Journal Publication of International Research for Engineering and Management (JOIREM)
Volume: 04 Issue: 4 | April-2026
ISSN (O) 3107-6696

Awareness and Empowerment: A Review of Deceptive Design and Interventions

Nimish Gupta, Vasudha

School Of Design, World University of Design

Abstract - This paper discusses the progress the User Experience(UX) design community has made in making users aware of dark patterns and what actions it has taken to remove them. It also looks at the legal side of the conversation to see how laws are implemented to tackle this problem. We explore how these patterns affect everyone from young children to adults, and why simple awareness or higher prices do not stop them from working. The review looks at the shift toward active empowerment, such as new technical tools that let users block deceptive designs themselves. It also looks at how mild tricks allow companies to avoid a negative reputation while still manipulating users. Finally, we discuss how internal company goals drive these designs and why we need a legal and a digital system where fairness is built into the design from the start to protect people from these manipulative tactics including the new, hyper-personalized artificial intelligent(AI) traps.

Keywords: Dark Patterns, User Awareness, User Empowerment, UX Design, AI Ethics, Digital Rights

Introduction

UX dark patterns are design tactics used to manipulate users into taking actions that benefit a company, often at a financial or privacy cost(Mathur et al., 2019). These tricks exploit human psychology to provoke harmful actions, such as signing up for services by mistake or sharing excessive data(Gray et al., 2018). While the term was coined by Harry Brignull in 2010 through personal observation, his findings captured the core reality of how these interfaces work and remain the foundation for academic study today(European Commission (2022).

The scale of this problem has moved from isolated tricks to a widespread industry standard (Dayananda (2025). Mathur et al. (2019) conducted a large-scale analysis of 11,000 shopping websites and discovered over 1,800 instances of dark patterns, proving their ubiquity across the digital economy. This growth has shifted academic interest from simply identifying tricks to creating complex systems for understanding the ethics and long-term impact of these designs.

While designers should aim for user-centered design, they often face pressure to prioritize company profit instead. Kallioniemi (2022) demonstrates that in large corporations, internal research into harmful design is often ignored because those same designs are responsible for keeping users engaged and generating revenue. This creates a growth at all costs culture where business goals are placed ahead of user well-being.

To standardize how we study these tactics, Gray et al. (2018) categorized dark patterns into five primary types. Nagging involves repeatedly provoking users to change their action, such as constant premium pop-ups. Obstruction is the act of intentionally creating friction for actions that hurt business, like making account deletion difficult. Sneaking refers to hiding or delaying information, such as burying costs in fine print. Interface interference uses UI decisions to manipulate information, like placing a giant accept button next to a tiny, hidden reject link. Finally, forced action requires users to surrender data, such as a weather app demanding contact lists to function.

Since this original research, these patterns have become common appearing in 95-97% of mobile games and

websites(European Commission, 2022). This review is critical because understanding these mechanisms is the first step toward identifying how they affect different people and determining which tools can actually stop them.

Literature Review

Recent research has moved past naming dark patterns to testing how they hurt specific groups. For children, studies show that those with low self-regulation are much more likely to get stuck in persuasive design(PD) traps(Mallawaarachchi et al., 2025). This happens because children are still learning how to control their impulses and stop doing something that feels rewarding. When an app or game uses PD, like showing constant progress bars, it creates a loop that is very hard to break.

The research found that kids who struggle to manage their own behavior find it almost impossible to walk away from these nudges. This leads to them spending much more time on screens than they intended and feeling frustrated when they are finally forced to stop (Mallawaarachchi et al., 2025, Monge et al., 2022) . This shows that PD is especially dangerous for younger users because it targets the fact that their brains aren't fully ready to say no to this digital trap.

In the world of casual mobile games, Dahlan and Susanty (2022) identified how these traps specifically target players looking for quick distractions. They categorized these into monetary patterns, which trick users into spending real money or virtual currency for pay-to-skip advantages, and social capital patterns. Social Capital tricks force players to invite friends or share game achievements on social media just to continue playing or unlock rewards. These exploit the user's social network to market the game for free.

