Between Threat and Tool: When Users Are Asked To Design Their Competitors

ANONYMOUS AUTHOR(S)*

AI foundation models are discussed and promoted in contradictory ways: as supportive tools that augment human work, e.g. helpful coding assistants, and as autonomous systems capable of human replacement, e.g. autonomous world-class coders. This paradox exemplifies a broader challenge in human-AI interaction design: we are asking people to help create 'human-centered' systems built on technologies that are simultaneously being developed into what users perceive as existential threats to their professional roles, regardless of their actual capabilities.

This position paper examines how concerns around professional replacement shape user experience, particularly through the lens of AI systems presented as both potential replacements and supportive tools, often based on the same foundation models. We argue that this contradiction creates significant challenges for design, as it can affect user acceptance, interaction patterns, and trust in AI-based systems in ways that current design approaches inadequately address. Drawing on examples from existing systems, we highlight how systems can simultaneously serve as workflow enhancements and be perceived as threats to professional identity and economic livelihood. Current interface paradigms, particularly chat-based interactions, can amplify these perceptions by implying cognitive equality between human and AI agents. This dynamic is further complicated by a tension that is faced by users: their participation is welcomed for improving specific applications but remains limited in influencing the foundation models that determine these applications' core capabilities.

To address these challenges, we propose shifting from design approaches that emphasize AI capability and autonomy, offering recommendations for design patterns that explicitly position AI as a tool and emphasize human agency and expertise – an approach that engages with, rather than ignores, the threat to professional identities.

CCS Concepts: • **Human-centered computing** → **Human computer interaction (HCI)**; **Natural language interfaces**; • **Computing methodologies** → **Intelligent agents**; **Natural language generation**.

Additional Key Words and Phrases: Human-Centered AI, Human-AI Collaboration, Human-Computer Interaction, Artificial Intelligence (AI), AI Anxiety, Large Language Models (LLMs), Participatory Design, Design Values

ACM Reference Format:

Anonymous Author(s). 2025. Between Threat and Tool: When Users Are Asked To Design Their Competitors. In *Proceedings of GenAICHI: CHI 2025 Workshop on Generative AI and HCI (GenAICHI '25).* ACM, New York, NY, USA, 10 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

The current discourse around generative AI systems presents designers with a paradoxical challenge: developing interfaces and interaction patterns for users who may view these systems as threats to their professional roles [65, 96]. This tension is particularly acute because the same foundation models [40, 83, 85] often power both assistive tools and autonomous systems marketed as potential human replacements. A coding assistant like Microsoft's CoPilot [13] and foundation models evaluated as autonomous coders [66] rely on identical underlying technology, yet their interfaces frame the AI's role in fundamentally different ways.

Manuscript submitted to ACM

The duality creates a crisis of authenticity in human-centered design: How can designers genuinely claim to empower users while building on systems whose training and deployment often aim toward automation of professional work domains, creative processes, and knowledge work – even promising to exceed human capabilities in these areas.

This tension becomes more pressing given the increasing development of agentic AI systems [3, 5, 8, 26, 28, 28, 78] – systems designed to operate with increasing autonomy across domains – with goals including the performance of economically valuable work [41, 68].

Companies increasingly present and market AI systems as both collaborative tools [1, 4, 11, 15, 75], and autonomous agents [5, 17, 78], creating an inherent tension in how users perceive and interact with these systems. Consider the duality of coding-focused AI systems again: The same foundation models simultaneously power both coding assistants that claim to augment developer workflows [13] and autonomous coding agents marketed as capable of independent software development [29]. A developer might use GitHub Copilot [18] as a collaborative assistant while being aware that its underlying technology is simultaneously being developed into systems that promise to generate, test, and deploy complete applications autonomously [27, 66]. The HCI community also shows increasing interest in developing collaborative coding tools [62, 94]. The challenge extends beyond interface design to the broader practice of human-centered design, where we must reconcile the development of 'collaborative' [37, 75, 91] or 'supportive' [63, 73, 95] tools with their underlying technologies being actively developed towards human capability replication.

The field of human-AI interaction must confront a critical question: How do we design systems that acknowledge this fundamental duality rather than obscuring it? Traditional approaches to human-centered design [25] may prove insufficient when users recognize the underlying technology as a potential threat to their professional identity [65] and standing – encompassing both immediate economic security [49] and deeper concerns about skill relevance, knowledge assets, and workplace authority [47, 98]. While these perceptions may not always align with current technical capabilities, they significantly influence how users interact with AI systems. Professional identity, as a core component of self-concept, shapes responses to AI tools, affecting the quality and authenticity of human input, trust in outputs, willingness to incorporate AI suggestions, engagement with designers and researchers during system development, and the overall effectiveness of human-AI collaboration.

We examine how the current media and marketing environment lends itself to rising displacement anxiety that can, in turn, shape user experience and trust in AI systems. For this, we particularly focus on cases where foundation models serve both assistive and autonomous functions. Based on this analysis, we propose design approaches that recenter AI as a tool while giving end-users appropriate authority and agency through specific interface decisions.

2 Dual Narratives: AI as Both Threat and Tool

The duality of the same technologies being positioned as collaborative tools [4, 10] and marketed as autonomous replacements for human workers [9, 29] manifests across multiple professional domains. This duality creates tension between the narrative of augmentation and the reality of displacement concerns. We illustrate this tension through concrete examples.

The creative industries provide a stark illustration of this tension [45]: While companies and researchers advance AI image and video generators as tools to enhance artistic workflows [43, 79, 89], foundation models drive systems marketed as autonomous creative agents [14]. This tension became particularly visible during the 2023 SAG-AFTRA and WGA strikes, where actors and writers confronted the possible impacts of AI systems on their industry [16].

Similar concerns emerge in scientific research, where AI systems are designed as collaborative research tools [7, 87, 92, 93], while also being presented as autonomous 'deep research' assistants [6, 8, 12, 39].

The pattern repeats in enterprise settings, with customer service providing a clear example: Companies promote AI as enhancing human agent capabilities [2] through better information retrieval and response suggestions. However, these same technologies are already being deployed to replace human customer service agents [88], creating a disconnect between marketing narratives and implementation reality. Sales departments also exemplify this contradiction: while AI tools are marketed to augment human sales representatives with data-driven insights and automated follow-ups [57], companies are simultaneously developing autonomous sales agents [9]. Government agencies demonstrate this duality through initiatives like ChatGPTGov [10], which markets AI as a tool to enhance worker productivity while AI is also being developed to automate analysis and auditing tasks traditionally performed by human analysts [74, 97].

This consistent pattern of dual positioning — tools for augmentation versus agents of automation — shapes how users approach and interact with AI systems in professional contexts. While current human-AI collaboration research emphasizes the benefits of system augmentation [24, 100, 103], it often overlooks how this fundamental contradiction affects user engagement.

