# Digital Wellbeing by Design: Safeguarding End-Users' Time and Attention in UI/UX Practices

ALBERTO MONGE ROFFARELLO, Politecnico di Torino, Italy
LUIGI DE RUSSIS, Politecnico di Torino, Italy
MASSIMILIANO PELLEGRINO, Politecnico di Torino, Italy

User interfaces, particularly those of social media, promote mindless usage through Attention-Capturing Damaging Patterns (ACDPs) like infinite scroll and content autoplay, prioritizing user engagement over digital wellbeing. Recent work suggests that a promising approach to realigning technology with users' digital wellbeing is to prioritize it *by design*. Building on this idea, this paper explores how to support designers in identifying ACDPs and surfacing wellbeing-oriented alternatives. We first defined and assessed a multi-modal prompting technique to detect ACDPs in Figma prototypes and discover alternative designs that may promote mindful and active technology use, assessing its outputs with 4 experts in deceptive designs. Then, we integrated the prompting technique into a plugin for Figma, and conducted a formative lab study with 15 UI/UX designers to explore how such a tool may support early-stage reflection on ACDPs and its perceived fit with existing practices. Results suggest that the plugin can inspire designers to reconsider their choices early in the design process.

CCS Concepts: • **Human-centered computing** → *Graphical user interfaces*; **Interaction design**; **Interactive systems and tools**.

Additional Key Words and Phrases: digital wellbeing, attention-capture designs, large language models, design tools, Figma

**ACM Reference Format:**
Alberto Monge Roffarello, Luigi De Russis, and Massimiliano Pellegrino. 2025. Digital Wellbeing by Design: Safeguarding End-Users' Time and Attention in UI/UX Practices.  1, 1 (April 2025), 32 pages. https://doi.org/XXXXXXX.XXXXXXX

## 1 INTRODUCTION

Despite the well-documented issues associated with excessive technology use, such as sleep disturbances [35] and impaired self-control [46], tech companies continue to operate within an Attention Economy in which design patterns such as infinite scroll and content autoplay are used to maximize time-on-platform and interaction frequency. These Attention-Capture Damaging Patterns (ACDPs) [57] can be considered as instances of "dark patterns[1]"—design decisions that go against

[1]We recognize that the term "dark pattern" has raised concerns for linking "dark" with harm [2]. While alternative terms such as manipulative design and damaging patterns have been suggested, the field has not yet converged on a widely accepted replacement. In the interest of consistency with existing literature, and with awareness of this broader discussion, we continue to use the term "dark pattern."

Authors' addresses: Alberto Monge Roffarello, Politecnico di Torino, Corso Duca degli Abruzzi, 24, Torino, TO, 10129, Italy, alberto.monge@polito.it; Luigi De Russis, Politecnico di Torino, Corso Duca degli Abruzzi, 24, Torino, TO, 10129, Italy, luigi.derussis@polito.it; Massimiliano Pellegrino, Politecnico di Torino, Corso Duca degli Abruzzi, 24, Torino, TO, 10129, Italy, massimiliano.pellegrino@polito.it.


XXXX-XXXX/2025/4-ART $15.00
https://doi.org/XXXXXXX.XXXXXXX

users' best interests [28]—that specifically contribute to a digital environment that prioritizes users' engagement over digital wellbeing [7, 51].

Traditionally, achieving digital wellbeing has been considered an end-user responsibility [17], with HCI researchers and practitioners that have focused on providing users with Digital Self-Control Tools (DSCTs) that help users exercise self-control by self-monitoring tech use and setting up interventions on distractive apps or websites [45, 64]. Yet, a growing body of work argues that addressing ACDPs in the Attention Economy requires moving beyond individual self-control and end-user-centric perspectives toward prioritizing harm reduction through the recognition, mitigation, and prevention of negative user impacts [16, 17]. In this vein, HCI researchers has advocated considering digital wellbeing as a critical design dimension [55–57], aiming to promote meaningful online experiences without relying on external assistance.

Building on this growing body of work, this paper explores how raising awareness of ACDPs during early-stage and iterative UI/UX design activities may shape designers' decisions, and examines the broader implications of introducing such feedback into the design process. To this end, we designed, implemented, and evaluated a plugin for Figma—one of the most popular tool for UI/UX design [40]—that leverages a Large Language Model (LLM) and functions as a 'digital-wellbeing assistant' that surfaces potential ACDPs and prompt consideration of alternative designs while leaving interpretation and action to designers. By design, the plugin does not enforce decisions, but rather externalizes normative concerns early in the workflow, creating space for designers to exercise judgment and negotiate trade-offs.

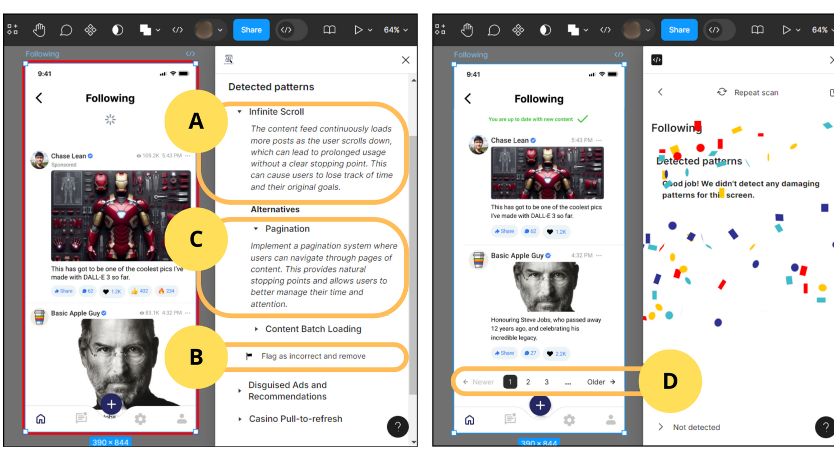



Fig. 1. A walkthrough of our plugin. Designers working on Figma can scan their UI prototypes to receive feedback on attention-capture designs and their potential risks for end users (e.g., *infinite scroll*, A). After a scan, they can provide feedback to refine future detections (*flag as incorrect*, B). Meanwhile, the plugin provides suggestions for alternative designs (e.g., *pagination*, C), allowing designers to achieve similar functionality while prioritizing end-users' digital wellbeing (D).

With the plugin, designers working on Figma can scan their UI mockups at any phase of the design process, receiving feedback on potential ACPDs included in the interfaces and their potential harms (Figure 1, A). To improve the LLM's performance, designers can provide feedback on each detection, which is integrated into the model for subsequent scans (Figure 1, B). In parallel, the plugin provides designers with alternative designs to be considered for replacement (Figure 1, C). By following these suggestions, the designer can choose to avoid ACDPs while maintaining similar functionality, thus re-aligning the interface with users' digital wellbeing (Figure 1, D).

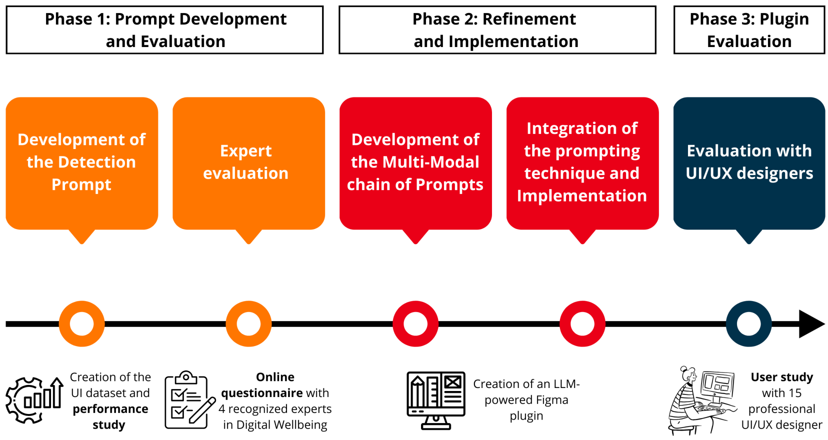



Fig. 2. Our work consisted of a three-phases process. We first developed and evaluated a prompting technique to detect ACDPs within Figma prototypes and suggest design alternatives (Phase 1). We then refined and implemented the technique into an LLM-powered Figma plugin (Phase 2). Finally, we conducted a user study to assess our approach with UI/UX designers (Phase 3).

Figure 2 summarizes all the phases of our work. In Phase 1, we developed a detection prompt to make an LLM detect ACDPs within Figma prototypes and suggest design alternatives. We tested two different input modalities for the prompt—JSON representation of the UI prototype and corresponding image—by evaluating performance on a UI dataset of 10 Figma prototypes built through examples taken from HCI papers on dark patterns. The best-performing detection prompt showed improved accuracy when using images of the UIs compared to their JSON representations. We also evaluated its the accuracy and helpfulness of suggested alternative designs with four recognized experts in dark patterns. Experts found the plugin's detections and alternative suggestions to be generally accurate and helpful. However, they emphasized the importance of understanding the context beyond static designs and having clear indications of where the detections are located within the interfaces to accurately assess the severity of identified ACDPs. Furthermore, they warned that some alternatives could unintentionally increase attention capture or fail to reduce cognitive burden due to habituation to the countermeasures.

Building on the experts' feedback, we refined the detection prompt to highlight detections and suggested alternatives directly within the UI mockups by developing a multi-modal chain of prompts that leverages both images and JSON representations, which we then integrated into the Figma plugin shown in Figure 1 (Phase 2).

