/10.1145/3711000

1 Introduction

The rapid development and adoption of large language models (LLMs) is reshaping the research and education landscape, sparking interest and concerns across many disciplines [28, 61, 113]. Large language models have also presented opportunities to complement research workflows in Human-Computer Interaction research, including the analysis of qualitative data [47, 123] and quantitative data [74, 80], replicating human-subject experiments in social sciences [3], and facilitating ways to simulate emerging social dynamics at scale [91].

However, in contrast to the excitement towards the potential of LLMs, a growing body of work has surfaced risks associated with these models [61, 119], including misinformation, discrimination and exclusion, malicious use, and more. Cheng *et al.* [23] demonstrated how GPT-4 simulations of

*Both authors contributed equally to this research.

Authors' Contact Information: Shivani Kapania, skapania@andrew.cmu.edu; Ruiyi Wang, ruiyiwan@andrew.cmu.edu, Carnegie Mellon University, Pittsburgh, PA, USA; Toby Jia-Jun Li, University of Notre Dame, Notre Dame, USA, toby.j.li@nd.edu; Tianshi Li, Northeastern University, Boston, USA, tia.li@northeastern.edu; Hong Shen, Carnegie Mellon University, Pittsburgh, USA, hongs@andrew.cmu.edu.

This work is licensed under a Creative Commons Attribution 4.0 International License.
© 2025 Copyright held by the owner/author(s).
ACM 2573-0142/2025/4-ARTCSCW102
https://doi.org/10.1145/3711000

specific demographics (*e.g.,* marginalized race/ethnicity groups) and topics are highly susceptible to caricature. Participants' data privacy is also at risk, with evidence showing how LLM-based tools might leak sensitive information with or without malicious prompting [17, 131].

Examining the use of LLMs holds particular relevance for CSCW research, given our frequent interactions with human subjects, engagement in community-collaborative efforts [72], and interest in designing human-AI collaboration systems with a socio-technical approach [11]. LLMs are being deployed in several interactive settings within CSCW research (*e.g.,* writing assistance [60, 110], software development [117]). The HCI and CSCW communities have also demonstrated a longstanding commitment to understanding the impacts and ethical considerations of emerging technologies in research practices, from the use of videotapes in the early 1990s [72] to more recent re-examination of the use of Reddit data in social computing research [39]. Over the years, the SIGCHI research ethics committee has been facilitating open conversations about ethical challenges in our communities through research ethics town halls and panels at conference venues such as CHI [37, 43, 77] and CSCW [12, 35]. Researchers within the community have also organized several workshops and meetings to discuss the ethical challenges of HCI research, such as how to conduct large-scale user trials [20], how to engage with vulnerable populations [2, 66] and how to conduct research with public data [38].

Despite HCI's rich tradition of centering the discourse on research ethics, the rapid uptake of emerging LLM applications has brought renewed urgency to examine and collaboratively shape norms for LLM use [102]. However, there is a gap in our understanding of HCI researchers' current practices surrounding LLMs, and uncovering them can offer a critical view into how they navigate ethical considerations. In this research, we ask: (1) How do HCI researchers integrate large language models in their projects? (2) What ethical concerns, if any, do they have regarding using LLMs? (3) How do HCI researchers approach and navigate those ethical concerns?

We report our results from 50 survey responses and 16 in-depth semi-structured interviews with HCI researchers using LLMs in their work. Across our participants, we find that LLMs were utilized throughout the entire HCI research process, from ideation to system development and paper writing. Large language models were perceived to open new possibilities for building tools and interactions, generating research ideas, and simplifying workflow for analysis and writing. We also came to see how researchers increasingly integrated LLMs into their everyday practice.

Our participants anticipated a wide range of potential ethical issues associated with LLMs, such as potential harms in interacting with LLM outputs, privacy concerns, violation of intellectual integrity, and overtrust & overreliance. While HCI researchers acknowledged these ethical considerations, in many cases, they were either unable to or only partially able to identify and address those ethical concerns in their projects. Many participants highlighted their perceived lack of control with their position in the LLM supply chain [121], a lack of established best practices, and competing priorities that took precedence over addressing ethical concerns.

We begin by situating our study within prior work on LLM-based tools to support research workflows, scholarship on research ethics in HCI, and literature on the ethical issues with large language models. After presenting a detailed description of our methods, we present our findings on HCI research practices surrounding the ethics of LLM-based tools. Finally, we reflect on these results and present implications for the HCI research community. Taken together, our research underscores the importance of foregrounding research ethics if we are to continue integrating LLMs into our work practices. We call for engaging with IRBs and other regulatory institutions, redesigning effective informed consent processes, developing tools and processes to interrupt the LLM supply chain, providing learning opportunities for ethics of LLM use in HCI, and shifting existing academic incentive structures to foreground ethical considerations in research.

2 Related Work

2.1 The Use of LLMs in HCI Research

As the capabilities of generative AI in understanding context and generating natural language responses continue to advance [84], large language models are increasingly used by researchers as tools in Human-Computer Interaction (HCI) research. Emerging work has explored the potential of leveraging LLMs to support the process of brainstorming and identifying research topics [49, 96, 109]. Tools like CoQuest [68] have been developed for research question development through human-agent co-creation. In addition to research ideation, LLMs have been used as tools in human-AI co-creative design ideation [71, 114] and writing support [130].

Other members of the community have made efforts to develop LLM-powered applications for data generation and data analysis [90]. Hämäläinen *et al.* [50] explored the potential of utilizing LLMs to produce synthetic user interview transcripts for piloting research ideas and designing interview protocols. Wei *et al.* [118] created LLM-based chatbots for generating synthesized user self-reported data. Researchers have also developed LLM-based applications for qualitative analysis, including identifying themes and generating codebooks [16, 47, 112]. Prior work has also explored using LLMs for deductive coding [25] and human-AI collaborative qualitative analysis [46, 47, 124]. In terms of quantitative analysis, tools such as Github Copilot[1] and GPT-4 [84] can enable researchers to transform natural language instructions into programming codes that assist quantitative data analysis and visualization. LLMs have been used as tools for quantitative data analysis in sampling, filtering, and analyzing survey data [56], gaining insights from large corpus [85] and crowdsourcing social data [34].

Researchers have also used LLMs as the underlying technology for system design and development Lu *et al.* [69], and Petridis *et al.* [93] explored infusing prompt-based prototyping enabled by LLM into functional user interface design. Researchers also utilized LLMs to build task-oriented social simulations for LLM agents, aiming to study the intricate social dynamics of societal systems [45, 65, 92]. Other applications of integrating LLMs into the design and development of HCI systems have focused on leveraging LLMs for efficient prototyping [64, 122] that HCI researchers and designers can potentially use. Our research extends this scholarship by empirically examining practices and ethical challenges of integrating large language models into HCI research projects.

2.2 Research Ethics in HCI

The Human-Computer Interaction (HCI) community has long been engaged in discussions regarding ethics [36, 95], revolving around responsible conduct in human subjects studies [11, 98], including privacy, informed consent, and institutional review boards (IRBs). These concerns are integral to ensuring ethical conduct and preventing harm to research participants [9]. Existing HCI privacy guidelines focus on protecting participants' sensitive information, the anonymity of the participants, and the confidentiality of collected data [15, 132]. The rights of autonomy and self-determination of participants require HCI researchers to establish transparent procedures for informed consent on collecting and analyzing personal information [42]. IRBs are critical in ensuring that HCI researchers meet their ethical obligations when conducting studies with human subjects [5, 14]. Moreover, prior scholarship underscores the need to mitigate potential biases in research design, data collection, and data analysis by refining existing ethical guidelines and protocols [78].

Creating ethical guidelines for emerging technologies within the HCI community has drawn inspiration from ethical theories and professional standards. Mackay [72] highlighted the need to go beyond legal requirements and develop comprehensive guidelines for responsible behavior by learning from other disciplines. Vitak *et al.* [115] encouraged interdisciplinary collaboration in building new

[1]https://github.com/features/copilot

ethics frameworks, such as developing ethics heuristics for online data research to ensure responsible research. Despite the rich tradition of considering and iterating ethical guidelines in the HCI community, Computer Science (CS) researchers have demonstrated uncertainty about the applicability of ethical concerns about human subject studies to their research [13], highlighting the necessity of educating CS and HCI researchers about research ethics. Recent endeavors of integrating ethics in CS education, such as adding systematic literature review on ethics in computing courses [99, 105] have illustrated both the barriers and opportunities of educating CS researchers in ethics.

Recently, advances in AI have brought renewed urgency of research ethics to HCI and CS communities. Clark *et al.* [26] proposed approaches to assessing ethical risks of research involving digital data, informing the development of ethical standards guidelines for research. Amershi *et al.* [1] provided guidelines for human-AI interaction design, emphasizing the importance of offering explanations of the systems and conveying the consequences of user actions. Another line of research strives to bridge the gap between ethics and AI practices by establishing guidelines for safe, transparent, and trustworthy AI systems [104]. With the rapid development of generative AI, the Association of Computing Machinery (ACM) and the Association for Computational Linguistics (ACL) have established ongoing efforts to develop guidelines [19, 40]. The ACM policy on authorship requires the entire disclosure of generative AI in the paper [40], while the ACL policy on AI-assisted tools on paper writing requires authors to elaborate on the scope and nature of their use [19].

Even though underlying ethical guidelines may be broadly applicable, emerging technologies might challenge existing ethical review processes [77]. Our goal is not to create a taxonomy of all possible ethical concerns with the use of LLMs in HCI research; instead, we hope to extend the discourse on research ethics in HCI by documenting the ways in which researchers are responding to ethical considerations of LLMs in-situ and reflecting on potential ways forward.

2.3 Ethics of LLM use

Extensive prior research has explored the ethical risks and harms related to language models [119, 120]. The discrimination & exclusion harms arise from the biased and unjust text in the training data of LLMs. Recent work identified the tendency of LLMs to display discrimination related to users' sensitive characteristics [125] and demonstrated the gender and racial biases of LLM-generated content [33]. The information hazards are the consequences of LLMs remembering private information in training data, posing the risk of privacy leaks [17]. Privacy violation has been observed in LLM-based AI assistants where Personally Identifiable Information (PII) can be revealed by employing adversarial privacy-inducing prompts [70, 75]. The detection and mitigation of hallucinations in LLM-generated text [129] remains a challenging problem [21]. This might lead to the spread of misinformation, as LLM-based chatbots are often treated as fact-checking tools [129]. There is a growing concern about malicious activities arising from the generation of scams and phishing using LLMs [76]. Recent research has also pointed out how the anthropomorphization of LLMs has the potential for manipulation and negative influence [30]. Finally, researchers have also documented the likelihood of LLMs increasing inequality and their adverse effects on job quality, undermining creative economies, and more [120].

In response to these concerns, recent research has started to look into how to assess [29] and mitigate [22, 76] the harms of LLMs. For example, researchers are exploring privacy-preserving strategies for language models in pre-processing, training, and post-training approaches [106]. For combating misinformation in LLMs [22], researchers have proposed several defense strategies, including integrating a misinformation detector [89] and redesigning prompting methods to guide LLMs [89, 128]. Mozes *et al.* [76] presented prevention measures for when LLMs were used for illicit purposes, arguing that red-teaming, safeguarding with RLHF and instruction following, and avoiding memorization and poisoning are potential solutions to the misuse of LLMs. Regarding human-computer interaction harms of LLMs, Liao and Vaughan [67] provided a roadmap for improving the transparency and explainability of LLMs.

While recent efforts have begun to mitigate the broader ethical concerns of LLMs, the application of LLMs in HCI research presents unique ethical challenges. For instance, biases in the training data can adversely affect research practices, leading to issues with internal and external validity, reproducibility, efficiency, and the risk of proliferating low-quality research [3]. Despite this, there is a notable gap in understanding how HCI researchers perceive and manage these ethical challenges in their day-to-day research activities. This study aims to fill this gap by exploring the unique ethical issues posed by using LLMs in HCI research and examining how researchers currently address these challenges.

