Understanding, Designing, and Governing ICT/AI Experience for Vulnerable Populations

1 Research Overview

Human-centered ICT/AI depends on the integration of deep understanding of human and society, value-centered design, and governance. I aim to understand, design, and govern ICT/AI experience for social good, with a focus on vulnerable populations (e.g., older adults, people with disabilities, people in authoritarian regimes), from a human-centered perspective. My ultimate goal is to make ICT/AI experience more equitable, accessible, beneficial, and ethical for all.

In my research, I combine qualitative and quantitative methods to gain deeper insights into the interaction between humans and technologies as well as how technologies incur societal impact. I conduct research through interview, qualitative content analysis, survey, design, and NLP/ML approach. I'm also keen on proposing policy recommendations to turn research insights into practice.

My research has been published in premier conferences in computer security, human-computer interaction (HCI), and information sciences, e.g., IEEE S&P [5], USENIX Security [6], SOUPS [10, 5], CSCW [12, 7], iConference [1], HICSS [8, 14, 9], and IC2S2 [13]. My work on understanding real world AI ethics incidents received a Best Paper Award at HICSS [8]. My research on understanding security practices and perceptions of smart contract developers [6] has been reported in press [2], which has the potential to inform security practices of real world software engineers.

2 Understanding ICT/AI Experience of Vulnerable Populations

How do people in authoritarian regimes experience ICT/AI?

More censorship and surveillance are imposed on people in authoritarian regimes. We approached the user experience and challenges of multiple ICT use cases in China, such as social media and blockchain. Through an online survey with Sina Weibo[1] users, we uncovered how users were censored by the platform, how and why they practiced self-censorship on different topics, and their various anti-censorship strategies [1]. While blockchain is known to be decentralized and free from governmental regulation, it might not be the case in authoritarian regimes. Through semi-structured interviews, we found that crypto investors in China tended to use centralized exchanges over decentralized exchanges, since they did not and were not allowed to have a mindset of decentralization [11].

How do women in conservative cultures use ICT/AI for expression and value realization?

More traditional gender roles are assigned to women and more gender oppression exists in relatively conservative cultures. While publicly discussing gender issues is hard for many of them, e.g., due to face saving in East Asian cultures, we approached an anonymous female community on Sina Weibo with a qualitative content analysis, identifying 20 issues commonly discussed such as workplace harassment and domestic violence [14]. With this taxonomy, we aim to spur more research in understanding and combating these gender issues. In a followup study, we approached a more specific gender issue: women often find difficulty in re-employment after giving birth due

[1] Sina Weibo is a Chinese microblogging website.

to a lack of legal protection in China. Through interviews with mom vloggers on short-form video sharing platforms (SVSPs), we found that mom vlogging helped them realize their values, but also brought challenges such as being stretched by both intensive motherhood and heavy digital work, competition among moms, and privacy concerns [12].

How do older adults become active contributors on social media platforms?

Older adults are conventionally perceived as later technology adopters and passive online recipients. We provided a counterexample with a case study of the significant surge in older adult content creators in SVSPs [7]. By conducting interviews with older adult content creators, we revealed the reasons why they were initially attracted to SVSPs (i.e., perceived ease of participation, enjoyment) and why they were kept engaged (i.e., equitable attention, reciprocal support). SVSPs offered a low-barrier and equitable platform through their near-automatic use and relatively equal opportunities for recommendations, allowing everyone to reach audiences. Based on the findings, we reflected on how SVSPs' technical affordance supported older adults in the transition from lurkers to contributors, and advocated for participation equity and supportive environments to promote more inclusive social media.

3 Designing ICT/AI Experience for Vulnerable Populations

How can researchers address accessibility challenges through inclusive design?

People with disabilities often face difficulty in using emergent technologies, which pay little attention to accessibility. We investigated crypto wallets as a case study and uncovered inter-related accessibility, learnability, and security issues with MetaMask, one of the most popular crypto wallets [10]. We further presented an iterative redesign of MetaMask to make it more accessible for blind users. Our evaluation results showed notable improvements for accessibility after two rounds of design iterations. Based on the results, we discussed design implications for creating more accessible and secure crypto wallets for blind users, such as adhering to accessibility standards and best practices, and understanding and addressing blind users' unique challenges and needs than sighted users. These design principles can easily generalize to ICT/AI design in general.

4 Governing ICT/AI to Make Them Work Ethically for All

How do ICT/AI ethics incidents happen in real world?