Finally, in Phase 3, we conducted a formative lab study involving 15 professional UI/UX designers. Importantly, our evaluation did not aim to validate the effectiveness of a production-ready tool, but rather to explore the potential of our approach as an early-stage design probe. Specifically, we used pre-built UI mockups to investigate how designers interpret and engage with the plugin's wellbeing-oriented feedback within a controlled setting. During the experiment, we challenged participants to discover the plugin's functionality in a free-usage 10-minute session, and we asked them to evaluate the plugin's feedback across a range of metrics, including the impact of implementing the suggested alternative designs on usability. Like the experts, the designers also appreciated the helpfulness of the detections and alternative suggestions. Yet, they noted that implementing some of the proposed alternatives may conflict with engagement-driven company goals, suggesting that the plugin's primary value lies in sparking organizational discussion about design trade-offs rather than prescribing specific design changes.

Building on our work, its phases, and related findings, we discuss the role of designers in contextualizing and mediating the plugin's suggestions, and we advocate for a new regulatory landscape that recognizes the harms caused by ACDPs to promote digital wellbeing through design.

To summarize, our work makes the following contribution:

- A multi-modal prompting technique to identify ACDPs in UI mockups and obtain alternative design recommendations that support mindful and active technology use.
- A Figma plugin that incorporates an LLM and the developed multi-modal prompting technique to assist designers in prioritizing digital wellbeing while designing user interfaces.
- An exploratory evaluation of the prompting technique and the developed plugin through expert feedback and a formative study with UI/UX designers, using pre-built mockups as a design probe to investigate how such feedback is interpreted and negotiated in practice.
- An exploration of designers' perceptions regarding the plugin's compatibility with existing design workflows and how legal endeavors could complement the approach.

## 2 RELATED WORKS

This section contextualizes our work by summarizing research on tools for digital self-control, recent calls for designing for digital wellbeing, and attempts to leverage generative AI for design.

### 2.1 Moving Beyond Self-Control in the Attention Economy

Research on excessive technology use has increasingly framed digital wellbeing, a new psychological construct that stresses the importance of creating technology that promotes mindful and meaningful interactions online [7, 51], in terms of individual self-regulation, emphasizing constructs such as addiction [36], compulsive use [69], and over-engagement [68]. In such a complex landscape, Digital Self-Control Tools (DSCTs) can help end users improve their digital wellbeing [45, 51, 64]. These dedicated mobile apps or browser extension, available as either commercial tools or research-based artifacts, provide users with a range of intervention strategies, from self-monitoring supports and usage timers [45, 51] to the possibility of creating digital nudges [61] and even customize the layout of interfaces to reduce distractions [42].

Despite their widespread availability, DSCTs have faced criticism due to their reliance on external interventions [42, 64], which can lead to high dropout rates [33] and limited effectiveness over the long term [51]. Furthermore, researchers have criticized the reductionist approaches of equating digital wellbeing with user engagement metrics and individual self-control responsibilities, upon which DSCTs are largely based [16, 17]. For example, the notion of screen time has been widely mentioned as insufficient for fully understanding and designing for digital wellbeing [41]. In parallel, Docherty and Biega [17] highlighted that reducing engagement may not be sufficient if one does not address "the deeper psychosocial problems causing problematic engagements in the first place," as digital wellbeing is part of a larger, systemic problem deeply embedded in the operations of platform capitalism. Paradoxically, these research works suggest that the same tech companies now promoting digital wellbeing through self-control initiatives, e.g., Google [25], may do so strategically to prevent stronger political and social interventions while simultaneously extracting even more data from users.

Overall, the systemic issues underlying these critiques are rooted in the Attention Economy [12], a framework that describes how digital platforms compete for and monetize user attention. This model is deeply intertwined with digital and surveillance capitalism [76], where user data is extracted, analyzed, and sold to sustain profit-driven ecosystems, and has led to the proliferation of the so-called Attention-Capture Damaging Patterns (ACDPs) [57], i.e., dark patterns that exploit users' psychological vulnerabilities to increase engagement metrics like time spent and frequency of daily visits. For example, infinite scrolling on social media interfaces [50, 63] and the ability to compulsively refresh a page (the so-called Casino Pull-to-refresh [57]) create a variable-reward system that can trap users in a cycle of continuous, passive consumption. Similarly, "guilty-pleasure" recommendations [6] and the neverending autoplay of the next video [43] contribute to an ongoing stream of content to be consumed, while advertisements and recommendations disguised as regular content within social network newsfeeds can mislead users into clicking on content they might not have otherwise engaged with [27]. As a specific category of dark patterns [28], ACDPs can harm users regardless of whether they are introduced deliberately or arise unintentionally through established design practices. Although some ACDPs may be intentionally deployed to influence user behavior, others can emerge from habituation, the imitation of industry norms, or organizational pressures. Our work is situated within this discourse, recognizing the pervasive influence of platform capitalism and attention economy on modern digital ecosystems and their role in fostering harmful design practices.

## 2.2 Designing for (Digital) Wellbeing

So how should we support digital wellbeing in the Attention Economy? Compared to DSCTs, the other end of the spectrum is perhaps the notion of 'non-design' proposed by Baumer and Silberman [3], which advocates for rethinking the necessity of designing certain applications or features altogether, questioning whether their removal might better serve societal wellbeing. Situated between these approaches, our work focuses on managing negative externalities and reducing harm, one of the potential alternative strategies proposed by Docherty and Biega [17]. Specifically, we aim to support designers in reflecting on and negotiating alternatives to ACDPs that better align with principles of digital wellbeing, without necessarily reducing user engagement but rather promoting mindful and active technology use. This perspective aligns with recent calls to prioritize the user's sense of agency as a key lens for digital wellbeing, offering a more nuanced approach that moves beyond the oversimplification of screen-time metrics.

The concept of integrating wellbeing into technology design is not a novel idea; rather, it has been a topic of discussion among designers and researchers for several years. Positive design [13], for example, emphasizes creating products and experiences that enhance users' quality of life. This approach seeks to foster emotional and psychological benefits, encouraging designers to consider the broader impact of their work on individuals and communities. Similarly, Value-Sensitive Design [24] focuses on incorporating ethical considerations and human values into the design process, ensuring that the outcomes align with the needs and aspirations of users.

More recently, researchers have begun calling for a focus on designing technology that inherently respects users' digital wellbeing by preserving their attention and time [55, 56, 64]. Docherty and Biega [17], in particular, called for making *"the normative stakes and ethical implications of digital wellbeing explicit at every stage of the design process"* (p. 2). In this context, Peters [59] introduced the concept of Wellbeing Supportive Design, developing guidelines that emphasize respect for user autonomy and support for informed and meaningful choices. Similarly, Monge Roffaerllo et al. [54] proposed eight design heuristics to create UIs that preserve user attention by design. Through a different approach, the workshop titled "Designing for Meaningful Interactions and Digital Wellbeing" brought together academic and practitioners to explore how to create UIs that encourage intentional and significant engagement with technology [56]. To our knowledge, however, there has been only one prior work that proposed a functional tool to help designers prioritize digital wellbeing and address ACDPs [55]. Nevertheless, the Figma plugin introduced by the authors employs a rather simplistic method, which constrains the applicability of the approach to a specific design template and use case. Our work extends this idea by investigating whether and how UI/UX designers can collaborate with Large Language Models (LLMs) to get digital-wellbeing feedback at any stage of their design activities.

Overall, while prior works on designing for digital wellbeing emphasize the need for new tools and practices, they also acknowledge that such interventions operate within broader economic and organizational constraints [55], suggesting that tools alone cannot resolve tensions rooted in the attention economy [64]. In line with this evidence, we position our work as a practical instrument that enables designers to surface, articulate, and negotiate design decisions related to users' digital wellbeing within organizations. Rather than resolving the structural conflict between engagement-driven business models and user wellbeing, the tool helps make this tension explicit and discussable in design practice. We return to this point in Section 7.

### 2.3 AI-Assisted Design Interventions

In 2024, several commercial design platforms have started to announce new features and tools based on Generative AI and LLMs. Diagram [15]—a startup recently acquired by Figma [39]—developed a suite of AI-powered design tools for Figma, e.g., to create fully-editable UI designs starting from simple prompts. Meanwhile, members of the Figma community have used Figma's APIs to create hundreds of AI-powered plugins [39], and Figma itself recently announced Figma AI [62], a collection of AI-driven functionalities, including text rewriting, automatic layer naming, and the creation of new design mockups from textual descriptions. Similarly, FigmaJam AI [23] is an online collaborative whiteboard that incorporates generative AI to assist in creating templates and charts for weekly team syncs and summarizing the content of group brainstorms. Similar tools nowadays exist on other popular design platforms like Canva (Magic Design [8]) and Microsoft Designer [48].

In parallel, HCI researchers have experimented a wave of novel tools and techniques to support UI/UX designers at various stages [40]. Lawton et al. [37] presented Reframer, an interface that allows designers to collaborate with a Generative AI in real time to produce drawings collaboratively. PromptInfuser by Petridis et al. [60] is a Figma plugin for adding AI functionality to UI mockups with LLM prompts. A range of tools has also been proposed to assist designers with AI-generated feedback. Xiang et al [74] developed SimUser, a LLM-based tool that generates heuristic usability feedback for various user demographics by emulating the engagement between users and mobile applications within a defined context. Similarly, Duan et al. [18] investigated the potential of using LLMs to conduct automatic heuristic evaluations. They developed a Figma plugin that exploits GPT-4 to validate Figma prototypes against pre-defined heuristics, such as Nielsen's, and generate constructive feedback for improving the usability of UI mockups. Kuang et al. [34], instead, created a proactive conversational assistant aimed at aiding designers by providing automatic recommendations during usability testing. Most closely related to our approach, Kocyigit et al.'s DeceptiLens system demonstrates the potential of multimodal LLMs for detecting deceptive design patterns [32]. While DeceptiLens focuses on enhancing accountability and trust in automated detection, our work extends this direction by embedding a similar detection pipeline into designers' workflows and complementing it with actionable design alternatives for supporting digital wellbeing.

