

Design Transparency Labels: Fostering User Trust and Well-Being Through Design Transparency

Original

Design Transparency Labels: Fostering User Trust and Well-Being Through Design Transparency / Graziosi, R., Passarella, L.. - STAMPA. - 4217:(2026), pp. 1-8. (BOTW 2026 Bridge Over Troubled Water: Aligning Commercial Incentives With Ethical Design Practice To Combat Deceptive Patterns Barcelona (ESP) April 13-17, 2026).

Availability:

This version is available at: 11583/3012553 since: 2026-06-30T12:37:18Z

Publisher:

CEUR-WS

Published

DOI:

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

Design Transparency Labels: Fostering User Trust and Well-Being Through Design Transparency

Riccardo Graziosi¹, Luca Passarella¹

¹*Polytechnic of Turin, Turin, Italy*

Abstract

Over the past decade, there has been growing recognition of unethical design practices aimed at deceiving users or persuading them beyond their original intentions. Businesses employ these techniques to compete within the advertising-based attention economy, whereas users often bear the cost in the form of reduced control over their attention and increased time spent online. To ease this tension between business incentives and user well-being, we propose the concept of Design Transparency Labels (DTLs). DTLs are in-app, contextual interface notices that make persuasive design features visible, explain how specific design choices may affect users' digital well-being, and offer mechanisms to regain control. We argue that DTLs can simultaneously promote more mindful and intentional technology use, while also helping service providers align with existing regulatory requirements and build trust capital that may foster user satisfaction and long-term retention.

Keywords

Persuasive Design, Digital Well-Being By Design, Regulation, Attention Economy

1. Introduction

Over the past decade, both scholars and regulators have expressed increasing concern about unethical design practices adopted by digital service providers that may be detrimental to users. While researchers have categorized harmful designs under terms such as deceptive design and attention-capture damaging patterns [1, 2] and have studied their effects on users [3, 4], regulators have proposed legislation to prohibit the most harmful practices, with notable examples including the European Digital Services Act (DSA) [5] and the Artificial Intelligence Act (AI Act) [6]. While some overtly deceptive practices such as sneaking items into a shopping basket or making the unsubscribing process unnecessarily complex are relatively easy to identify and regulate, other questionable practices (e.g. attention-capture damaging patterns) are more difficult to frame and regulate. These latter strategies, hereinafter referred to as *persuasive design*, persuade rather than deceive, making it challenging to draw a clear line between fair and unfair use.

As service providers deploy these persuasive strategies at scale to compete for users' attention and advertising revenue, a structural tension becomes evident: maximizing engagement to sustain advertising income may come at the expense of respecting users' time and attention. To ease this tension, we propose the concept of Design Transparency Labels (DTLs). DTLs are contextual interface notices that inform users about the (persuasive) design strategies being employed and their potential impact on user well-being, while also providing practical calls to action that enable users to regain control. DTLs are not external interventions, but are intended to be integrated directly within the target interface. In this respect, they are inspired by transparency mechanisms such as cookie notices, but they extend the concept beyond data collection to interface-level persuasion. For example, a DTL could raise awareness about the use of infinite scroll by explaining its potential effect on time spent and attentional engagement, and offering users the option to disable it.

DTLs may be beneficial for both users and service providers. Unlike existing external tools for digital self-control [7], DTLs aim to empower users by increasing awareness and fostering understanding of digital well-being, while also offering contextual calls to action to get back control. From the service

Bridge Over Troubled Water: Aligning Commercial Incentives With Ethical Design Practice To Combat Deceptive Patterns. Workshop at the 2026 CHI Conference on Human Factors in Computing Systems (CHI EA '26), April 13-17, 2026, Barcelona, Spain
✉ riccardo.graziosi@polito.it (R. Graziosi); s333229@studenti.polito.it (L. Passarella)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

providers' perspective, integrating DTLs can be understood as an investment in trust capital. As users become increasingly aware of the techniques used to sustain engagement, it is plausible that, in the near future, they may begin to prefer services that respect their attention by design. From this perspective, DTLs can function as an ethical differentiation strategy. Moreover, DTLs could play a role for interface design analogous to that of nutrition labels in the food domain: rather than eliminating unhealthy or processed foods, such labels reframed competition around more informed consumer choice. Finally, DTLs may help companies operationalize alignment with emerging regulatory frameworks such as the DSA and the AI Act.