This research emphasizes the importance of participatory engagement in system development [23, 32, 71, 99, 104], with approaches often focusing narrowly on improving system functionality — making systems more usable or efficient within their defined scope — rather than **addressing fundamental power dynamics**. While some approaches address power dynamics for specific, limited parts of the system [44, 50, 90], they rarely confront how users are positioned within broader power structures, or how they might meaningfully influence the trajectory of AI systems that have far-reaching implications for their professional domains and society at large [21, 60, 81]. This gap between localized interventions and systemic influence remains an area needing research attention.

The responsible AI community has begun exploring contestability in AI systems [48, 102], recognizing that meaningful user engagement requires more than just soliciting feedback for system improvement. Recent work in critical computing suggests that community engagement should extend beyond accepting or improving existing systems to include the power to fundamentally reshape or reject them [34, 46, 56, 82, 84]. This perspective aligns with broader discussions about democratizing AI development [22] and ensuring that technological advancement serves rather than supplants human agency [31, 58, 59].

These frameworks for user agency and system contestability provide context for understanding how users respond to fundamental concerns about professional autonomy and agency [65]. When confronted with technologies that can be perceived as both threat and tool, users may resist [34, 82], adapt [54, 101], or withdraw [46, 55] from interaction. Current approaches to human-AI collaboration must therefore address not just system usability but the fundamental tension between augmentation [42, 64, 70, 72, 76] and automation [20, 51, 52, 61] that underlies these systems.

3 Recentering AI as a Tool

The dual positioning of AI systems as both tools and potential replacements demands a fundamental shift in how we approach system and interaction design. Current approaches prioritize enhancing system performance and workflow integration without examining how this dual role shapes user participation. When system improvements increase both capability and autonomy, they can undermine user agency [67] and authority despite their intended benefits [30, 69].

Specific design elements can illustrate these challenges. For example, conversational AI systems often adopt interaction patterns traditionally reserved for human-to-human communication. When AI systems engage users through chat interfaces—a medium historically used for human conversation – they can create misleading expectations about the AI's capabilities and role [36]. Recent research demonstrates how these interfaces shape both AI behaviour and user

expectations: models exhibit sycophantic tendencies, adjusting their responses to agree with user corrections regardless of accuracy [33, 77], while users treat them as conversational platforms for information retrieval and 'search' [53, 86].

This challenge extends beyond interface design to core research values in AI development. Rather than focusing on enhancing narrow capabilities — which can inadvertently advance replacement potential and work against user interest through the duality — we need to reimagine the purposes of engagement. Eryk Salvaggio's analysis of NotebookLM [19] demonstrates how interfaces could instead center collaborative discovery and learning through valuing 'interestingness' [38], offering alternative central design values in AI research and design.

To address these challenges, we propose four design strategies particularly relevant to human-computer interaction research, focusing on how system design can reshape user engagement with AI tools:

- Creating asymmetric visual and interaction hierarchies that position AI as a tool rather than a professional peer
- Implementing transparent communication about system limitations and capabilities
- Developing interface patterns that require meaningful human oversight and input
- Incorporating design elements that highlight human expertise and judgment

Fundamentally, this approach requires shifting our design language and values from **automation to augmentation**. Interface patterns, interaction flows, and even descriptive language should emphasize AI's role in enhancing rather than replacing human capabilities. This reframing requires careful attention to transparency challenges. While the explainable AI (XAI) community has made progress in communicating system limitations [35, 80], designers must balance technical disclosure with maintaining user confidence in system utility.

The tool-threat duality introduces complex considerations, where even incremental improvements in AI capabilities can also work against user interests. By acknowledging these tensions explicitly in our design approaches, we can create interfaces that support genuine human-AI collaboration while maintaining appropriate engagement dynamics that preserve user agency and professional identity.

4 Conclusion

The examples examined in this paper illustrate how generative AI systems simultaneously occupy roles as collaborative tools and perceived threats to professional identity. This duality creates fundamental challenges for participatory design approaches that aim to enhance human-centered AI applications. Specifically, the challenge lies in how to acknowledge and account for this duality rather than obscuring it. Current interface paradigms, e.g. conversational AI, can inadvertently amplify these tensions rather than address them.

We argue that addressing these tensions requires expanding our design focus beyond technical capabilities to consider how users' professional identity and agency shape their engagement with AI systems. By developing interfaces that explicitly acknowledge these dynamics while emphasizing human expertise and judgment, we can create more responsible systems.

This work directly addresses rising uncertainties in AI system design, particularly the gap between design intentions and user expectations. These uncertainties call for new design approaches that can effectively position AI systems as tools while acknowledging and addressing users' concerns about professional displacement. Given these uncertainties, we call for reshaping how the HCI community designs and develops AI systems to establish their role as tools, not threats.

Through this workshop, we aim to facilitate dialogue between HCI researchers, practitioners, and affected professionals to develop concrete design patterns that better navigate these complex dynamics.

References

[1] [n. d.]. AI assistants | IBM. https://www.ibm.com/ai-assistants

[2] [n. d.]. AI for Customer Service | IBM. https://www.ibm.com/ai-customer-service

[3] [n. d.]. Building effective agents. https://www.anthropic.com/research/building-effective-agents

[4] [n. d.]. ChatGPT for enterprise. https://openai.com/chatgpt/enterprise/

[5] [n. d.]. Computer-Using Agent. https://openai.com/index/computer-using-agent/

[6] [n. d.]. Gemini: Try Deep Research and Gemini 2.0 Flash Experimental. https://blog.google/products/gemini/google-gemini-deep-research/

[7] [n. d.]. Google NotebookLM | Note Taking & Research Assistant Powered by AI. https://notebooklm.google/

[8] [n. d.]. Grok 3 Beta — The Age of Reasoning Agents. https://x.ai/blog/grok-3

[9] [n. d.]. Hire Ava, the AI SDR | the Best AI Sales Agent. https://www.artisan.co/ai-sales-agent

[10] [n. d.]. Introducing ChatGPT Gov. https://openai.com/global-affairs/introducing-chatgpt-gov/

[11] [n. d.]. Introducing Claude. https://www.anthropic.com/news/introducing-claude

[12] [n. d.]. Introducing deep research. https://openai.com/index/introducing-deep-research/

[13] [n. d.]. Microsoft Copilot: Ihr KI-Begleiter. https://copilot.microsoft.com

[14] [n. d.]. Sora | OpenAI | OpenAI. https://openai.com/sora/

[15] 2020. Collaboration.AI | Unleash the power of relationships. https://collaboration.ai/

[16] 2024. AI and Hollywood: 5 questions for SAG-AFTRA's chief negotiator. https://www.weforum.org/stories/2024/03/ai-hollywood-strike-sag-aftra-technology/

[17] 2024. Introducing Gemini 2.0: our new AI model for the agentic era. https://blog.google/technology/google-deepmind/google-gemini-ai-update-december-2024/

[18] 2025. GitHub Copilot · Your AI pair programmer. https://github.com/features/copilot

[19] 2025. It's Interesting Because. https://mail.cyberneticforests.com/its-interesting-because/

[20] Daron Acemoglu and Pascual Restrepo. 2018. Artificial intelligence, automation, and work. In *The economics of artificial intelligence: An agenda*. University of Chicago Press, 197–236.