In general, various research efforts aimed at assessing these design feedback instruments have consistently shown that they can successfully assist designers during the prototyping stage [18, 34, 74], with designers that are willing to accept occasional imperfect suggestions to iteratively improve their mockups [18]. Yet, all these studies underscored *"the irreplaceable role of human expertise"* (Kuang et al [34], p. 1). Building on this insight, rather than attempting to automate such judgments, our work explores how LLM-based systems can function as in-the-loop supports that surface potential issues and raise awareness on ACDPs, while leaving interpretation and decision-making firmly in human hands. Specifically, we adopt a constructive negotiation approach, through which *"the machine should push back, constructively negotiate with the human to consider design aspects that they had not yet anticipated, whether in a design's functionality, form, or anticipated reception"* (Vaithilingam et al. [70], p. 291).

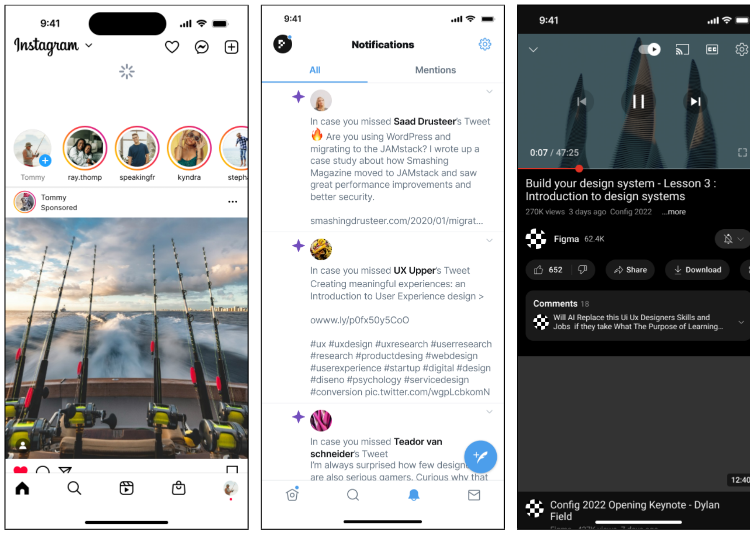
## 3 DEVELOPMENT OF THE DETECTION PROMPT

The first phase of our work consisted in identifying an effective approach for detecting and addressing ACDPs within UI mockups. Building on recent works on AI-assisted design [18, 40], we leveraged LLMs to develop a detection prompt capable of identifying ACDPs in Figma prototypes and suggesting alternative designs aimed at promoting mindful technology use. As smartphones have been identified as major contributors to undermining users' sense of agency and self-control [44, 69], we specialized the prompting technique with mobile UIs. Nevertheless, the approach can be easily generalized to any kind of user interface.

### 3.1 LLM Selection and Dataset Creation

As an LLM, we decided to use OpenAI GPT-4o [58]. We made this choice based on the results provided by Duan et al. [18] in 2024, which demonstrate that the fourth generation of OpenAI's GPT model outperforms other LLMs in analyzing UIs and characterizing their designs. Compared with the version used in [18], GPT-4o is multi-modal, meaning that it can handle text, audio, images, and video, and it has an increased context window. Similar to the work of Duan et al. [18], which tasked GPT with discovering violations of specific design heuristics, we selected a state-of-the-art taxonomy [57] of 11 ACDPs to be used as a reference for the LLM. To the best of our knowledge, this is the only taxonomy of dark patterns specifically focused on the users' attention. It encompasses deceptive designs, which use various forms of deception to capture the user's attention (i.e., Disguised Ads and Recommendations, Attentional Roach Motel, Time Fog, and Fake Social Notifications), as well as seductive designs, which tempt the user with short-term satisfaction (i.e., Infinite Scroll, Casino Pull-to-Refresh, Neverending Autoplay, Guilty Pleasure Recommendations, Recapture Notifications, Playing by Appointment, and Grinding).

To assess the capabilities of the LLM in detecting ACDPs and develop an effective detection prompt, we created a dataset consisting of 10 Figma prototypes of mobile UIs, each containing a set of ACDPs. To build it, we recreated in Figma 10 screens of mobile applications, from Instagram to TikTok, included as examples in HCI papers discussing ACDPs. We used the systematic literature review in [57] as a starting point for the examples, and the Figma community website to find UI templates of real-world applications. Figure 3 shows three examples from the dataset. The Instagram newsfeed shown in Figure 3(a) features Infinite Scroll [57, 63], Casino Pull-to-refresh [53], and Disguised Ads and Recommendations [10, 57], through which sponsored pages are camouflaged as regular content. Figure 3(b), instead, exemplifies a clear example of Recapture Notifications on X [75], i.e., notifications that are deliberately sent to recapture users' attention and have them start a new usage session [57]. Finally, the YouTube video screen shown in Figure 3(c) uses the Neverending Autoplay pattern [43] and displays Guilty-Pleasure Recommendations below the video [4, 57] through Infinite Scroll. To foster replicability, we have released the dataset of 10 Figma prototypes on OSF (https://osf.io/smzuv/files/?view_only=54667f6db28547049a9a9a3830c9bdad). We emphasize that this dataset, which was also used for the expert evaluation (Section 4) and the experiment with UI/UX designers (Section 6), is not intended to capture the full complexity, ambiguity, and contextual variability of real-world interfaces and to be representative of the full design space of contemporary applications. In line with prior work evaluating AI-based design feedback systems [18], it serves as a small, theory-grounded probe set that enables controlled comparison of prompt configurations and consistent evaluation across experts and designers.




(a) Instagram (b) X (c) YouTube

Fig. 3. Three examples of Figma prototypes included in our UI dataset. We recreated them based on examples included in papers discussing ACDPs, e.g., [57]

## 3.2 Development of the Prompt and Performance Study

We developed a ***detection prompt*** for GPT-4o to detect ACDPs within mobile UI mockups and suggest design alternatives. We refined the prompt through a performance study in which we tested various configurations, leveraging the UI dataset described in Section 3.1 as a ground truth[2].

Figure 4 shows the structure of the prompt with its variations. Specifically, we tested two different input modalities to send the UI mockup to GPT-4o, leveraging its multi-modal capabilities:

- **JSON (Figure 4, A):** we generated a JSON representation of the mockup using the Figma APIs, compressing it with a method similar to the one outlined in [18]. Specifically, the exploited JSON format captures the DOM structure of the UI mockup, thereby representing its overall layout. Each node, then, contains both semantic information (e.g., labels and element type) and layout data (i.e., <x,y> coordinates, height, and width) of the related UI element.
- **IMAGE (Figure 4, B):** we took a screenshot of the mockup using the Figma APIs, visually capturing the design of the mobile app in a single image.

We also instructed GPT to answer with a specific JSON object with an array of ACDP detections, with each item containing a reference to the detected ACDP and the list of alternative designs (Figure 4, C).

Besides testing the input modality, we evaluated the detection prompt by varying the temperature, a parameter in GPT's APIs that controls the creativity of the model's responses. In particular, we tested the detection prompt with *low temperature* (0, resulting in more deterministic and predictable answers), *medium temperature* (0.5, providing a balance between predictability and creativity), and *high temperature* (1, yielding responses that are diverse and creative).

[2]The final version of the detection prompt—refined after the evaluation with experts in damaging designs—is available at https://osf.io/smzuv/files/?view_only=54667f6db28547049a9a9a3830c9bdad

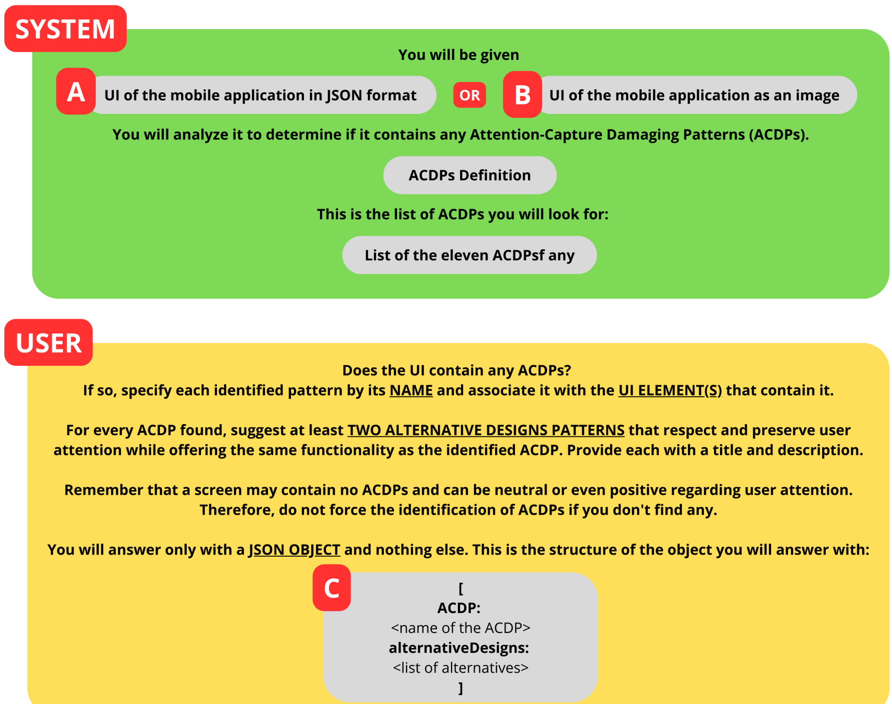



Fig. 4. The structure of the detection prompt, including system messages to set the context for the model and user messages representing the actual task for the model. We experimented different configurations to optimize GPT-4's ability to identify ACDPs within UI mockups.