3 Methods

In this research, we focus on examining the ways in which HCI researchers apply large language models (LLMs) across their research workflows and their ethical considerations for using LLM-based tools. For a holistic view of LLM use practices, we employed a mixed-method approach with a sequential explanatory design [55]. This involved conducting a survey to elicit broad-brushed and higher-level perspectives from a wide audience. After closing the survey and analyzing the results, we developed several questions for our semi-structured interview guide to better understand the observed phenomena. These questions included, for example, *"How do you address ethical concerns when using LLMs?"* and *"What strategies do you employ to mitigate risks when using LLMs?"* The interviews then focused on investigating, in more detail, how researchers approach the ethical considerations of using LLMs as part of their research activities.

The qualitative data served as the foundation for our inquiry and analysis [55]. The survey responses were integrated to support and contextualize the interview findings. Specifically, the survey provided quantitative insights into the relative degree to which participants are using large language models across various stages of their work, which were then further explored and explained through the qualitative interviews. Our sequential explanatory study design ensured that both data sources complement each other. Our research study was reviewed and approved by the IRB at our institution. We present our approaches to the survey and interviews in the following subsections.

3.1 Survey

The goal of the survey was to identify the ways in which HCI researchers are using LLMs and any ethical considerations they have encountered in their projects. We conducted the survey using an online questionnaire implemented in Qualtrics and analyzed responses from 50 respondents.

Participant recruitment. We recruited survey participants through multiple channels: advertising on social media networks such as Twitter and LinkedIn, emailing direct contacts, and leveraging university distribution lists to which we had access. We began the survey by eliciting informed consent from respondents. No personally identifiable information was recorded about the respondents following our organization's research privacy and ethics guidelines. The inclusion criteria for our survey were similar to the interviews, where we recruited researchers who are currently studying or working in areas related to Human-Computer Interaction (HCI) and have used large language models (LLMs) in their research.

After the screening questions, we were left with n = 77 participants. Out of the 77 respondents, 50 completed all sections except for the demographics (which was optional). Among the 43 survey respondents who filled in the demographic questions, most reported working in academia (34), industry (6), and non-profit (3) organizations. Researchers worked on projects within Human-AI Interaction (32), Design (13), Understanding People: Theory, Concepts, and Methods (12), Collaborative and Social Computing (10), and User Experience and Usability (10). In our sample, respondents were located in the United States (20), Afghanistan (5), Germany (3), Algeria (2), Hong Kong (4), China (1), Spain (1), Nigeria (1), Australia (1), Japan (1). On average, respondents had 4 years of experience working on HCI research projects.

Questionnaire. Our questionnaire consisted of 18 questions in total, with a mixture of multiple-choice (14) and open-text questions (4). We began the survey by describing LLMs as *"a subset of generative language (and multimodal) models with increasing size as measured by the number of parameters and size of training data" [4]* (e.g., *GPT-4, GPT-3.5, Llama 2, Vicuna, and more*). We then ask respondents to refer to their latest research project involving LLMs and respond to the following questions related to that project only. The survey was divided into three sections: (1) questions about their use of LLMs in their HCI research projects, (2) questions about how they engage with ethics of LLMs use in HCI research, and (3) demographic questions relevant to our study. In the first section, respondents were asked to describe the project in one sentence and share the primary research method, sub-area of HCI, and the stage of their research process where they incorporated LLMs in their project.

After understanding the HCI research project in which they used LLMs, we focused on their potential ethical considerations with the use of LLMs. We asked, *"have you encountered or observed any ethical challenges related to LLMs in your research project?"*. This was followed by the question asking what the ethical challenges are in the form of a close-ended (with choices such as security and privacy, consent, harmful outputs, copyright issues, authorship, prefer not to say) and an open-ended question, respectively. We asked how they identified, potentially mitigated, and reported these ethical challenges. Finally, we also included demographic questions (optional to answer) about the respondent's type of institution, country, and years of experience in HCI. Respondents were also invited to optionally share their email addresses if they were interested in participating in a follow-up interview study.

Analysis. We computed a range of descriptive statistics using SPSS to better understand researchers' approach to ethical concerns with LLMs. These included descriptive statistics to questions with multiple choice answers (*e.g.,* the ethical challenges in using LLMs). In cases where a subset of respondents completed questions, we report question-specific response rates and the percentage of respondents who answered that question. Finally, we conducted a qualitative analysis of open-ended questions following the same approach to the interviews (see the following Section 3.2 below). We performed multiple rounds of coding at the response level in conjunction with participants' survey ratings to surface high-level themes. We include direct quotes from our survey respondents in the Findings with the prefix 'S#' to differentiate them from our interview participants, prefixed with 'P#'.

3.2 Interviews

Between October and November 2023, we interviewed 16 HCI researchers who used LLMs in their research projects. Each interview had structured sub-sections beginning with the participants describing a recent project where they applied LLMs across any of their research activities to gain more context for the following questions. The survey responses informed the interview questionnaire's design (e.g., further exploring the ethical concerns raised in the quantitative data). Participants could take a few minutes to look at their history with the LLM tools/applications. Our interviews focused on (1) LLM use across the research workflow, (2) specific ethical considerations, (3) the process of navigating ethical considerations, (4) the role of IRBs, (5) the role of ethical frameworks and toolkits; (6) incentives and accountability. Each session focused on the researchers' practices and the associated ethical considerations.

Participant recruitment. We recruited participants through a combination of distribution lists, professional networks, and personal contacts, using snowball and purposive sampling [88] iteratively until saturation. Our sample included researchers located in the United States (13), China

Type	Count
Institution Type	Academia (14), Industry (2)
Gender	Female (6), Male (10)
Expertise	Human-AI Interaction (5), Privacy and Security (2), Computational Social Science (1), Educational Technology (1), Social Computing (1), Intelligent User Interface (1), UX Design (1), Creativity Tools (1), Mobile Computing (1), AI Ethics (1), Augmented and Virtual reality (1)
Location	USA (13), Singapore (1), China (1), Germany (1)
Experience	0-5 years (5), 6-10 years (3), 10-15 years (1), N/A (7)

Table 1. Summary of interview participants' institution types, domain expertise, and demographics.

(1), Singapore (1), and Germany (1). We interviewed 10 male and 6 female researchers. Most of our participants work in academia (14), with two participants from industry. Refer to Table 1 for details on participant domains, institution types, and demographics.

Interview moderation. Given the geographical spread of our participants, the interviews were conducted online using video conferencing. We scheduled sessions based on the participants' convenience and conducted all interviews in English. During recruitment, participants were informed of the purpose of the study. Informed written consent was obtained electronically for all interview participants. Participants were informed that they could refuse to answer any questions and/or ask for the recording to be paused at any time. Each session lasted about 40–60 minutes each. We recorded interview notes through field notes and video recordings and transcribed them verbatim for analysis. Each participant received a gift card of 30 USD for their participation.

Analysis. To analyze the qualitative data, we followed the reflexive thematic analysis approach by Braun and Clarke [7, 8]. Reflexive thematic analysis foregrounds the researcher's role in knowledge production, with *'themes actively created by the researcher at the intersection of data, analytic process, and subjectivity'* [8]. Two research team members independently read each interview transcript multiple times, starting with open coding instantiations of perceptions, ethical considerations, and challenges with using LLMs. They met regularly to discuss diverging interpretations or ambiguities. The entire research team met frequently to define themes based on our initial codes iteratively. We transcribed 664 minutes of audio recording and obtained 336 first-level codes. As we generated themes from the codes, we also identified categories (units of analysis) with descriptions and examples of each category. These categories included (1) the use of LLMs across the research workflows, (2) ethical concerns, (3) approaches towards navigating ethical concerns, (4) barriers in addressing ethical issues, and (5) organizational structures. These categories were discussed and iteratively refined through meeting, diverging, and synthesizing into three top-level categories, presented in our Findings. Since codes are our process, not product, IRR was not used [73, 107].

4 Findings

We begin by describing how HCI researchers are using large language models across various stages of their research process (Section 4.1). We then describe their ethical concerns (Section 4.2) and the four ways in which researchers engage with potential ethical concerns (Section 4.3). In our interviews, the most commonly used LLM-based tools were general-purpose offerings such as ChatGPT, GPT-4 API, Bard, and Bing Chat.

Fig. 1. LLMs are used in various ways for ideation and project scoping, study design and execution, and analysis and paper writing. The figure illustrates a typical HCI study across our research participants. Not all HCI research projects would include all activities listed above (e.g., critical theoretical contributions).

4.1 Where do HCI researchers use LLMs in their everyday work?

HCI researchers referred to various parts of their research workflow where they integrated large language models, such as for ideation, literature review, study design, data analysis, system building and evaluation, and paper writing (Figure 1). Overall, they perceived that LLMs open up new possibilities in their research, such that *"if we can leverage LLMs the right way, it will enable us to do new cool things that will be genuinely empowering"* (P4). Our survey revealed that the stages where LLMs are most frequently used are paper writing (25) and study design (24), followed by project scoping (17) and system development (16), and then data generation (15), data collection (14) and data analysis (13). Below, we present the ways in which interview participants incorporate LLMs in their research practices.

Interview participants frequently used large language models in the **ideation and project scoping** phase for tasks such as reviewing and synthesizing literature, discovering new research questions in their sub-field of HCI, and defining their research problems. P11 would input broad topic areas into the LLM to generate HCI research questions and refine them into concrete research objectives. Similarly, P10 would probe the LLM to *"pretend that it is a career coach for [participant name]. What would [the LLM] recommend if [participant name] is writing their NSF career grants? This is a big thing for early career HCI academic researchers. What should [participant name] explore at the intersection of AI and cybersecurity?"* During brainstorming, HCI researchers found value in using large language models for a breadth-first search approach, enabling them to generate a diverse range of ideas quickly.

LLMs were also helpful in **data generation, collection, and analysis**. Many participants mentioned how LLMs were incredibly productive in synthesizing information from web sources that would otherwise require significant time and effort. P1 prompted the model to generate multiple arguments for and against hypothetical scenarios for use in a classroom setting. They noted how creating these research artifacts for data collection would have otherwise taken weeks. Additionally, HCI researchers integrated LLMs into their data analysis process, utilizing them for tasks such as qualitative data coding, creating plots, and data visualizations. P7 applied LLMs for multiple aspects of open coding interview data, including (1) proposing new codes, (2) acting as a mediator between the coding team consisting of two members, and (3) generating the primary code groups.

HCI researchers described how they increasingly relied on LLMs in the **paper writing** stage. Participants shared their experiences of leveraging LLMs for iteratively refining their paper drafts, including searching for synonyms and fixing grammatical issues. LLMs offered a distinct perspective on aspects that the researcher *"can hardly think about from [their] vantage point. Interacting with the*

LLM is like talking to other researchers and getting their feedback" (P11). Researchers also mentioned using LLMs to generate paper reviews (P14) and provide suggestions for their writing content and style (*e.g.,* as an alternative to Grammarly[2] (P4)). P14 spoke of using GPT-4 to *"provide critical paper reviews, check grammar, the style, and the logic consistency"*.

Finally, many HCI researchers used LLMs as a design material for **system development**, including system building and evaluation. Participants shared many use cases, such as developing new tools and interactions for library services, science communication, and AI-assisted writing. P4 expressed interest in exploring *"what do large language models enable us to do in HCI with which our field has struggled for many decades."* Finally, researchers in our study also discussed instances where they used large language models for software development, including day-to-day debugging and creating web interfaces with LLMs.