[21] Abeba Birhane, William Isaac, Vinodkumar Prabhakaran, Mark Diaz, Madeleine Clare Elish, Iason Gabriel, and Shakir Mohamed. 2022. Power to the people? Opportunities and challenges for participatory AI. In *Proceedings of the 2nd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*. 1–8.

[22] Daniel James Bogiatzis-Gibbons. 2024. Beyond Individual Accountability: (Re-)Asserting Democratic Control of AI. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Rio de Janeiro Brazil, 74–84. https://doi.org/10.1145/3630106.3658541

[23] Elizabeth Bondi, Lily Xu, Diana Acosta-Navas, and Jackson A Killian. 2021. Envisioning communities: a participatory approach towards AI for social good. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*. 425–436.

[24] Amanda Buddemeyer, Jennifer Nwogu, Jaemarie Solyst, Erin Walker, Tara Nkrumah, Amy Ogan, Leshell Hatley, and Angela Stewart. 2022. Unwritten Magic: Participatory Design of AI Dialogue to Empower Marginalized Voices. In *Proceedings of the 2022 ACM Conference on Information Technology for Social Good*. ACM, Limassol Cyprus, 366–372. https://doi.org/10.1145/3524458.3547119

[25] Tara Capel and Margot Brereton. 2023. What is Human-Centered about Human-Centered AI? A Map of the Research Landscape. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–23. https://doi.org/10.1145/3544548.3580959

[26] Alan Chan, Rebecca Salganik, Alva Markelius, Chris Pang, Nitarshan Rajkumar, Dmitrii Krasheninnikov, Lauro Langosco, Zhonghao He, Yawen Duan, Micah Carroll, Michelle Lin, Alex Mayhew, Katherine Collins, Maryam Molamohammadi, John Burden, Wanru Zhao, Shalaleh Rismani, Konstantinos Voudouris, Umang Bhatt, Adrian Weller, David Krueger, and Tegan Maharaj. 2023. Harms from Increasingly Agentic Algorithmic Systems. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (FAccT '23)*. Association for Computing Machinery, New York, NY, USA, 651–666. https://doi.org/10.1145/3593013.3594033

[27] Jun Shern Chan, Neil Chowdhury, Oliver Jaffe, James Aung, Dane Sherburn, Evan Mays, Giulio Starace, Kevin Liu, Leon Maksin, Tejal Patwardhan, Lilian Weng, and Aleksander Mądry. 2024. MLE-bench: Evaluating Machine Learning Agents on Machine Learning Engineering. https://doi.org/10.48550/arXiv.2410.07095 arXiv:2410.07095 [cs].

[28] Nazli Cila. 2022. Designing Human-Agent Collaborations: Commitment, responsiveness, and support. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–18. https://doi.org/10.1145/3491102.3517500

[29] cognition.ai. [n. d.]. Cognition | Introducing Devin, the first AI software engineer. https://cognition.ai/

[30] Grant Cooper, Kok-Sing Tang, and Natasha Anne Rappa. 2025. Generative Artificial Intelligence as Epistemic Authority?: Perspectives from Higher Education. In *Artificial Intelligence Applications in Higher Education*. Routledge, 106–122.

[31] Vedant Das Swain, Lan Gao, William A Wood, Srikruthi C Matli, Gregory D. Abowd, and Munmun De Choudhury. 2023. Algorithmic Power or Punishment: Information Worker Perspectives on Passive Sensing Enabled AI Phenotyping of Performance and Wellbeing. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–17. https://doi.org/10.1145/3544548.3581376

[32] Kerstin Denecke, Elia Gabarron, Rebecca Grainger, Stathis Th Konstantinidis, Annie Lau, Octavio Rivera-Romero, Talya Miron-Shatz, and Mark Merolli. 2019. Artificial intelligence for participatory health: applications, impact, and future implications. *Yearbook of medical informatics* 28, 01 (2019), 165–173.

[33] Carson Denison, Monte MacDiarmid, Fazl Barez, David Duvenaud, Shauna Kravec, Samuel Marks, Nicholas Schiefer, Ryan Soklaski, Alex Tamkin, Jared Kaplan, Buck Shlegeris, Samuel R. Bowman, Ethan Perez, and Evan Hubinger. 2024. Sycophancy to Subterfuge: Investigating Reward-Tampering in Large Language Models. https://doi.org/10.48550/arXiv.2406.10162 arXiv:2406.10162 [cs].

[34] Alicia DeVrio, Motahhare Eslami, and Kenneth Holstein. 2024. Building, Shifting, & Employing Power: A Taxonomy of Responses From Below to Algorithmic Harm. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24)*. Association for Computing Machinery, New York, NY, USA, 1093–1106. https://doi.org/10.1145/3630106.3658958

[35] Rudresh Dwivedi, Devam Dave, Het Naik, Smiti Singhal, Rana Omer, Pankesh Patel, Bin Qian, Zhenyu Wen, Tejal Shah, Graham Morgan, et al. 2023. Explainable AI (XAI): Core ideas, techniques, and solutions. *Comput. Surveys* 55, 9 (2023), 1–33.

[36] Andrea Ferrario, Alessandro Facchini, and Alberto Termine. 2024. Experts or authorities? The strange case of the presumed epistemic superiority of artificial intelligence systems. *Minds and Machines* 34, 3 (2024), 30.

[37] Fiona Fui-Hoon Nah, Ruilin Zheng, Jingyuan Cai, Keng Siau, and Langtao Chen. 2023. Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration. , 277–304 pages.

[38] Google DeepMind. 2024. Inside NotebookLM with Raiza Martin and Steven Johnson. https://www.youtube.com/watch?v=mccQdu5afZw

[39] Juraj Gottweis, Wei-Hung Weng, Alexander Daryin, Tao Tu, Anil Palepu, Petar Sirkovic, Artiom Myaskovsky, Felix Weissenberger, Keran Rong, Ryutaro Tanno, Khaled Saab, Dan Popovici, Jacob Blum, Fan Zhang, Katherine Chou, Avinatan Hassidim, Burak Gokturk, Amin Vahdat, Pushmeet Kohli, Yossi Matias, Andrew Carroll, Kavita Kulkarni, Nenad Tomasev, Vikram Dhillon, Eeshit Dhaval Vaishnav, Byron Lee, Tiago R D Costa, José R Penadés, Gary Peltz, Yunhan Xu, Annalisa Pawlosky, Alan Karthikesalingam, and Vivek Natarajan. [n. d.]. Towards an AI co-scientist. ([n. d.]).

