Clicks, Screens, and Control: Dark Patterns in Teen Privacy and Safety Settings on Social Media

Jingxin Dong
UXP2 Lab
Indiana University
Bloomington, Indiana, USA
dong11@iu.edu

Lingyun Chen
Luddy School of Informatics, Computing, and Engineering
Indiana University Bloomington
Bloomington, Indiana, USA
lch2@iu.edu

Chen Ling
Indiana University
Bloomington, Indiana, USA
ccling@iu.edu

Colin M. Gray*
UXP2 Lab
Indiana University
Bloomington, Indiana, USA
comgray@iu.edu

Abstract

Dark patterns in social media privacy and safety settings can quietly shape what teenagers see as possible when they try to protect themselves online. While prior work has documented manipulative design in individual features or single platforms, we know less about how privacy and safety controls compare across the services teens use every day. This paper presents an expert evaluation of seven privacy and safety management tasks on four social media platforms. For each task, we mapped task flows to complete each task, identifying the number of clicks and screens required. We then analyzed these flows using the dark patterns ontology to identify where setting discoverability was compromised or unnecessary friction was employed. All settings employed dark patterns on at least one platform, and some settings, such as account deletion and managing in-app notifications, were particularly onerous. We argue that safety settings often look like control, but the presence of dark patterns makes the settings difficult to use and easy to circumvent. Building on our findings, we outline implications for the evaluation and design of teen-focused privacy and safety features and identify opportunities for policy intervention.

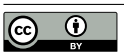
CCS Concepts

• **Human-centered computing** → **Empirical studies in HCI**; **Social media**; **HCI design and evaluation methods**; • **Security and privacy** → **Usability in security and privacy**.

Keywords

social media, dark patterns, privacy and safety settings, teenagers

*Colin has engaged in paid expert witness and consulting work on cases relating to social media. This expert witness work is independent from the data and methods presented in this paper.



CHI EA '26, Barcelona, Spain

ACM ISBN 979-8-4007-2281-3/26/04
https://doi.org/10.1145/3772363.3798556

ACM Reference Format:
Jingxin Dong, Lingyun Chen, Chen Ling, and Colin M. Gray. 2026. Clicks, Screens, and Control: Dark Patterns in Teen Privacy and Safety Settings on Social Media. In *Extended Abstracts of the 2026 CHI Conference on Human Factors in Computing Systems (CHI EA '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3772363.3798556

1 Introduction

Social media plays a major role in teenagers' emotional and social development [8]. Research on social media harms shows that teens' experiences are often mixed, with benefits and risks frequently co-occurring [3, 40]. For example, teens use messaging and online communities to express vulnerability and seek peer help [1], while also facing risks such as harassment, impersonation, and unsafe contact [2, 5]. Teens also demonstrate the potential for agency by managing visibility and boundaries through tactics such as selective audience controls and strategic hashtags [34, 39]. This makes privacy and safety settings a central site of struggle: whether teens can locate controls, enact protective actions, and interpret what changed afterward [3].

Importantly, these risks are not shaped by content alone, but how platforms structure user choice and control through interface design [28, 38]. Teens often report distress, time loss, and privacy concerns related to manipulative design features such as infinite scroll, fake security cues, and social comparison, with family and friends playing a critical role in helping them recognize these risks [13, 35]. One significant factor that shapes these experiences is the presence of *dark patterns*, sometimes known as *"deceptive patterns"* [15, 17]. These patterns exploit cognitive biases and shape user behavior in ways that primarily benefit shareholders, representing one of the main structural barriers to user autonomy online [7, 16, 25, 31]. In this paper, we focus on *interaction cost* [6, 26] as a proxy for this structural barrier, which we operationalize as the number of taps (clicks) and distinct interface states (screens, including modals) required to locate and complete a privacy or safety task.

Prior research has documented dark patterns in social media and privacy-related flows, including evidence that platforms deploy obstructive or manipulative tactics to sustain engagement, collect data, or delay exit [28, 30, 36, 37]. Analyses of Instagram's interface,

for instance, reveal shifts from overt manipulation to more subtle forms of control, reinforcing user engagement while maintaining the appearance of regulatory compliance [37]. Scholars have found similar use of dark patterns across Facebook, TikTok, and Twitter/X, where patterns of interface interference, bad defaults, attention capture patterns, and obstructive tactics prove effective even if users are aware [28, 30]. Yet there is little knowledge about how the effort required to engage with core privacy and safety settings [24, 37], including how these user flows compare across the services teens use every day.

