Hauke Sandhaus

Research Statement

Ph.D. candidate in Information Science, Cornell Tech
sandhaus@duck.com · hauke.haus
Generic version, November 2026

I study whether people are still deciding when AI acts for them, and whether the designers who build with AI are still deciding the design. My research investigates how to design with and for AI interaction that supports the autonomy of users and workers, ethically and productively.

"How to ethically and productively design *with* AI" means studying how to use AI conscientiously to design interactive systems, so that the designer shapes the system in ways that follow from value-based deliberation rather than whatever the tool makes easy. "How to design AI" means studying the concrete design decisions that go into building data-driven interactive systems and their interfaces, by understanding both the explicit handcrafting that goes into them and the automated optimization processes that shape them.

All my work is motivated by the question of how to develop a society that nourishes AI's capabilities while not just respecting but also supporting human autonomy in a deeply relational world. Autonomy underlies my care for privacy [1, 2] and my push to preserve design intent when designing with AI [3, 4]. My framework for designing AI for human autonomy, under submission at CHI 2027 [5], responds to the long-standing negative impact of technologies on autonomy (dark patterns, which I have measured with users [6] and with designers [7]) and to new challenges, such as the sycophancy that reinforcement learning from human feedback produces [8]. It is equally motivated by the opportunities AI opens: helping people become more autonomous through bright patterns that prompt reflection [9], and enabling new capacities. And it untangles the complex context that distinguishes a design that undermines someone's autonomy from one that appropriately limits it.

The same question applies to the people who build these systems. Lessig argued that code is law: architecture regulates behavior most thoroughly because it is the least visible of the forces that regulate us, and the individual at the center of law, markets, norms, and architecture is a "pathetic dot," moved by forces it may take for its own choice [10]. Designers today are moved by the metrics they are handed, the deadlines they face, the norms of their teams, and the AI tools that now sit in every step of their work [4, 11]. Through research and teaching, I want to produce visionary, conscientious designers rather than pathetic ones shaped by external forces.

I have published 8 full research papers and 11 workshop papers and extended abstracts (13 as first or co-first author) at CHI, CSCW, FAccT, DIS, AutomotiveUI, UIST, IEEE ITSC, and Zeitschrift für Psychologie, and have three further first-author papers under review at CHI 2027 [4, 5, 20]. My work has received a Best Paper Honorable Mention at CHI 2024 [15] and the Best Late-Breaking Work Award at CHIWORK 2026 [7], and has been supported by a Digital Life Initiative Doctoral Fellowship, a Siegel Public Interest Technology Impact Fellowship, and a Public Interest Responsible Technology Fund grant. It also reaches beyond academia: brightpatterns.org, UEEQ.org, and a public interactive autonomy rubric put my instruments in practitioners' hands.

A. Foundation: Autonomy by Design

Should this feature ask the user, persuade the user, restrict the user, or simply act? Teams building AI answer that question weekly and settle it by instinct or by what the law requires. In *Autonomy by Design* [5], I took a first set of guidelines from the literature through twelve recorded versions in 24 co-design interviews with 8 philosophers of autonomy, 13 AI designers from 13 products, and 3 legal scholars. The framework evaluates one feature on three criteria (authentic values, the value–action link, and the user as informed decider) at four levels: undermining, appropriately limiting, respecting, and supporting. It unifies major tensions between philosophical traditions in a pluralistic framework. These distinctions become more urgent as models increasingly purport to interpret users' "intent" without distinguishing needs from momentary wants. Practitioners building AI systems for mental health, creativity support, hiring, and work automation used the framework to score features, understand their implications, and change them. The

framework is published as an interactive rubric at human-ai-autonomy-rubric.surge.sh, with participant evidence for every co-design decision.

B. End-User Autonomy: Resisting Manipulation and Protecting Information

My first thread gives users, designers, and regulators instruments for recognizing manipulation, and gives designers an antonym to design toward. I defined bright patterns (persuasive designs that serve users' goals and wellbeing over their momentary desires and business objectives) and started brightpatterns.org [9]. With my mentee Doris Rhomberg, I extended the User Experience Questionnaire with autonomy-focused items to distinguish appropriate from indecent persuasion [13]. In *Deception Is Ugly* [6], 126 social media users rated 15 dark patterns: perceived deception co-occurred with unattractiveness at $r = .93$, while the standard UEQ-S flagged one of the 15 patterns as problematic; our autonomy items flagged six as deceptive and eight as pressuring — indicating that user experience metrics need to extend beyond aesthetics and pragmatic qualities to capture manipulation.