[40] Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, Amy Yang, Angela Fan, Anirudh Goyal, Anthony Hartshorn, Aobo Yang, Archi Mitra, Archie Sravankumar, Artem Korenev, Arthur Hinsvark, Arun Rao, Aston Zhang, Aurelien Rodriguez, Austen Gregerson, Ava Spataru, Baptiste Roziere, Bethany Biron, Binh Tang, Bobbie Chern, Charlotte Caucheteux, Chaya Nayak, Chloe Bi, Chris Marra, Chris McConnell, Christian Keller, Christophe Touret, Chunyang Wu, Corinne Wong, Cristian Canton Ferrer, Cyrus Nikolaidis, Damien Allonsius, Daniel Song, Danielle Pintz, Danny Livshits, Danny Wyatt, David Esiobu, Dhruv Choudhary, Dhruv Mahajan, Diego Garcia-Olano, Diego Perino, Dieuwke Hupkes, Egor Lakomkin, Ehab AlBadawy, Elina Lobanova, Emily Dinan, Eric Michael Smith, Filip Radenovic, Francisco Guzmán, Frank Zhang, Gabriel Synnaeve, Gabrielle Lee, Georgia Lewis Anderson, Govind Thattai, Graeme Nail, Gregoire Mialon, Guan Pang, Guillem Cucurell, Hailey Nguyen, Hannah Korevaar, Hu Xu, Hugo Touvron, Iliyan Zarov, Imanol Arrieta Ibarra, Isabel Kloumann, Ishan Misra, Ivan Evtimov, Jack Zhang, Jade Copet, Jaewon Lee, Jan Geffert, Jana Vranes, Jason Park, Jay Mahadeokar, Jeet Shah, Jelmer van der Linde, Jennifer Billock, Jenny Hong, Jenya Lee, Jeremy Fu, Jianfeng Chi, Jianyu Huang, Jiawen Liu, Jie Wang, Jiecao Yu, Joanna Bitton, Joe Spisak, Jongsoo Park, Joseph Rocca, Joshua Johnstun, Joshua Saxe, Junteng Jia, Kalyan Vasuden Alwala, Karthik Prasad, Kartikeya Upasani, Kate Plawiak, Ke Li, Kenneth Heafield, Kevin Stone, Khalid El-Arini, Krithika Iyer, Kshitiz Malik, Kuenley Chiu, Kunal Bhalla, Kushal Lakhotia, Lauren Rantala-Yeary, Laurens van der Maaten, Lawrence Chen, Liang Tan, Liz Jenkins, Louis Martin, Lovish Madaan, Lubo Malo, Lukas Blecher, Lukas Landzaat, Luke de Oliveira, Madeline Muzzi, Mahesh Pasupuleti, Mannat Singh, Manohar Paluri, Marcin Kardas, Maria Tsimpoukelli, Mathew Oldham, Mathieu Rita, Maya Pavlova, Melanie Kambadur, Mike Lewis, Min Si, Mitesh Kumar Singh, Mona Hassan, Naman Goyal, Narjes Torabi, Nikolay Bashlykov, Nikolay Bogoychev, Niladri Chatterji, Ning Zhang, Olivier Duchenne, Onur Çelebi, Patrick Alrassy, Pengchuan Zhang, Pengwei Li, Petar Vasic, Peter Weng, Prajjwal Bhargava, Pratik Dubal, Praveen Krishnan, Punit Singh Koura, Puxin Xu, Qing He, Qingxiao Dong, Ragavan Srinivasan, Raj Ganapathy, Ramon Calderer, Ricardo Silveira Cabral, Robert Stojnic, Roberta Raileanu, Rohan Maheswari, Rohit Girdhar, Rohit Patel, Romain Sauvestre, Ronnie Polidoro, Roshan Sumbaly, Ross Taylor, Ruan Silva, Rui Hou, Rui Wang, Saghar Hosseini, Sahana Chennabasappa, Sanjay Singh, Sean Bell, Seohyun Sonia Kim, Sergey Edunov, Shaoliang Nie, Sharan Narang, Sharath Raparthy, Sheng Shen, Shengye Wan, Shruti Bhosale, Shun Zhang, Simon Vandenhende, Soumya Batra, Spencer Whitman, Sten Sootla, Stephane Collot, Suchin Gururangan, Sydney Borodinsky, Tamar Herman, Tara Fowler, Tarek Sheasha, Thomas Georgiou, Thomas Scialom, Tobias Speckbacher, Todor Mihaylov, Tong Xiao, Ujjwal Karn, Vedanuj Goswami, Vibhor Gupta, Vignesh Ramanathan, Viktor Kerkez, Vincent Gonguet, Virginie Do, Vish Vogeti, Vítor Albiero, Vladan Petrovic, Weiwei Chu, Wenhan Xiong, Wenyin Fu, Whitney Meers, Xavier Martinet, Xiaodong Wang, Xiaofang Wang, Xiaoqing Ellen Tan, Xide Xia, Xinfeng Xie, Xuchao Jia, Xuewei Wang, Yaelle Goldschlag, Yashesh Gaur, Yasmine Babaei, Yi Wen, Yiwen Song, Yuchen Zhang, Yue Li, Yuning Mao, Zacharie Delpierre Coudert, Zheng Yan, Zhengxing Chen, Zoe Papakipos, Aaditya Singh, Aayushi Srivastava, Abha Jain, Adam Kelsey, Adam Shajnfeld, Adithya Gangidi, Adolfo Victoria, Ahuva Goldstand, Ajay Menon, Ajay Sharma, Alex Boesenberg, Alexei Baevski, Allie Feinstein, Amanda Kallet, Amit Sangani, Amos Teo, Anam Yunus, Andrei Lupu, Andres Alvarado, Andrew Caples, Andrew Gu, Andrew Ho, Andrew Poulton, Andrew Ryan, Ankit Ramchandani, Annie Dong, Annie Franco, Anuj Goyal, Aparajita Saraf, Arkabandhu Chowdhury, Ashley Gabriel, Ashwin Bharambe, Assaf Eisenman, Azadeh Yazdan, Beau James, Ben Maurer, Benjamin Leonhardi, Bernie Huang, Beth Loyd, Beto De Paola, Bhargavi Paranjape, Bing Liu, Bo Wu, Boyu Ni, Braden Hancock, Bram Wasti, Brandon Spence, Brani Stojkovic, Brian Gamido, Britt Montalvo, Carl Parker, Carly Burton, Catalina Mejia, Ce Liu, Changhan Wang, Changkyu Kim, Chao Zhou, Chester Hu, Ching-Hsiang Chu, Chris Cai, Chris Tindal, Christoph Feichtenhofer, Cynthia Gao, Damon Civin, Dana Beaty, Daniel Kreymer, Daniel Li, David Adkins, David Xu, Davide Testuggine, Delia David, Devi Parikh, Diana Liskovich, Didem Foss, Dingkang Wang, Duc Le, Dustin Holland, Edward Dowling, Eissa Jamil, Elaine Montgomery, Eleonora Presani, Emily Hahn, Emily Wood, Eric-Tuan Le, Erik Brinkman, Esteban Arcaute, Evan Dunbar, Evan Smothers, Fei Sun, Felix Kreuk, Feng Tian, Filippos Kokkinos, Firat Ozgenel, Francesco Caggioni, Frank Kanayet, Frank