In adults, Luguri and Strahilevitz proved that mild patterns can more than double sign-up rates. One major finding they found was that the price does not stop these tricks from working. Even when a service was made three times more expensive, people still fell for the patterns at the same rate. This is a big deal because it shows that dark patterns bypass our System 2, which is logical thinking, the part of our brain that slowly weighs pros and cons and instead targets our System 1 automatic reactions, which is our fast, autopilot brain (Kahneman, 2012). Because these tricks hit our autopilot, we don't even stop to think about the cost, we just follow the path the designer laid out for us. Furthermore, people with less schooling or lower digital literacy are much more vulnerable to these subtle tricks.

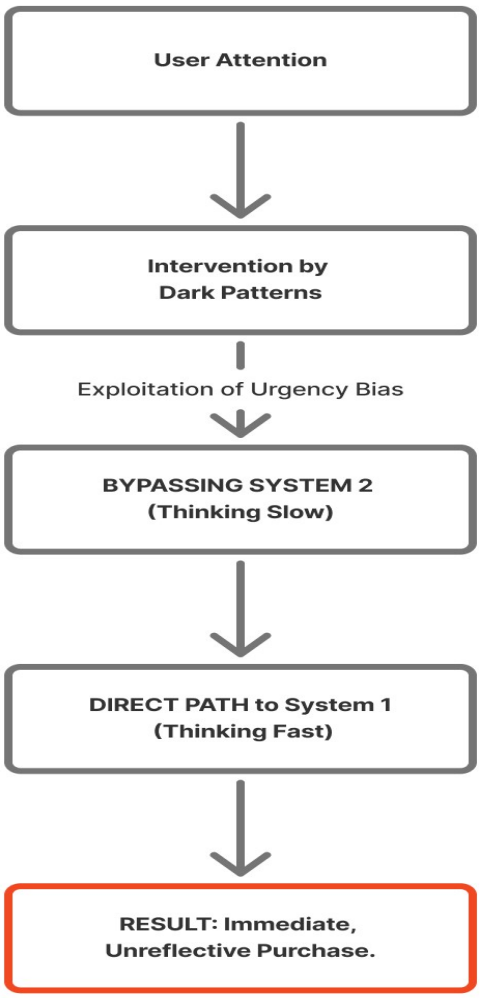

Fig: How dark patterns hack the logical thinking of a brain to go to quick thinking.

This connects to the backlash effect. The backlash effect is simply when a user realizes they are being tricked and gets angry enough to quit the service or delete the app. If a dark pattern is too aggressive or loud, the user fights back by leaving. To avoid this, companies use mild tricks. These are dangerous because they work just as well as aggressive ones but are subtle enough that the user never feels that anger or backlash. They stay signed up and keep paying without ever realizing they are being manipulated. This allows companies to keep their brand reputation while still being dishonest.

These tricks often cost users their privacy. Acquisti(2015) argues that making decisions about privacy isn't a logical process. Users are usually unsure about the future risks, and they are easily influenced by the design or context of a website. Therefore they could be pushed into sharing more personal data than they originally planned. This shows that dark patterns aren't just design mistakes, they are tools built to exploit the natural way the human brain works. These privacy risks often happen because of three main reasons. First is uncertainty. People do not really know what will happen to their data in five years, so they usually ignore the long-term risk and just focus on the reward they are getting at the moment, like a discount or access to an app. Second is the power of context. This means that users tend to share more information if a website looks friendly or professional, even if the site is actually dangerous. The design of a website could trick users into feeling safe. Finally, there is malleability, which means people's privacy preferences aren't set in stone. A small change in a button's color or a pre-checked box can completely change their decision without them even realizing it.

Weaponised Progress: AI

AI has transformed traditional design tricks into a specialized system of continuous manipulation(Deligöz, 2025). Unlike static designs that appear the same for everyone, AI utilizes a Learning Loop that observes how users interact with an interface(European Data Protection Board, 2024). It identifies patterns in user's mistakes and refines its deceptive tactics in real-time. For instance, these systems can predict the exact moment users intend to cancel a subscription and immediately deploy hurdles to prevent them from leaving. Users are often unable to keep up with the speed at which these digital traps are generated, because AI can modify a website in seconds.

One of the most significant limitations identified in recent research is what scholars call the Black Box effect(Gunawan et al., 2025)

A black box is a system where the inputs (the data users provide) and the outputs (a specific price or recommendation) are visible, but the internal logic remains hidden. Neither users nor legal regulators can see how the machine made its decision. This creates a state of Information Asymmetry, where a company understands a user's weaknesses perfectly, but the user has no way to see or challenge the system being used against them(Acquisti et al, 2015).

