

Counter-Documenting Surveillance-Relevant Deceptive Designs: A Work-in-Progress Between Usable Privacy and Surveillance Studies

Helen Bader

*Science and Technology for Peace and Security (PEASEC),
Technical University of Darmstadt
Darmstadt, Germany
bader@peasec.tu-darmstadt.de*

Abstract—This extended abstract presents work in progress at the intersection of usable privacy and surveillance studies. It introduces *surveillance-relevant deceptive designs* as a sensitizing concept for interface patterns that steer users while depending on, obscuring, or intensifying data extraction, profiling, prediction, or behavioral intervention. Drawing on Marwick’s critique of “privacy without power” and Gray’s Dark Patterns Knowledge Stack, the concept shifts attention from identifying deceptive interfaces to asking how visible interface encounters point to platform incentives, data infrastructures, refusal burdens, and collective harms. Conceptually, the contribution is to frame deceptive designs as empirical entry points into surveillance power. Methodologically, the abstract proposes citizen-science counter-documentation as a way to support awareness, build a long-term evidence base, and enable subsequent surveillance-sensitive layered analysis.

I. FROM DECEPTIVE DESIGNS TO DATA POWER

Digital systems increasingly combine data collection, profiling, personalization, and manipulative interface design. These systems do not only steer individual choices. They also create conditions for tracking, prediction, targeting, and future behavioral intervention. This creates a tension for usable privacy research. Many privacy interfaces are organized around individual choice and user control. Surveillance studies, however, shows that privacy problems are often structural. They involve power, inequality, marginalization, and collective harms. Marwick’s critique of “privacy without power” is central here because it shows how privacy research can obscure systemic surveillance when it focuses too narrowly on individual responsibility [7].

Prior work has already examined manipulative interfaces, privacy “dark patterns”¹, large-scale prevalence, and platform business models [1], [8]. More recent work also calls for socio-technical and layered approaches to deceptive design [5], [10]. Yet fewer studies directly connect specific deceptive interface patterns to observable data-extraction practices, such as tracking, profiling, third-party data flows, or behavioral prediction.

¹I use the term “dark pattern” only when referring to established terms or publication titles to avoid reproducing associations of darkness with harm.

This extended abstract responds to that gap. It asks how deceptive designs can be studied not only as manipulative interface encounters, but also as empirical access to data power. A deceptive design becomes surveillance-relevant when it helps produce, maintain, intensify, or obscure data-extractive relations. The term data-extractive relations describes how interface actions become linked to tracking, profiling, user modeling, ranking, targeting, hypernudging, or behavioral intervention. This shifts the analytic focus from asking whether a user was manipulated into a specific choice to asking what data relation is created, authorized, or hidden through that choice.

II. SURVEILLANCE-RELEVANT DECEPTIVE DESIGNS

The term *surveillance-relevant deceptive designs* is used as a sensitizing concept, not as a new taxonomy. It helps identify cases in which deceptive design and surveillance power intersect. This distinction matters because privacy dark patterns already provide an important foundation. Bösch et al. show how interface design can exploit cognitive tendencies and structure privacy choices against users’ interests [1]. A surveillance-sensitive perspective does not replace this concept, but extends it beyond consent and disclosure to the infrastructures of data collection, inference, and retention. It asks how behavior is shaped so that continuous observation, profiling, and later data use become stabilized.

The proposed concept is therefore broader than “privacy dark patterns” in one respect, but more cautious in another. It includes designs that may not initially appear to be privacy interfaces, such as recommendation systems, engagement prompts, personalization flows, retention interfaces, loyalty mechanisms, or commercial nudges, when they depend on data extraction or profiling. At the same time, it does not imply that all deceptive designs are surveillance-relevant. A deceptive design becomes surveillance-relevant when its persuasive force, optimization, or harm depends on data extraction, user modeling, profiling, behavioral prediction, or the production of future intervention capacity. For example, a cookie banner is directly privacy-related because it asks users to accept or reject tracking. A recommendation widget such as “you