The paper is structured as follows. Section 2 situates the contribution within the existing literature. Section 3 presents the Design Transparency Label framework, detailing some operational guidelines to help designing effective DTLs. Section 4 discusses the potential impact of DTLs both from the business and user perspectives. Finally, Section 5 concludes the paper.

2. Related Works

Concerns over persuasive design strategies have contributed to the emergence of the concept of digital well-being, which refers to the impact of digital technologies on individuals' ability to live good, meaningful, and fulfilling lives [8]. Digital well-being have been originally seen as an individual responsibility, i.e. users exercise self-control in the face of persuasive technologies using digital self-control tools [9].

More recently, however, researchers have begun to move the responsibility of digital well-being from users to platforms and designers of user interfaces. Peters [10] introduced the notion of Well-being Supportive Design, proposing guidelines for designing with respect for user autonomy and the support of informed, meaningful choices. Despite these advances, online platforms often struggle to implement such principles in practice, as engagement-driven business models may conflict with the goal of promoting user well-being.

DTLs occupy an intermediate position within this landscape. They are not digital self-control tools, as their primary aim is not to restrict use or impose hard limits, but rather to foster awareness and gradual learning. At the same time, they do not require platforms to radically redesign their services or abandon engagement-based features. Instead, DTLs preserve existing interface structures while making persuasive mechanisms more visible, understandable, and accountable, thereby enabling users to better grasp how design choices operate and how they may influence behavior.

In parallel, a body of HCI research on explainable AI has studied how to make algorithmic systems more transparent and intelligible to end users, particularly in the context of recommender systems and social media [11, 12, 13, 14]. For example, prior work has examined explanation interfaces framed around questions such as "Why am I [not] seeing this?" [14, 15]. These approaches suggest that well-designed explanations can enhance users' understanding of how algorithmic systems curate and prioritize content [16]. However, simply making algorithmic processes transparent does not necessarily promote agency, as users may understand how an algorithm works yet remain unable to steer it to their needs [17].

Moreover, researchers have proposed label-based mechanisms to ensure transparency. Cookie notices, for example, now ubiquitous due to regulations such as the GDPR [18], aim to inform users about data collection practices. Other proposals, such as Model Cards [19] and FactSheets [20], seek to summarize the properties, risks, and performance characteristics of AI systems in standardized and comparable formats. While these approaches contribute to documentation and accountability, recent studies suggest that they are often too technical to meaningfully support non-expert users during actual use [21].

DTLs build on insights from these works but differ in both scope and intent. Rather than explaining at a general level or summarizing system-wide characteristics, DTLs focus on specific persuasive design strategies as they are enacted through interface design. In this sense, DTLs complement rather than replace existing explanation and labeling approaches by shifting transparency from static documentation or one-time prompts such as cookie notices to contextual, in-situ interface cues embedded directly