4.2 What are HCI researchers' ethical concerns with using LLMs?

Researchers expressed a wide range of ethical concerns regarding the use of LLMs. Across our survey respondents, 30 reported observing ethical challenges associated with using LLMs, 10 expressed uncertainty, and 10 indicated no awareness of such concerns. Among the self-reported ethical challenges, the most common issues were data privacy (19) (including secondary use of participants' data and data leak), authorship (16) (including disclosure of the use of LLMs in publications), harmful outputs (14), copyright issues (11) and consent (10). These concerns were more prevalent across research design, system development, and the research's analysis and paper writing stages (see Figure 1). Below, we describe the specific ethical concerns highlighted by HCI researchers in our interviews.

4.2.1 Harms of engaging with LLM outputs. Our interviews revealed a shared concern of research subjects engaging with potentially harmful outputs, especially if LLMs were integrated into interactive systems and tools enabling real-time user interaction and an immediate feedback loop. The risk of harm (from discriminatory or hateful responses) might be amplified in settings where HCI scholars are working with and centering vulnerable populations. For example, P3 used large multimodal models to create a mental well-being and self-reflection tool and articulated their concern about LLMs generating uncontrolled outputs. In such contexts, LLM-generated content could disproportionately harm marginalized groups through exposure to socially harmful biases and stereotyping behaviors. Hate speech and exclusionary and discriminatory language could cause psychological and representational harm to research participants. P16, developing LLM-based roleplay robots, articulated how *"[LLM] may exaggerate or diminish some of the capacity with which the robot has already been equipped. It may introduce another layer of bias [toward people with disabilities]. "*

When deploying LLM-based services in unmoderated field user studies, participants expressed concerns beyond issues with biased LLM outputs. They worried about controlling LLMs so the tools were not misused or abused. Users might circumvent the constraints and use LLMs for tasks outside of the original purpose of the studies. Such uncontrolled use could even cause harm to other users. For example, P14 worried about the possibility of LLM facilitating the *"surveillance and monitoring of intimate partners"* by providing links to various spy tools.

Many researchers expressed concerns about large language models impacting their research pipeline by generating seemingly authoritative but fabricated information. LLM hallucinations during paper writing could mislead authors and, if published, would undermine trust in knowledge production processes. P10 used LLMs for literature review and worried that *"ChatGPT might make up titles of things that simply do not exist."* Participants highlighted the need for vigilance in identifying these hallucinations, especially when LLMs produced fake citations or mismatched paper references. Beyond paper writing, researchers indicated the possibility of inheriting biases

[2] https://app.grammarly.com/

from LLMs in ideation, research design, and evaluation. P2 articulated how LLMs could introduce *"gender bias in the entire computational pipeline– in places that we've never had gender bias before because we've never used LLMs before, and because we used to come up with our own solutions"*

Participants pointed out the tendency of LLMs towards generating homogenous content, where the model would prioritize generalization and converge diverse perspectives into standardized outputs. P8, working on sensemaking tools for product selection using LLMs, described how they *"want the selection criteria to encompass a diverse set of opinions rather than just focus on the perspective of a single demographic."* This tendency of LLMs towards *"flattening human diversity and nuance"* (P9) raised concerns among researchers who emphasized the importance of capturing the complexity of lived experiences within the context of their research. Participants also noted concerns that using LLMs to edit their paper drafts could influence researchers' writing styles, streamlining towards homogeneous and *"bland way of writing"* (P9). Participants also discussed how LLMs' training data, often reflecting Western morals and values, could contribute to this homogeneity.

4.2.2 Threats to privacy of participant data. HCI research studies often involve the collection of sensitive personal or behavioral data. This includes collecting and analyzing participants' backgrounds and lived experiences or their interactions with a technology probe. Researchers in our study expressed anxiety about how LLM providers are using participant data input into LLMs and the potential for privacy violations. They worried about breaches of private or sensitive information involving various forms of participant data, such as transcripts, audio recordings, and application-specific log data. P7 used LLMs for qualitative data analysis and discussed how *"privacy issue is the main concern for us because in many cases, the audio transcripts are not supposed to be put into ChatGPT".* P16 also shared the concern that users' navigation data is extremely sensitive and could lead to material physical harm if uploaded to a model endpoint.

Many participants emphasized their concern about confidential and personally identifiable information leakage due to sharing user data with LLM providers. P8 mentioned the risk of confidential information being incorporated into the LLM training corpus and the potential for security leaks, where *"backend messes up and user's private information shows up in someone else's chat history".* Relatedly, some participants expressed concerns about the possibility of unconsented data, for example, *"explicit sexual content"* (P2), being a part of the training corpus for large multimodal models. While some believed that manually erasing sensitive information could mitigate these concerns, others relied on the efficacy of APIs in safeguarding user data. The complexity and opacity of 'LLMs as a service' make it challenging to ensure that user study participant data is handled appropriately and securely throughout the research process.

4.2.3 Violations of intellectual integrity. Interview participants raised ethical concerns about the intellectual integrity of using LLMs in HCI research. A central theme of these concerns was the ambiguity of ownership of LLM-generated text and visuals. Many participants who co-created with LLMs also highlighted the difficulty in attributing what portion of the content was the researcher's original work and what was generated by the LLMs when they were refining and co-creating with the LLM outputs. Participants discussed the ambiguity related to plagiarism when LLM outputs are part of the research contribution. For example, interviewee P10 would consider crediting the LLM *"where a substantial amount ends up in a publication".* While most participants questioned the extent to which one can claim ownership of LLM-generated content, especially in the paper writing stage, some believed that *"personhood is important for attribution"* (P2).

In addition, HCI researchers also raised concerns about the reproducibility of the research results obtained via LLMs, highlighting the potential for an illusion of efficacy if LLMs work well in some cases but fail to generalize to others. P2 described using LLMs in HCI research as a *"computational Wizard of Oz"* method and continued to point out how the quick and opaque updates of LLMs are another barrier to reproducibility, noting that researchers *"don't really have control of what version of GPT they are talking to and something that might work in a previous version won't work so much in future versions"*.

4.2.4 Overtrust and overreliance on LLMs. Participants also noted ethical considerations of overtrust and overreliance on LLMs. They expressed concerns that research participants directly interact with LLMs but are unaware of model biases (due to limited AI literacy), which could overestimate LLM capabilities and place unwarranted trust in model outputs. This can be particularly problematic in LLM-based HCI applications offering decision-making support. For example, in a healthcare application, an LLM might provide medical advice that appears credible but is incorrect or inappropriate for the user's specific condition. P13, who built LLM-powered creativity support tools, was concerned that *"[research participants] may be misled by the content, even if [the content] contains wrong information or wrong references."*

Participants also expressed concern about researchers' overreliance on LLMs, which could compromise the quality and creativity of HCI research. P3, using LLMs for creative ideation for their study, warned about the risk that researchers might uncritically accept the information provided as factual and integrate it into their experiment or user interface design. Researchers raised validity concerns about using LLM-based generative agents as proxies to simulate user behavior and social interactions in designing new technologies. They argue that the behaviors generated by LLMs may fail to accurately capture the complexities and unique positionality of real research subjects, leading to findings that oversimplify the nuanced lived realities of individuals. Overreliance on LLMs in the iterative prototyping process could result in designs that are optimized based on model outputs rather than diverse user needs and experiences, ultimately compromising the effectiveness of the system. Participants mentioned how the overuse of LLMs among HCI researchers could impede paradigm shift (*e.g., "LLMs are useful but not creative"* (P15)) and raise questions about the long-term impact of such overreliance on LLMs on trust and integrity within the academic community.

4.2.5 Environmental and societal impacts. Interview participants identified a range of other higher-level ethical considerations that extended beyond their specific research project or pipeline. Some researchers were concerned about the environmental degradation due to the extensive electricity usage and hardware deployment of building *larger* LLMs. For instance, P2 mentioned *"we have all these models competing against each other, that's like millions of dollars of electricity and computer components, [which is] really bad for the environment"*. Participants articulated anxieties that the wide adoption of LLMs could lead to inequitable distribution of benefits– *"we certainly have not addressed the social issues. I think a lot of people will, if not losing their jobs, be substantially diminished in terms of their utility to their workplace"*.

4.3 How do HCI researchers approach the ethical concerns of using LLMs?

Our study focused on understanding how HCI researchers are currently navigating ethical considerations about using LLMs in their projects. For most HCI researchers in our interview study, the ethical concerns they discussed (Section 4.2) remained largely speculative. Although our participants reported their awareness of a diverse set of potential ethical concerns around using LLMs, they were either unable or only partially able to identify or address those ethical concerns in their projects. Researchers described their current strategies towards ethical concerns– often manifesting as forms of *inactions or workarounds*– such as engaging LLM ethics in a conditional and reactive manner, inconsistent disclosure practices, limiting their LLM use and reflecting on their co-creation process, as well as delaying responsibility in research accountability. In most cases, these strategies did not directly address the underlying ethical challenges; instead, they served as temporary fixes to manage immediate concerns. Below, we capture these varied strategies for LLM ethics and return to their implications in section 5.

4.3.1 Conditional and reactive engagement with LLM ethics. HCI researchers often emphasized that the need to and approach towards addressing ethical concerns was conditional on various factors such as the use case or domains of study. Many HCI researchers invoked the specifics of their research domain to justify why they did not need to foreground ethical concerns. For example, when participants categorized their research as low-stakes, they spoke of how commonly anticipated ethical issues associated with LLMs did not apply to their work and did not need to take proactive measures. For instance, P11, who focused on creativity support in writing, considered their research non-sensitive and did not find it necessary to intervene to prevent unsafe generated text. Some participants believed that the potential harm caused by LLMs was comparable to that caused by social media. Therefore, they felt that engaging in ethical considerations was unnecessary if LLMs were not being used in high-stakes domains, such as providing medical advice: *"typical LLM like ChatGPT was created to support people browsing the internet. People already see toxic content online, right? So we always claim this tool is no more harmful than typical social media"* (P13). If researchers explored topics deemed safe or if users did not directly prompt the model, there was a perception that the LLM was unlikely to generate unsafe content[3].

Such 'conditional engagement' sometimes resulted in a more reactive, rather than proactive, approach to ethical considerations. For example, when deliberating whether LLMs should be attributed and held accountable for model-generated content, P2 emphasized *"I think many people are very hasty to say yes or no. And I think that's not the answer. The answer is always in a gray area."* They continued to emphasize this "wait and see" approach, discussing the potential benefits of model hallucinations and highlighting their role in promoting divergent thinking by introducing specificity that the user may not have considered. Some participants mentioned being aware of frameworks and toolkits (such as Perspective API [58]) designed to address ethical issues. Still, they never incorporated these tools into their research processes. P3 justified this by discussing how a significant portion of their HCI studies were conducted in a laboratory setting. The ethical concern related to participants encountering harmful outputs generated by LLMs was less probable in a short usability test. At the same time, serious issues could occur in longitudinal studies when the participants heavily use the system.

4.3.2 Limited disclosure practices. In our study, HCI researchers positioned large language models as everyday tools within their research practice. As a result, participants did not believe reporting their usage of LLMs to study participants formally, the Institutional Review Board (IRB), and/or the broader academic community was necessary. In particular, participants described a shift towards the tacit incorporation of LLMs: they slowly transitioned from research tools that were a part of their regular practice to using LLMs. This change was partly due to the perception that LLMs are *'fancier'* (P11) and more advanced versions of previously used tools. P10 explained, *"I advise my students that they are allowed to use any generative AI tool just like they would use other productivity tools. So I'm categorizing LLMs as a productivity tool."*

When using LLMs in paper writing, participants drew a parallel between LLM reporting practices and their approach to previous tools. They emphasized that if, for example, tools like *"grammar assistance by Grammarly, word check by Google Docs, or accessibility checking by Acrobat pro"* (P10) were not explicitly reported, the same should also hold true for large language models. Researchers expressed reservations, suggesting that reporting of LLM-based tools might call into question the validity of their work. P10 captured this perspective:

[3] In contrast, prior work has demonstrated that LMs can generate unsafe content from seemingly innocuous prompts [48].