Toward obtaining a general and practical understanding of how AI ethics incidents occur in the real world as well as their social impact, we analyzed the AI Incident Database with a qualitative content analysis [8]. We categorized application areas that often involved ethical issues of AI (e.g.,language/vision models, autonomous driving) and AI risks (e.g., physical safety, racial/gender bias), and inspected AI risks associated with each application area. With the taxonomies and analysis, we aimed to provide a perspective for policy makers to formulate more contextual and operable AI guidelines and regulatory frameworks. Drawing on this research as well as my colleagues' research, we wrote to National Telecommunications and Information Administration to encourage contextual federal rule-making regarding anti-discrimination and contextual AI audits [4].

How can design play a role in addressing ethical issues of ICT/AI?

Design can play a role in nudging users to make more ethical decisions during ICT/AI use. The energy consumption of cryptocurrencies varies greatly: cryptocurrencies based on proof-of-work (e.g. Bitcoin) consume much more electricity than their counterparts that use alternative consensus mechanisms, such as proof-of-stake (e.g. Ethereum). Nevertheless, proof-of-work cryptocurrencies dominate the market. We adopted the design of energy labelling, i.e., displaying electricity consumption information on centralized exchanges, to successfully influence consumers' product preferences [3]. Exposing users to energy labels resulted in their decreased preference for energy-inefficient cryptocurrencies, showcasing the possibility of leveraging designs to address risks brought by ICT/AI.

How can AI ethics and cybersecurity be taught to K-12 students?

Teaching AI ethics to today's youth is a promising direction to prepare them for ICT/AI which are often fraught with ethical issues. To better deliver AI ethics education to youth, we interviewed US high school teachers and students about their experience of teaching and learning AI ethics. We identified relevant topics, such as epistemic norms, privacy, and digital citizenship as well as teaching strategies, such as discussions around current events, gamified activities, and content creation [5]. Challenges remained. For example, cyber hygiene instruction appeared ineffective at educating youth and promoting safer online behavior. Based on the study results, we offered practical suggestions for educators, school administrators, and cybersecurity practitioners to improve youth education on cybersecurity and AI ethics.

5 Future Research

Large language models (LLMs) are increasingly popular given their broad applicability and potential in assisting a wide range of tasks such as information seeking, content creation, and coding. Understanding how people perceive their use, how researchers can mindfully design them, and how to govern them to avoid potential risks is vital for their wide and ethical adoption. Below, I outline proposed research projects focusing on understanding, designing, and governing, respectively, toward creating human-centered, value-encoded LLMs.

Understanding How Professionals Perceive Benefits and Risks of LLMs

LLMs such as ChatGPT are associated with a lot of negative social sentiments such as fear and anxiety mainly for the ethical concerns brought by them, e.g., generating harmful content, replacing human workers, etc. However, there has been limited research in understanding how professionals who are extensively exposed to AI or even have worked with AI perceive the pros and cons of working with LLMs. Toward gaining a more realistic understanding of the perceived benefits and risks of LLMs in workplaces, we will conduct a mixed-methods study with professionals who have more informed opinions about AI.

Methodologically, we will conduct interviews and distribute a large-scale survey to understand (1) people's exposure to LLMs/AI, (2) how they worked with LLMs/AI if they had any experience, and (3) their perceived benefits and risks of working or co-existing with LLMs. We will target survey respondents and interview participants who have used LLMs/AI in their work, e.g., (1) financial practitioners who have used BloombergGPT to help with business tasks, (2) software engineers who have used ChatGPT to help with coding, (3) journalists who have used ChatGPT to help with writing, and (4) researchers who have used ChatGPT to help with research.

Our contribution will be (1) understanding the perceived benefits and risks of working with LLMs from people who have relevant experience, (2) imagining how people can co-exist with LLMs and how fruitful human-LLM collaboration can be formed, and (3) informing governance models, policy, and legislation to regulate the potential risks of LLMs.

Designing Value-encoded Health LLMs to Deliver Inclusive Healthcare

Designing LLMs applicable to healthcare is important for democratizing healthcare for people with limited access, such as people who are in disadvantaged socioeconomic status or people who live in developing countries with limited medical resources. In this project, we seek to design a value-encoded health LLM to serve people in needs.

In the first phase of the project (formative study), we seek to understand people's motivations for using LLMs in the realm of healthcare and their experienced challenges and concerns, as well as how the public adoption of LLMs for healthcare affects doctors. We will specifically explore the values people expect in such an LLM, such as inclusiveness, privacy, and accessibility. An online survey will be distributed, and semi-structured interviews will be conducted with 20 people who have used LLMs for healthcare purposes and 10 doctors in practice. The use cases, motivations, and challenges expressed by the participants will be leveraged to design a more inclusive, private, and accessible health LLM with relevant features in the second phase of the project (design).