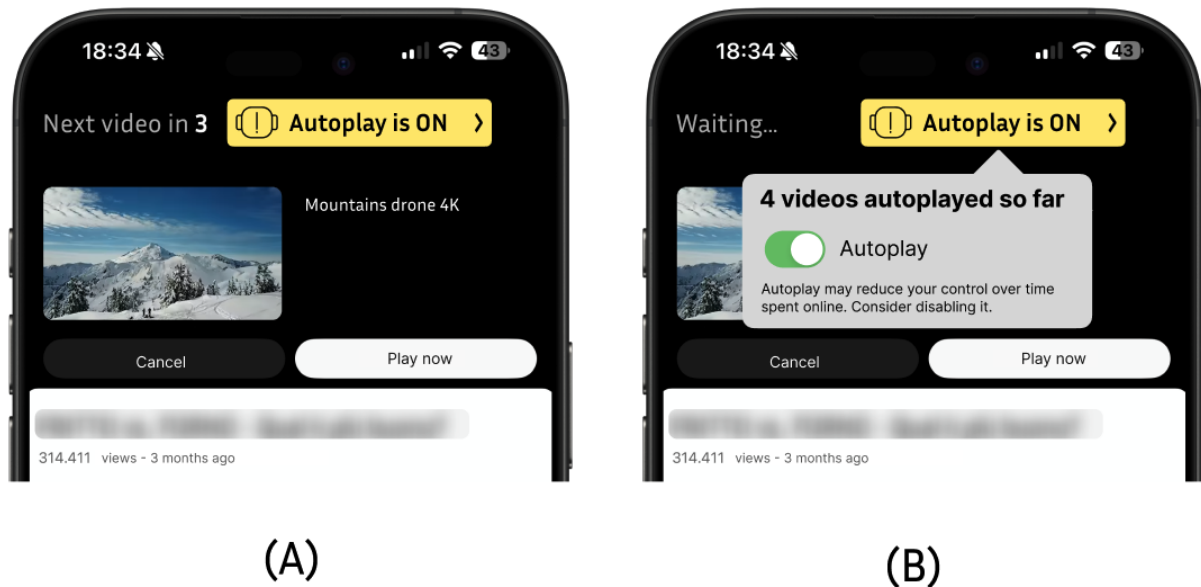


Figure 1: A DTL for YouTube that supports mindful reflection over continuous autoplay. The DTL warns that autoplay is currently activated (A), while tapping it opens a an explanatory panel showing how many videos in a row the user has watched, plus a call to action to disable autoplay (B).

within the flow of use.

3. Design Transparency Labels

Design Transparency Labels (DTLs) are in-app, contextual notices that inform users about a platform's design choices and their potential impact on well-being. Just as cookie consent banners promote transparency around data collection, DTLs extend this concept to interface and algorithmic persuasion. An example DTL for YouTube is depicted in Figure 1. After a video finishes playing, a yellow box is displayed to inform users that autoplay is activated. If the box is tapped, an informative pop-up appears showing the number of consecutive videos played, raising awareness of the potential loss of control associated with this feature, and offering a toggle to disable autoplay.

Because DTLs are intended to be integrated directly into interfaces by platform designers, we do not propose a standardized library of DTLs. Instead, we outline an initial set of guidelines to support designers in implementing their own DTLs:

1. **Language:** DTLs should adopt a concise and to-the-point tone while remaining informative enough to support meaningful reflection. The wording should be clear, concrete, and free of technical jargon, so that users can easily understand which design feature is being highlighted and why it matters. Importantly, DTLs should avoid blaming or shaming users for their behavior. Rather than implying a lack of self-control (e.g., "You have been watching for too long"), they should frame the information in neutral or supportive terms (e.g., "4 videos autoplayed so far"). The tone should be non-judgmental, empathetic, and oriented toward user empowerment. Whenever possible, DTLs should emphasize agency by explicitly stating that users can modify the feature or adjust their settings. In this way, language becomes not only a vehicle for transparency, but also a mechanism for fostering autonomy and trust.
2. **Design:** DTLs should respect the platform's visual identity by adhering to its brand style, including typography, color palette, and interaction patterns. This consistency can prevent DTLs from appearing as intrusive warnings imposed from outside the system. At the same time, DTLs must remain visually distinguishable from surrounding content to ensure that they are noticeable and not mistaken for regular interface elements. Achieving this balance may involve subtle contrast in color, the use of borders or background shading.