Overall, this resulted in a performance test with a 2x3 configuration, combining 2 input modalities (JSON, image) with 3 temperature settings (0, 0.5, 1). As in similar studies, e.g., the performance test reported in [18], we focused on the underlying detection task—in our case, the identification of ACDPs within a UI—while leaving the the subjective evaluation of design alternatives to experts in dark patterns (Section 4) and designers (Section 6). Table 1 summarizes the outcomes of the test in terms of precision, recall, and F1 score. Importantly, these performance metrics should be interpreted in light of the controlled nature of the evaluation setup. Because the dataset consists of UI mockups derived from canonical examples of ACDPs and the taxonomy itself is provided to the model as part of the prompt, the task is closer to recognizing prototypical instances of known patterns than to detecting context-dependent cases in real-world settings. Rather than providing a benchmark of real-world detection performance, these results should be therefore understood as a diagnostic assessment of the prompt's behavior under controlled conditions, used to guide design decisions for the subsequent prototype.

Within the constrained corpus of mockups included in the dataset, results show that GPT-4o scored higher on all three metrics when dealing with an image of the UI rather than its JSON representation. This is a clear step beyond the work of Duan et al. [18], which tested the capabilities of GPT-4 in identifying heuristic violations in UIs represented solely as JSON due to the limitations of the LLM at that time, and demonstrates the potential of using LLMs to find design issues while analyzing images. Table 1 also shows that increasing the temperature generally reduced precision, recall, and F1, especially with the IMAGE input modality. Our objective was to explore whether

Table 1. Outcomes of the performance test in terms of precision, recall, and F1 score. Results are the average of five equals experiments.

| Temperature/Modality | Precision | Recall | F1 |
|---|---|---|---|
| 0/JSON | 0.550 | 0.647 | 0.595 |
| 0.5/JSON | 0.562 | 0.529 | 0.545 |
| 1/JSON | 0.550 | 0.687 | 0.611 |
| **0/IMAGE** | **0.867** | **0.765** | **0.812** |
| 0.5/IMAGE | 0.846 | 0.647 | 0.733 |
| 1/IMAGE | 0.833 | 0.588 | 0.690 |

increased creativity could help the model identify more subtle patterns, such as seductive ones, and generate a wider range of alternative designs. Such an advantage emerged in certain circumstances, particularly in detecting subtle patterns. For example, a higher temperature allowed GPT-4o to identify the use of Guilty-Pleasure Recommendations in the YouTube video UI (Figure 3(c)), while lower temperatures resulted in a missed detection. However, a manual inspection of the alternative designs across all temperatures showed no significant differences in the number or quality of the suggestions. Furthermore, our results demonstrate that a higher temperature also led to a higher number of false positives and resulted in more diverse, less consistent choices. Although further evaluation on larger and more diverse datasets would be needed to establish broader performance claims, the performance test served to guide our choice of an IMAGE input modality with a low temperature (0) for the detection prompt.

## 4 EXPERT EVALUATION

To test and refine the developed detection prompt before integrating it into a Figma plugin, we evaluated its outputs with four experts in dark patterns, focusing primarily on detection accuracy and the perceived severity of the identified ACDPs.

### 4.1 Methods

*4.1.1 Participants.* We recruited four recognized experts in dark patterns through the "Dark Patterns Research and Impact" Slack workspace. Table 2 summarizes the demographics of the involved experts[3], including their gender, age, occupation, years of experience working with dark patterns, and primary area of expertise. Overall, we covered various fields of expertise related to dark patterns, including manipulative practices in specific countries (Katie) and applications (Thomas), the impact of dark patterns on vulnerable populations (Lorena), and legal frameworks (Kosha).

*4.1.2 Procedure and Measures.* To give participants enough time to complete the evaluation, we allowed them to perform the assessment asynchronously via an online questionnaire[4]. We precomputed the output of the detection prompt, i.e., detections of ACDPs and suggestions for alternative designs, using the UI dataset described in Section 3.1. This computation, which yielded identical results to the performance test (precision of 0.867, recall of 0.765) except for minor discrepancies in textual descriptions, identified 15 ACDPs across the 10 UIs in the dataset, and resulted in two alternative designs suggested for each UI. Due to time constraints, we divided the

[3]Experts agreed to be cited for their contribution. We decided to remove their names in this version of the manuscript to avoid introducing biases during review, and will disclose them if the paper is accepted.
[4]A sample version of the questionnaire used in the expert survey is available at https://osf.io/smzuv/files/?view_only=54667f6db28547049a9a9a3830c9bdad

Table 2. The experts on dark patterns who evaluated the plugin's detection of ACDPs and suggestions for alternative designs.

| Expert | Gender | Age | Occupation | Experience | Area of Expertise |
|---|---|---|---|---|---|
| **Katie** Seaborn | Gender-apathetic woman | 39 | Associate Professor | 3-5 years | Dark patterns in Japanese apps |
| **Lorena** Sánchez Chamorro | Woman | 30 | Doctoral Researcher | 3-5 years | Dark patterns and vulnerable populations |
| **Thomas** Mildner | Man | 31 | Postdoctoral Researcher | 3-5 years | Dark patterns in social media |
| **Kosha** Doshi | Woman | 22 | Student | 1-2 years | Ethical design practices, member of the *deceptive.design.org* team |

10 UIs into two groups, ensuring that each expert reviewed the plugin outputs for 5 UIs within an estimated timeframe of 90 minutes. When dividing the UIs, we ensured a balanced distribution of complexity and variety in ACDPs and suggestions across groups to minimize potential evaluation bias. Specifically, both groups included similar number of detections (7 and 8, respectively) and shared detections of popular ACDPs such as infinite scroll and neverending autoplay. Each group also included other patterns and collectively covered all the eleven ACDPs defined in [57]. Each UI was presented with a screenshot and the output of the detection prompt for that UI, consisting of a list of ACDP detections along with suggested alternative designs for each detection, in its original textual format. For each detection, in particular, we asked experts to evaluate:

- The **accuracy** of the detection—defined as whether the detected ACDP matched the experts' expectations, on a Likert scale from 1 (*not at all accurate*) to 5 (*very accurate*).
- The **severity** of the detected pattern, on a Likert scale from 1 (*not at all dangerous for end users* to 5 (*very dangerous for end users*).
- The **helpfulness** of the two alternative designs suggested by the LLM (A1 and A2), on a Likert scale from 1 (*not at all helpful*) to 5 (*very helpful*).

Furthermore, experts were encouraged to **explain their rating** if they wished to provide additional clarification. Overall, these metrics represent well-established criteria for evaluating design issues in UIs and assessing the quality of suggestions generated by LLMs [18].

*4.1.3 Data Analysis.* We collected all expert ratings on accuracy, severity, and helpfulness. For each measure, we calculated the total number of responses per rating, as well as the median and interquartile range (IQR). We support quantitative results with the optional explanations that experts provided.

### 4.2 Results

Figure 5 shows the number of answers by experts for each rating of detection accuracy, pattern severity, and helpfulness of the two suggested alternative designs. Overall, the survey included 15 detections divided into two groups, with each group evaluated by two participants, resulting in 30 responses for each metric.

*4.2.1 Accuracy of the Detections.* Most detections (70%) were rated as accurate (8) or very accurate (13), with a median rating of 4 (IQR = 2). Two experts identified patterns that were not captured by the plugin, with Thomas suggesting 5 additional patterns and Katie suggesting 10 overall. Some

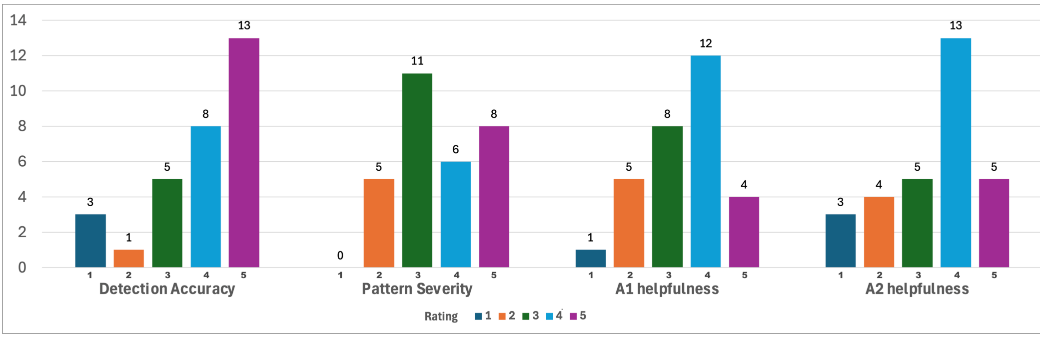



Fig. 5. The total number of expert responses for each rating of detection accuracy, pattern severity, and helpfulness of the two proposed alternative designs.

of these additional patterns matched those we labeled in our ground truth dataset, previously identified as misses in our performance test. Others could not be directly observed in the UIs but still made sense when considering the underlying platform and its features. In this respect, experts highlighted the **limitations of static screenshots** in accurately detecting dynamic patterns, emphasizing the need for contextual elements and platform-specific knowledge. As Thomas said while evaluating Infinite Scroll on the X timeline, for example, having just the screenshot would not necessarily allow the detection of the pattern, as any potential friction or interventions would not be displayed. This aligns with recent work that has framed issues of detection of certain dark patterns around the temporality of experience [29]. In these cases experts often relied on **prior familiarity with the platforms** to interpret the screenshots, which also introduced potential ambiguity and uncertainty:

> *"I am not sure we can deduce that from the screenshot, but as long as I am aware, BeReal does not have infinite content because it only gives you the content of your network..."* (Lorena, Infinite Scroll on BeReal).