> *I would not feel it is appropriate to say Bing Chat was used to define the initial structure of the X and Y sections of the paper. To me, it is not gonna be helpful in researchers assessing the credibility or validity of the work. It is just like a meta issue about how the actual document was formed and refined. And what matters in that case is the output. Is it easy to read? Did you find it useful? You know, that's what I care about. So, in those cases, I do not credit the LLM.*

The complexity of describing LLMs to audiences without technical background further impacted participants' willingness to disclose the specific uses of large language models in research. In some cases, researchers chose to characterize their LLM use simply as *'AI models'* to their research participants so as not to *"confuse them with what is a language model."* The rationale behind this approach was the perception that the broader public (target population in this case) tends to view AI in a homogeneous manner, and *"to them, there isn't much difference between how different AI systems work, or which is a large language model, right?"* (P8). Researchers also justified this approach by highlighting their intention not to burden participants with the need to modify their behavior and " *think about how to interact with an LLM"* (P15). Finally, participants also noted cases where explicit disclosure about LLMs was unnecessary if research participants or system users were not directly exposed to LLMs. For example, if LLMs were used for ideation or analysis, participants felt that explicit communication about their use was not imperative.

4.3.3 Restricting LLM use and reflecting on the co-creation process. A recurring sentiment among the researchers in our study was a perceived lack of control in addressing ethical concerns related to large language models. As a result, they developed various workarounds, such as restricting the use of LLMs to a limited set of tasks, avoiding directly integrating LLM-generated inputs into their work, and hosting group reflection sessions. The limited visibility and decision-making capability were particularly evident in privacy and data leaks, where the reliance on LLMs provided by large companies diminished individual control. Simultaneously, participants expressed a mistrust towards LLM providers. For instance, P12 discussed how LLM providers' claim to protect data privacy and not using it for their training *"is very problematic because such claims don't articulate what they mean by not using the data. How can an external researcher validate this claim?"* Participants' mistrust towards LLM providers was exacerbated by the lack of clarity and transparency in data usage policies.

To navigate the *'unknown territory'* of LLMs, where HCI researchers were not fully aware of the capabilities and risks, they would often restrict their usage of LLMs to a limited set of tasks. Most HCI researchers avoided directly integrating LLM-generated outputs into their work. Rather than accepting the initial output without scrutiny, they would iteratively refine and carefully verify it before incorporating it into their research artifacts. In the context of paper writing, researchers would ensure the text aligns with the draft they had input into the system. Many participants elaborated on their practice of co-creating with LLMs as a precautionary measure. This would involve relying on their judgment on how much to use the LLM suggestions. This cautious and iterative co-creation process was a strategy without explicit usage norms.

Indeed, many participants shared the view that LLMs are still evolving, and there is a lack of comprehensive guidelines on navigating ethical considerations. P9, primarily using LLMs for paper writing, mentioned how *"LLMs are still an unknown territory, so people don't know how to react, I assume."* Our participants expressed discomfort using new LLM-generated content in their papers. For example, P15 was hesitant to use the LLM as anything more than a spell checker. Researchers highlighted the importance of applying LLMs in a way consistent with their practices with previously used tools to mitigate potential unintended consequences. P9 spoke of their experience as a reviewer for major HCI conferences like CHI, where they did not encounter any guidelines regarding the disclosure of using ChatGPT for writing. Finally, they emphasized that the cost and accountability of using LLMs was still *"ad-hoc and unclear."*

In cases where participants recognized the possibility of ethical concerns with LLMs, they often struggled to identify specific issues to address. As P16 pointed out: *"the main problem is that I don't know what bias it has, and I don't know how to figure it out."* P4 discussed how it is crucial to *"not pretend that this ethical consideration does not exist, which some people do. We are trained in human-computer interaction, so we are well equipped to reason about it or at least understand that these concerns exist, but addressing it is very difficult."* In many cases, when people did not have mechanisms to identify or address ethical concerns, they felt it was essential to at least acknowledge them and spread awareness about these issues.

Some researchers chose to host group reflection sessions to account for the uncertainties with the responsible use of LLMs. This involved discussing with collaborators or advisors to determine the appropriate approach for navigating any ethical considerations. P10 exemplified this approach by holding regular team meetings to address questions or concerns about using LLM-based tools or methods. After an internal discussion with their team, P9, too, decided not to use LLMs for qualitative analysis. A collaborative decision-making process was a common practice to determine where LLMs are appropriate.

4.3.4 Delaying responsibility in research accountability. Finally, our participants indicated that determining who bears responsibility for the ethical implications of using LLMs is challenging. At times, they expressed reluctance towards stringent regulatory guidelines and opted to postpone dealing with ethical issues in their projects.

In discussions about accountability with using LLMs within HCI research, participants noted the ambiguity in determining where responsibility lies regarding the ethical concerns of using LLMs. They highlighted the notion of *distributed responsibility* across the AI supply chain, emphasizing that multiple stakeholders share the responsibility for ensuring that the use of LLMs within research is ethical. Participants highlighted the LLM provider's responsibility to implement safeguards preventing users from sending sensitive personal data to the model. For instance, P8 discussed their confidence in LLM providers, sharing how they *"believe OpenAI has done a decent job in making a model safe in general."* P2 presented a similar perspective:

> *"When professionally something wrong happens, it needs to follow the chain of command. So, who is the person most directly responsible? It's actually the user, right? The user did something with my system, and it created a harmful output, and then the user will say, well, the system wasn't designed well enough for me to know it's gonna create this harmful output. I should have been warned. In that case, the responsibility falls on me. And then I could move it up the chain of command to say my advisor shouldn't have let me release this system if it was gonna produce harmful output, or I could say OpenAI is irresponsible for releasing a project that they are publicizing as being broadly accessible and safe."*

Partially due to the difficulty in determining "distributed responsibility" for some HCI researchers, addressing ethical concerns could be relegated to future work. They viewed it as imperative, as HCI researchers focused on system-building, to explore emerging technologies and develop new tools and interactions with them. As articulated by P6, *"we are the system builders, and we think about the unique benefits the new technology can bring to the users and enhance the user capabilities."* This perspective underscored the desire to continue building tools, even as researchers acknowledged the biases associated with LLMs.

Sometimes, participants even expressed resistance towards prescriptive regulations, which they perceived would suppress innovation in HCI research. They emphasized that external constraints on LLM usage, especially given its role in everyday research practices, would slow down their research. LLMs were also perceived to lack the stability necessary for regulation. As P12 described, *"the challenge is the regulation needs to come much later rather than early. If you come up with regulations too early, you may kill innovation, and on the other hand, you don't know what you want to regulate because the representation of the product hasn't been stabilized yet."*

5 Discussion

Our results highlight the various ways in which HCI researchers integrated LLMs in their research projects for the purpose of research ideation (*e.g.,* finding novel research areas), data collection and analysis (*e.g.,* qualitative coding), preparing artifacts (*e.g.,* drafting research publications), and more. Many participants were aware of the potential ethical concerns related to using LLMs in HCI research, including the harms of engaging with LLM outputs, threats to the privacy of participant data, intellectual integrity, and environmental and societal impacts. However, the sociotechnical assemblage of LLMs mediated researchers' diverse approaches to these ethical concerns.

While HCI researchers increasingly use LLM-based tools within their 'everyday' practices, our research community still lacks well-established guidelines and best practices, contributing to the complexity of navigating ethical considerations. In addition, researchers noted a perceived lack of control over the functionality and outputs of LLMs. Unlike previous tools where researchers had more predictability and influence over the behavior, LLMs presented outputs that may not always align with explicit instructions. Finally, many participants spoke of the challenges of navigating ethical concerns within the broader ecosystem of the 'LLM supply chain' [27]. Below, we present the implications of researchers' approaches and opportunities for HCI researchers to better engage with ethical considerations of LLMs as part of their projects to support the formation of emerging ethical norms in LLM-impacted HCI research.

Proactively engaging with IRB and other regulatory institutions. Within the U.S. context, Institutional Review Boards (IRBs) are responsible for reviewing and monitoring research activities to protect the rights and welfare of human research subjects [82]. One of the primary responsibilities of the IRB office is to assess whether "risks to subjects are reasonable in relation to anticipated benefits" [81]. Given their expertise in ensuring ethical conduct in research practices, the IRB office may be well-positioned to advise on identifying ethical concerns and potential prevention strategies. To establish whether a research study will undergo a comprehensive review, they determine if it is a minimal risk study where the "probability and magnitude of harm anticipated in the research are not greater than those ordinarily encountered in daily life" [81].

Our findings reveal that most HCI researchers did not consider it necessary to report their usage of LLMs to the IRB, partly due to their perception of LLMs as everyday tools. According to some, their scope of use did not justify including additional details, while others did not anticipate their use of LLMs at the time of submitting the IRB application. For those who disclosed their use of LLMs to the IRB, they used the minimal risk rule [81] to indicate that research subjects would typically encounter harmful outputs in their daily life as well. However, these practices can have short-term and long-term adverse implications. In the short term, the opacity in research practices (*e.g.,* which tools are used to generate research artifacts or analyze data) might lead to challenges in replicating studies [53] and understanding the motivation and effects of methodological decisions [98]. In the long term, the consequences of limited disclosure might extend beyond individual studies to shape the trajectory of the community, for example, by informing us which research topics we pursue.

We invite HCI researchers to proactively engage with the IRB at the time of study design to unpack the likelihood and ways in which any potential LLM use might harm our participants. This includes careful reflection and documentation of any implicit or anticipated use during the project planning stage. Furthermore, researchers should actively implement mechanisms to monitor LLM outputs and any adverse impacts on participants throughout the project life-cycle, mainly if used for system building. In addition, openly communicating and collaborating with IRB members to uncover potential harms is also important to increase transparency. Embracing guidelines, policies, and oversight is essential for ensuring responsible development and use of these technologies. Bockting et al. [6] urged scientists to oversee the use of generative AI because *"controlling developments in AI will require a continuous process that balances expertise and independence."* We call for collaborative efforts between researchers, policymakers, and generative AI companies to create a set of 'living guidelines' for the responsible use of generative AI in research [6].

Re-examining the informed consent process. In our study, researchers acknowledged the limited transparency to their study participants, either by not disclosing LLM use (*e.g.,* considering it irrelevant unless directly involved in system building) or by characterizing the underlying technology as Artificial Intelligence (AI). This was often rationalized to avoid overwhelming participants with seemingly unnecessary technical details. However, terminologies can be powerful too: they shape narratives around capabilities and set expectations for use [18]. AI can appear *'magical'*, where systems are considered reliable and free of bias [59]. In contrast, LLMs have specific evidence-based, known risks [120] that must be clearly communicated to research subjects. For example, in March 2023, a bug allowed some ChatGPT users to see the titles of other users' chat histories [10]. It is important to recognize how research practices might place participants' safety at risk [103].

Researchers should carefully attend to the informed consent process for projects involving LLMs across different research stages. We describe above how participants interacting with LLM outputs can lead to several risks, including amplifying socially harmful biases, creating convincing yet false information, and exposure to toxic or discriminatory language. The consent process must also attend to scenarios where subjects may not directly interact with the LLMs, but their data is sent to the model (*e.g.,* for analysis or writing), posing similar threats to privacy and agency. A well-established body of research on informed consent in research advocates for moving beyond obtaining consent as 'instrumental in nature' [51] only to satisfy regulatory and reporting obligations [62]. Research participants need to have a clear understanding of the objectives of the research, methods, and procedures, as well as any foreseeable risks that may arise from direct interaction with LLMs or if their data is processed by these models [41]. How might informed consent facilitate a collective, continuous sense-making process between the researcher and participant to understand the implications and ethical concerns of using LLMs? Nonetheless, it is crucial to highlight that documentation, transparency, and adequate informed consent might help identify and prevent issues. Still, they cannot substitute for efforts to mitigate ethical concerns using LLMs.