To address this important area of research, we answer the following research questions:

- RQ1: What baseline interaction cost (i.e., clicks and screens) does a teen user encounter when completing seven common privacy and safety tasks on major Social Networking Services (SNSs)?
- RQ2: Which dark patterns potentially shape teenagers' experience of these privacy and safety settings, and how might these patterns impact teens' sense of control across tasks and platforms?

The contributions of this poster are two-fold. First, we identify the task complexity of common privacy and safety related tasks on social media platforms, revealing whether the presence of these settings is sufficient to support user engagement and wellbeing. Second, we reveal how dark patterns subvert user goals of control and agency that relate to these settings, underscoring the need for more robust policy responses that account for the user experience.

2 Methodology

We conducted an expert evaluation study leveraging principles of qualitative content analysis [9, 22] to examine how privacy and safety settings are structured on teen accounts across four platforms: TikTok, Instagram, Snapchat, and YouTube. We selected these SNSs due to their popularity among teenagers and varied design logics [4, 7, 19, 20, 27]. We focused on seven representative tasks that frequently appear in teen safety narratives and prior privacy settings research [10, 18, 21, 28, 33, 36, 42]: Onboarding, setting a screen-time limit, making an account private, managing in-app notifications, reporting content, downloading account data, and deleting an account.

Our research team, each of which had prior experience studying dark patterns, completed each task in November 2025. We created teen accounts configured via each platform's age-based settings (birth date under 18)[1]. For each task, we recorded all interactions required to locate and complete the setting and counted the number of clicks and screens along an ideal or optimized completion path. We operationalized *clicks* as discrete tap actions and *screens* as distinct UI states (including modal dialogs), then counted the minimum clicks and screens needed to successfully complete the task. All screenshots were organized on Miro to visualize the completion path and facilitate further analysis.

We then used the Gray et al. [16] dark patterns ontology to identify where manipulative or obstructive design elements appeared in these tasks. Following best practices for the evaluation of dark patterns using qualitative content analysis [14, 16, 28], we coded each task flow as an analytic unit, using the ontology to identify the presence of high-level dark pattern strategies (i.e., obstruction, sneaking, interface interference, forced action, social engineering). Two researchers coded the task flows independently and resolved disagreements through discussion; an additional reviewer audited the final mappings and we maintained a coding memo for edge cases.

[1]TikTok Version 42.2.0; Instagram Version 406.1.0; Snapchat Version 13.59.0; YouTube Version 20.50.9; iOS 26

3 Findings

Across all privacy and safety tasks, we observed consistent structural friction in how teens would need to locate and complete key settings across TikTok, Instagram, Snapchat, and YouTube. We measured baseline interaction cost as the number of clicks and screens in the shortest successful path to completion, excluding mistakes, backtracking, and exploratory navigation. We then interpreted these structural costs through the lens of user control, asking how the design of each flow might shape teens' sense of agency and their willingness to complete each task. In this section, we provide an analytic overview of our findings for all seven tasks in Figure 1 and describe three varied tasks that illustrate a range of issues relating to task difficulty and the presence of dark patterns.

3.1 Task 2: Set a Screen Time Limit

Platforms frame screen time settings as a way for teens to stay in control of their attention and routines, but our analysis suggests that this promise is often obscured and complicated by the task framing. Setting screen time limits across platforms revealed a clear asymmetry between availability and configuration. For instance, setting a same limit everyday was relatively straightforward (requiring between four and eight clicks across three to seven screens), but configuring custom limit per day took many more clicks and screens than setting a single overall limit, especially on TikTok and YouTube. This asymmetry matters because it creates two different experiences of "control": (1) an immediate form of control that is easy to turn on but is easily ignored or changed; (2) a more malleable form of control that becomes costly once teens try to match limits to real schedules (e.g., school, weekends, hobbies). The platforms can claim that the feature exists, while a durable and teen-relevant version of control is absent.