The resulting product has the potential to ethically deliver healthcare for people who otherwise have no access and change the current, centralized, exclusive healthcare model.

Governing LLMs with A New Privacy Paradigm

LLMs are trained on a huge amount of data and interact with users intensively after deployment, naturally leading to privacy concerns. Differential privacy (DP) is a state-of-the-art mechanism to computationally operationalize privacy. Simply put, it adds randomized noise to the data of a group of people, so that no one's individual information can be inferred. Despite the fruitful research outcomes in differential privacy, this mechanism has not been widely deployed in applications and information systems. One key challenge is that the trade-off between privacy and data utility in different contexts, e.g., education, healthcare, etc., cannot be mathematically formulated. Enhancing DP with socially aware theories such as contextual integrity (CI) can potentially make it more socially meaningful in different contexts.

Toward this end, we will theoretically synthesize DP with CI into a contextual integrated differential privacy framework (CI-DP), especially regarding (1) the mapping between natural language privacy rules to numerical privacy budget and (2) the mapping between contextual conditions to DP variants. To assess the framework, we will distribute a survey to understand multiple stakeholders' (e.g., developers-app users, doctors-patients) choices of these mappings in different contexts.

The resulting privacy paradigm, i.e., CI-DP, has the potential to govern data usage and privacy risks of data-intensive applications such as LLMs.

References

[1] Chen, X., Xie, J., Wang, Z., Shen, B., and Zhou, Z. How we express ourselves freely: Censorship, self-censorship, and anti-censorship on a chinese social media. In *iConference 2023* (2023), Springer Nature Switzerland, pp. 93–108.

[2] Claburn, T. Smart contract developers not really focused on security. who knew? *The Register* (2022).

[3] Dragnoiu, A.-E., Platt, M., Wang, Z., and Zhou, Z. The more you know: Energy labelling enables more sustainable cryptocurrency investments. In *Workshop on Fintech and Decentralized Finance* (2023), IEEE.

[4] Kilhoffer, Z., Nlkolich, A., Sanfilippo, M. R., and Zhou, Z. Ai accountability policy. *NTIA-2023-0005-0810* (2023).

[5] Kilhoffer, Z., Zhou, Z., Wang, F., Tamton, F., Huang, Y., Kim, P., Yeh, T., and Wang, Y. "how technical do you get? i'm an english teacher"': Teaching and learning cybersecurity and ai ethics in high school. In *44th IEEE Symposium on Security and Privacy* (2023).

[6] Sharma, T., Zhou, Z., Miller, A., and Wang, Y. A mixed-methods study of security practices of smart contract developers. In *32nd USENIX Security Symposium (USENIX Security 23)* (2023), pp. 2545–2562.

[7] Tang, X., Ding, X. S., and Zhou, Z. Towards equitable online participation: A case of older adult content creators' role transition on short-form video sharing platforms. In *26th ACM Conference On Computer-Supported Cooperative Work And Social Computing* (2023).

[8] Wei, M., and Zhou, Z. Ai ethics issues in real world: Evidence from ai incident database. In *56th Hawaii International Conference on System Sciences* (2022).

[9] Zhou, Z., Ding, X., Tang, X., and Chen, Y. "i prefer an everyday style and dislike big food fighters": Integrating foodshow into everyday life. In *55th Hawaii International Conference on System Sciences* (2022), Hawaii International Conference on System Sciences.

[10] Zhou, Z., Sharma, T., Emano, L., Das, S., and Wang, Y. Iterative design of an accessible crypto wallet for blind users. In *19th Symposium on Usable Privacy and Security* (2023).

[11] Zhou, Z., and Shen, B. Toward understanding the use of centralized exchanges for decentralized cryptocurrency. In *13th International Conference on Applied Human Factors and Ergonomics* (2022).

[12] Zhou, Z., Shen, B., Zimmer, F., Xia, C., and Tong, X. More than a wife and a mom: A study of mom vlogging practices in china. In *26th ACM Conference On Computer-Supported Cooperative Work And Social Computing* (2022).

[13] Zhou, Z., Sun, J., Pei, J., Peng, N., and Xiong, J. A moral- and event- centric inspection of gender bias in fairy tales at a large scale. In *9th International Conference on Computational Social Science* (2022).

[14] Zhou, Z., Wang, Z., and Zimmer, F. Anonymous expression in an online community for women in china. In *56th Hawaii International Conference on System Sciences* (2022).