might also like” is different. It may be presented as helpful, convenient, or caring, yet it can depend on clicks, purchases, dwell time, inferred preferences, or similarity clusters. The analytic value of the concept lies in this shift of evidence. It does not only classify an interface as deceptive. It asks what kind of data relation the interface helps create, hide, or stabilize. This makes it possible to connect visible design choices to less visible practices of tracking, profiling, ranking, and behavioral intervention. Socio-technical work supports this broader lens. Baroni and Pereira show that deceptive patterns are discussed across privacy, consent, ethics, consequences, and surveillance capitalism, but that many studies still examine isolated aspects of the phenomenon [10]. This supports an analytic lens that does not stop at the interface, but also considers infrastructures, data lifecycle harms, and power mediated through business models and platforms.

Gray’s work helps explain why this layered view is useful. Gray et al.’s ontology frames deceptive designs as practices that can extract profit, harvest data, and limit consumer choice across domains [4]. Gray’s Dark Patterns Knowledge Stack then connects interface-level evidence to user goals, organizational incentives, legal contexts, business metrics, and socio-technical infrastructures [5]. This matters because the surveillance relevance of a deceptive design is often not visible from the interface alone. It emerges when an interface encounter is read together with data practices, optimization goals, and infrastructural relations. A surveillance-sensitive lens further foregrounds collective harms and political implications beyond individual losses of autonomy and control. Deceptive designs based on data extraction, profiling, and hypernudging may undermine privacy, autonomy, and opinion formation [3]. They can also normalize data extraction as a condition of digital participation and matter because privacy protects autonomy, dignity, and fairness [3]. This builds on feminist critiques of digital consent as authorizing exploitative adtech ecosystems while shifting responsibility for surveillance onto users [2]. From this perspective, deceptive designs do not encounter users under neutral conditions. They operate through opacity, dependency, time pressure, affective cues, cognitive limits, and personalized vulnerabilities [4], [8]. To sum up, surveillance-relevant deceptive designs describe socio-technical arrangements that organize data extraction, behavioral steering, and asymmetrical visibility. The concept sharpens privacy dark patterns along scope, harm, mechanism, and level: from individual consent to infrastructural surveillance, from UI tricks to data practices, and from isolated decisions to collective and structural harms. This motivates collective counter-documentation as a bridge from visible interface encounters to layered analysis.

III. COUNTER-DOCUMENTATION AS WORK IN PROGRESS

This work-in-progress asks how deceptive design documentation can move beyond recording individual manipulative encounters toward producing collective evidence of surveillance power. The methodological starting point is a tech-assisted citizen-science approach for documenting deceptive designs in everyday browsing contexts. The proposed tool is a browser

extension. Users can report suspected deceptive designs by submitting screenshots, short descriptions, pattern labels, and contextual information. At this stage, the primary aim is low-threshold documentation and awareness-building. The tool should help users recognize and differentiate manipulative interfaces. At the same time, it must avoid becoming another source of unnecessary data extraction.

Citizen science is relevant here because it can support learning, participation, and data quality when participants are treated as contributors to knowledge production rather than merely as data sources [6]. This is important for counter-documentation as users should not only provide examples for researchers. Their reports should help create a shared database of interface encounters that are difficult to recognize, preserve, and analyze individually. Counter-documentation serves three purposes. First, it can raise awareness during the initial encounter by supporting users in recognizing and classifying deceptive designs. Second, it contributes to a long-term database of reports across websites, services, and contexts. Third, it creates empirical entry points into surveillance-sensitive analysis. A single screenshot does not prove surveillance harm. However, it can mark an interface encounter that deserves further analysis. Following Gray’s Dark Patterns Knowledge Stack, selected reports can be examined in relation to user goals, organizational incentives, business metrics, legal contexts, and socio-technical infrastructures [5]. Where appropriate, selected cases can also be linked to browser-observable indicators such as tracker presence, third-party requests, cookie behavior, or permission flows. This responds to a methodological limitation in existing work. Automated and semi-automated approaches have made important progress in detecting deceptive interface elements. Yet detection alone does not explain whether, when, or how a detected pattern enables data extraction. Regulatory-oriented work similarly emphasizes that useful evidence must go beyond pattern matching and support transparent links between interface observations and underlying system behavior [9].