These control issues are reinforced by multiple dark pattern strategies that constrain teen agency while preserving the appearance of choice. These flows exemplify **Obstruction**[2] through *Adding Steps*, because meaningful control requires navigating multiple nested menus, even under ideal conditions. The lack of a consolidated grouping of relevant settings also indicates the presence of a *Privacy Maze*, since teens must move across separate pages in the settings panel to locate and finalize a single limit configuration, which makes it harder to maintain a clear sense of progress and completion. Differences in nomenclature across platforms can add another layer of friction beyond navigation. Instagram places this setting under *"Time Management"*, while YouTube places it under *"Time Watched"*, which can make it harder for teens to recognize

[2]We use the following notation to indicate the level of dark pattern being referenced: **Bold** for High-Level Pattern, *Italic* for Meso-Level Pattern, *Italic Underline*. for Low-Level Pattern

Task	High-Level Dark Pattern	T	I	S	Y
Task 1:Onboarding / Account Set up	Sneaking	○	●	○	○
	Obstruction	○	●	●	●
	Interface Interference	○	●	●	○
	Forced Action	●	●	●	●
	Social Engineering	●	●	●	●
	Number of Clicks	8	15	15	14
	Number of Screens	7	15	8	9
Task 2: Set a Screen-Time Limit	Sneaking	●	●	-	●
	Obstruction	●	●	-	●
	Interface Interference	●	●	-	●
	Forced Action	○	○	-	○
	Social Engineering	○	○	-	○
	Number of Clicks	8	5	-	4
	Number of Screens	7	5	-	3
Task 3: Set Private Account	Sneaking	○	○	-	-
	Obstruction	●	●	-	-
	Interface Interference	○	○	-	-
	Forced Action	○	○	-	-
	Social Engineering	○	○	-	-
	Number of Clicks	4	4	-	-
	Number of Screens	3	3	-	-
Task 4: Manage In-App Notifications	Sneaking	○	○	○	○
	Obstruction	●	●	●	●
	Interface Interference	●	●	●	●
	Forced Action	○	○	○	○
	Social Engineering	●	●	●	●
	Number of Clicks	40	83	27	16
	Number of Screens	6	29	6	3
Task 5: Report Content	Sneaking	○	○	○	○
	Obstruction	●	●	●	●
	Interface Interference	●	●	●	●
	Forced Action	○	○	○	○
	Social Engineering	○	●	○	○
	Number of Clicks	6	5	5	4
	Number of Screens	5	4	4	4
Task 6: Download Your Data	Sneaking	○	○	○	○
	Obstruction	●	●	●	●
	Interface Interference	○	○	○	○
	Forced Action	○	○	○	○
	Social Engineering	○	○	○	○
	Number of Clicks	7	9	5	4
	Number of Screens	7	9	5	4
Task 7: Delete Account	Sneaking	○	●	○	-
	Obstruction	●	●	●	-
	Interface Interference	●	●	○	-
	Forced Action	●	●	●	-
	Social Engineering	○	○	○	-
	Number of Clicks	5	13	4	-
	Number of Screens	5	11	3	-

Legend: **T** - TikTok, **I** - Instagram, **S** - Snapchat, **Y** - YouTube

Figure 1: Each privacy and safety task is presented alongside the number of clicks and screens required to complete the setting. Additionally, the presence of high-level dark patterns is indicated per task and platform, leveraging the Gray et al. ontology [16]. TikTok (T), Instagram (I), Snapchat (S), and YouTube (Y).

that these labels are pointing to similar "self control" goals across apps and to locate the feature when they are uncertain where to start.

Finally, the interface framing itself can soften the teen's intention to set a firm boundary. Screen time flows on TikTok and YouTube exhibit **Interface Interference** through *Emotional or Sensory Manipulation*, where friendly illustrations and calm visual framing reduce the perceived urgency of limit setting. This framing could makes disengagement feel low stakes even when a teen is trying to protect themselves or fight back against addiction or compulsive use. TikTok also relies on *False Hierarchy* by making the path into screen time settings more prominent than the path to exit. These patterns position screen time limits as a visible tool for "control," while nudging teens toward a lighter, more reversible form of control that is easy to dismiss and difficult to customize.