Seide, Gabriela Medina Florez, Gabriella Schwarz, Gada Badeer, Georgia Swee, Gil Halpern, Grant Herman, Grigory Sizov, Guangyi, Zhang, Guna Lakshminarayanan, Hakan Inan, Hamid Shojanazeri, Han Zou, Hannah Wang, Hanwen Zha, Haroun Habeeb, Harrison Rudolph, Helen Suk, Henry Aspegren, Hunter Goldman, Hongyuan Zhan, Ibrahim Damlaj, Igor Molybog, Igor Tufanov, Ilias Leontiadis, Irina-Elena Veliche, Itai Gat, Jake Weissman, James Geboski, James Kohli, Janice Lam, Japhet Asher, Jean-Baptiste Gaya, Jeff Marcus, Jeff Tang, Jennifer Chan, Jenny Zhen, Jeremy Reizenstein, Jeremy Teboul, Jessica Zhong, Jian Jin, Jingyi Yang, Joe Cummings, Jon Carvill, Jon Shepard, Jonathan McPhie, Jonathan Torres, Josh Ginsburg, Junjie Wang, Kai Wu, Kam Hou U, Karan Saxena, Kartikay Khandelwal, Katayoun Zand, Kathy Matosich, Kaushik Veeraraghavan, Kelly Michelena, Keqian Li, Kiran Jagadeesh, Kun Huang, Kunal Chawla, Kyle Huang, Lailin Chen, Lakshya Garg, Lavender A, Leandro Silva, Lee Bell, Lei Zhang, Liangpeng Guo, Licheng Yu, Liron Moshkovich, Luca Wehrstedt, Madian Khabsa, Manav Avalani, Manish Bhatt, Martynas Mankus, Matan Hasson, Matthew Lennie, Matthias Reso, Maxim Groshev, Maxim Naumov, Maya Lathi, Meghan Keneally, Miao Liu, Michael L. Seltzer, Michal Valko, Michelle Restrepo, Mihir Patel, Mik Vyatskov, Mikayel Samvelyan, Mike Clark, Mike Macey, Mike Wang, Miquel Jubert Hermoso, Mo Metanat, Mohammad Rastegari, Munish Bansal, Nandhini Santhanam, Natascha Parks, Natasha White, Navyata Bawa, Nayan Singhal, Nick Egebo, Nicolas Usunier, Nikhil Mehta, Nikolay Pavlovich Laptev, Ning Dong, Norman Cheng, Oleg Chernoguz, Olivia Hart, Omkar Salpekar, Ozlem Kalinli, Parkin Kent, Parth Parekh, Paul Saab, Pavan Balaji, Pedro Rittner, Philip Bontrager, Pierre Roux, Piotr Dollar, Polina Zvyagina, Prashant Ratanchandani, Pritish Yuvraj, Qian Liang, Rachad Alao, Rachel Rodriguez, Rafi Ayub, Raghotham Murthy, Raghu Nayani, Rahul Mitra, Rangaprabhu Parthasarathy, Raymond Li, Rebekkah Hogan, Robin Battey, Rocky Wang, Russ Howes, Ruty Rinott, Sachin Mehta, Sachin Siby, Sai Jayesh Bondu, Samyak Datta, Sara Chugh, Sara Hunt, Sargun Dhillon, Sasha Sidorov, Satadru Pan, Saurabh Mahajan, Saurabh Verma, Seiji Yamamoto, Sharadh Ramaswamy, Shaun Lindsay, Shaun Lindsay, Sheng Feng, Shenghao Lin, Shengxin Cindy Zha, Shishir Patil, Shiva Shankar, Shuqiang Zhang, Shuqiang Zhang, Sinong Wang, Sneha Agarwal, Soji Sajuyigbe, Soumith Chintala, Stephanie Max, Stephen Chen, Steve Kehoe, Steve Satterfield, Sudarshan Govindaprasad, Sumit Gupta, Summer Deng, Sungmin Cho, Sunny Virk, Suraj Subramanian, Sy Choudhury, Sydney Goldman, Tal Remez, Tamar Glaser, Tamara Best, Thilo Koehler, Thomas Robinson, Tianhe Li, Tianjun Zhang, Tim Matthews, Timothy Chou, Tzook Shaked, Varun Vontimitta, Victoria Ajayi, Victoria Montanez, Vijai Mohan, Vinay Satish Kumar, Vishal Mangla, Vlad Ionescu, Vlad Poenaru, Vlad Tiberiu Mihailescu, Vladimir Ivanov, Wei Li, Wenchen Wang, Wenwen Jiang, Wes Bouaziz, Will Constable, Xiaocheng Tang, Xiaojian Wu, Xiaolan Wang, Xilun Wu, Xinbo Gao, Yaniv Kleinman, Yanjun Chen, Ye Hu, Ye Jia, Ye Qi, Yenda Li, Yilin Zhang, Ying Zhang, Yossi Adi, Youngjin Nam, Yu, Wang, Yu Zhao, Yuchen Hao, Yundi Qian, Yunlu Li, Yuzi He, Zach Rait, Zachary DeVito, Zef Rosnbrick, Zhaoduo Wen, Zhenyu Yang, Zhiwei Zhao, and Zhiyu Ma. 2024. The Llama 3 Herd of Models. https://doi.org/10.48550/arXiv.2407.21783 arXiv:2407.21783 [cs].

[41] Kunal Handa, Alex Tamkin, Miles McCain, Saffron Huang, Esin Durmus, Sarah Heck, Jared Mueller, Jerry Hong, Stuart Ritchie, Tim Belonax, Kevin K Troy, Dario Amodei, Jared Kaplan, Jack Clark, and Deep Ganguli. [n. d.]. Which Economic Tasks are Performed with AI? Evidence from Millions of Claude Conversations. ([n. d.]).