Developing tools, methods, and processes to interrupt the LLM supply chain. Over the last few years, progress in NLP has been characterized by the development of *larger* language models [4] requiring substantial computational resources and access to extensive datasets that have been challenging to obtain outside big companies and well-funded research labs. As a result, large language models are increasingly organized within a supply chain and produced by several interconnected actors. Cobbe et al. [27] describe the algorithmic supply chain as a pipeline where *'several actors contribute towards the production, deployment, use, and functionality of AI technologies.'* In the case of large language models, companies providing LLMs 'as a service' offer access to pre-built general-purpose models. As a result, LLM providers hold a lot of control over the distribution and development of the underlying technologies, including how to identify and address ethical concerns. Researchers in our study, situated downstream within the algorithmic supply chain, described a lack of control to examine and mitigate ethical issues. While some expressed a lack of desire, others struggled to pay off ethical debt [94] accrued through components developed upstream in the supply chain. How might we then interrupt the supply chain and shift control downstream: could we imagine tools where researchers can set up and maintain the infrastructures needed to integrate large language models in their workflow? How might such tools center a proactive approach to identifying ethical concerns?

We recognize the barriers to navigating ethical concerns when researchers depend on AI service providers to access these tools. A significant practical cost is associated with choosing a language model that attends to ethical concerns. While our participants only used general-purpose AI models at the time of this study, it is important to note that several customized LLM-based research tools (*e.g.,* academic paper writing [57, 100], conducting literature reviews [32, 111], or assisting with data analysis [54, 97]) are now available. These specialized tools offer several opportunities to address the ethical concerns raised by our participants. First, customized LLMs can be designed to implement stricter data privacy protocols [127] that are aligned with established research best practices, ensuring that data handling complies with institutional standards. Second, these tools can incorporate advanced safety filters and monitoring mechanisms to prevent exposure to harmful or discriminatory language [52]. This can help create a safer environment for research subjects interacting with LLM-based applications. Third, specialized LLMs could also allow for greater control over the model's behavior, preventing malicious use outside the scope of the research project [116]. Finally, specialized LLMs can also offer greater transparency compared to general-purpose models, specifically in terms of the data they are trained on and how the model processes queries. This transparency can help ensure the research findings are reliable and valid.

There have been established efforts to develop shared guidelines for a broader community of computer science, such as ACM authorship policy on using AI tools[4] and conference-level policy from CHI 2025[5] on the use of large language models. Specifically, these policies set clear boundaries on how AI tools can generate paper content, emphasizing the importance of transparency and proper attribution. We argue that there remains a significant gap in guidance for other phases of research, such as data collection, analysis, and participant interaction. We need guidelines to sufficiently address risks that surface in interactive settings, which is standard in HCI/CSCW research. We need procedures that, first and foremost, help researchers assess if, when, and how to use LLMs in their workflow. Suppose LLMs are integrated at any stage where research subjects or their data interact with these models. In that case, researchers must have resources to address transparency, user consent and refusal, privacy, and methodological validity considerations.

Our findings suggest opportunities to create tools, methods, and processes to support HCI researchers in better navigating the ethical considerations of integrating language models into their projects. For example, known evaluation frameworks for the use of LLM simulations, such as CoMPosT [24], make LLMs' limitations transparent by helping researchers measure models' susceptibility to caricature. Existing privacy guidelines for advanced AI assistants powered by LLMs cover norms on input and output privacy of data use and data disclosure privacy when communicating with second parties [44]. Evaluation frameworks of human-language model interactions provide researchers with design metrics when constructing interactive LLM systems involving users [63]. However, evaluating the use of LLMs in research requires a multi-faceted approach extending beyond capability, reliability, stereotyping harms, and alignment metrics. In addition to evaluation tools for LLMs for research-specific purposes, there is a growing call for auditing LLMs across technology providers, model development, and downstream applications, covering the LLM supply chain ecosystem [79]. Auditing tools of LLMs will be crucial for overseeing whether technology providers maintain ethical standards and ensure responsible development of LLMs, including transparent policies on user data handling.

[4]https://www.acm.org/publications/policies/new-acm-policy-on-authorship
[5]https://chi2025.acm.org/for-authors/papers/

A potential issue of depending on the black-box LLMs from service providers, which was also discovered in our findings, is the risk of inheriting the biases and misinformation from upstream in the LLM supply chain. To address these ethical challenges, it is essential to develop frameworks and tools that empower researchers to host and manage their LLMs for research. Prior studies have provided a comprehensive and practical guide for practitioners and end-users to work with LLMs [126], focusing on closed-source LLMs such as ChatGPT. New technological support in streamlining that process is needed to offer more autonomy and control for researchers in integrating LLMs into their research workflow. By providing these capabilities, researchers would be able to actively engage with and mitigate ethical concerns.

Creating learning opportunities and materials for ethics of LLM use in HCI. Researchers' awareness of ethical issues related to LLMs influenced how they addressed these considerations in their projects. Many researchers described LLMs as unfamiliar territory, lacking the training and well-established guidelines to approach ethical concerns with this emerging technology. We could draw on lessons from scholarship in FAccT, STS, NLP and other allied domains, as well as industry resources, to develop learning opportunities, such as toolkits [101], guidebooks [86], or frameworks [29, 83], for understanding and addressing ethical concerns with using LLMs. Achieving this requires particular attention from the HCI community to collaboratively explore how LLMs are used in different stages of HCI research.

Conferences are a valuable starting point to promote cross-institutional learning. We propose organizing workshops and panels (such as [102]) with interdisciplinary experts in sociotechnical understanding of LLMs to raise awareness on the impacts of using LLMs. Additionally, we call for creating and disseminating case studies (*e.g.,* [87]), potentially included in newsletters, illustrating how HCI researchers actively address ethical concerns related to LLMs. A repository of case studies covering diverse HCI domains and epistemologies might also offer resources for preventing and mitigating ethical concerns. Education on research ethics for LLM use should be a necessary component of HCI research and practice, given LLMs' well-documented risks and adverse impacts [120]. One avenue is incorporating such case studies into the HCI curriculum, mainly if targeted toward research students, to provide real-world examples that help build their awareness of these issues.

Shifting academic incentives to foreground ethical concerns. Throughout the interviews, we observed that researchers could generally foresee potential ethical concerns with LLMs, often drawing from their familiarity with the literature or discussions with other researchers. Nonetheless, some participants went on to articulate their reason for not prioritizing these ethical considerations in their projects. Researchers described how constraints such as limited funding, pressure to publish, and conference deadlines often came in the way of focusing on ethical concerns. The need to address ethical issues could then be relegated to the limitations or future work section, reflecting broader perceptions within Computer Science research that ethics are often viewed as secondary considerations. Indeed, Do and Pang *et al.* [31] discussed how academics might have a lower incentive to examine unintended consequences of their research (also observed in our study) in contrast with industry practices.

There is a renewed urgency to reconsider research ethics practices within HCI and how we navigate challenges presented by emerging technologies such as LLMs. We take inspiration from Do and Pang *et al.* [31] and Soergel *et al.* [108] to propose starting points for changing structural incentives within academia. Publications and citations are the currency and means for upward mobility in research. Could we shift the criteria for recognition and funding to explicitly recognize attention to ethical considerations in research practices? Publishing organizations, funding agencies, and regulatory institutions will play a crucial role in reshaping incentives. Most importantly, HCI researchers must acknowledge how they actively shape the discourse around associated risks and ethical considerations through their research and work actively towards a cultural shift, demonstrating a commitment to LLMs' ethical use.

6 Limitations and Future Work

Our study sheds light on the emerging practices regarding LLM ethics among HCI researchers, but it has limitations due to its exploratory nature. Firstly, while we sought to offer a deep understanding of how HCI researchers navigate LLM ethics, our sample was constrained by our snowball sampling method. HCI research encompasses many diverse research traditions that we could not include in our study. This limitation highlights the need for more comprehensive and systematic future studies. Our interview sample primarily consists of researchers from the USA, and future research may want to explore further the ethical challenges and practices related to using LLMs by researchers from other regions. Secondly, our study has a potential selection bias: we may have primarily attracted respondents who are conscious of their LLM usage and are open to discussing their experiences in a research setting. To gain a broader perspective, future research should explore how ethical practices with LLMs vary across different research methodologies, domains, and settings, including both industry and academia.

7 Conclusion

In this paper, we drew empirical data from a survey and interviews to explore how HCI researchers have currently integrated LLMs into their research practices, what ethical concerns they have encountered, and how they've navigated those concerns. Our results suggested that although HCI researchers have used LLMs across their research processes and are aware of a wide variety of ethical considerations, in many cases, they have challenges in effectively identifying and navigating those concerns in their projects. Reflecting on these findings, we discuss potential approaches to support the formation of emerging ethical norms for using LLMs in HCI research. We encourage HCI researchers to proactively engage with IRB and collaborate with policymakers and generative AI companies to create guidelines for the responsible use of LLMs. We also identify the need to re-examine the informed consent process and provide technological support to interrupt the LLM supply chain. In addition, we discuss the importance of creating learning opportunities for the ethics of LLMs use in HCI and shifting academic incentives to prioritize ethical concerns.

Acknowledgments

We express gratitude to our participants who contributed valuable time and shared their research practices for this study. We are grateful to the external reviewers, who sharpened our contribution with their feedback and critical comments. We would like to thank Anubhav Jangra and Ningjing Tang for their generous feedback on the paper.

References

[1] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N. Bennett, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, and Eric Horvitz. 2019. Guidelines for Human-AI Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3290605.3300233

[2] Alissa N Antle. 2017. The ethics of doing research with vulnerable populations. *Interactions* 24, 6 (2017), 74–77.

[3] Christopher Bail. 2023. Can generative artificial intelligence improve social science.

[4] Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (Virtual Event, Canada) *(FAccT '21)*. Association for Computing Machinery, New York, NY, USA, 610–623. doi:10.1145/3442188.3445922

[5] ACM Publications Board. 2021. ACM Publications Policy on Research Involving Human Participants and Subjects. https://www.acm.org/publications/policies/research-involving-human-participants-and-subjects. (Accessed on 01/16/2024).

[6] Claudi L. Bockting, Eva A. M. van Dis, Robert van Rooij, Willem Zuidema, and Johan Bollen. 2023. Living guidelines for generative AI — why scientists must oversee its use. *Nature* 622, 7984 (Oct. 2023), 693–696. doi:10.1038/d41586-023-03266-1

[7] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[8] Virginia Braun and Victoria Clarke. 2019. Reflecting on reflexive thematic analysis. *Qualitative research in sport, exercise and health* 11, 4 (2019), 589–597.

[9] Barry Brown, Alexandra Weilenmann, Donald McMillan, and Airi Lampinen. 2016. Five Provocations for Ethical HCI Research. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 852–863. doi:10.1145/2858036.2858313

[10] Ryan Browne. 2023. OpenAI CEO says a bug allowed some ChatGPT to see others' chat titles. https://www.cnbc.com/2023/03/23/openai-ceo-says-a-bug-allowed-some-chatgpt-to-see-others-chat-titles.html. (Accessed on 01/15/2024).

[11] Amy Bruckman. 2014. Research ethics and HCI. 449–468 pages.

[12] Amy S. Bruckman, Casey Fiesler, Jeff Hancock, and Cosmin Munteanu. 2017. CSCW Research Ethics Town Hall: Working Towards Community Norms. In *Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (Portland, Oregon, USA) *(CSCW '17 Companion)*. Association for Computing Machinery, New York, NY, USA, 113–115. doi:10.1145/3022198.3022199

[13] Elizabeth Buchanan, John Aycock, Scott Dexter, David Dittrich, and Erin Hvizdak. 2011. Computer Science Security Research And Human Subjects: Emerging Considerations For Research Ethics Boards. *Journal of empirical research on human research ethics : JERHRE* 6 (06 2011), 71–83. doi:10.1525/jer.2011.6.2.71

[14] Elizabeth A. Buchanan and Charles M. Ess. 2009. Internet Research Ethics and the Institutional Review Board: Current Practices and Issues. *SIGCAS Comput. Soc.* 39, 3 (dec 2009), 43–49. doi:10.1145/1713066.1713069

[15] Elizabeth A. Buchanan and Michael Zimmer. 2023. Internet Research Ethics. https://plato.stanford.edu/archives/win2023/entries/ethics-internet-research/.