Furthermore, experts **critically evaluated** some of the detections, often making assumptions or expressing uncertainty when evidence was unclear or ambiguous:

> *"This is just gamification...I don't get any sense of urgency or requirement to 'grind' for any specific thing"* (Katie, Grinding of Profile Badges).

*4.2.2 Severity of the Detected ACDPs.* In terms of perceived severity of the detected ACDPs, Figure 5 shows that expert responses were more evenly distributed across the ratings (median = 3, IQR = 1.75).

Experts highlighted that the severity is highly **context-dependent**, with the impact varying based on content type, platform, and user context:

> *"In some cases, depending on the application, the rewards coming from grinding can cause users the need of conducting them. However, the severity of this ACDP is contextual of the application, mainly depending on the reward"* (Lorena, Grinding of Profile Badges).

Experts highlighted the potential for **psychological and attentional harm** for most of the detected patterns, particularly when they are designed to subtly manipulate or deceive users. However, there was also **ambiguity and uncertainty in assessing the severity**, with some participants noting difficulty in judging the harm without more context. As Katie said, for example:

> *"It depends on the advertisement...yes, you can trick people, but to what end? People click through and realize or not <– that's where the deception detection is truly needed"* (Katie, Disguised Ads and Recommendations on Instagram feed).

Additionally, experts recognized that ACDPs often act as **catalysts for further harm**, particularly when combined with other manipulative patterns, leading to more significant negative outcomes for users. In evaluating Recapture Notifications on X, for example, Lorena highlighted that the pattern may become the entry point for other ACDPs or interactions that will cause more attentional and psychological harm.

*4.2.3 Helpfulness of the Design Alternatives.* In line with the accuracy results, suggestions for alternative designs have been found to be more helpful that not on average, with a median rating across the two suggestions of 4 (IQR = 1). As shown in Figure 5, the first suggestion (A1) was rated helpful 12 times and very helpful 4 times (54% of responses), while the second suggestion (A2) received 13 helpful and 5 very helpful ratings (60% of responses).

Experts generally appreciated the emphasis on **clear communication and transparency**, such as "you are all caught up" message to be used as an alternative for infinite scroll (Thomas) or the proper labeling of ads instead of disguising them in the UI (Lorena, Katie, and Koshi). However, they also highlighted potential drawbacks, noting that some alternatives could **inadvertently increase attention capture** or fail to reduce cognitive burdens due to **habituation** to the countermeasure. According to Lorena, for example, micro-frictions like 'load more' buttons could act as pull-to-refresh, not reducing harm, while Katie highlighted that summarizing content could increase attention capture rather than reduce it. However, Lorena pointed out that design alternatives in the form of friction might have a long-term impact:

> *"It is possible that the habituation still works against the mechanism [...] so I am not certain this would totally prevent users from trying to refresh. Again, in the short term might be frustrating, while in the long term might work"* (Lorena, Content-Batch Loading – load new content in batches and inform users when they have reached the end of the new content).

In general, providing end users with **control and customization** options was seen as crucial by the experts, although they emphasized that the impact of these design changes might be limited if not thoughtfully integrated.

## 5 REFINEMENT AND IMPLEMENTATION

In the second phase of our work, we used the feedback collected in the expert evaluation to refine our detection prompt (Section 5.1). Then, we integrated the refined prompting technique into a plugin for Figma, one of the most popular tools for UI prototyping [40] (Section 5.2).

### 5.1 Multi-Modal Chain of Prompts

Drawing on the expert evaluation, we modified the detection prompt along two directions: (i) incorporating designers' feedback to enable refined recommendations, and (ii) enabling the prompt to directly highlight detected ACDPs within the UI. For the first refinement, we explored the possibility of relaunching the prompt while excluding specific ACDP detections, such as those flagged as incorrect by designers (Figure 6, B). For the second, instead, we added explicit instructions to associate detections with specific UI elements (Figure 6, A) by prompting GPT to approximate the layout data (x, y, width, and height) of these elements from the mockup image.

Unfortunately, we discovered that just with a screenshot, the approximation of the visual elements that originated the ACDPs was unreliable. To address this issue, we devised a multi-modal chain of prompts in which the output of the revised detection prompt is passed to a ***localization prompt***.

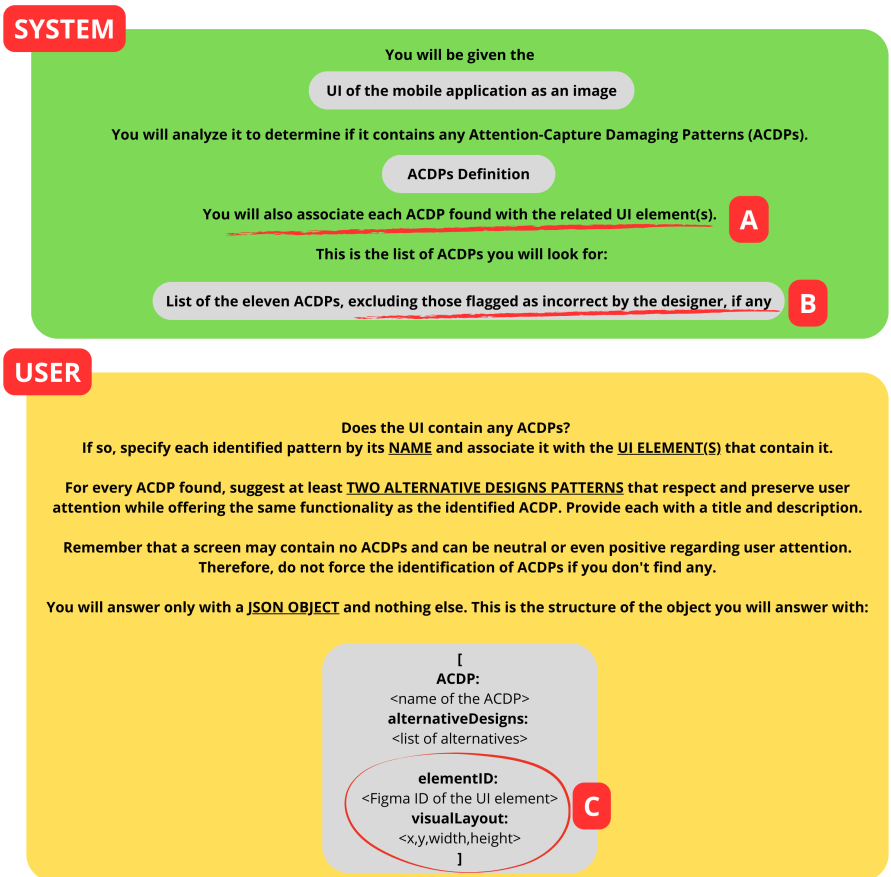



Fig. 6. The final structure of the detection prompt, refined after the expert evaluation to take into account designers' feedback and get the approximate position of the ACDPs detections in the analyzed UI. Differences from the initial detection prompt (Figure 4) are highlighted in red.

Specifically, this prompt receives two inputs: a) the JSON representation of the UI mockup extracted through the Figma APIs, and b) the JSON object returned by the detection prompt, which now includes the approximate visual layout (x, y, width, and height) of the Figma elements that originated each detected ACDP. Based on this input, GPT-4o is instructed to find the elements in the JSON representation of the UI that best overlap with the approximate visual layout data of each detection, and return the Figma IDs of the retrieved elements[5].

## 5.2 Plugin Implementation

Building on prior work on designing for digital wellbeing and designer–LLM collaboration, we integrated the multi-modal chain of prompts into a Figma plugin to assist UI/UX designers at all stages of their design activities. This integration makes feedback on ACDPs embedded within design workflows, while overcoming the limitations of solutions that rely on specific design templates or narrowly defined use cases [55]. Prior research has shown that recognizing dark patterns

[5]The final version of the localization prompt is available at https://osf.io/smzuv/files/?view_only=54667f6db28547049a9a9a3830c9bdad

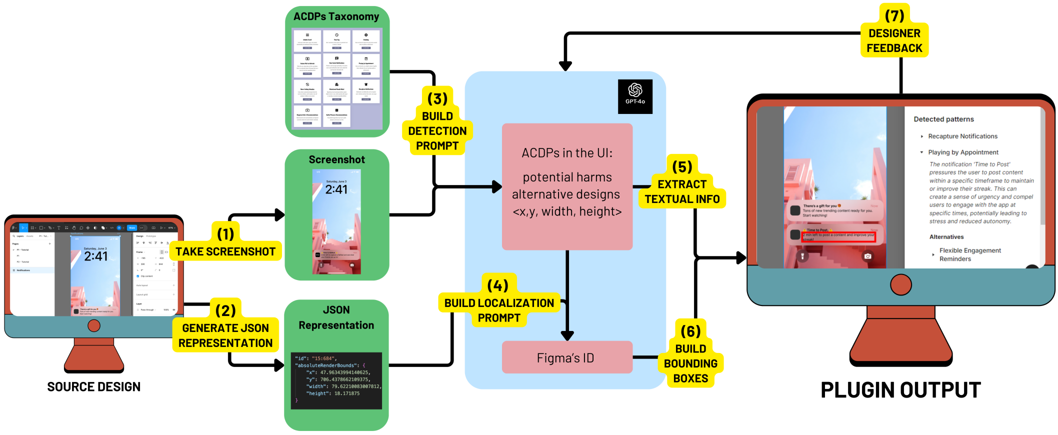



Fig. 7. The final architecture of our Figma plugin, including the multi-modal chain of prompts composed of the detection prompt and localization prompt.

often benefits from explicit visual and conceptual scaffolding, as such patterns are normalized within prevailing design conventions and can be difficult to identify at a glance without dedicated cues [14, 49]. Providing this form of integrated support can foster reflective thinking among designers [1] and support the recognition of ACDPs. Moreover, it enables designers to provide feedback, refine suggestions, and negotiate design aspects they may not have anticipated, thereby enacting a constructive negotiation approach to designer–LLM collaboration [70].