3. **Timing and Position:** The effectiveness of DTLs depends not only on what they communicate, but also on when and where they are presented. DTLs should be displayed at meaningful interaction points, usually when a persuasive design feature is activated or begins to shape user behavior. For example, a label addressing autoplay is most impactful immediately after a video ends and the next one begins automatically; similarly, a notice about infinite scroll may be triggered after prolonged scrolling. By aligning the appearance of DTLs with the moment of influence, designers can ensure that the information is contextually grounded and easier for users to interpret.
In addition to timing, spatial placement within the interface is critical. DTLs should appear close to the relevant design element they refer to, thereby minimizing ambiguity and supporting users in understanding the connection between the notice and the underlying feature. Poor placement—such as positioning a DTL far from the persuasive mechanism it explains—may reduce clarity and limit its effectiveness. Strategic placement should therefore prioritize proximity, visibility, and coherence with the user’s current focus of attention.
4. **Personalization:** DTLs are dynamic notices that should adapt based on multiple variables. Both their tone and visual design may progressively change as users spend more time in the app, conveying the increasing criticality of a situation. For example, the YouTube autoplay DTL shown in Figure 1 could gradually change color (e.g., from informative blue to alerting red) as the number of consecutive videos increases. In addition, session-based metrics should be integrated into DTLs to better align them with the user’s current context. In this regard, the proposed YouTube DTL displays the number of consecutive videos watched. Other possible sources of personalization include user demographics, patterns from past sessions and behavior, as well as the time of day or day of the week.
5. **Interactivity:** Beyond providing information, DTLs should almost always offer users a clear and immediate way to regain control. Therefore, DTLs should include well-labeled buttons or toggles that allow users to modify the relevant feature directly (e.g., disabling autoplay, pausing recommendations, steering feed’s algorithm). The available actions should be simple, reversible, and clearly explained to reduce friction and avoid unintended consequences. Designers may also consider graduated interaction options, such as “Remind me later”, “Pause for today”, or “Turn off permanently”, thereby supporting different levels of commitment.

4. Discussion

4.1. Build User Trust

We argue that DTLs could foster a win-win dynamic for both businesses and users. As individuals become increasingly aware of the persuasive strategies employed to capture and monetize their attention, they may grow more sensitive to how different platforms shape their behavior. In such a context, users may begin to factor the extent to which a platform actively safeguards or prioritizes their well-being into their usage decisions. Transparency about persuasive design features can therefore function as a signal of accountability and ethical commitment. From this perspective, integrating DTLs may be understood not merely as a compliance effort, but as a strategic investment in trust capital. This trust capital can strengthen user satisfaction and long-term retention.

This dynamic may become particularly important as alternative and decentralized social networking protocols, such as Bluesky and Mastodon, gain broader mainstream adoption. These platforms promote different governance and moderation models, potentially lowering switching costs and fostering more open competition. In such an evolving ecosystem, differentiation may no longer revolve solely around features, network effects, or monetization strategies, but also around perceived alignment with users’ values and well-being. As user sensitivity to persuasive design increases, competition could gradually be reframed to include well-being, transparency, and responsible attention design as key dimensions of value. In this context, the DTL framework provides a practical mechanism for incorporating these emerging dimensions into platform design.