3.2 Task 4: Manage In-App Notifications

Platforms present notification settings as a mechanism for teens to control when and how they get notified, but our expert review suggests that this promise of control relies on obstructive tactics to suppress teens' ability to get notified less frequently. This task was the most interaction-intensive across platforms, requiring 40 clicks and 6 screens on TikTok, 83 clicks and 29 screens on Instagram, 27 clicks and 6 screens on Snapchat, and 16 clicks and 3 screens on YouTube. TikTok and YouTube largely keep notification changes within a single primary notification settings page, so teens can carry out adjustments without needing to negotiate nested category screens. In contrast, Instagram and Snapchat frame notification control as a distributed process, requiring teens to understand and navigate multiple categories of notifications to reach a meaningful "quiet" state. For instance, on Instagram, settings are fragmented across posts, messages, and system alerts, forcing the user to navigate multiple screens. Snapchat similarly distributes notification options across different locations, which increases the chance that teens miss a category or assume they have already turned something off. These notification controls also leverage teen social norms, such as framing notifications as "recommended," "personalized," or tied to socially meaningful cues (e.g., friends, messages, and activity). In these flows, the interface position notifications as the "right" or "normal" choice for staying connected, which leverages teens' social motivations to keep more alerts enabled than they would otherwise choose. This pressure operates alongside the structural friction: even when a teen intends to reduce interruptions, personalization framing can make turning categories off feel like opting out of connection rather than simply changing a setting.

The notification settings are reinforced by dark pattern strategies that constrain teen agency while preserving the appearance of choice. The flows exemplify **Obstruction** through *Adding Steps*, since teens cannot meaningfully tailor notifications through a single consolidated control (i.e., turn off all notifications), especially on Instagram where the sheer number of granular settings result in *Choice Overload*. Default configurations also align with **Interface Interference** through *Bad Defaults*, because most notification categories begin in an "on" state and require teens to opt out one toggle at a time. Finally, fragmented pathways and long, visually uniform toggle lists contribute to *Hidden Information*, particularly on Instagram, where the lack of a comprehensive overview of settings makes it difficult for teens to identify which notification controls matter most and whether their intended changes have actually been completed.

3.3 Task 7: Delete Account

All platforms present account deletion as a clear form of user choice: if a user wants to leave, they can delete their account. However, our analysis shows a consistent gap between this promise and the actual experience of carrying deletion through to completion. While TikTok requires a relatively short path (5 clicks, 5 screens) and Snapchat presents the option in a straightforward way with a single extra identity confirmation step, Instagram extends the flow to 13 clicks and 11 screens, including multiple decision points that can slow progress and redirect attention.

This gap is reinforced through interface structures that make "leaving" feel possible while making permanent deletion harder than it appears. Across platforms, the need to move through multi-page

flows rather than a single consolidated control reflects **Obstruction**, *Adding Steps* and increasing friction once users express an interest in leaving the platform. TikTok, Instagram and Snapchat also require account verification before deletion can proceed. In our coding process, we count password re-entry and similar verification prompts as part of the overall interaction cost, and we therefore evaluate the overall deletion pathway as obstructive even when individual steps may be framed as protective. Instagram combines these strategies by utilizing **Interface Interference**: the delete control is buried inside the Account Center menu, and the interface repeatedly introduces more prominent alternatives such as deactivation or other options that discourage account deletion, relying on *False Hierarchy* and *Visual Prominence* to steer choices. Instagram uses **Sneaking** through *Bait and Switch*, because a user who enters the flow intending to delete is repeatedly routed into deactivation and alternative paths are provided, increasing the likelihood that they complete a less final action than the one they originally chose. Instagram also prompts users to download or transfer data and to select reasons for leaving by giving alternative options rather than deleting the account, using ambiguous wording and *Trick Questions* that can create uncertainty about whether deletion will actually occur.

Importantly, account deletion is shaped not only by steps and screens, but also by what happens after the request is made. On all platforms, deletion is not immediate and requires a 30-day retention or waiting period (an example of **Forced Action**), which can be interpreted as a protective mechanism that gives users time to recover an account or respond to suspicious activity. However, this design also makes deletion feel reversible in practice, because during that period, logging back in cancels the deletion process entirely. This means that a user who returns even briefly (e.g., in a moment of weakness) can lose all prior progress and must restart the deletion process and wait an additional 30 days [29]. From a teen centered lens, this reversibility can increase the chance of returning to the platform, especially in attention capture systems where habit and relapse are common. In other words, the same retention structure that can protect users can also function as deceptive friction by leveraging predictable return behavior, allowing platforms to claim that deletion is supported while still benefiting from canceled deletion attempts.