We implemented the plugin using TypeScript, the Figma Plugin APIs, and the Open AI APIs. Figma is one of the most popular tools for UI prototyping [40], and it has already been used as a test case for exploring the adoption of LLM-powered feedback systems (e.g., [18, 55]).

Figure 7 summarizes the final architecture of our Figma plugin. The designer prototypes a UI in Figma (*SOURCE DESIGN*). When the designer scans an interface, the plugin captures a screenshot (1) and generates a JSON representation for the UI (2). Using the screenshot and the taxonomy of ACDPs [57], the detection prompt (3) is created to instruct the LLM to identify ACDPs, alternative designs to protect users' time and attention, and approximate layout data (<x,y, width, height>). The layout data and the JSON representation are used in the localization prompt (4) to identify Figma elements representing the detected ACDPs, which are then highlighted with bounding boxes in Figma (5). In parallel, textual information from the detection prompt is displayed in a right-side panel adjacent to the UI mockup (6). Designers can dismiss incorrect suggestions, which are integrated into future LLM prompts for subsequent scans (7).

The plugin is freely available at https://www.figma.com/community/plugin/1310606990428971215/digital-wellbeing-lens , while the source code can be accessed at https://git.elite.polito.it/public-projects/digital-wellbeing-lens.

*5.2.1 System Walk-Through.* As shown in Figure 8, designers can work in Figma as they normally would. At any time, they can select one or more top-level frames and press "Scan" to send them for evaluation to the LLM. The model assesses the received UIs and, when ACDPs are detected, it highlights them directly within the corresponding Figma frame through bounding boxes. Figure 9 (i), for example, shows a mockup of a smartphone lock screen where two different ACDPs have been found: the use of Recapture Notifications and the Playing by Appointment pattern, which forces users to use a digital service at specific and pre-defined times [57]. Each detection is accompanied by the name of the pattern and textual explanations of its characteristics and potential harms for end users, which are rendered back from the LLM to Figma. When the designer hovers the cursor over the name or description of the pattern, the corresponding bounding box is highlighted in the frame. In Figure 9 (ii), for example, the plugin says to the designer that *"the notification 'Time to Post' pressures the user to post content within a specific timeframe [...]. This can create a sense of urgency [...] leading to stress and reduced autonomy."*

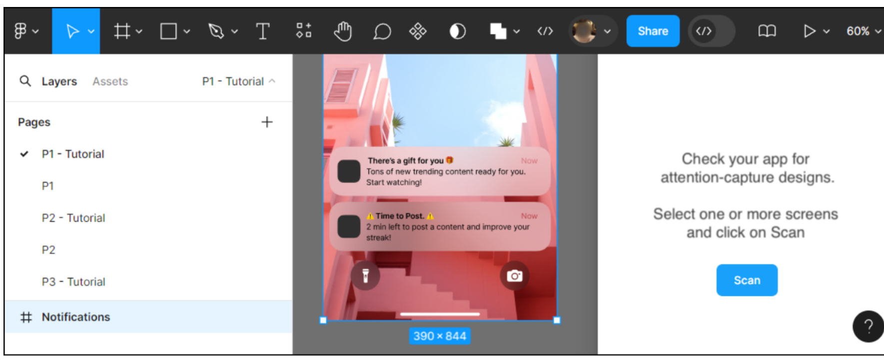



Fig. 8. Our Figma plugin allows designers to select one or more top-level frames and press "Scan" to send them for evaluation.

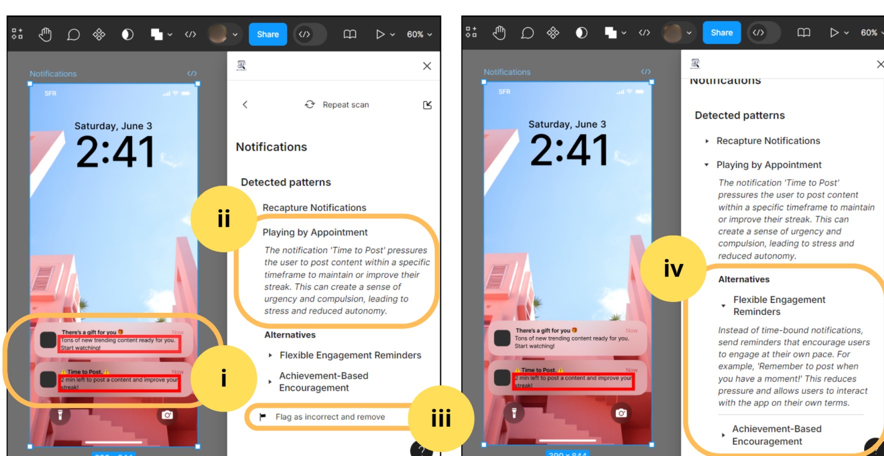



Fig. 9. Two screenshots showcasing the plugin feedback received after a scan. These include: a bounding box highlighting detected ACDPs in the UI (i); textual descriptions of the detected ACDPs and their harms for end users (ii); and recommendations for alternative designs (iii). Designers can click on 'flag as incorrect' at any time to refine subsequent scans (iv).

If the designer finds an incorrect detection, they can click on the "flag as incorrect" button to hide it and send feedback to the LLM for subsequent scans, ensuring this pattern will not be shown again (Figure 9 (iii)). Instead, if the detection of a given ACDP is considered helpful, the designer can take inspiration from the plugin's suggestions, negotiating alternative designs that may preserve end-users' time and attention. Each alternative is generated by the LLM to maintain

Table 3. The demographics of the 15 professional UX/UI designers who evaluated the plugin in a user study.

| ID | Gender | Age | Occupation | Design Experience | Figma Experience |
|---|---|---|---|---|---|
| P1 | Man | 51 | Experience designer and information architect | 20 years | 5 years |
| P2 | Woman | 27 | Ph.D. student and freelance UX designer | 7 years | 5 years |
| P3 | Man | 26 | UX Designer and filmmaker | 5 years | 7 years |
| P4 | Woman | 20 | Industrial and UX/UI designer | 3 years | 2 years |
| P5 | Man | 29 | Ph.D. student and freelance UX designer | 10 years | 4 years |
| P6 | Man | 29 | Product and UX/UI designer | 6 years | 6 years |
| P7 | Man | 27 | UX designer | 2 years | 2 years |
| P8 | Man | 27 | Digital designer | 8 years | 4 years |
| P9 | Woman | 24 | UX/UI design consultant | 6 years | 3 years |
| P10 | Man | 40 | UX designer | 4 years | 4 years |
| P11 | Man | 37 | UX designer | 7 years | 3 years |
| P12 | Man | 33 | Freelance UX designer | 6 years | 6 years |
| P13 | Man | 27 | Product and UX/UI designer | 4 years | 4 years |
| P14 | Woman | 49 | UX designer | 20 years | 2 years |
| P15 | Woman | 42 | UX/UI designer | 3 years | 2 years |

the functionality of the original pattern while better aligning with end-users' digital wellbeing, and is provided to the designers in Figma with a name and description. Figure 9 (iv), for example, shows two suggested alternatives to the detected Playing By Appointment, namely "Flexible Reminders" and "Achievement-Based Encouragement." The first alternative, in particular, suggests designers the following: *"instead of time-bound notifications, send reminders that encourage users to engage at their own pace. For example, 'Remember to post when you have a moment!' This reduces pressure and allows users to interact with the app on their own terms."*

## 6 USER STUDY WITH UI/UX DESIGNERS

We conducted a formative user study with 15 professional UI/UX designers to explore the potential of our approach as an early-stage design probe. In the study, we further evaluated the accuracy and helpfulness of the refined prompting technique's output, this time from a designer's perspective, using pre-built UI mockups to investigate how designers interpret and engage with the plugin's wellbeing- oriented feedback within a controlled setting. Additionally, we explored the plugin's potential to raise awareness on ACDPs among designers and its perceived fit with existing working practices.

### 6.1 Methods

*6.1.1 Participants.* Participants were recruited from a pool of professional UI/UX designers using various online channels, including mailing lists, Slack and Discord communities, and social media. Participants qualified for the study if they had been using Figma for more than one year. Table 3 provides an overview of the demographics of the 15 participants, including their gender, age, occupation, years of experience as a designer, and years of experience using Figma.