The choice of instrument also changes what designers decide. In *Making Indecent Persuasion Visible* [7, 20], 141 UX professionals who saw our persuasion-focused data rejected manipulative interfaces at nearly twice the baseline rate (56.2% vs. 30.0%), and their 1,313 written justifications shifted from business and aesthetic framings to manipulation, trust, and autonomy. Importantly, their confidence stayed equally high without data: designers were not cautious; they were confidently uncalibrated.

My privacy research is motivated by autonomy over information, which I study through the lens of contextual integrity [16]. In *Privacy of Groups in Dense Street Imagery* [1], we ran a penetration test on 25 million dashcam images from New York City and showed that blurring faces and license plates does not prevent inference of sensitive group affiliations; we built a typology of identifiable groups and recommendations for researchers using such data. In *Regaining Trust* [2], a transparent walkthrough increased trust and reduced perceived creepiness of camera-based in-car health monitoring without lengthening onboarding. Under a standing IRB protocol with Wendy Ju, I am currently conducting field studies on privacy notices for autonomous robots in New York City, comparing privacy expectations across demographics and jurisdictions.

C. Designer Autonomy: From Pathetic to Conscientious Designers

My second thread studies the people who build technology while AI takes part in their work. In *"I Think We Did This"* [4], I interviewed eighteen interaction design students with their repositories and chat logs open: the artifact walk-through, because the record can contradict the self-report where AI's influence would otherwise go undetected. Design intent eroded along two paths: shared accounts and AI-merged work left no record of who decided what, and the tool offered directions so readily that they became the design. Students pushed back through design-intent cosmetics (redrawing a generated diagram by hand, restoring the feeling of authorship without genuine design intent) and through design-intent recapture, in which they repaired, steered, or guarded the idea. We observe this not as a short-term phenomenon but as a longitudinal trend across four course offerings, in which grades increasingly top out as quality proxies no longer discriminate once AI fills them. Across 203 repositories and 23,065 commits from three cohorts, AI-use disclosure rose from 0% to 66% while attribution of what the tool contributed stayed rare [11]: disclosure norms built for episodic acts cannot capture ambient co-creation. While AI cannot yet do meaningful design, with AI it becomes harder to see what design has meaning.

The forces on designers also live in infrastructure. In *My Precious Crash Data* [14], twelve employees of autonomous-vehicle companies revealed two previously unknown barriers to sharing safety-critical data: datasets embed costly safety knowledge, so sharing is political even within a company, and safety knowledge is treated as a private competitive advantage rather than a public good. To initiate industry–academia discussion around these pressures, I co-organized *Bridge Over Troubled Water* at CHI 2026 on aligning commercial incentives with ethical design [17], *Interrogating GenAI Augmentation* at CHIWORK 2026 toward a professional code of conduct for AI-augmented work [18], and *"Nobody Did This"* at CSCW 2026 on contribution and accountability in agent-mediated collaboration [19].

Together, my work develops an account of human autonomy as a value that must increasingly be designed for actively rather than merely respected, as AI and its incentive structures mediate our digital and physical lives.

D. Future Directions

1. Autonomy audits for AI products. Whether a product changes after a team has applied our framework is an open question I plan to answer through longitudinal field studies with product teams. The framework's structure also invites AI-assisted pre-screening of features against the rubric, which I want to develop and validate against expert judgments. Working with additional legal scholars, I want to map the framework onto already-binding regulations to identify the minimum level of autonomy that the law currently requires. I also plan to return to the driver–vehicle handover with the framework in hand, examining whether automation-level transitions meet the framework's criteria for appropriately limiting, rather than undermining, driver control.

2. Fine-tuned nudging: targeted manipulation. With Helen Nissenbaum, I have begun studying the minimum signals an AI system needs to send to change what a person does. Frontier AI companies present themselves as benevolent societal actors; we investigate how far a commercially or politically motivated adversary could get with personal data and customized open models, as dark-pattern tactics combined with AI personalization can be targeted to individual vulnerabilities [21].