4 Discussion and Provocations

4.1 Platforms Create a Sense of Teen Control while Quietly Undermining Agency

A core expectation in HCI is that people should feel they can make choices, recover from mistakes, and steer the system toward their own goals, captured in Nielsen's heuristic of *user control and freedom* [32]. Viewed through the lens of control, the dark patterns we identified suggest that teen privacy and safety settings often operate as "control surfaces" that look legitimate, but are structured in ways that may reduce a teen's ability to confidently locate, enact, and verify protective choices. We discuss further the implication of these control surfaces and the balance between user and shareholder assumptions of agency.

First, our analysis revealed that **control is technically possible, but not easily reachable**. Platforms can claim they offer control because settings *exist*, yet teens still may not be able to use them in practice. When controls are buried in nested menus, split across categories, or labeled inconsistently, teens may not be able to find them and may not even realize the setting exists. Even after locating a setting, completing a protective change can require repeated steps, redundant screens, and back-and-forth navigation. This raises the interaction cost of following through, so the default state becomes the easiest path. In this way, *having* control may not guarantee that teens can readily activate it when they need it. Second, we discovered that **soft framing can weaken protective intentions**. Some teen safety and wellbeing features are framed with friendly language and calming visual cues that position them as reflective tools rather than firm boundaries. This framing can subtly reshape what "control" means, shifting it from enforcing limits that promise better wellbeing to merely monitoring behavior. As a result, teens may be guided toward self management (assuming *they* are the problem) rather than protective action (where the platform is placing the user in control, even when they are responsible for "hooking" the user). Third, we found that **reversibility makes control unnecessarily brittle**. Control can be rapidly undermined when settings are easy to dismiss, override, or reverse. If protective actions can be undone quickly or with little consequence, teens may learn that these controls are easy to reverse. This could potentially make the experience of control feel brittle rather than absent: teens can take action, but the outcome may not last, which can reduce confidence and make them less likely to rely on these settings again.

4.2 Privacy and Safety Settings Act as a Façade of Control

Privacy and safety settings also relate to contemporary policy conversations. Platforms publicize these controls in an attempt to show regulators that they are protecting minors and following the rules. However, our analysis suggests a gap between claiming a setting exists and the practical question of whether a teen can reliably use it. When a setting is hard to locate or a task is difficult to complete, it is quietly limiting the chance that teens will complete protective actions. In that sense, the interface can function as a façade: visible enough to cite, but structured in ways that reduce real uptake and positive impact on wellbeing. This framing points to an evaluation shift that is relevant for HCI-informed platform governance and policy [12, 41]: assessment should not stop at whether a setting exists. Instead, we should ask whether teens can (1) form a clear protective intention; (2) carry it through to completion without undue friction or reversal; and (3) understand the outcome and its durability. This approach also aligns with policy-oriented research that treats digital harm as something that is defined and enforced differently across contexts, and argues that design details matter because they shape what protections mean in practice [11]. A control-based evaluation stance would therefore treat usability, durability, and ultimate impact as central to compliance.

4.3 Provocations for Future Work

These findings and provocations reveal the need for future work, including how teens actually interpret, weigh, and respond to the frictions we have identified through our expert review. In addition, future work should explore how HCI perspectives might inform more impactful policies that take usability and impact into account, building on the work of Yang et al. [41] and Jackson et al. [23]. There are numerous facets of teenager use of these platforms that are also important to consider, including the relationships between the structural patterns we identified and teenagers' self control and agency. For example, indicating the socio-technical complexity of cases where a teen might want to limit their use of social media or leave entirely, but may feel pulled back toward engagement due to social or psychological factors outside of their control. By considering both usability issues and the ways that experiences are shaped by dark patterns, future scholarship should seek to link interface level analysis with teen reported strategies and tensions, and to inform future design and policy work that better supports teen privacy and safety as platforms evolve.