Counter-documentation can therefore operate as a methodological bridge between HCI-style interface documentation and surveillance-sensitive analysis. It begins with visible interface encounters and then asks how these encounters relate to data extraction, profiling, platform interests, and collective privacy harms. The central open question is how counter-documentation can move from awareness-raising and database building toward collective evidence production without reproducing the individualizing logic it critiques.

ACKNOWLEDGMENT

This work was developed in the context of the AIDA project which is funded by the National Research Center for Applied Cybersecurity ATHENE. ATHENE is funded jointly by the German Federal Ministry of Research, Technology and Space and the Hessian Ministry of Science and Research, Arts and Culture. I thank Prof. Dr. Christian Reuter and Dr. Ephraim Zimmer for their feedback and discussions on this work.

REFERENCES

- [1] Christoph Bösch, Benjamin Erb, Frank Kargl, Henning Kopp and Stefan Pfattheicher (2016). Tales from the dark side: Privacy dark strategies and privacy dark patterns. In *Proceedings on Privacy Enhancing Technologies*, 237-254. <https://doi.org/10.1515/popets-2016-0038>
- [2] Elinor Carmi. 2021. A feminist critique to digital consent. *Seminar.net*, 17(2). <https://doi.org/10.7577/seminar.4291>
- [3] Ann Kristin Glenster. 2025. Surveillance Capitalism, Deceptive Designs, and Epistemic Rights. In *Carr-Ryan Center for Human Rights*. Harvard Kennedy School, Carr-Ryan Center Discussion Paper, Issue 10.
- [4] Colin M. Gray, Cristiana Teixeira Santos, Nataliia Bielova, and Thomas Mildner. 2024. An Ontology of Dark Patterns Knowledge: Foundations, Definitions, and a Pathway for Shared Knowledge-Building. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 289, 1–22. <https://doi.org/10.1145/3613904.3642436>
- [5] Colin M. Gray. 2026. The Dark Patterns Knowledge Stack: Exploring New Ways to Negotiate Context, Law, and Design. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26)*. Association for Computing Machinery, New York, NY, USA, Article 1283, 1–11. <https://doi.org/10.1145/3772318.3791264>
- [6] Christothea Herodotou, Eileen Scanlon and Mike Sharples. 2021. Methods of Promoting Learning and Data Quality in Citizen and Community Science. *Frontiers in Climate*, 3:614567. <https://doi.org/10.3389/fclim.2021.614567>
- [7] Alice Marwick. 2022. Privacy Without Power: What Privacy Research Can Learn from Surveillance Studies. In *Surveillance & Society*, 20(4), 397-405. <https://doi.org/10.24908/ss.v20i4.16009>
- [8] Arunesh Mathur, Gunes Acar, Michael J. Friedman, Eli Lucherini, Jonathan Mayer, Marshini Chetty, and Arvind Narayanan. 2019. Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites. In *Proceedings of the ACM on Human-Computer Interaction, CSCW 3*, 1-32. <https://doi.org/10.1145/3359183>
- [9] Arianna Rossi and Simon Parkin. 2026. “What I’m interested in is something that violates the law”: Regulatory practitioner views on automated detection of deceptive design patterns. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26)*. Association for Computing Machinery, New York, NY, USA, Article 42, 1–18. <https://doi.org/10.1145/3772318.3790832>
- [10] Luiz Adolpho Baroni and Roberto Pereira. 2024. Deceptive Patterns under a Sociotechnical View. In *Proceedings of the XXIII Brazilian Symposium on Human Factors in Computing Systems (IHC '24)*. Association for Computing Machinery, New York, NY, USA, Article 42, 1–13. <https://doi.org/10.1145/3702038.3702081>