3. How AI-assisted design practices harm vulnerable populations. My work shows that designers can unwittingly lose control over their design when using AI. An open question is whether these practices translate to real harms downstream. Designers increasingly replace real user research with AI-generated personas and synthetic user data. These proxies encode majority norms and demographic assumptions, potentially erasing the people least represented in training data (users with non-standard abilities, lower digital literacy, or marginal social positions) before any product decision is made. I want to study whether this is empirically the case, and whether AI-assisted design can alternatively be structured to surface vulnerable users' needs more reliably than individual designers alone.

4. Infrastructure for conscientious design. I will carry lessons from our workshops with UX practitioners and HCI researchers into developing a code of conduct for AI use in design: not just disclosure that AI was used, but specification of how it was used and what steps ensured it did not shape design thinking in unintended ways. I will also continue building ethical user experience metrics, including a validated persuasion scale as part of a modular UEQ extension. A parallel infrastructure thread extends the AV safety-data work [14] toward shared data infrastructure for the industry: the organizational and competitive barriers identified in that paper point to the need for neutral intermediaries and governance structures, a conversation I began in *Bridge Over Troubled Water* [17].

5. Values in vibe coding. A project nearing completion studies what value defaults vibe coding tools encode. Leaked system prompts reveal that some tools instruct the model to prioritize accessible and ethical design, while others push for engagement-maximizing or visually showy output [23]. Beyond explicit prompts, many defaults likely emerge from training data: given that much of the commercial web employs deceptive design, LLMs may have absorbed these patterns [12]. We are investigating whether this is empirically the case and what it means for design practice.

References

1. Matt Franchi*, **Hauke Sandhaus***, Madiha Zahrah Choksi, Severin Engelmann, Wendy Ju, and Helen Nissenbaum. 2025. Privacy of Groups in Dense Street Imagery. In FAccT '25, 2874–2891. https://doi.org/10.1145/3715275.3732185 (*equal contribution)
2. **Hauke Sandhaus**, Madiha Zahrah Choksi, and Wendy Ju. 2024. Regaining Trust: Impact of Transparent User Interface Design on Acceptance of Camera-Based In-Car Health Monitoring Systems. In Adjunct Proceedings of AutomotiveUI '24, 203–208. https://doi.org/10.1145/3641308.3685048