5 Conclusion

In this paper, we analyzed seven privacy and safety settings across four major SNSs, describing how these settings are impacted by dark patterns that complicate teen efforts to manage risk. Through expert evaluation, we identified the number of clicks and screens to complete each task and mapped the flows to relevant dark patterns [16, 28]. Our findings indicate that privacy and safety settings are present, but often difficult to find and require undue friction to complete, disrupting teens' sense of control. We provide provocations for platform governance and policy, indicating that evaluation should not stop at determining whether settings *exist*, but rather assessing whether teens can make informed choices to protect themselves in robust and durable ways.

References

[1] Naima Samreen Ali, Sarvech Qadir, Ashwaq Alsoubai, Munmun De Choudhury, Afsaneh Razi, and Pamela J Wisniewski. 2024. "I'm gonna KMS": From imminent risk to youth joking about suicide and self-harm via social media. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, Vol. 15. ACM, New York, NY, USA, 1–18. doi:10.1145/3613904.3642489

[2] Shiza Ali, Afsaneh Razi, Seunghyun Kim, Ashwaq Alsoubai, Chen Ling, Munmun De Choudhury, Pamela J Wisniewski, and Gianluca Stringhini. 2023. Getting meta: A multimodal approach for detecting unsafe conversations within Instagram direct messages of youth. *Proc. ACM Hum. Comput. Interact.* 7, CSCW1 (April 2023), 1–30. doi:10.1145/3579608

[3] Abdulmalik Alluhidan, Mamtaj Akter, Ashwaq Alsoubai, Jinkyung Katie Park, and Pamela Wisniewski. 2024. Teen talk: The good, the bad, and the neutral of adolescent social media use. *Proc. ACM Hum. Comput. Interact.* 8, CSCW2 (Nov. 2024), 1–35. doi:10.1145/3686961

[4] Monica Anderson and Jingjing Jiang. 2018. Teens, Social Media & Technology 2018. *Pew research center* (May 2018). https://www.pewresearch.org/internet/2018/05/31/teens-social-media-technology-2018/

[5] Arezou Behfar, Ankit Shrestha, and Mahdi Nasrullah Al-Ameen. 2024. A first look into fake profiles on social media through the lens of victim's experiences. In *Companion Publication of the 2024 Conference on Computer-Supported Cooperative Work and Social Computing*. ACM, New York, NY, USA, 444–450. doi:10.1145/3678884.3681889

[6] Raluca Budiu. 2013. Interaction Cost. https://www.nngroup.com/articles/interaction-cost-definition/. Accessed: 2026-1-21.

[7] Yixin Chen, Yue Fu, Zeya Chen, Jenny Radesky, and Alexis Hiniker. 2024. The engagement-prolonging designs teens encounter on Very Large Online Platforms. *arXiv [cs.HC]* (Nov. 2024). https://arxiv.org/abs/2411.12083

[8] Jenna Palermo Christoferson. 2016. How is Social Networking Sites Efecting Teen's Social and Emotional Development: A Systemic Review. *Master of Social Work Clinical Research Papers* (2016), Paper 650.

[9] Satu Elo and Helvi Kyngäs. 2008. The qualitative content analysis process. *Journal of advanced nursing* 62, 1 (2008), 107–115. doi:10.1111/j.1365-2648.2007.04569.x

[10] Emma Fahlman, Thomas Mejtoft, and Helen Cripps. 2018. Evaluation of Push Notifications for Social Media Applications. (2018). doi:10.18690/978-961-286-170-4.21

[11] Ritika Gairola. 2025. Centering harm in Socio-technical systems: Connecting design, user experience, and policy. In *Companion Publication of the 2025 Conference on Computer-Supported Cooperative Work and Social Computing*. ACM, New York, NY, USA, 11–14. doi:10.1145/3715070.3747333

[12] Ritika Gairola and Colin M Gray. 2025. How is "Public Policy" Used in HCI Scholarship?. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25)*. ACM Press. doi:10.1145/3706599.3719997

[13] Ritika Gairola, Colin M Gray, Jingxin Dong, Kyung Jin Jeong, Ege Otenen, and Juan J Sarria. 2026. "Social Media Killed Our Generation": Teenagers' Felt Experiences of Harm on Social Media. In *CHI '26: CHI Conference on Human Factors in Computing Systems Proceedings*. Association for Computing Machinery. doi:10.1145/3772318.3791519