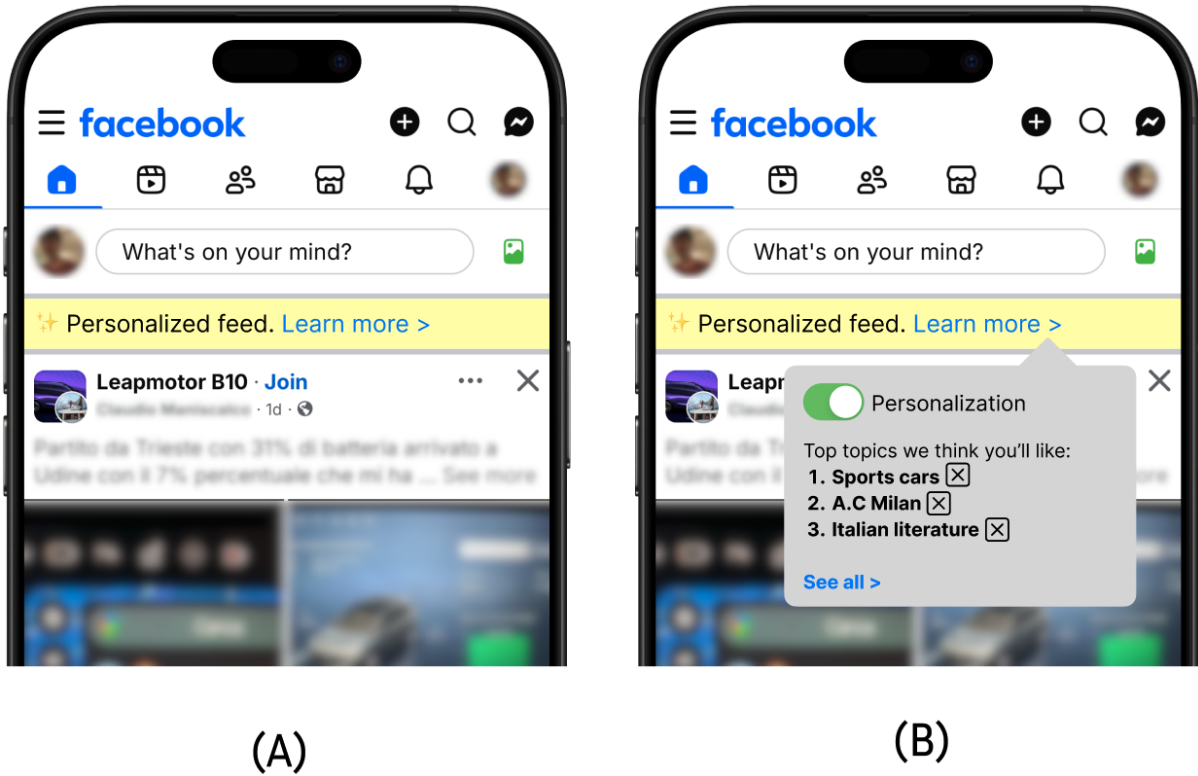


Figure 2: A DTL for Facebook that informs users about the personalized nature of the feed (A). If tapped, an explanatory panel is opened providing information about the main parameters driving recommendations as well as ways to edit them; moreover, it is possible to disable personalized recommendations completely (B).

4.2. Facilitate Regulatory Alignment

In recent years, several regulatory frameworks have been introduced to govern the operation of online platforms, including the European Digital Services Act (DSA) [5] and Artificial Intelligence Act (AI Act) [6], the United Kingdom’s Digital Markets, Competition and Consumers Act [22], the California Consumer Privacy Act [23] and Colorado Privacy Act [24] in the United States, and the Consumer Protection Act [25] in India. We argue that DTLs are well aligned with the underlying philosophy of these frameworks and may serve as practical instruments to support compliance.

Transparency constitutes a common thread across these regulatory initiatives. In general, since the DTL framework is grounded in the principle of transparency, it is particularly well suited to supporting compliance in this direction. In particular, for example, the DSA requires online platforms that deploy recommender systems to disclose the main parameters shaping those systems. While the regulation mandates that such information be made available in terms and conditions, prior research suggests that users prefer transparency information when it is presented in an interactive and personalized manner [26]. This finding indicates that contextual, in-interface cues such as DTLs could complement formal disclosures by making algorithmic transparency more accessible, comprehensible, and actionable. Figure 2 depicts an example DTL that enables users to learn more about Facebook’s recommendation algorithm parameters and steer its functioning.

Psychological well-being is also explicitly addressed in the DSA, where negative effects on physical and mental well-being are framed as systemic risks requiring mitigation. The regulation states that such risks may arise “from online interface design that may stimulate behavioural addictions of recipients of the service” (DSA, Recital 83). In this context, DTLs can be deployed to inform users when specific interface features are likely to encourage prolonged engagement and to provide options to regain control. For instance, the YouTube DTL shown in Figure 1 may be interpreted as a mechanism to enhance user awareness and agency, thereby contributing to the promotion of mental well-being.