This technological power is used to target users when they are at their most vulnerable moment(Lechevalier et al., 2025). AI waits until it detects that users are fatigued, stressed, or in a rush before requesting intrusive data permissions. This undermines the user's ability to make autonomous decisions, allowing the machine to make the choice for them even before they had a chance to think. This is not merely a psychological issue. Experiments show that AI-driven forced actions cause real neurophysiology stress, increasing users' heart rates and anxiety levels during digital interactions(European Commission, 2022).

The industry has reached a point where the speed and sophistication of AI-driven manipulation have surpassed the biological limits of human attention. When users can no longer see the traps and the law cannot see the math behind them, the idea that users can protect themselves simply by paying attention is officially over (Bongard et al., 2021). This raises a critical question for the future of UX: what happens when users can no longer trust their own choices?

Comparative Analysis of different Approaches

Researchers proposed three main strategies to solve this problem:

1. The Education Approach: This method focused on teaching users how to identify dark patterns through digital literacy. The idea was that if a user is taught that a countdown timer is fake, they won't feel pressured to buy. However, research consistently shows that awareness does not equal protection.

Even when people are taught to recognize these tricks, they still fall for them because the psychological nudges are designed to trigger our System 1 autopilot brain, which is faster than our System 2 logical thinking(Bongard et al., 2021). This supports the finding that simply knowing that a trap exists isn't enough to avoid it. The emotional pressure is often too strong to ignore(Luguri, 2021; Mildner 2021). Essentially, the education approach fails because it puts the entire burden on the user to be perfectly rational at all times, which is psychologically impossible.

2. The Technical Approach: Since education and awareness often fail to protect users in real-time, research shifted towards Empowerment Interventions. Lu et al. (2024) explored this by moving past passive awareness, just knowing a trick exists, to active action. They developed a browser extension called Dark Pita, which gives users the power to manually modify or remove dark patterns directly from a website's interface.

Users in the study reported a strong sense of autonomy and power when they could actively block a sneaking or nagging element. However, this technical solution revealed three major gaps. Some users experienced a Fear of Missing Out (FOMO) when the tool hid information like limited-time offers. Even if the offer was a manipulative urgency trap, users still wanted to see it just in case. In addition, Users found it annoying to manually edit every dark pattern they encountered. There was a clear demand for the tool to just know what was deceptive and fix it automatically without user input. A critical limitation of this specific case study was the small sample size of only 15 users. This makes the results hard to generalize. We need future research with much larger, more diverse groups to see if these technical tools actually work for people with different levels of digital literacy.

While effective, these tools often require a lot of effort from the user and might not work on all platforms(Gunawan et al., 2025).

3. The Legal Approach: The burden shouldn't be on the user to defend themselves. Recent legal updates like the EU's Digital Services Act (DSA) and the 2024 AI Act have started to legally ban specific manipulative designs. This shift is moving the industry toward Fairness by Design, where the law forces companies to build honest interfaces from the beginning (Lechevalier et al., 2025). Instead of relying on users to spot tricks, this approach requires designers to use Fair Patterns that prioritize transparency and user choice by default.

Many now argue that the burden shouldn't be on the user at all. The EU's Digital Services Act (DSA) and the new AI Act (2024) have started to legally ban certain manipulative designs. This leads to the idea of Fairness by Design, where the law forces companies to be honest from the very start (Lechevalier et al., 2025; Gunawan et al., 2025).

Limitations

A major challenge is that dark patterns are part of a company's internal culture(Dayananda,2025;). Designers often use them to hit specific Key Performance Indicators (KPIs) even if they feel it is unethical. This creates a value tension where designers are stuck between their own ethics and the business goals of the company. Another issue is that as AI gets smarter, it can create personalized traps. These AI-driven patterns can target a person's specific weaknesses in real-time, making them much harder to detect than standard patterns(Kadir Deligoz, 2025; Lechevalier, 2025).