[42] Hossein Hassani, Emmanuel Sirimal Silva, Stephane Unger, Maedeh TajMazinani, and Stephen Mac Feely. 2020. Artificial Intelligence (AI) or Intelligence Augmentation (IA): What Is the Future? *AI* 1, 2 (2020), 143–155. https://doi.org/10.3390/ai1020008

[43] Jimpei Hitsuwari, Yoshiyuki Ueda, Woojin Yun, and Michio Nomura. 2023. Does human–AI collaboration lead to more creative art? Aesthetic evaluation of human-made and AI-generated haiku poetry. *Computers in Human Behavior* 139 (Feb. 2023), 107502. https://doi.org/10.1016/j.chb.2022.107502

[44] Yoyo Tsung-Yu Hou, Wen-Ying Lee, and Malte Jung. 2023. "Should I Follow the Human, or Follow the Robot?" — Robots in Power Can Have More Influence Than Humans on Decision-Making. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3544548.3581066

[45] Nanna Inie, https://orcid.org/0000-0002-5375-9542, View Profile, Jeanette Falk, https://orcid.org/0000-0001-7278-9344, View Profile, Steve Tanimoto, https://orcid.org/0000-0002-8175-7456, and View Profile. 2023. Designing Participatory AI: Creative Professionals? Worries and Expectations about Generative AI. *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (April 2023), 1–8. https://doi.org/10.1145/3544549.3585657

[46] Nari Johnson, https://orcid.org/0009-0008-3180-3582, View Profile, Sanika Moharana, https://orcid.org/0009-0007-4899-5309, View Profile, Christina Harrington, https://orcid.org/0000-0003-1850-6459, View Profile, Nazanin Andalibi, https://orcid.org/0000-0002-3257-2527, View Profile, Hoda Heidari, https://orcid.org/0000-0003-3710-4076, View Profile, Motahhare Eslami, https://orcid.org/0000-0002-1499-3045, and View Profile. 2024. The Fall of an Algorithm: Characterizing the Dynamics Toward Abandonment. *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency* (June 2024), 337–358. https://doi.org/10.1145/3630106.3658910

[47] 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 (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–18. https://doi.org/10.1145/3491102.3517533

[48] Naveena Karusala, Sohini Upadhyay, Rajesh Veeraraghavan, and Krzysztof Z. Gajos. 2024. Understanding Contestability on the Margins: Implications for the Design of Algorithmic Decision-making in Public Services. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–16. https://doi.org/10.1145/3613904.3641898

[49] Atoosa Kasirzadeh. 2025. Two Types of AI Existential Risk: Decisive and Accumulative. arXiv:2401.07836 [cs.CY] https://arxiv.org/abs/2401.07836

[50] Dajung Kim, Niko Vegt, Valentijn Visch, and Marina Bos-De Vos. 2024. How Much Decision Power Should (A)I Have?: Investigating Patients' Preferences Towards AI Autonomy in Healthcare Decision Making. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–17. https://doi.org/10.1145/3613904.3642883

[51] Seth Lazar. 2024. Automatic Authorities: Power and AI. arXiv:2404.05990 [cs.CY] https://arxiv.org/abs/2404.05990

[52] Michael Leyer and Sabrina Schneider. 2021. Decision augmentation and automation with artificial intelligence: threat or opportunity for managers? *Business Horizons* 64, 5 (2021), 711–724.

[53] Alice Li and Luanne Sinnamon. 2024. Generative AI Search Engines as Arbiters of Public Knowledge: An Audit of Bias and Authority. *Proceedings of the Association for Information Science and Technology* 61, 1 (2024), 205–217.

[54] Xuedong Liang, Gengxuan Guo, Lingli Shu, Qunxi Gong, and Peng Luo. 2022. Investigating the double-edged sword effect of AI awareness on employee's service innovative behavior. *Tourism Management* 92 (2022), 104564.

[55] Yunshuo Liu, Yanbin Li, Keni Song, and Fulei Chu. 2024. The two faces of Artificial Intelligence (AI): Analyzing how AI usage shapes employee behaviors in the hospitality industry. *International Journal of Hospitality Management* 122 (2024), 103875.

[56] Chiara Longoni, Andrea Bonezzi, and Carey K Morewedge. 2019. Resistance to medical artificial intelligence. *Journal of consumer research* 46, 4 (2019), 629–650.

[57] Xueming Luo, Marco Shaojun Qin, Zheng Fang, and Zhe Qu. 2021. Artificial Intelligence Coaches for Sales Agents: Caveats and Solutions. *Journal of Marketing* 85, 2 (March 2021), 14–32. https://doi.org/10.1177/0022242920956676 Publisher: SAGE Publications Inc.

[58] Henrietta Lyons, Tim Miller, and Eduardo Velloso. 2023. Algorithmic Decisions, Desire for Control, and the Preference for Human Review over Algorithmic Review. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (FAccT '23)*. Association for Computing Machinery, New York, NY, USA, 764–774. https://doi.org/10.1145/3593013.3594041

[59] Shuai Ma, Xinru Wang, Ying Lei, Chuhan Shi, Ming Yin, and Xiaojuan Ma. 2024. "Are You Really Sure?" Understanding the Effects of Human Self-Confidence Calibration in AI-Assisted Decision Making. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–20. https://doi.org/10.1145/3613904.3642671

[60] Jonne Maas and Aarón Moreno Inglés. 2024. Beyond Participatory AI. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, Vol. 7. 932–942.

[61] James Manyika, Michael Chui, Mehdi Miremadi, Jacques Bughin, Katy George, Paul Willmott, and Martin Dewhurst. 2017. A future that works: AI, automation, employment, and productivity. *McKinsey Global Institute Research, Tech. Rep* 60 (2017), 1–135.

[62] 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 (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–16. https://doi.org/10.1145/3544548.3580940

[63] Jingbo Meng and Yue Dai. 2021. Emotional support from AI chatbots: Should a supportive partner self-disclose or not? *Journal of Computer-Mediated Communication* 26, 4 (2021), 207–222.

[64] Steven M. MILLER. 2018. AI: Augmentation, more so than automation. *Asian Management Insights (Singapore Management University)* 5, 1 (May 2018), 1–20. https://ink.library.smu.edu.sg/ami/83

[65] Milad Mirbabaie, Felix Brünker, Nicholas R. J. Möllmann Frick, and Stefan Stieglitz. 2022. The rise of artificial intelligence – understanding the AI identity threat at the workplace. *Electronic Markets* 32, 1 (March 2022), 73–99. https://doi.org/10.1007/s12525-021-00496-x

[66] Samuel Miserendino, Michele Wang, Tejal Patwardhan, and Johannes Heidecke. 2025. SWE-Lancer: Can Frontier LLMs Earn $1 Million from Real-World Freelance Software Engineering? https://doi.org/10.48550/arXiv.2502.12115 arXiv:2502.12115 [cs].

[67] Margaret Mitchell, Avijit Ghosh, Alexandra Sasha Luccioni, and Giada Pistilli. 2025. Fully Autonomous AI Agents Should Not be Developed. arXiv:2502.02649 [cs.AI] https://arxiv.org/abs/2502.02649

[68] Meredith Ringel Morris, Jascha Sohl-dickstein, Noah Fiedel, Tris Warkentin, Allan Dafoe, Aleksandra Faust, Clement Farabet, and Shane Legg. 2024. Levels of AGI for Operationalizing Progress on the Path to AGI. https://doi.org/10.48550/arXiv.2311.02462 arXiv:2311.02462 [cs].