[14] Colin M Gray, Yubo Kou, Bryan Battles, Joseph Hoggatt, and Austin L Toombs. 2018. The dark (patterns) side of UX design. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–14. doi:10.1145/3173574.3174108

[15] Colin M Gray, Cristiana Santos, Nataliia Bielova, Michael Toth, and Damian Clifford. 2021. Dark patterns and the legal requirements of consent banners: An interaction criticism perspective. In *Proceedings of the 2021 CHI conference on human factors in computing systems*. 1–18. 10.48550/arXiv.2009.10194

[16] 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 CHI Conference on Human Factors in Computing Systems (CHI '24, Article 289)*. Association for Computing Machinery, New York, NY, USA, 1–22. doi:10.1145/3613904.3642436

[17] Colin M Gray, Cristiana Teixeira Santos, Nicole Tong, Thomas Mildner, Arianna Rossi, Johanna T Gunawan, and Caroline Sinders. 2023. Dark patterns and the emerging threats of deceptive design practices. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–4. doi:10.1145/3544549.3583173

[18] Anatoliy Gruzd and Ángel Hernández-García. 2018. Privacy concerns and self-disclosure in private and public uses of social media. *Cyberpsychol. Behav. Soc. Netw.* 21, 7 (July 2018), 418–428. doi:10.1089/cyber.2017.0709.

[19] Alexis Hiniker, Sharon S Heung, Sungsoo (ray) Hong, and Julie A Kientz. 2018. Coco's videos: An empirical investigation of video-player design features and children's media use. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA. doi:10.1145/3173574.3173828

[20] Alexis Hiniker, Sarita Y. Schoenebeck, and Julie A. Kientz. 2016. Not at the Dinner Table: Parents' and Children's Perspectives on Family Technology Rules. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (San Francisco, California, USA) *(CSCW '16)*. Association for Computing Machinery, New York, NY, USA, 1376–1389. doi:10.1145/2818048.2819940

[21] Jihyeong Hong, Eun-Young Ko, Juho Kim, and Jeong-Woo Jang. 2025. Why social media users press 'Not Interested': Motivations, anticipated effects, and result interpretation. *Proc. ACM Hum. Comput. Interact.* 9, 7 (Oct. 2025), 1–25. doi:10.1145/3757706

[22] Hsiu-Fang Hsieh and Sarah E Shannon. 2005. Three approaches to qualitative content analysis. *Qualitative health research* 15, 9 (2005), 1277–1288. doi:10.1177/1049732305276687

[23] Steven J Jackson, Tarleton Gillespie, and Sandy Payette. 2014. The policy knot: re-integrating policy, practice and design in cscw studies of social computing. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing (CSCW '14)*. ACM, New York, NY, USA, 588–602. doi:10.1145/2531602.2531674

[24] Dominique Kelly and Jacquelyn Burkell. 2023. Documenting privacy dark patterns: How social networking sites influence users' privacy choices. (2023). https://ir.lib.uwo.ca/fimspub/376/

[25] Yubo Kou, Colin M Gray, Austin L Toombs, and Robin S Adams. 2018. Understanding social roles in an online community of volatile practice: A study of user experience practitioners on reddit. *ACM Trans. Soc. Comput.* 1, 4 (Dec. 2018), 1–22. doi:10.1145/3283827

[26] Heidi Lam. 2008. A framework of interaction costs in information visualization. *IEEE Trans. Vis. Comput. Graph.* 14, 6 (Nov. 2008), 1149–1156. doi:10.1109/TVCG.2008.109

[27] Rotem Landesman, Jina Yoon, Jaewon Kim, Daniela E Munoz Lopez, Lucia Magis-Weinberg, Alexis Hiniker, and Katie Davis. 2024. "I just don't care enough to be interested": Teens' moment-by-moment experiences on Instagram. In *Proceedings of the 23rd Annual ACM Interaction Design and Children Conference*. ACM, New York, NY, USA, 14–29. doi:10.1145/3628516.3655812