Beyond internal culture, research and enforcement are heavily concentrated in the European Union and the United States, leading to a geographic bias toward Western markets(Gray et al., 2023). Most studies fail to report specific demographics, meaning we do not yet fully understand how these patterns adapt to different local design norms. Furthermore, a significant regulatory gap exists because legislators often focus on the substance of the law while neglecting the actual design of the interface(Lechevalier, 2025). Even when laws mandate privacy, the designs remain so easy to accept that they bypass the intent of the regulation.

There are also major methodological and technical constraints. Most large-scale research focuses only on text-based patterns, ignoring visual manipulation like font size, colour, or button hierarchy. Studies also struggle with ecological validity, making it difficult to see how different patterns interact in real-world settings over time. Additionally, current research is limited to specific domains like shopping or social media, leaving many other digital task areas unexamined(Lu et al., 2024).

Finally, the lack of a shared language between academics and legislators limits the ability to enforce legal sanctions. There is a fundamental evidentiary gap because regulators rely on internal company records that are completely unavailable to researchers(Gunawan, 2025). Behavioral studies have also proven that current remedies like informational notices or cool-down periods are ineffective. Consumers have become so used to certain tactics, a phenomenon known as habituation, that they no longer perceive them as stressors, even as the designs continue to successfully manipulate their choices(European Commission, 2022).

Conclusion

This review shows that dark patterns have become a massive structural problem that user education alone cannot fix(European Commission, 2022). While early research named these tactics, the data now shows that simple awareness does not stop System 1 autopilot reactions from falling for deceptive designs(Luguri, 2021). The education gap and the success of mild tricks, which doubles conversion rates without causing the backlash seen in aggressive patterns. It proves that even when users feel manipulated, they often cannot resist(Luguri, 2021; Bongard et al., 2021)

The industry is currently shifting toward technical tools like Dark Pita and legal frameworks like the EU AI Act (2024)(Lu et al., 2024; Gunawan et al., 2025; Lechevalier, 2025). These are critical because they move the burden of protection away from the individual and mandate Fairness by Design. By forcing companies to prioritize transparency from the start, these laws aim to pierce the "black box" effect and fix the digital asymmetry where users are treated as targets rather than participants. Ultimately, protecting user autonomy requires a system where digital interfaces are legally required to be as honest as the physical products we use every day.

Future scope

The next phase of dark patterns research must move beyond just identifying tricks to address the learning loop of automated manipulation(European Commission, 2022; Lechevalier, 2025). Future studies should focus on algorithmic invisibility, investigating how real-time AI refinements can be made transparent to both users and regulators to break information asymmetry(Lechevalier, 2025; Gunawan et al., 2025). To solve the limit of ecological validity, researchers must conduct longitudinal field studies in real-world settings to see how these patterns affect diverse populations over several years, rather than just in lab experiments. Furthermore, to bridge the evidentiary gap, future work should focus on interdisciplinary collaboration between legal auditors and UX researchers(Gunawan et al., 2025). This collaboration is necessary to create shared languages and retro-design programs that allow regulators to access internal company records that are currently hidden from academic view. Finally, the community must work to standardize fair patterns, creating a universal repository of legally compliant design features that ensure all websites

follow a baseline level of honesty and user empowerment. This shift from identifying problems to creating problem-solving frameworks is essential for a fair digital ecosystem(Potel-Saville, 2023).

References

1. Brignull, H. (2010, July 8). *Dark patterns: Dirty tricks designers use to make people do stuff*. 90 Percent of Everything. https://90percentofeverything.com/2010/07/08/dark-patterns-dirty-tricks-designers-use-to-make-people-do-stuff/

2. Lewis, F. B., & Vassileva, J. (2024, May). Seeing in the Dark: Revealing the Relationships, Goals, and Harms of Dark Patterns. In *DDPCHI@ CHI*.

3. Acquisti, A., Brandimarte, L., & Loewenstein, G. (2015). Privacy and human behavior in the age of information. *Science*, *347*(6221), 509–514. https://doi.org/10.1126/science.aaa1465

4. Kallioniemi, P. (2022). Facebook's dark pattern design, public relations and internal work culture. *Journal of Digital Media & Interaction*, *5*(12).

5. Gray, C. M., Kou, Y., Battles, B., Hoggatt, J., & Toombs, A. L. (2018, April). The dark (patterns) side of UX design. In *Proceedings of the 2018 CHI conference on human factors in computing systems* (pp. 1-14).