[69] Carl Öhman. 2024. We are building gods: AI as the anthropomorphised authority of the past. *Minds and Machines* 34, 1 (2024), 8.

[70] Souren Paul, Lingyao Yuan, Hemant K Jain, Lionel P Robert Jr, Jim Spohrer, and Hila Lifshitz-Assaf. 2022. Intelligence augmentation: Human factors in ai and future of work. *AIS Transactions on Human-Computer Interaction* 14, 3 (2022), 426–445.

[71] Organizers Of Queerinai, Anaelia Ovalle, Arjun Subramonian, Ashwin Singh, Claas Voelcker, Danica J. Sutherland, Davide Locatelli, Eva Breznik, Filip Klubicka, Hang Yuan, Hetvi J, Huan Zhang, Jaidev Shriram, Kruno Lehman, Luca Soldaini, Maarten Sap, Marc Peter Deisenroth, Maria Leonor Pacheco, Maria Ryskina, Martin Mundt, Milind Agarwal, Nyx Mclean, Pan Xu, A Pranav, Raj Korpan, Ruchira Ray, Sarah Mathew, Sarthak Arora, St John, Tanvi Anand, Vishakha Agrawal, William Agnew, Yanan Long, Zijie J. Wang, Zeerak Talat, Avijit Ghosh, Nathaniel Dennler, Michael Noseworthy, Sharvani Jha, Emi Baylor, Aditya Joshi, Natalia Y. Bilenko, Andrew Mcnamara, Raphael Gontijo-Lopes, Alex Markham, Evyn Dong, Jackie Kay, Manu Saraswat, Nikhil Vytla, and Luke Stark. 2023. Queer In AI: A Case Study in Community-Led Participatory AI. In *2023 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Chicago IL USA, 1882–1895. https://doi.org/10.1145/3593013.3594134

[72] Sebastian Raisch and Sebastian Krakowski. 2021. Artificial intelligence and management: The automation–augmentation paradox. *Academy of management review* 46, 1 (2021), 192–210.

[73] Alejandro Rodríguez-Ruiz, Elizabeth Krupinski, Jan-Jurre Mordang, Kathy Schilling, Sylvia H Heywang-Köbrunner, Ioannis Sechopoulos, and Ritse M Mann. 2019. Detection of breast cancer with mammography: effect of an artificial intelligence support system. *Radiology* 290, 2 (2019), 305–314.

[74] Scott Rosenberg. 2025. Musk's DOGE wants to overhaul the government with AI. https://www.axios.com/2025/02/05/musk-doge-ai-government-efficiency-safeguards

[75] Emma Schleiger, Claire Mason, Claire Naughtin, Andrew Reeson, and Cecile Paris. 2024. Collaborative Intelligence: A scoping review of current applications. *Applied Artificial Intelligence* 38, 1 (2024), 2327890.

[76] Yiduo Shao, Chengquan Huang, Yifan Song, Mo Wang, Young Ho Song, and Ruodan Shao. 2024. Using augmentation-based AI tool at work: A daily investigation of learning-based benefit and challenge. *Journal of Management* (2024), 01492063241266503.

[77] Mrinank Sharma, Meg Tong, Tomasz Korbak, David Duvenaud, Amanda Askell, Samuel R. Bowman, Newton Cheng, Esin Durmus, Zac Hatfield-Dodds, Scott R. Johnston, Shauna Kravec, Timothy Maxwell, Sam McCandlish, Kamal Ndousse, Oliver Rausch, Nicholas Schiefer, Da Yan, Miranda Zhang, and Ethan Perez. 2023. Towards Understanding Sycophancy in Language Models. https://doi.org/10.48550/arXiv.2310.13548 arXiv:2310.13548 [cs].

[78] Yonadav Shavit, Cullen O'Keefe, Tyna Eloundou, Paul McMillan, Sandhini Agarwal, Miles Brundage, Steven Adler, Rosie Campbell, Teddy Lee, Pamela Mishkin, Alan Hickey, Katarina Slama, Lama Ahmad, Alex Beutel, Alexandre Passos, and David G Robinson. [n. d.]. Practices for Governing Agentic AI Systems. ([n. d.]).

[79] Yang Shi, Tian Gao, Xiaohan Jiao, and Nan Cao. 2023. Understanding Design Collaboration Between Designers and Artificial Intelligence: A Systematic Literature Review. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW2 (Oct. 2023), 368:1–368:35. https://doi.org/10.1145/3610217

[80] Timo Speith. 2022. A review of taxonomies of explainable artificial intelligence (XAI) methods. In *Proceedings of the 2022 ACM conference on fairness, accountability, and transparency*. 2239–2250.

[81] Cella M Sum and Alicia DeVrio. 2024. Shifting Power Through Resistance: Social Justice in Practice. , 7–9 pages.

[82] Cella M Sum, Caroline Shi, and Sarah E Fox. 2024. " It's Always a Losing Game": How Workers Understand and Resist Surveillance Technologies on the Job. *arXiv preprint arXiv:2412.06945* (2024).

[83] Harini Suresh, Emily Tseng, Meg Young, Mary Gray, Emma Pierson, and Karen Levy. 2024. Participation in the age of foundation models. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24)*. Association for Computing Machinery, New York, NY, USA, 1609–1621. https://doi.org/10.1145/3630106.3658992

[84] Ningjing Tang, Jiayin Zhi, Tzu-Sheng Kuo, Calla Kainaroi, Jeremy J. Northup, Kenneth Holstein, Haiyi Zhu, Hoda Heidari, and Hong Shen. 2024. AI Failure Cards: Understanding and Supporting Grassroots Efforts to Mitigate AI Failures in Homeless Services. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24)*. Association for Computing Machinery, New York, NY, USA, 713–732. https://doi.org/10.1145/3630106.3658935

[85] Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, Dan Bikel, Lukas Blecher, Cristian Canton Ferrer, Moya Chen, Guillem Cucurull, David Esiobu, Jude Fernandes, Jeremy Fu, Wenyin Fu, Brian Fuller, Cynthia Gao, Vedanuj Goswami, Naman Goyal, Anthony Hartshorn, Saghar Hosseini, Rui Hou, Hakan Inan, Marcin Kardas, Viktor Kerkez, Madian Khabsa, Isabel Kloumann, Artem Korenev, Punit Singh Koura, Marie-Anne Lachaux, Thibaut Lavril, Jenya Lee, Diana Liskovich, Yinghai Lu, Yuning Mao, Xavier Martinet, Todor Mihaylov, Pushkar Mishra, Igor Molybog, Yixin Nie, Andrew Poulton, Jeremy Reizenstein, Rashi Rungta, Kalyan Saladi, Alan Schelten, Ruan Silva, Eric Michael Smith, Ranjan Subramanian, Xiaoqing Ellen Tan, Binh Tang, Ross Taylor, Adina Williams, Jian Xiang Kuan, Puxin Xu, Zheng Yan, Iliyan Zarov, Yuchen Zhang, Angela Fan, Melanie Kambadur, Sharan Narang, Aurelien Rodriguez, Robert Stojnic, Sergey Edunov, and Thomas Scialom. 2023. Llama 2: Open Foundation and Fine-Tuned Chat Models. https://doi.org/10.48550/arXiv.2307.09288 arXiv:2307.09288 [cs].