[16] Courtni Byun, Piper Vasicek, and Kevin Seppi. 2023. Dispensing with Humans in Human-Computer Interaction Research. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (<conf-loc>, <city>Hamburg</city>, <country>Germany</country>, </conf-loc>) *(CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, Article 413, 26 pages. doi:10.1145/3544549.3582749

[17] Nicholas Carlini, Florian Tramer, Eric Wallace, Matthew Jagielski, Ariel Herbert-Voss, Katherine Lee, Adam Roberts, Tom Brown, Dawn Song, Ulfar Erlingsson, Alina Oprea, and Colin Raffel. 2021. Extracting Training Data from Large Language Models. arXiv:2012.07805 [cs.CR]

[18] Stephen Cave, Claire Craig, Kanta Dihal, Sarah Dillon, Jessica Montgomery, Beth Singler, and Lindsay Taylor. 2018. Portrayals and perceptions of AI and why they matter. doi:10.17863/CAM.34502

[19] ACL 2023 Program Chairs. 2023. ACL 2023 Policy on AI Writing Assistance - ACL 2023. https://2023.aclweb.org/blog/ACL-2023-policy/. (Accessed on 01/11/2024).

[20] Matthew Chalmers, Donald McMillan, Alistair Morrison, Henriette Cramer, Mattias Rost, and Wendy Mackay. 2011. Ethics, logs and videotape: ethics in large scale user trials and user generated content. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI EA '11)*. Association for Computing Machinery, New York, NY, USA, 2421–2424. doi:10.1145/1979742.1979574

[21] Canyu Chen and Kai Shu. 2023. Can LLM-Generated Misinformation Be Detected? arXiv:2309.13788 [cs.CL]

[22] Canyu Chen and Kai Shu. 2023. Combating Misinformation in the Age of LLMs: Opportunities and Challenges. arXiv:2311.05656 [cs.CY]

[23] Myra Cheng, Tiziano Piccardi, and Diyi Yang. 2023. CoMPosT: Characterizing and Evaluating Caricature in LLM Simulations. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, Houda Bouamor, Juan Pino, and Kalika Bali (Eds.). Association for Computational Linguistics, Singapore, 10853–10875. doi:10.18653/v1/2023.emnlp-main.669

[24] Myra Cheng, Tiziano Piccardi, and Diyi Yang. 2023. CoMPosT: Characterizing and Evaluating Caricature in LLM Simulations. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, Houda Bouamor, Juan Pino, and Kalika Bali (Eds.). Association for Computational Linguistics, Singapore, 10853–10875. doi:10.18653/v1/2023.emnlp-main.669

[25] Robert Chew, John Bollenbacher, Michael Wenger, Jessica Speer, and Annice Kim. 2023. LLM-Assisted Content Analysis: Using Large Language Models to Support Deductive Coding. arXiv:2306.14924 [cs.CL]

[26] Karin Clark, Matt Duckham, Marilys Guillemin, Assunta Hunter, Jodie McVernon, Christine O'Keefe, Cathy Pitkin, Steven Prawer, Richard Sinnott, Deborah Warr, and Jenny Waycott. 2018. Advancing the ethical use of digital data in

human research: challenges and strategies to promote ethical practice. *Ethics and Information Technology* 21, 1 (Nov. 2018), 59–73. doi:10.1007/s10676-018-9490-4

[27] Jennifer Cobbe, Michael Veale, and Jatinder Singh. 2023. Understanding accountability in algorithmic supply chains. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (Chicago, IL, USA) *(FAccT '23)*. Association for Computing Machinery, New York, NY, USA, 1186–1197. doi:10.1145/3593013.3594073

[28] Dorottya Demszky, Diyi Yang, David S Yeager, Christopher J Bryan, Margarett Clapper, Susannah Chandhok, Johannes C Eichstaedt, Cameron Hecht, Jeremy Jamieson, Meghann Johnson, et al. 2023. Using large language models in psychology. *Nature Reviews Psychology* 2, 11 (2023), 688–701.

[29] Leon Derczynski, Hannah Rose Kirk, Vidhisha Balachandran, Sachin Kumar, Yulia Tsvetkov, M. R. Leiser, and Saif Mohammad. 2023. Assessing Language Model Deployment with Risk Cards. arXiv:2303.18190 [cs.CL]

[30] Ameet Deshpande, Tanmay Rajpurohit, Karthik Narasimhan, and Ashwin Kalyan. 2023. Anthropomorphization of AI: Opportunities and Risks. arXiv:2305.14784 [cs.AI]

[31] Kimberly Do, Rock Yuren Pang, Jiachen Jiang, and Katharina Reinecke. 2023. "That's important, but...": How Computer Science Researchers Anticipate Unintended Consequences of Their Research Innovations. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 602, 16 pages. doi:10.1145/3544548.3581347

[32] Elicit. 2024. Elicit Help Center. https://support.elicit.com/en/categories/146369-about-elicit. (Accessed on 07/16/2024).

[33] Xiao Fang, Shangkun Che, Minjia Mao, Hongzhe Zhang, Ming Zhao, and Xiaohang Zhao. 2023. Bias of AI-Generated Content: An Examination of News Produced by Large Language Models. arXiv:2309.09825 [cs.AI]

[34] Yunhe Feng, Sreecharan Vanam, Manasa Cherukupally, Weijian Zheng, Meikang Qiu, and Haihua Chen. 2023. Investigating Code Generation Performance of ChatGPT with Crowdsourcing Social Data. 876-885 pages. doi:10.1109/COMPSAC57700.2023.00117

[35] Casey Fiesler, Melissa Densmore, Michael Muller, and Cosmin Munteanu. 2021. SIGCHI Research Ethics Committee Town Hall. In *Companion Publication of the 2021 Conference on Computer Supported Cooperative Work and Social Computing* (Virtual Event, USA) *(CSCW '21 Companion)*. Association for Computing Machinery, New York, NY, USA, 232–233. doi:10.1145/3462204.3483283

[36] Casey Fiesler, Christopher Frauenberger, Michael Muller, Jessica Vitak, and Michael Zimmer. 2022. Research Ethics in HCI: A SIGCHI Community Discussion. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI EA '22)*. Association for Computing Machinery, New York, NY, USA, Article 169, 3 pages. doi:10.1145/3491101.3516400

[37] Casey Fiesler, Jeff Hancock, Amy Bruckman, Michael Muller, Cosmin Munteanu, and Melissa Densmore. 2018. Research Ethics for HCI: A Roundtable Discussion. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI EA '18)*. Association for Computing Machinery, New York, NY, USA, 1–5. doi:10.1145/3170427.3186321

[38] Casey Fiesler and Nicholas Proferes. 2018. "Participant" perceptions of Twitter research ethics. *Social Media+ Society* 4, 1 (2018), 2056305118763366.

[39] Casey Fiesler, Michael Zimmer, Nicholas Proferes, Sarah Gilbert, and Naiyan Jones. 2024. Remember the human: A systematic review of ethical considerations in reddit research. *Proceedings of the ACM on Human-Computer Interaction* 8, GROUP (2024), 1–33.

[40] Association for Computing Machinery. 2023. ACM Policy on Authorship. https://www.acm.org/publications/policies/new-acm-policy-on-authorship. (Accessed on 01/11/2024).

[41] Office for Human Research Protections. 1998. Informed Consent Checklist (1998) | HHS.gov. https://www.hhs.gov/ohrp/regulations-and-policy/guidance/checklists/index.html. (Accessed on 01/15/2024).

[42] ACM Code 2018 Task Force. 2018. Code of Ethics. https://www.acm.org/code-of-ethics. (Accessed on 01/16/2024).

[43] Christopher Frauenberger, Amy S. Bruckman, Cosmin Munteanu, Melissa Densmore, and Jenny Waycott. 2017. Research Ethics in HCI: A Town Hall Meeting. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI EA '17)*. Association for Computing Machinery, New York, NY, USA, 1295–1299. doi:10.1145/3027063.3051135

[44] Iason Gabriel, Arianna Manzini, Geoff Keeling, Lisa Anne Hendricks, Verena Rieser, Hasan Iqbal, Nenad Tomašev, Ira Ktena, Zachary Kenton, Mikel Rodriguez, Seliem El-Sayed, Sasha Brown, Canfer Akbulut, Andrew Trask, Edward Hughes, A. Stevie Bergman, Renee Shelby, Nahema Marchal, Conor Griffin, Juan Mateos-Garcia, Laura Weidinger, Winnie Street, Benjamin Lange, Alex Ingerman, Alison Lentz, Reed Enger, Andrew Barakat, Victoria Krakovna, John Oliver Siy, Zeb Kurth-Nelson, Amanda McCroskery, Vijay Bolina, Harry Law, Murray Shanahan, Lize Alberts, Borja Balle, Sarah de Haas, Yetunde Ibitoye, Allan Dafoe, Beth Goldberg, Sébastien Krier, Alexander Reese, Sims Witherspoon, Will Hawkins, Maribeth Rauh, Don Wallace, Matija Franklin, Josh A. Goldstein, Joel Lehman, Michael Klenk, Shannon Vallor, Courtney Biles, Meredith Ringel Morris, Helen King, Blaise Agüera y Arcas, William Isaac, and James Manyika. 2024. The Ethics of Advanced AI Assistants. arXiv:2404.16244 [cs.CY] https://arxiv.org/abs/2404.16244

[45] Chen Gao, Xiaochong Lan, Zhihong Lu, Jinzhu Mao, Jinghua Piao, Huandong Wang, Depeng Jin, and Yong Li. 2023. S3: Social-network Simulation System with Large Language Model-Empowered Agents. arXiv:2307.14984 [cs.SI]

[46] Jie Gao, Yuchen Guo, Gionnieve Lim, Tianqin Zhang, Zheng Zhang, Toby Jia-Jun Li, and Simon Tangi Perrault. 2023. CollabCoder: A GPT-Powered Workflow for Collaborative Qualitative Analysis. arXiv:2304.07366 [cs.HC]

[47] Simret Araya Gebreegziabher, Zheng Zhang, Xiaohang Tang, Yihao Meng, Elena L. Glassman, and Toby Jia-Jun Li. 2023. PaTAT: Human-AI Collaborative Qualitative Coding with Explainable Interactive Rule Synthesis. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 362, 19 pages. doi:10.1145/3544548.3581352

[48] Samuel Gehman, Suchin Gururangan, Maarten Sap, Yejin Choi, and Noah A. Smith. 2020. RealToxicityPrompts: Evaluating Neural Toxic Degeneration in Language Models. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, Trevor Cohn, Yulan He, and Yang Liu (Eds.). Association for Computational Linguistics, Online, 3356–3369. doi:10.18653/v1/2020.findings-emnlp.301

[49] Karan Girotra, Lennart Meincke, Christian Terwiesch, and Karl T. Ulrich. 2023. Ideas are dimes a dozen: Large language models for Idea generation in innovation. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4526071#

[50] Perttu Hämäläinen, Mikke Tavast, and Anton Kunnari. 2023. Evaluating Large Language Models in Generating Synthetic HCI Research Data: A Case Study. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 433, 19 pages. doi:10.1145/3544548.3580688

[51] Jennifer A. Hamilton. 2011. On the Ethics of Unusable Data. 73–88 pages. https://doi.org/10.7591/9780801463594-005

[52] Seungju Han, Kavel Rao, Allyson Ettinger, Liwei Jiang, Bill Yuchen Lin, Nathan Lambert, Yejin Choi, and Nouha Dziri. 2024. WildGuard: Open One-Stop Moderation Tools for Safety Risks, Jailbreaks, and Refusals of LLMs. arXiv:2406.18495 [cs.CL] https://arxiv.org/abs/2406.18495

[53] Transparent Statistics in Human–Computer Interaction Working Group. 2019. Transparent Statistics Guidelines. doi:10.5281/zenodo.1186169 (Available at https://transparentstats.github.io/guidelines).