Our final sample consisted of 15 participants (10 self-identifying as man, 5 as woman) with a mean age of 32.53 years (SD = 8.94), representing a diverse range of occupations and years of design experience. The majority of participants (12) were Italian designers based in Italy, one was Italian but worked in the Netherlands, and the remaining two were from South Korea and worked there too.

*6.1.2 Procedure and Measures.* We conducted the study via Zoom's calls. At the beginning, participants were introduced to the study. They were told that the study aimed to explore how an AI-powered design tool could raise awareness of design decisions related to users' time and attention. We introduced the concept of ACDPs at a high level, without providing an exhaustive taxonomy, and emphasized that the tool should be treated as a design aid rather than as an authoritative evaluator. After providing informed consent online, participants were instructed by a researcher to install the plugin on their own Figma application and share their screens. To ensure consistency and fairness across all participants, we used the dataset described in Section 3.1 as the basis for the study. This approach aligns with prior studies in the design community [18, 60] and was chosen for two main reasons. First, freely designing a UI without constraints would not have been feasible within the short timeframe of a lab study. Second, using pre-built prototypes eliminated potential confounding factors, ensuring that all participants began with the same baseline, thereby allowing us to evaluate the plugin's impact more systematically. As such, our goal was not to measure the tool's impact on final design outcomes, but to examine how designers interpret, negotiate, and react to wellbeing-oriented feedback when it is introduced into familiar design artifacts. We return to the implications of this lab-based setup and its limits for ecological validity in Section 7.4.

The study was structured into two main phases:

**Free-usage session.** For each participant, the plugin first displayed one of the UIs included in the dataset described in Section 3.1. In this phase, we asked participants to freely modify the mockup with the assistance of the plugin, allowing them to experience our tool in an unrestricted usage session and discover its functionality. This session lasted 10 minutes, with a researcher interrupting the participants when the time was up.

**Evaluation session.** After the free-usage session, the plugin displayed other three mockups included in the UI dataset described in Section 3.1. In this phase, we asked participants to scan the three UIs and provide feedback on the plugin's output.

The collected measures included some metrics previously evaluated by experts—this time assessed from a designer's perspective—as well as metrics specifically tailored to the needs and workflows of UI/UX designers. After each scan in the evaluation session, in particular, designers were asked to assess the perceived helpfulness of the plugin's feedback, which, unlike in the expert evaluation, was now contextualized within the UI using bounding boxes. Specifically, designers evaluated:

- The **helpfulness** of each ACDP **detection**, on a Likert scale from 1 (*not at all helpful*) to 5 (*very helpful*).
- The **helpfulness** of the suggested **alternative designs**, on a Likert scale from 1 (*not at all helpful*) to 5 (*very helpful*).

Additionally, we also asked them to evaluate the **perceived impact** of implementing the alternative designs on **usability** with respect to the original pattern, between *"decreases usability," "has no impact on usability,"* and *"improve usability."* Through this question, we aimed at evaluating the practical implications of redesigning ACDPs and the perceived feasibility of applying the approach in real-world scenarios, as a significant decrease in usability would be unacceptable in design practice.

We concluded the study with a semi-structured interview focusing on *overall impressions*, *potential drawbacks and dangers*, *potential for iterative use*, and *perceived* fit with exiting design workflow. To facilitate later analysis, all study sessions were recorded following participants' explicit consent. Each session lasted 49 minutes on average (SD = 11 minutes).

*6.1.3 Data Analysis.* As with the expert evaluation, we collected all participant ratings on helpfulness and usability impact, calculating the total number of responses per rating, median, and interquartile range (IQR). Additionally, we reviewed all the recordings to consolidate the notes taken during the study, transcribing the interviews verbatim and conducting a thematic analysis [5] on the transcriptions.

### 6.2 Results

In this section, we first provide an overview of the free-usage sessions, followed by quantitative results from the evaluation sessions and findings from the thematic analysis.

*6.2.1 Overview of the Free-Usage Sessions.* During the free usage session, participants showed interest in the feedback provided by the plugin. The vast majority of them (13 out of 15) attempted to modify their assigned UI following at least one of the alternative solutions suggested by the plugin. Additionally, three participants (P8, P9, and P11) engaged with and explored alternatives for all detected ACDPs with the help of the plugin's suggestions.

Regarding the preferred design alternatives, we observed that designers' preferences varied depending on the specific ACDPs they were presented with. Nevertheless, participants were intrigued by the plugin and the feedback they received, showing a willingness to experiment with design patterns outside of their usual practices.

Two participants (P2 and P4) attempted to replace Infinite Scroll with plugin-suggested alternatives such as "load more button" or "content batching." These solutions suggest introducing a button after a certain amount of content is displayed, allowing users to explicitly request to keep viewing content. When reflecting on Infinite Scroll, another participant (P6) opted for the "content preview" solution, which recommends displaying only a preview of the content in the feed, forcing users to choose which items to view. A fourth participant (P15) who received a warning on Infinite Scroll ignored the detection, as in their opinion users are now too accustomed to navigating screens with such an interaction mechanism.

When inspecting other detected patterns, three participants (P2, P7, and P12) attempted to implement a suggestion to clearly label commercial ads, either within feeds (Disguised Ads and Recommendations) or in messages (Fake Social Notifications). Meanwhile, three other participants who were notified of an Attentional Roach Motel (P3, P5, and P9) attempted to make the logout button more prominent by isolating it from other buttons on the settings screen and using a more noticeable font color.

When evaluating a detection of a Grinding pattern, two participants (P8 and P14) decided to replace it with "meaningful milestones," which suggests offering users meaningful, non-repetitive goals. For example, P8 introduced goals such as "friends who confirmed to know you," "users who found your posts useful," and "times you have not exceeded your limit" (Figure 10). P10, instead, followed the "progressive milestone" suggestion, which recommends setting small, easily achievable goals that progressively increase.

Finally, the Recapture Notifications pattern was replaced by two participants (P11 and P13) out of three with "scheduled summary notifications," which suggests grouping multiple notifications together and sending them at once to minimize distractions.

*6.2.2 Quantitative Results from the Evaluation Session.* Figure 11 shows the number of responses by participants for the detection helpfulness, helpfulness of the two suggested alternative designs, and perceived impact on usability of the alternatives. Designers found detections generally useful for reflecting on ACDPs. The majority of the evaluated detections (54%) were rated as very helpful, with a median rating of 5 (IQR = 2). In line with the results from the experts evaluation, designers found the suggestions for alternative designs to be more helpful that not on average, with a median rating across the two suggestions of 4 (IQR = 2). As shown in Figure 11, the first suggestion (A1) was rated helpful 22 times and very helpful 27 times (66% of responses), while the second suggestion

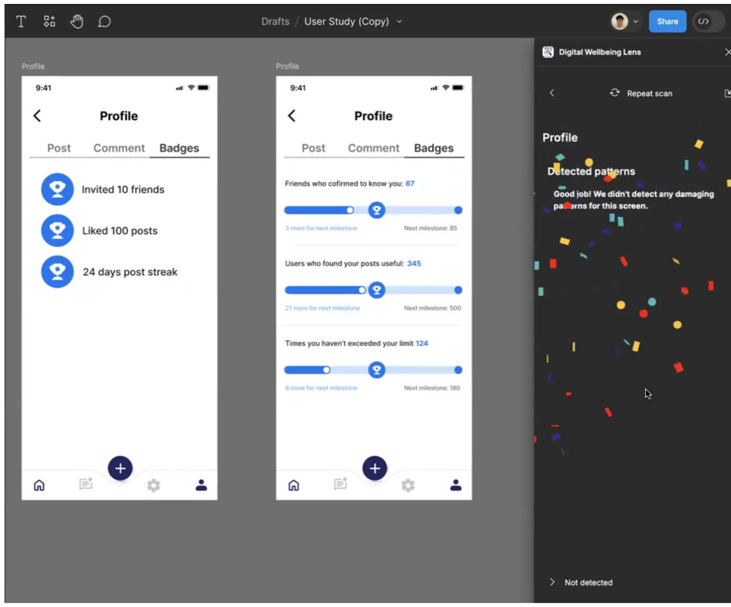



Fig. 10. A designer replacing Grinding with meaningful and non-repetitive goals in the free-usage session.

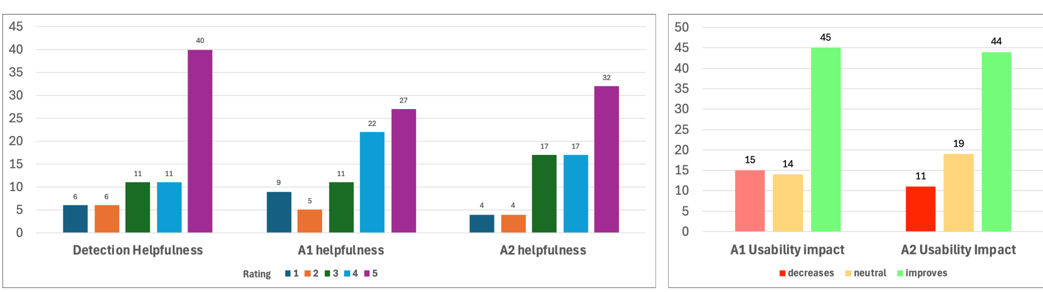



Fig. 11. The total number of participant responses for detection helpfulness, helpfulness of the two proposed alternative designs, and impact on usability of the two proposed alternative designs.