[28] Thomas Mildner, Merle Freye, Gian-Luca Savino, Philip R Doyle, Benjamin R Cowan, and Rainer Malaka. 2023. Defending against the dark arts: Recognising

dark patterns in social media. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*. ACM, New York, NY, USA, 2362–2374. doi:10.1145/3563657.3595964

[29] Thomas Mildner, Gian-Luca Savino, Philip R Doyle, Benjamin R Cowan, and Rainer Malaka. 2023. About Engaging and Governing Strategies: A Thematic Analysis of Dark Patterns in Social Networking Services. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23, Article 192)*. Association for Computing Machinery, New York, NY, USA, 1–15. doi:10.1145/3544548.3580695

[30] Thomas Mildner, Gian-Luca Savino, Susanne Putze, and Rainer Malaka. 2024. Finding a way through the social media labyrinth: Guiding design through user expectations. In *Proceedings of the International Conference on Mobile and Ubiquitous Multimedia*. ACM, New York, NY, USA, 157–171. doi:10.1145/3701571.3701605

[31] Alberto Monge Roffarello, Kai Lukoff, and Luigi De Russis. 2023. Defining and identifying attention capture deceptive designs in digital interfaces. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA. doi:10.1145/3544548.3580729

[32] Jakob Nielsen. 1994. Enhancing the explanatory power of usability heuristics. In *Proceedings of the SIGCHI conference on Human factors in computing systems celebrating interdependence - CHI '94*. ACM Press, New York, New York, USA. doi:10.1145/191666.191729

[33] Luci Pangrazio and Neil Selwyn. 2018. "it's not like it's life or death or whatever": Young people's understandings of social media data. *Soc. Media Soc.* 4, 3 (July 2018), 205630511878780. doi:10.1177/2056305118787808

[34] Rizu Paudel and Mahdi Nasrullah Al-Ameen. 2025. Understanding Teen's Expectations and Challenges to Seek Help in Social Media. In *Proceedings of the 24th Interaction Design and Children*. Association for Computing Machinery, New York, NY, USA, 1056–1063. https://doi.org/10.1145/3713043.3731530

[35] Lorena Sanchez Chamorro, Carine Lallemand, and Colin M Gray. 2024. "my mother told me these things are always fake" - understanding teenagers' experiences with manipulative designs. In *Designing Interactive Systems Conference*. ACM, New York, NY, USA. doi:10.1145/3643834.3660704

[36] Brennan Schaffner, Neha A Lingareddy, and Marshini Chetty. 2022. Understanding account deletion and relevant dark patterns on social media. *Proc. ACM Hum. Comput. Interact.* 6, CSCW2 (Nov. 2022), 1–43. doi:10.1145/3555142

[37] Shamim Seyson and Wesley Willett. 2025. Exploring the evolution of dark patterns and manipulative design on Instagram. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–8. doi:10.1145/3706599.3719771

[38] Yike Shi, Qing Xiao, Qing Hu, Hong Shen, and Hua Shen. 2025. The siren song of llms: How users perceive and respond to dark patterns in large language models. *arXiv preprint arXiv:2509.10830* (2025).

[39] Ruyuan Wan, Lingbo Tong, Tiffany Knearem, Toby Jia-Jun Li, Ting-Hao 'kenneth' Huang, and Qunfang Wu. 2025. Hashtag re-appropriation for audience control on recommendation-driven social media xiaohongshu (rednote). In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–25. doi:10.1145/3706598.3713379

[40] Pamela Wisniewski. 2018. The privacy paradox of adolescent online safety: A matter of risk prevention or risk resilience? *IEEE Secur. Priv.* 16, 2 (March 2018), 86–90. doi:10.1109/MSP.2018.1870874

[41] Qian Yang, Richmond Y Wong, Steven Jackson, Sabine Junginger, Margaret D Hagan, Thomas Gilbert, and John Zimmerman. 2024. The future of HCI-policy collaboration. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24, Vol. 72)*. ACM, New York, NY, USA, 1–15. doi:10.1145/3613904.3642771

[42] Mingrui Ray Zhang, Kai Lukoff, Raveena Rao, Amanda Baughan, and Alexis Hiniker. 2022. Monitoring screen time or redesigning it?: Two approaches to supporting intentional social media use. In *CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA. doi:10.1145/3491102.3517722