3. **Hauke Sandhaus**, Qiuquan Gu, Maria Teresa Parreira, and Wendy Ju. 2025. Co-Designing with Transformers: Unpacking the Complex Role of GenAI in Interactive System Design Education. In DIS '25, 1228–1243. https://doi.org/10.1145/3715336.3735805
4. **Hauke Sandhaus** and Wendy Ju. "I Think We Did This": Erosion and Reassertion of Students' Design Intention in AI-Saturated Work. Under review at CHI 2027.
5. **Hauke Sandhaus**, Jen Semler, Helen Nissenbaum, and Qian Yang. Autonomy by Design: Co-Designing a Framework to Evaluate Whether AI Interfaces Support User Autonomy. Under review at CHI 2027.
6. **Hauke Sandhaus** and Doris Maria Rhomberg. 2026. Deception Is Ugly: Linking Aesthetic Judgment to Perceived Manipulation in Dark Patterns. Zeitschrift für Psychologie. https://doi.org/10.1027/2151-2604/a000635
7. **Hauke Sandhaus**, Doris Maria Rhomberg, and Helen Nissenbaum. 2026. Making Indecent Persuasion Visible: How Evaluation Metrics Shape UX Designers' Ethical Reasoning. In CHIWORK '26 Late-Breaking Work. https://doi.org/10.1145/3805029.3818279 Best Late-Breaking Work Award.
8. Mrinank Sharma, Meg Tong, Tomasz Korbak, et al. 2024. Towards Understanding Sycophancy in Language Models. ICLR 2024.
9. **Hauke Sandhaus**. 2023. Promoting Bright Patterns. In CHI '23 Workshop: Designing Technology and Policy Simultaneously. arXiv:2304.01157
10. Lawrence Lessig. 2006. Code: Version 2.0. Basic Books.
11. Maria Teresa Parreira, Pranav Prabhat Sinha, **Hauke Sandhaus**, and Wendy Ju. 2026. Visible Adoption, Untracked Contribution: GitHub Evidence of the Accountability Gap Across Three Cohorts of an HCI Prototyping Course. In CHIWORK '26 Adjunct Proceedings. arXiv:2606.14054
12. Veronika Krauß, Mark McGill, Thomas Kosch, Yolanda Thiel, Dominik Schön, and Jan Gugenheimer. 2025. "Create a Fear of Missing Out": ChatGPT Implements Unsolicited Deceptive Designs in Generated Websites Without Warning. In CHI '25. https://doi.org/10.1145/3706598.3713083
13. Doris Maria Rhomberg* and **Hauke Sandhaus***. 2024. Towards Quantifying Ethical User Experience: Evaluating User Perceptions of Dark Patterns in Social Media. In CHI '24 Workshop: Mobilizing Research and Regulatory Action on Dark Patterns and Deceptive Design Practices. CEUR-WS Vol. 3720. (*equal contribution)
14. **Hauke Sandhaus**, Angel Hsing-Chi Hwang, Wendy Ju, and Qian Yang. 2025. My Precious Crash Data: Barriers and Opportunities in Encouraging Autonomous Driving Companies to Share Safety-Critical Data. Proceedings of the ACM on Human-Computer Interaction 9, 7 (CSCW), Article 348. https://doi.org/10.1145/3757493
15. Emma Harvey, **Hauke Sandhaus**, Abigail Z. Jacobs, Emanuel Moss, and Mona Sloane. 2024. The Cadaver in the Machine: The Social Practices of Measurement and Validation in Motion Capture Technology. In CHI '24, Article 768. https://doi.org/10.1145/3613904.3642004 Best Paper Honorable Mention.
16. Helen Nissenbaum. 2010. Privacy in Context: Technology, Policy, and the Integrity of Social Life. Stanford University Press.
17. Thomas Eßmeyer, Johanna Gunawan, Colin M. Gray, Lorena Sánchez Chamorro, Alberto Monge Roffarello, Katie Seaborn, **Hauke Sandhaus**, Kai Lukoff, Gian-Luca Savino, Shruthi Sai Chivukula, and Satoshi Nakamura. 2026. Bridge Over Troubled Water: Aligning Commercial Incentives With Ethical Design Practice To Combat Deceptive Patterns. In Extended Abstracts of CHI '26. https://doi.org/10.1145/3772363.3778758
18. **Hauke Sandhaus**, Kashif Imteyaz, Mohammed Almutairi, Pooja Prajod, Divya Ramesh, Saiph Savage, Qian Yang, and Michael Muller. 2026. Interrogating GenAI Augmentation for CHIworkers: Strategies for Professional Autonomy and Accountability. In CHIWORK '26 (Workshop). https://doi.org/10.1145/3805029.3818271
19. Kashif Imteyaz, Mohammad Rashidujjaman Rifat, Divya Ramesh, Steven R. Rick, Simo Hosio, **Hauke Sandhaus**, Advait Sarkar, Christoph Riedl, and Saiph Savage. 2026. "Nobody Did This":

Contribution, Originality, and Accountability in Agent-Mediated Collaboration. In Companion Publication of CSCW ‘26. arXiv:2607.26387

20. **Hauke Sandhaus**, Doris Maria Rhomberg, and Helen Nissenbaum. Persuading Metrics: How User Evaluation Data Shapes Professional Designers’ Willingness to Reject Dark Patterns. Under review at CHI 2027.
21. Thomas Kosch, Veronika Krauß, Christopher Katins, Dominik Schön, Mark McGill, and Jan Gugenheimer. 2026. The AI Accomplice: Exploring Generative Artificial Intelligence in Facilitating and Amplifying Deceptive Designs. In Extended Abstracts of CHI ‘26, Article 971. https://doi.org/10.1145/3772363.3778770
22. Martin Schrepp. 2021. Measuring User Experience with Modular Questionnaires. In International Conference on Advanced Computer Science and Information Systems (ICACSIS). https://doi.org/10.1109/ICACSIS53237.2021.9631321
23. asgeirtj. 2026. System prompts leaks. GitHub repository. https://github.com/asgeirtj/system_prompts_leaks (accessed February 2026)
24. Cornell Tech Vibe Coding Summer 2026. Ethical Vibe Coding for High School Students. https://cornell-tech-vibe-coding-summer-2026.github.io/website/ (accessed October 2026)