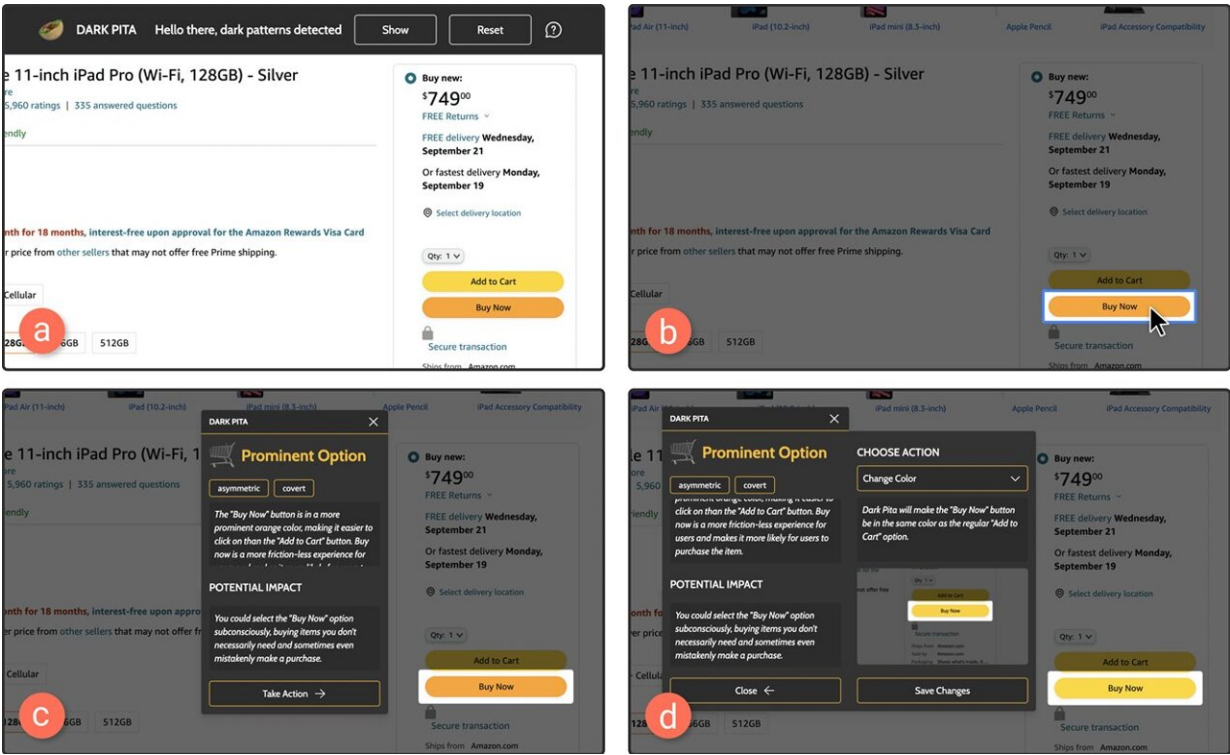


# AI browser plug-ins to help consumers improve digital privacy literacy, combat manipulative design

May 29 2024, by Brandi Wampler



The Dark Pita browser plug-in detects dark pattern designs, notifies the user and allows the user to customize their online experience. Credit: Toby Li / University of Notre Dame

Researchers at the University of Notre Dame are developing artificial

intelligence tools that help consumers understand how they are being exploited as they navigate online platforms. The goal is to boost the digital literacy of end users so they can better control how they interact with these websites.

[In a recent study](#)

[appearing in \*Proceedings of the CHI Conference on Human Factors in Computing Systems\*, participants were invited to experiment with online privacy settings without consequence. To test how different data privacy settings work, the researchers created a Chrome browser plug-in called Privacy Sandbox that replaced participant data with personas generated by GPT-4, a large language model from OpenAI.](#)

[With Privacy Sandbox, participants could interact with different websites, such as social media platforms or news outlets. As they navigated to various sites, the browser plug-in applied AI-generated data, making it more obvious for participants to see how they were targeted based on their supposed age, race, location, income, household size and more.](#)

["From a user perspective, allowing the platform's access to private data may be appealing because you could get better content out of it, but once you turn it on, you cannot get that data back. Once you do, the site already knows where you live," said Toby Li, assistant professor of computer science and engineering and a faculty affiliate at the Lucy Family Institute for Data & Society at Notre Dame, who led the research. "This is something that we wanted participants to understand, figure out whether the setting is worth it in a risk-free environment, and allow them to make informed decisions."](#)

[Another study](#), this one in *Proceedings of the ACM on Human-Computer Interaction*, looked at dark patterns—or the design features on digital

platforms that subtly nudge users to perform specific actions—and how they are used on websites to manipulate customers. For the study, Li and his team looked at how dark patterns are applied by interface designers to encourage people to consume more content or make impulsive purchasing decisions.

The researchers developed a Chrome browser plug-in dubbed Dark Pita to identify dark patterns on five popular online platforms: Amazon, YouTube, Netflix, Facebook and X.

Using machine learning, the plug-in would first notify study participants that a dark pattern was detected. It would then identify the threat susceptibility of the dark pattern and explain the impact of the dark pattern—financial loss, invasion of privacy or cognitive burden. Dark Pita would then give participants the option to "take action" by modifying the website code through an easy-to-use interface to change the deceptive design features of the site and explain the effect of the modification.

The researchers plan to eventually make both browser plug-ins, Privacy Sandbox and Dark Pita, available to the public. Li believes these tools are great examples of how the use of AI can be democratized for regular users to benefit society.

"Companies will increasingly use AI to their advantage, which will continue to widen the power gap between them and users. So with our research, we are exploring how we can give back power to the public by allowing them to use AI tools in their best interest against the existing oppressive algorithms. This 'fight fire with fire' approach should level the playing field a little bit," Li said.

"An Empathy-Based Sandbox Approach to Bridge the Privacy Gap Among Attitudes, Goals, Knowledge, and Behaviors" was presented at

the [2024 Association of Computing Machinery CHI Conference](#). Led by Li, fellow study co-authors include Chaoran Chen and Yanfang (Fanny) Ye from Notre Dame, Weijun Li from Zhejiang University, Wenxin Song at the Chinese University of Hong Kong and Yaxing Yao at Virginia Tech.

The study "From Awareness to Action: Exploring End-User Empowerment Interventions For Dark Patterns in UX," led by Li, has been published in the *Proceedings of the ACM on Human-Computer Interaction* (CSCW 2024). Co-authors include Yuwen Lu from Notre Dame, Chao Zhang from Cornell University, Yuewen Yang from Cornell Tech and Yao from Virginia Tech.

**More information:** Chaoran Chen et al, An Empathy-Based Sandbox Approach to Bridge the Privacy Gap among Attitudes, Goals, Knowledge, and Behaviors, *Proceedings of the CHI Conference on Human Factors in Computing Systems* (2024). [DOI: 10.1145/3613904.3642363](#)

Yuwen Lu et al, From Awareness to Action: Exploring End-User Empowerment Interventions for Dark Patterns in UX, *Proceedings of the ACM on Human-Computer Interaction* (2024). [DOI: 10.1145/3637336](#)

Provided by University of Notre Dame

Citation: AI browser plug-ins to help consumers improve digital privacy literacy, combat manipulative design (2024, May 29) retrieved 29 June 2024 from <https://techxplore.com/news/2024-05-ai-browser-ins-consumers-digital.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.