The protection of minors represents another foundational principle within contemporary regulatory

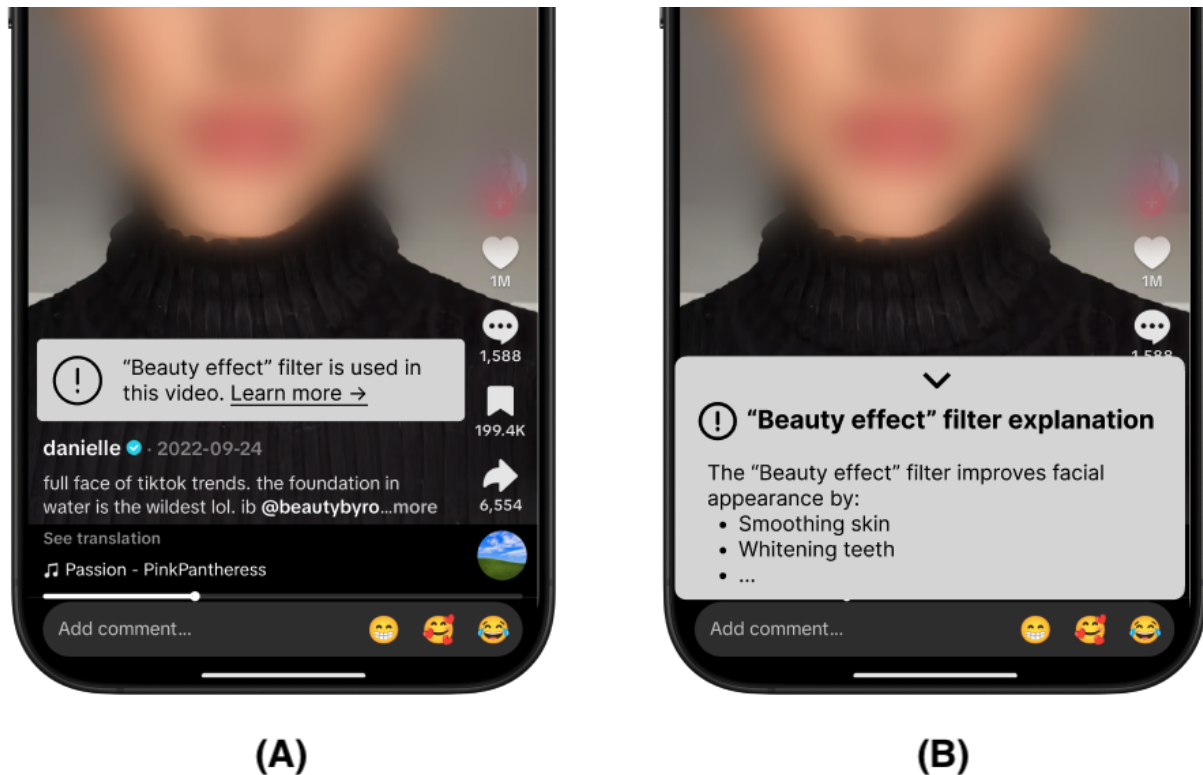


Figure 3: A DTL for TikTok that informs users about filter effects applied to the current video. The DTL alerts about the presence of a filter effect and presents minimal information (A), while tapping it opens a an explanatory panel with in-depth information about the filter (B).

frameworks. DTLs could adapt to users’ age by employing communication styles and visual designs that are appropriate and more effective for younger audiences. For example, social media platforms such as TikTok allow users to apply AI-based beauty filters to their photos and videos. Because such filters may promote unrealistic beauty standards, platforms could implement DTLs that inform when a filter has been applied and explain how the image has been altered. By explicitly distinguishing between unmodified and enhanced content, these labels could support media literacy and help minors better differentiate between reality and digitally manipulated representations. An example of such a DTL is depicted in Figure 3.