[86] Johanne R. Trippas, Sara Fahad Dawood Al Lawati, Joel Mackenzie, and Luke Gallagher. 2024. What do Users Really Ask Large Language Models? An Initial Log Analysis of Google Bard Interactions in the Wild. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, Washington DC USA, 2703–2707. https://doi.org/10.1145/3626772.3657914

[87] Himanshu Verma, Jakub Mlynar, Roger Schaer, Julien Reichenbach, Mario Jreige, John Prior, Florian Evéquoz, and Adrien Depeursinge. 2023. Rethinking the Role of AI with Physicians in Oncology: Revealing Perspectives from Clinical and Research Workflows. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–19. https://doi.org/10.1145/3544548.3581506

[88] Pranshu Verma. 2023. ChatGPT provided better customer service than his staff. He fired them. *Washington Post* (Oct. 2023). https://www.washingtonpost.com/technology/2023/10/03/ai-customer-service-jobs/

[89] Florent Vinchon, Todd Lubart, Sabrina Bartolotta, Valentin Gironnay, Marion Botella, Samira Bourgeois-Bougrine, Jean-Marie Burkhardt, Nathalie Bonnardel, Giovanni Emanuele Corazza, Vlad Glăveanu, Michael Hanchett Hanson, Zorana Ivcevic, Maciej Karwowski, James C. Kaufman, Takeshi Okada, Roni Reiter-Palmon, and Andrea Gaggioli. 2023. Artificial Intelligence & Creativity: A Manifesto for Collaboration. *The Journal of Creative Behavior* 57, 4 (2023), 472–484. https://doi.org/10.1002/jocb.597 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jocb.597.

[90] Torben Volkmann, Markus Dresel, and Nicole Jochems. 2023. Balancing Power Relations in Participatory Design: The Importance of Initiative and External Factors. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, Article 50, 6 pages. https://doi.org/10.1145/3544549.3585864

[91] Dakuo Wang, Elizabeth Churchill, Pattie Maes, Xiangmin Fan, Ben Shneiderman, Yuanchun Shi, and Qianying Wang. 2020. From human-human collaboration to Human-AI collaboration: Designing AI systems that can work together with people. In *Extended abstracts of the 2020 CHI conference on human factors in computing systems*. 1–6.

[92] Dakuo Wang, Justin D Weisz, Michael Muller, Parikshit Ram, Werner Geyer, Casey Dugan, Yla Tausczik, Horst Samulowitz, and Alexander Gray. 2019. Human-AI collaboration in data science: Exploring data scientists' perceptions of automated AI. *Proceedings of the ACM on human-computer interaction* 3, CSCW (2019), 1–24.

[93] Hanchen Wang, Tianfan Fu, Yuanqi Du, Wenhao Gao, Kexin Huang, Ziming Liu, Payal Chandak, Shengchao Liu, Peter Van Katwyk, Andreea Deac, et al. 2023. Scientific discovery in the age of artificial intelligence. *Nature* 620, 7972 (2023), 47–60.

[94] Ruotong Wang, Ruijia Cheng, Denae Ford, and Thomas Zimmermann. 2024. Investigating and Designing for Trust in AI-powered Code Generation Tools. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24)*. Association for Computing Machinery, New York, NY, USA, 1475–1493. https://doi.org/10.1145/3630106.3658984

[95] Xinghua Wang, Qian Liu, Hui Pang, Seng Chee Tan, Jun Lei, Matthew P Wallace, and Linlin Li. 2023. What matters in AI-supported learning: A study of human-AI interactions in language learning using cluster analysis and epistemic network analysis. *Computers & Education* 194 (2023), 104703.

[96] Justin D. Weisz, Jessica He, Michael Muller, Gabriela Hoefer, Rachel Miles, and Werner Geyer. 2024. Design Principles for Generative AI Applications. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–22. https://doi.org/10.1145/3613904.3642466

[97] Charles Rollet Whittaker, Zack. 2025. Exclusive: Elon Musk staffer created a DOGE AI assistant for making government "less dumb". https://techcrunch.com/2025/02/18/elon-musk-staffer-created-a-doge-ai-assistant-for-making-government-less-dumb/

[98] Yue You, Yubo Kou, Xianghua(Sharon) Ding, and Xinning Gui. 2021. The Medical Authority of AI: A Study of AI-enabled Consumer-Facing Health Technology. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–16. https://doi.org/10.1145/3411764.3445657

[99] Meg Young, Upol Ehsan, Ranjit Singh, Emnet Tafesse, Michele Gilman, Christina Harrington, and Jacob Metcalf. 2024. Participation versus scale: Tensions in the practical demands on participatory AI. *First Monday* (2024).

[100] Angie Zhang, Olympia Walker, Kaci Nguyen, Jiajun Dai, Anqing Chen, and Min Kyung Lee. 2023. Deliberating with AI: Improving Decision-Making for the Future through Participatory AI Design and Stakeholder Deliberation. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW1, Article 125 (April 2023), 32 pages. https://doi.org/10.1145/3579601

[101] Shuai Zhou, Ni Yi, Rajah Rasiah, Haipeng Zhao, and Zile Mo. 2024. An empirical study on the dark side of service employees' AI awareness: behavioral responses, emotional mechanisms, and mitigating factors. *Journal of Retailing and Consumer Services* 79 (2024), 103869.

[102] Marta Ziosi and Dasha Pruss. 2024. Evidence of What, for Whom? The Socially Contested Role of Algorithmic Bias in a Predictive Policing Tool. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24)*. Association for Computing Machinery, New York, NY, USA, 1596–1608. https://doi.org/10.1145/3630106.3658991

[103] Douglas Zytko, Pamela J. Wisniewski, Shion Guha, Eric P. S. Baumer, and Min Kyung Lee. 2022. Participatory Design of AI Systems: Opportunities and Challenges Across Diverse Users, Relationships, and Application Domains. 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 154, 4 pages. https://doi.org/10.1145/3491101.3516506

[104] Douglas Zytko, Pamela J. Wisniewski, Shion Guha, Eric PS Baumer, and Min Kyung Lee. 2022. Participatory design of AI systems: opportunities and challenges across diverse users, relationships, and application domains. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. 1–4.

Received 20 February 2025