6. Dahlan, R., & Susanty, M. (2022). Finding dark patterns in casual mobile games using heuristic evaluation. *PETIR: Jurnal Pengkajian dan Penerapan Teknik Informatika*, *15*(2), 185-195.

7. Zagal, J. P., Björk, S., & Lewis, C. (2013). Dark patterns in the design of games. In *Foundations of digital games 2013*.

8. Mallawaarachchi, S., Cliff, D. P., Neilsen-Hewett, C., White, S. L., Radesky, J., Horwood, S., ... & Howard, S. J. (2025). Effects of Persuasive App Design and Self-Regulation on Young Children's Digital Disengagement. *Human Behavior and Emerging Technologies*, *2025*(1), 8187768.

9. Monge Roffarello, A., & De Russis, L. (2022, April). Towards understanding the dark patterns that steal our attention. In *Chi conference on human factors in computing systems extended abstracts* (pp. 1-7).

10. Luguri, J., & Strahilevitz, L. J. (2021). Shining a light on dark patterns. *Journal of Legal Analysis*, *13*(1), 43-109.

11. Kahneman, D. (2012). Two systems in the mind. *Bulletin of the American Academy of Arts and Sciences*, *65*(2), 55-59.

12. Bongard-Blanchy, K., Rossi, A., Rivas, S., Doublet, S., Koenig, V., & Lenzini, G. (2021). "I am definitely manipulated, even when I am aware of it. It's ridiculous!": Dark patterns from the end-user perspective. *Proceedings of the 2021 ACM Designing Interactive Systems Conference (DIS '21)*, 763–776. https://doi.org/10.1145/3461778.3462086

13. Mildner, T., & Savino, G. L. (2021, May). Ethical user interfaces: Exploring the effects of dark patterns on facebook. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1-7).

14. Gunawan, J., Gray, C.M., Santos, C., & Bielova, N. (2025). Leveraging interdisciplinary methods for evidence collection in enforcement: Dark patterns as a case study. Internet Policy Review, 14(4). https://doi.org/10.14763/2025.4.2047

15. Lechevalier, F., & Saville, M. P. (2025, January). Fairness by design: Combatting deceptive AI-driven

interfaces. In *Cambridge Forum on AI: Law and Governance* (Vol. 1, p. e31). Cambridge University Press.

16. European Data Protection Board. (2024). *EDPB annual report 2023*. https://www.edpb.europa.eu/our-work-tools/our-documents/annual-report/edpb-annual-report-2023_en

17. Deligöz, K. (2025). Consumer manipulation with artificial intelligence: Dark patterns and hidden techniques. https://doi.org/10.58830/ozgur.pub710.c3029

18. Gray, C. M., Santos, C. T., Bielova, N., & Mildner, T. (2024). An ontology of dark patterns knowledge: Foundations, definitions, and a pathway for shared knowledge-building. *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*, Article 289, 1–22. https://doi.org/10.1145/3613904.3642436

19. Arunesh, M., Acar, G., Friedman, M. J., Lucherini, E., Mayer, J., Chetty, M., & Narayanan, A. (2019). Dark patterns at scale: Findings from a crawl of 11K shopping websites. *Proceedings of the ACM on Human-Computer Interaction, 3*(CSCW), Article 81. https://doi.org/10.1145/3359183

20. Dayananda, L. (2025). *Dark patterns in user experience design: The erosion of user autonomy and trust—Introduction: Defining and contextualizing dark patterns in UX design*. https://doi.org/10.5281/zenodo.15278819

21. Lu, Y., Zhang, C., Yang, Y., Yao, Y., & Li, T. J.-J. (2024). Dark patterns in UX. *From awareness to action: Exploring end-user empowerment interventions for dark patterns in UX. Proceedings of the ACM on Human-Computer Interaction, 8*(CSCW1), Article 59. https://doi.org/10.1145/3637336

22. European Commission. (2022). *Behavioural study on unfair commercial practices in the digital environment: Dark patterns and manipulative personalisation* (Final report). Publications Office of the European Union. https://doi.org/10.2838/859030

23. Potel-Saville, M., & Francois, M. (2023). *From dark patterns to fair patterns? Usable taxonomy to contribute solving the issue with countermeasures*.