5. Conclusion

This paper introduced the concept of Design Transparency Labels (DTLs): contextual, interface-level notices designed to make persuasive design strategies visible, understandable, and steerable. As digital well-being gains societal and regulatory attention, DTLs can help align business incentives with user interests. By disclosing persuasive mechanisms, platforms can signal accountability and build trust, without having to redesign interfaces from scratch or abandoning engagement-driven features completely. At the same time, users can gain greater awareness and agency over their digital time. Finally, we outlined how the DTL framework can support the operationalization of existing and emerging regulatory frameworks.

Declaration on Generative AI

During the preparation of this work, the authors used Gemini for grammar and spelling check. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication’s content.

References

- [1] C. M. Gray, Y. Kou, B. Battles, J. Hoggatt, A. L. Toombs, The dark (patterns) side of ux design, in: Proceedings of the 2018 CHI conference on human factors in computing systems, 2018, pp. 1–14.
- [2] A. Monge Roffarello, K. Lukoff, L. De Russis, Defining and identifying attention capture deceptive designs in digital interfaces, in: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, 2023, pp. 1–19.
- [3] A. Mathur, M. Kshirsagar, J. Mayer, What makes a dark pattern... dark? design attributes, normative considerations, and measurement methods, in: Proceedings of the 2021 CHI conference on human factors in computing systems, 2021, pp. 1–18.
- [4] T. Naheyan, K. Oyibo, The effect of dark patterns and user knowledge on user experience and decision-making, in: International Conference on Persuasive Technology, Springer, 2024, pp. 190–206.
- [5] E. Commission, The digital services act package, 2022. URL: <https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package>, accessed: 2026-02-10.
- [6] European Commission, Artificial intelligence act, 2024. URL: <https://artificialintelligenceact.eu/>, accessed: 2026-02-10.
- [7] U. Lyngs, K. Lukoff, P. Slovak, R. Binns, A. Slack, M. Inzlicht, M. Van Kleek, N. Shadbolt, Self-control in cyberspace: Applying dual systems theory to a review of digital self-control tools, in: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, CHI '19, Association for Computing Machinery, New York, NY, USA, 2019, p. 1–18. URL: <https://doi.org/10.1145/3290605.3300361>. doi:10.1145/3290605.3300361.
- [8] C. Burr, M. Taddeo, L. Floridi, The ethics of digital well-being: A thematic review, Science and Engineering Ethics (2020) 2313–2343. doi:10.1007/s11948-020-00175-8.
- [9] A. M. Roffarello, L. De Russis, Achieving digital wellbeing through digital self-control tools: A systematic review and meta-analysis, ACM Transactions on Computer-Human Interaction 30 (2023) 1–66.
- [10] D. Peters, Wellbeing supportive design – research-based guidelines for supporting psychological wellbeing in user experience, International Journal of Human–Computer Interaction 0 (2022) 1–13. doi:10.1080/10447318.2022.2089812.
- [11] M. Eslami, A. Rickman, K. Vaccaro, A. Aleyasen, A. Vuong, K. Karahalios, K. Hamilton, C. Sandvig, “i always assumed that i wasn’t really that close to [her]”: Reasoning about invisible algorithms in news feeds, in: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI '15, Association for Computing Machinery, New York, NY, USA, 2015, p. 153–162. URL: <https://doi.org/10.1145/2702123.2702556>. doi:10.1145/2702123.2702556.
- [12] Y. Register, L. Qin, A. Baughan, E. S. Spiro, Attached to “the algorithm”: Making sense of algorithmic precarity on instagram, in: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, CHI '23, Association for Computing Machinery, New York, NY, USA, 2023. URL: <https://doi.org/10.1145/3544548.3581257>. doi:10.1145/3544548.3581257.
- [13] U. Ehsan, Q. V. Liao, M. Muller, M. O. Riedl, J. D. Weisz, Expanding explainability: Towards social transparency in ai systems, in: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, CHI '21, Association for Computing Machinery, New York, NY, USA, 2021. URL: <https://doi.org/10.1145/3411764.3445188>. doi:10.1145/3411764.3445188.
- [14] S. Arnórsson, F. Abeillon, I. Al-Hazwani, J. Bernard, H. Hauptmann, M. El-Assady, Why am I reading this? Explaining Personalized News Recommender Systems, in: M. Angelini, M. El-Assady (Eds.), EuroVis Workshop on Visual Analytics (EuroVA), The Eurographics Association, 2023. doi:10.2312/eurova.20231099.
- [15] R. Shang, K. J. K. Feng, C. Shah, Why am i not seeing it? understanding users’ needs for counterfactual explanations in everyday recommendations, in: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency, FAccT '22, Association for Computing Machinery, New York, NY, USA, 2022, p. 1330–1340. URL: <https://doi.org/10.1145/3531146.3533189>. doi:10.1145/3531146.3533189.