[54] Dovetail Inc. 2024. New in Dovetail: Search API, new editor, improved features. https://dovetail.com/blog/evolution-dovetail-magic/. (Accessed on 07/16/2024).

[55] Nataliya V Ivankova, John W Creswell, and Sheldon L Stick. 2006. Using mixed-methods sequential explanatory design: From theory to practice. *Field methods* 18, 1 (2006), 3–20.

[56] Bernard J. Jansen, Soon gyo Jung, and Joni Salminen. 2023. Employing large language models in survey research. *Natural Language Processing Journal* 4 (2023), 100020. doi:10.1016/j.nlp.2023.100020

[57] Jenni. 2024. Jenni AI. https://jenni.ai/. (Accessed on 07/16/2024).

[58] Jigsaw and Google's Counter Abuse Technology team. 2017. Perspective API. https://perspectiveapi.com/. (Accessed on 01/16/2024).

[59] Shivani Kapania, Oliver Siy, Gabe Clapper, Azhagu Meena SP, and Nithya Sambasivan. 2022. "Because AI is 100% right and safe": User Attitudes and Sources of AI Authority in India. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 158, 18 pages. doi:10.1145/3491102.3517533

[60] Elise Karinshak, Sunny Xun Liu, Joon Sung Park, and Jeffrey T Hancock. 2023. Working with AI to persuade: Examining a large language model's ability to generate pro-vaccination messages. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (2023), 1–29.

[61] Enkelejda Kasneci, Kathrin Seßler, Stefan Küchemann, Maria Bannert, Daryna Dementieva, Frank Fischer, Urs Gasser, Georg Groh, Stephan Günnemann, Eyke Hüllermeier, et al. 2023. ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and individual differences* 103 (2023), 102274.

[62] Fride Haram Klykken. 2022. Implementing continuous consent in qualitative research. *Qualitative Research* 22, 5 (2022), 795–810.

[63] Mina Lee, Megha Srivastava, Amelia Hardy, John Thickstun, Esin Durmus, Ashwin Paranjape, Ines Gerard-Ursin, Xiang Lisa Li, Faisal Ladhak, Frieda Rong, Rose E. Wang, Minae Kwon, Joon Sung Park, Hancheng Cao, Tony Lee, Rishi Bommasani, Michael Bernstein, and Percy Liang. 2024. Evaluating Human-Language Model Interaction. arXiv:2212.09746 [cs.CL] https://arxiv.org/abs/2212.09746

[64] Daniel Leiker, Sara Finnigan, Ashley Ricker Gyllen, and Mutlu Cukurova. 2023. Prototyping the use of Large Language Models (LLMs) for adult learning content creation at scale. arXiv:2306.01815 [cs.CY]

[65] Yuan Li, Yixuan Zhang, and Lichao Sun. 2023. MetaAgents: Simulating Interactions of Human Behaviors for LLM-based Task-oriented Coordination via Collaborative Generative Agents. arXiv:2310.06500 [cs.AI]

[66] Calvin A Liang, Sean A Munson, and Julie A Kientz. 2021. Embracing four tensions in human-computer interaction research with marginalized people. *ACM Transactions on Computer-Human Interaction (TOCHI)* 28, 2 (2021), 1–47.

[67] Q. Vera Liao and Jennifer Wortman Vaughan. 2023. AI Transparency in the Age of LLMs: A Human-Centered Research Roadmap. arXiv:2306.01941 [cs.HC]

[68] Yiren Liu, Si Chen, Haocong Cheng, Mengxia Yu, Xiao Ran, Andrew Mo, Yiliu Tang, and Yun Huang. 2023. How AI Processing Delays Foster Creativity: Exploring Research Question Co-Creation with an LLM-based Agent. arXiv:2310.06155 [cs.HC]

[69] Yuwen Lu, Chengzhi Zhang, Iris Zhang, and Toby Jia-Jun Li. 2022. Bridging the Gap Between UX Practitioners' Work Practices and AI-Enabled Design Support Tools. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI EA '22)*. Association for Computing Machinery, New York, NY, USA, Article 268, 7 pages. doi:10.1145/3491101.3519809

[70] Nils Lukas, Ahmed Salem, Robert Sim, Shruti Tople, Lukas Wutschitz, and Santiago Zanella-Béguelin. 2023. Analyzing Leakage of Personally Identifiable Information in Language Models. arXiv:2302.00539 [cs.LG]

[71] Kevin Ma, Daniele Grandi, Christopher McComb, and Kosa Goucher-Lambert. 2023. Conceptual Design Generation Using Large Language Models. arXiv:2306.01779 [cs.CL]

[72] Wendy E Mackay. 1995. Ethics, lies and videotape.... In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, USA, 138–145.

[73] Nora McDonald, Sarita Schoenebeck, and Andrea Forte. 2019. Reliability and inter-rater reliability in qualitative research: Norms and guidelines for CSCW and HCI practice. *Proceedings of the ACM on human-computer interaction* 3, CSCW (2019), 1–23.

[74] Andrew M Mcnutt, Chenglong Wang, Robert A Deline, and Steven M. Drucker. 2023. On the Design of AI-powered Code Assistants for Notebooks. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 434, 16 pages. doi:10.1145/3544548.3580940

[75] Niloofar Mireshghallah, Hyunwoo Kim, Xuhui Zhou, Yulia Tsvetkov, Maarten Sap, Reza Shokri, and Yejin Choi. 2023. Can LLMs Keep a Secret? Testing Privacy Implications of Language Models via Contextual Integrity Theory. arXiv:2310.17884 [cs.AI]

[76] Maximilian Mozes, Xuanli He, Bennett Kleinberg, and Lewis D. Griffin. 2023. Use of LLMs for Illicit Purposes: Threats, Prevention Measures, and Vulnerabilities. arXiv:2308.12833 [cs.CL]

[77] Cosmin Munteanu, Amy Bruckman, Michael Muller, Christopher Frauenberger, Casey Fiesler, Robert E. Kraut, Katie Shilton, and Jenny Waycott. 2019. SIGCHI Research Ethics Town Hall. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI EA '19)*. Association for Computing Machinery, New York, NY, USA, 1–6. doi:10.1145/3290607.3311742

[78] Cosmin Munteanu, Heather Molyneaux, Wendy Moncur, Mario Romero, Susan O'Donnell, and John Vines. 2015. Situational Ethics: Re-thinking Approaches to Formal Ethics Requirements for Human-Computer Interaction. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 105–114. doi:10.1145/2702123.2702481

[79] Jakob Mökander, Jonas Schuett, Hannah Rose Kirk, and Luciano Floridi. 2023. Auditing large language models: a three-layered approach. doi:10.1007/s43681-023-00289-2

[80] Zheng Ning, Zheng Zhang, Tianyi Sun, Yuan Tian, Tianyi Zhang, and Toby Jia-Jun Li. 2023. An Empirical Study of Model Errors and User Error Discovery and Repair Strategies in Natural Language Database Queries. In *Proceedings of the 28th International Conference on Intelligent User Interfaces* (Sydney, NSW, Australia) *(IUI '23)*. Association for Computing Machinery, New York, NY, USA, 633–649. doi:10.1145/3581641.3584067

[81] US Department of Health and Human Services. 2018. 2018 Requirements (2018 Common Rule) | HHS.gov. https://www.hhs.gov/ohrp/regulations-and-policy/regulations/45-cfr-46/revised-common-rule-regulatory-text/index.html. (Accessed on 01/15/2024).

[82] US Department of Health and Human Services. 2018. Policy for the Protection of Human Subjects. https://www.hhs.gov/ohrp/regulations-and-policy/regulations/45-cfr-46/index.html. (Accessed on 01/15/2024).

[83] National Institute of Standards and Technology. 2023. Artificial Intelligence Risk Management Framework (AI RMF 1.0).

[84] OpenAI. 2023. GPT-4 Technical Report. arXiv:2303.08774 [cs.CL]

[85] Jonas Oppenlaender and Joonas Hämäläinen. 2023. Mapping the Challenges of HCI: An Application and Evaluation of ChatGPT and GPT-4 for Mining Insights at Scale. arXiv:2306.05036 [cs.HC]

[86] Google PAIR. 2021. People + AI Guidebook. https://pair.withgoogle.com/guidebook. (Accessed on 01/15/2024).

[87] Google PAIR. 2022. People + AI Research Guidebook - Case Studies. https://pair.withgoogle.com/guidebook/case-studies. (Accessed on 01/15/2024).

[88] Lawrence Palinkas, Sarah Horwitz, Carla Green, Jennifer Wisdom, Naihua Duan, and Kimberly Hoagwood. 2013. Purposeful Sampling for Qualitative Data Collection and Analysis in Mixed Method Implementation Research. *Administration and policy in mental health* 42 (11 2013). doi:10.1007/s10488-013-0528-y

[89] Yikang Pan, Liangming Pan, Wenhu Chen, Preslav Nakov, Min-Yen Kan, and William Yang Wang. 2023. On the Risk of Misinformation Pollution with Large Language Models. arXiv:2305.13661 [cs.CL]

[90] Stefano De Paoli. 2023. Can Large Language Models emulate an inductive Thematic Analysis of semi-structured interviews? An exploration and provocation on the limits of the approach and the model. arXiv:2305.13014 [cs.CL]

[91] Joon Sung Park, Joseph O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. 2023. Generative Agents: Interactive Simulacra of Human Behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (San Francisco, CA, USA) *(UIST '23)*. Association for Computing Machinery, New York, NY, USA, Article 2, 22 pages. doi:10.1145/3586183.3606763

[92] Joon Sung Park, Lindsay Popowski, Carrie J. Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. 2022. Social Simulacra: Creating Populated Prototypes for Social Computing Systems. arXiv:2208.04024 [cs.HC]

[93] Savvas Petridis, Michael Terry, and Carrie Jun Cai. 2023. PromptInfuser: Bringing User Interface Mock-Ups to Life with Large Language Models. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (<conf-loc>, <city>Hamburg</city>, <country>Germany</country>, </conf-loc>) *(CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, Article 237, 6 pages. doi:10.1145/3544549.3585628

[94] Catherine Petrozzino. 2021. Who pays for ethical debt in AI? *AI and Ethics* 1, 3 (2021), 205–208.

[95] Lumpapun Punchoojit and Nuttanont Hongwarittorrn. 2015. Research ethics in human-computer interaction: A review of ethical concerns in the past five years. 180-185 pages. doi:10.1109/NICS.2015.7302187

[96] Riaz Qureshi, Daniel Shaughnessy, Kayden Gill, Karen Robinson, Tianjing Li, and Eitan Agai. 2023. Are ChatGPT and large language models "the answer" to bringing us closer to systematic review automation? *Systematic reviews* 12 (04 2023), 72. doi:10.1186/s13643-023-02243-z

[97] Zeeshan Rasheed, Muhammad Waseem, Aakash Ahmad, Kai-Kristian Kemell, Wang Xiaofeng, Anh Nguyen Duc, and Pekka Abrahamsson. 2024. Can Large Language Models Serve as Data Analysts? A Multi-Agent Assisted Approach for Qualitative Data Analysis. arXiv:2402.01386 [cs.SE] https://arxiv.org/abs/2402.01386

[98] Kavous Salehzadeh Niksirat, Lahari Goswami, Pooja S. B. Rao, James Tyler, Alessandro Silacci, Sadiq Aliyu, Annika Aebli, Chat Wacharamanotham, and Mauro Cherubini. 2023. Changes in Research Ethics, Openness, and Transparency in Empirical Studies between CHI 2017 and CHI 2022. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (<conf-loc>, <city>Hamburg</city>, <country>Germany</country>, </conf-loc>) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 505, 23 pages. doi:10.1145/3544548.3580848

[99] Jeffrey Saltz, Michael Skirpan, Casey Fiesler, Micha Gorelick, Tom Yeh, Robert Heckman, Neil Dewar, and Nathan Beard. 2019. Integrating Ethics within Machine Learning Courses. *ACM Trans. Comput. Educ.* 19, 4, Article 32 (aug 2019), 26 pages. doi:10.1145/3341164

[100] Kathrin Seßler, Tao Xiang, Lukas Bogenrieder, and Enkelejda Kasneci. 2023. PEER: Empowering Writing with Large Language Models. In *Responsive and Sustainable Educational Futures*, Olga Viberg, Ioana Jivet, Pedro J. Muñoz-Merino, Maria Perifanou, and Tina Papathoma (Eds.). Springer Nature Switzerland, Cham, 755–761.

[101] Hong Shen, Wesley H. Deng, Aditi Chattopadhyay, Zhiwei Steven Wu, Xu Wang, and Haiyi Zhu. 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* (Virtual Event, Canada) *(FAccT '21)*. Association for Computing Machinery, New York, NY, USA, 850–861. doi:10.1145/3442188.3445971

[102] Hong Shen, Tianshi Li, Toby Jia-Jun Li, Joon Sung Park, and Diyi Yang. 2023. Shaping the Emerging Norms of Using Large Language Models in Social Computing Research. In *Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing* (Minneapolis, MN, USA) *(CSCW '23 Companion)*. Association for Computing Machinery, New York, NY, USA, 569–571. doi:10.1145/3584931.3606955

[103] Irina Shklovski and Janet Vertesi. 2012. "Un-Googling": Research Technologies, Communities at Risk and the Ethics of User Studies in HCI. 4 pages.

[104] Ben Shneiderman. 2020. Bridging the Gap Between Ethics and Practice: Guidelines for Reliable, Safe, and Trustworthy Human-Centered AI Systems. *ACM Trans. Interact. Intell. Syst.* 10, 4, Article 26 (oct 2020), 31 pages. doi:10.1145/3419764

[105] Jessie J. Smith, Blakeley H. Payne, Shamika Klassen, Dylan Thomas Doyle, and Casey Fiesler. 2023. Incorporating Ethics in Computing Courses: Barriers, Support, and Perspectives from Educators. In *Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1* (<conf-loc>, <city>Toronto ON</city>, <country>Canada</country>, </conf-loc>) *(SIGCSE 2023)*. Association for Computing Machinery, New York, NY, USA, 367–373. doi:10.1145/3545945.3569855

[106] Victoria Smith, Ali Shahin Shamsabadi, Carolyn Ashurst, and Adrian Weller. 2023. Identifying and Mitigating Privacy Risks Stemming from Language Models: A Survey. arXiv:2310.01424 [cs.CL]

[107] Robert Soden, Austin Toombs, and Michaelanne Thomas. 2024. Evaluating Interpretive Research in HCI. *Interactions* 31, 1 (2024), 38–42.

[108] David Soergel, Adam Saunders, and Andrew McCallum. 2013. Open Scholarship and Peer Review: a Time for Experimentation. https://api.semanticscholar.org/CorpusID:14548845

[109] Sangho Suh, Meng Chen, Bryan Min, Toby Jia-Jun Li, and Haijun Xia. 2023. Structured Generation and Exploration of Design Space with Large Language Models for Human-AI Co-Creation. arXiv:2310.12953 [cs.HC]

[110] Lu Sun, Stone Tao, Junjie Hu, and Steven P Dow. 2024. MetaWriter: Exploring the Potential and Perils of AI Writing Support in Scientific Peer Review. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW1 (2024), 1–32.
[111] Teo Susnjak, Peter Hwang, Napoleon H Reyes, Andre LC Barczak, Timothy R McIntosh, and Surangika Ranathunga. 2024. Automating research synthesis with domain-specific large language model fine-tuning.
[112] Robert H. Tai, Lillian R. Bentley, Xin Xia, Jason M. Sitt, Sarah C. Fankhauser, Ana M. Chicas-Mosier, and Barnas G. Monteith. 2024. An Examination of the Use of Large Language Models to Aid Analysis of Textual Data. *International Journal of Qualitative Methods* 23 (2024), 16094069241231168. doi:10.1177/16094069241231168 arXiv:https://doi.org/10.1177/16094069241231168
[113] Arun James Thirunavukarasu, Darren Shu Jeng Ting, Kabilan Elangovan, Laura Gutierrez, Ting Fang Tan, and Daniel Shu Wei Ting. 2023. Large language models in medicine. *Nature medicine* 29, 8 (2023), 1930–1940.
[114] Jakob Tholander and Martin Jonsson. 2023. Design Ideation with AI - Sketching, Thinking and Talking with Generative Machine Learning Models. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference* (Pittsburgh, PA, USA) *(DIS '23)*. Association for Computing Machinery, New York, NY, USA, 1930–1940. doi:10.1145/3563657.3596014
[115] Jessica Vitak, Katie Shilton, and Zahra Ashktorab. 2016. Beyond the Belmont Principles: Ethical Challenges, Practices, and Beliefs in the Online Data Research Community. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (San Francisco, California, USA) *(CSCW '16)*. Association for Computing Machinery, New York, NY, USA, 941–953. doi:10.1145/2818048.2820078
[116] Jane Wakefield. 2016. Microsoft chatbot is taught to swear on Twitter - BBC News. https://www.bbc.com/news/technology-35890188. (Accessed on 07/16/2024).
[117] Thomas Weber, Maximilian Brandmaier, Albrecht Schmidt, and Sven Mayer. 2024. Significant Productivity Gains through Programming with Large Language Models. *Proceedings of the ACM on Human-Computer Interaction* 8, EICS (2024), 1–29.
[118] Jing Wei, Sungdong Kim, Hyunhoon Jung, and Young-Ho Kim. 2023. Leveraging Large Language Models to Power Chatbots for Collecting User Self-Reported Data. arXiv:2301.05843 [cs.HC]
[119] Laura Weidinger, John Mellor, Maribeth Rauh, Conor Griffin, Jonathan Uesato, Po-Sen Huang, Myra Cheng, Mia Glaese, Borja Balle, Atoosa Kasirzadeh, Zac Kenton, Sasha Brown, Will Hawkins, Tom Stepleton, Courtney Biles, Abeba Birhane, Julia Haas, Laura Rimell, Lisa Anne Hendricks, William Isaac, Sean Legassick, Geoffrey Irving, and Iason Gabriel. 2021. Ethical and social risks of harm from Language Models. arXiv:2112.04359 [cs.CL]
[120] Laura Weidinger, Jonathan Uesato, Maribeth Rauh, Conor Griffin, Po-Sen Huang, John Mellor, Amelia Glaese, Myra Cheng, Borja Balle, Atoosa Kasirzadeh, Courtney Biles, Sasha Brown, Zac Kenton, Will Hawkins, Tom Stepleton, Abeba Birhane, Lisa Anne Hendricks, Laura Rimell, William Isaac, Julia Haas, Sean Legassick, Geoffrey Irving, and Iason Gabriel. 2022. Taxonomy of Risks Posed by Language Models. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (Seoul, Republic of Korea) *(FAccT '22)*. Association for Computing Machinery, New York, NY, USA, 214–229. doi:10.1145/3531146.3533088
[121] David Gray Widder and Dawn Nafus. 2023. Dislocated accountabilities in the "AI supply chain": Modularity and developers' notions of responsibility. *Big Data & Society* 10, 1 (2023), 20539517231177620.
[122] Tongshuang Wu, Ellen Jiang, Aaron Donsbach, Jeff Gray, Alejandra Molina, Michael Terry, and Carrie J Cai. 2022. PromptChainer: Chaining Large Language Model Prompts through Visual Programming. arXiv:2203.06566 [cs.HC]
[123] Ziang Xiao, Xingdi Yuan, Q. Vera Liao, Rania Abdelghani, and Pierre-Yves Oudeyer. 2023. Supporting Qualitative Analysis with Large Language Models: Combining Codebook with GPT-3 for Deductive Coding. In *Companion Proceedings of the 28th International Conference on Intelligent User Interfaces* (Sydney, NSW, Australia) *(IUI '23 Companion)*. Association for Computing Machinery, New York, NY, USA, 75–78. doi:10.1145/3581754.3584136
[124] Ziang Xiao, Xingdi Yuan, Q. Vera Liao, Rania Abdelghani, and Pierre-Yves Oudeyer. 2023. Supporting Qualitative Analysis with Large Language Models: Combining Codebook with GPT-3 for Deductive Coding. In *Companion Proceedings of the 28th International Conference on Intelligent User Interfaces* (Sydney, NSW, Australia) *(IUI '23 Companion)*. Association for Computing Machinery, New York, NY, USA, 75–78. doi:10.1145/3581754.3584136
[125] Chen Xu, Wenjie Wang, Yuxin Li, Liang Pang, Jun Xu, and Tat-Seng Chua. 2023. Do LLMs Implicitly Exhibit User Discrimination in Recommendation? An Empirical Study. arXiv:2311.07054 [cs.IR]
[126] Jingfeng Yang, Hongye Jin, Ruixiang Tang, Xiaotian Han, Qizhang Feng, Haoming Jiang, Bing Yin, and Xia Hu. 2023. Harnessing the Power of LLMs in Practice: A Survey on ChatGPT and Beyond. arXiv:2304.13712 [cs.CL]
[127] Da Yu, Saurabh Naik, Arturs Backurs, Sivakanth Gopi, Huseyin A. Inan, Gautam Kamath, Janardhan Kulkarni, Yin Tat Lee, Andre Manoel, Lukas Wutschitz, Sergey Yekhanin, and Huishuai Zhang. 2021. Differentially Private Fine-tuning of Language Models. arXiv:2110.06500 https://arxiv.org/abs/2110.06500
[128] Xuan Zhang and Wei Gao. 2023. Towards LLM-based Fact Verification on News Claims with a Hierarchical Step-by-Step Prompting Method. arXiv:2310.00305 [cs.CL]
[129] Yue Zhang, Yafu Li, Leyang Cui, Deng Cai, Lemao Liu, Tingchen Fu, Xinting Huang, Enbo Zhao, Yu Zhang, Yulong Chen, Longyue Wang, Anh Tuan Luu, Wei Bi, Freda Shi, and Shuming Shi. 2023. Siren's Song in the AI Ocean: A

Survey on Hallucination in Large Language Models. arXiv:2309.01219 [cs.CL]

[130] Zheng Zhang, Jie Gao, Ranjodh Singh Dhaliwal, and Toby Jia-Jun Li. 2023. VISAR: A Human-AI Argumentative Writing Assistant with Visual Programming and Rapid Draft Prototyping. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (, San Francisco, CA, USA,) *(UIST '23)*. Association for Computing Machinery, New York, NY, USA, Article 5, 30 pages. doi:10.1145/3586183.3606800

[131] Zhiping Zhang, Michelle Jia, Hao-Ping (Hank) Lee, Bingsheng Yao, Sauvik Das, Ada Lerner, Dakuo Wang, and Tianshi Li. 2024. "It's a Fair Game", or Is It? Examining How Users Navigate Disclosure Risks and Benefits When Using LLM-Based Conversational Agents. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 156, 26 pages. doi:10.1145/3613904.3642385

[132] Michael Zimmer. 2010. "But the data is already public": on the ethics of research in Facebook. *Ethics and Information Technology* 12 (2010), 313–325. https://api.semanticscholar.org/CorpusID:24881139

Received January 2024; revised July 2024; accepted October 2024