- [16] F. Poursabzi-Sangdeh, D. G. Goldstein, J. M. Hofman, J. W. Wortman Vaughan, H. Wallach, Manipulating and measuring model interpretability, in: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, Association for Computing Machinery, New York, NY, USA, 2021. URL: <https://doi.org/10.1145/3411764.3445315>. doi:10.1145/3411764.3445315.
- [17] Z. C. Lipton, The mythos of model interpretability, *Queue* 16 (2018) 31–57. doi:10.1145/3236386.3241340.
- [18] European Union, Regulation (eu) 2016/679 of the european parliament and of the council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/ec (general data protection regulation), *Official Journal of the European Union*, L 119/1, 4 May 2016, 2016. URL: <https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng>, accessed: YYYY-MM-DD.
- [19] M. Mitchell, S. Wu, A. Zaldivar, P. Barnes, L. Vasserman, B. Hutchinson, E. Spitzer, I. D. Raji, T. Gebru, Model cards for model reporting, in: *Proceedings of the Conference on Fairness, Accountability, and Transparency*, FAT* '19, Association for Computing Machinery, New York, NY, USA, 2019, p. 220–229. URL: <https://doi.org/10.1145/3287560.3287596>. doi:10.1145/3287560.3287596.
- [20] M. Arnold, R. K. E. Bellamy, M. Hind, S. Houde, S. Mehta, A. Mojsilović, R. Nair, K. N. Ramamurthy, A. Olteanu, D. Piorkowski, D. Reimer, J. Richards, J. Tsay, K. R. Varshney, Factsheets: Increasing trust in ai services through supplier's declarations of conformity, *IBM Journal of Research and Development* 63 (2019) 6:1–6:13. doi:10.1147/JRD.2019.2942288.
- [21] E. P. Bogucka, M. Constantinides, S. Šćepanović, D. Quercia, Impact assessment card: Communicating risks and benefits of ai uses, *Proc. ACM Hum.-Comput. Interact.* 9 (2025). URL: <https://doi.org/10.1145/3757482>. doi:10.1145/3757482.
- [22] Parliament of the United Kingdom, Digital markets, competition and consumers act 2024 (c. 13), <https://www.legislation.gov.uk/ukpga/2024/13/contents>, 2024. Accessed: 2025-08-26.
- [23] California Department of Justice, Office of the Attorney General, California consumer privacy act (ccpa), <https://oag.ca.gov/privacy/ccpa>, 2024. Accessed 2025-01-24.
- [24] Colorado Department of Law, Colorado Attorney General's Office, Colorado privacy act resource, <https://coag.gov/resources/colorado-privacy-act/>, 2025. Accessed: 2025-08-26.
- [25] Department of Consumer Affairs, Ministry of Consumer Affairs, Food & Public Distribution, Government of India, Consumer protection act(s), <https://consumeraffairs.gov.in/pages/consumer-protection-acts>, 2025. Accessed 2025-01-24.
- [26] M. Luria, “this is transparency to me” user insights into recommendation algorithm reporting, 2022.