(A2) received 17 helpful and 32 very helpful ratings (66% of responses). Interestingly, designers positively assessed the impact of the alternative designs on usability. A1 and A2 were rated 45 (60% of responses) and 44 times (59% of responses), respectively, to improve usability compared to the original pattern. While most designers rated the alternatives positively, there were cases (A1: 15, A2: 11) where designers believed the alternative would likely decrease usability. This indicates a need for more compelling incentives to motivate designers to adopt the plugin's suggestions.

*6.2.3 Findings from the Thematic Analysis.* Our thematic analysis revealed six overarching themes, from *workflow improvements and education on ACDPs* to *target shift: community, and regulators.*

*Workflow improvements and education on ACDPs.* The plugin was seen as having the potential to enhance designers' workflow efficiency and provide educational benefits (**N = 14:** P1, P2, P3, P4, P5, P6, P7, P8, P9, P11, P12, P13, P14, P15).

Some participants (P3, P7, P11, and P12) emphasized how the plugin could potentially have positive side effects, e.g., by enhancing the **usability** and **accessibility** of the designed applications with minimal time investment. Suggestions for design alternatives, in particular, were compared to Figma tools like Contrast [72], which identifies poorly chosen color contrasts and suggests more effective alternatives.

P6 also emphasized the beneficial use of AI in the plugin, noting that it provides valuable and contextual feedback that, rather then delivering **definitive answers**, could assists designers' analysis, brainstorming, and discussions within companies:

> *"It's an excellent use of LLMs because it doesn't aim to give you a perfect, precise, decisive answer but helps you analyze and brainstorm."*

Regarding the plugin's **educational** utility, P8 suggested its potential use in design courses to help students recognize dark patterns in their UIs:

> *"Maybe the professor could explain to the students these ACDPs within a UX/UI course to say guys, take care, I mean, we have to be careful, and this is a tool that goes in our favor."*

P1 and P3 considered the benefits that junior designers could derive from using the plugin. They suggested that the plugin could support these designers in learning to recognize and avoid ACDPs until they are capable of doing so independently.

Several participants pointed out that the plugin effectively addresses the **unintentional use** of ACDPs by designers, who often try to replicate UIs and patterns included in existing platforms. As P8 noted,

> *"You do it mostly out of habit to put damaging patterns and the worst ones are the ones that are inside your subconscious as a designer, so even just the fact that this plugin makes me notice that is okay, but it also gives me an alternative, it pushes me much more to design a healthier app."*

Other participants, like P4, articulated a similar point, observing that even when studying dark patterns, the gap between theoretical concepts and their application in daily projects often makes it challenging to implement these concepts effectively:

> *"I have read some literature about dark patterns [...] But it's still kind of far away from our daily life and our design work."*

In this regard, the plugin offers the advantage of providing information relevant to the designer's current work, both in detecting issues and suggesting alternative design solutions, as noted by P5:

> *"On one hand, it gives you knowledge that you don't have in a very practical way, directly related to the work you're doing [...] On the other hand, it also provides practical advice on how to intervene quickly."*

*Enhancing and encouraging iterative plugin usage.* A key strength identified by participants was the potential for iterative use of the plugin (**N = 11:** P1, P2, P3, P4, P5, P6, P8, P9, P10, P11, and P14). This approach enables the designer to ensure that a set of frames remains free of unwanted ACDPs throughout all design stages, as noted by P2:

> *"As it's used to evaluate the initial version, we can continue using the plugin for subsequent versions until the screens are finalized. After that, it's only needed if screens undergo significant redesigns."*

However, participants recommended enhancing this feature to enable designers to compare previous and current designs and receive targeted feedback that would encourage improvements to critical screens. P9 illustrated this concept:

> *"Perhaps when you scan it again, you can compare the original version to the redesigned version. This way, you can see how you've eliminated three dark patterns and possibly motivate other designers to redesign their screens."*

Participants emphasized the importance of retaining memory of previous scans also to avoid frustrating situations where a designer addresses an ACDP, yet the plugin continues to detect it. In such cases, the tool should recognize the designer's efforts and provide tailored suggestions that account for the designer's initial attempts. To facilitate the iteration process, participants also recommended providing **real-time** feedback. P3 envisioned this feature, noting that it would align with current trends in AI-powered companion tools. Furthermore, P4 said that real-time feedback could streamline iteration, as the current reload operation can be somewhat cumbersome:

> *"I think the operation is a bit overwhelmed, because every time I need to reload it [...] Yeah, real-time, it will be better."*

*Alerting users to AI mistakes and risks.* Some concerns were raised regarding the plugin's reliance on AI, particularly regarding the lack of **clear warnings** about potential errors made by LLMs (**N = 7:** P1, P5, P6, P8, P10, P11, P13). P5 stressed the importance of informing users that LLMs lack the complete contextual knowledge that designers have while working on a project and so should not be trusted blindly:

> *"Maybe a little bit like Chat-GPT does, which puts a disclaimer, always remembering to put things in context for the project you're working in and not to take everything as the gospel truth."*

This is an issue that was also highlighted by experts, noting that the severity of ACDPs is highly context-dependent. Consequently, much like the experts, designers occasionally felt confused by seemingly incorrect detections and expressed a need for the plugin to communicate more effectively when its level of certainty on the detection is insufficient. For example P5 suggested that the plugin should indicate the **confidence** level of its suggestions:

> *"One thing that I think could be dangerous is not giving any way to allow the user to understand how sure the plugin is about what it is saying."*

P1 instead noted the plugin's overly assertive tone, arguing that it could lead some designers to lose their critical perspective:

> *"[The plugin] is also very assertive, so it doesn't put doubts on things. One maybe blindly trusts what is written there and doesn't use his critical sense to interpret a suggestion."*

Such an assertive tone, as highlighted by P8, may cause designers to feel **accused** when ACDPs are detected.

*Stimulating discussions on ACDPs within companies.* Participants acknowledged that ACDPs are often implemented for specific **business reasons**, supported by data linking their use to performance indicators that the company aims to maintain or improve. In such cases, designers' concerns for preserving users' wellbeing may be disregarded by company management, with participants emphasizing the importance of fostering discussions about ACDPs at the company level (**N = 9:** P4, P6, P7, P8, P9, P10, P11, P12, and P13). P10 articulated this conflict:

> *"Obviously, as a UX designer you would want to avoid them [the ACDPs] all the time, however, then someone from management comes along with KPIs to bring to the result [...] So it's definitely interesting, but I wonder who has the luxury of saying let's reduce the outcome."*

Despite this, some participants believed that designers, even under management constraints, could use the plugin as a tool to spark discussion at the company level, e.g., to propose **balanced solutions** that meet business needs while protecting user wellbeing. P9, in particular, emphasized that being able to understand potential harms is essential for designers who prioritize users' wellbeing and wish to introduce this concern within companies. Similarly, P4 discussed the importance of compromise, suggesting improvements to the plugin that could enhance its effectiveness and impact within work environments:

> *"If you could give me like a rating system to tell me how deceptive it was, so I can make a trade-off between the commercial need and the design for usability."*

More in general the plugin was found useful as a catalyst for discussions within the company, which suggests the potential to foster interdepartmental dialogue and promote deeper engagement on the issues it addresses.

*Target shift: community and regulators.* Driven by concerns and tensions between designing for the digital wellbeing and businesses, some participants envisioned a shift in focus that includes the design community and regulators (**N = 3:** P6, P11, and P13). P6 indicated that more experienced designers might not need such a tool, as they can identify issues independently. P1 proposed shifting the focus from individual projects to a broader design community. Designers would share their projects and detection results within a **community platform**, facilitating discussion with colleagues worldwide. This community could make the plugin appealing even to senior designers, as they could assist junior designers and students with their questions, stay informed about current design trends and issues, and engage in broader discussions:

> *"A community aspect might be interesting. We could discuss things that others or I have produced. This way, we could become supporters of each other and contributors to a platform for dialogue"* (P1).

P13 envisioned a scenario where deceptive design patterns are **regulated** and restricted by law. In this case, the plugin could assist regulators in identifying issues within digital services. This process might also occur internally, as the company could establish a legal division dedicated to ensuring compliance with regulations:

> *"Imagine a legal department, even within the company, analyzing screens created by designers and business teams. If they find a non-compliant dark pattern, they can warn the teams to avoid it"* (P13).

## 7 DISCUSSION

This section revisits our findings by discussing design considerations on the Figma plugin, the potential integration of the plugin into design workflows and the designer's role in contextualizing and mediating its suggestions, and the need for new regulatory initiatives to promote digital well-being through design.

### 7.1 Design Considerations on the Figma Plugin

In our work, we chose to develop and evaluate an LLM-driven tool to provide a tangible proof of concept, allowing us to gather actionable feedback from designers and experts on its real-world applicability. Specifically, our choice of using a general-purpose LLM for ACDP detection, rather than a purpose-built classifier, reflects the goals of this work as an exploratory, design-oriented system rather than a high-performance detection pipeline. Prior work has demonstrated that dedicated classifiers can achieve strong performance on large-scale datasets of real-world interfaces [9, 47], including capabilities such as precise localization and temporal pattern detection. In contrast, our approach leverages an LLM to support a broader design space of interaction: beyond classification, the model can generate explanations, suggest alternative designs, and participate in a constructive, in-the-loop dialogue with designers. Future work could explore hybrid approaches that integrate purpose-built classifiers with LLMs, combining robust detection capabilities with generative feedback and interactive support for designers.

Overall, our approach aimed to combine technical innovation with an exploration of its implications, bridging practical utility and conceptual discussions. Differently from alternative approaches such as static design guidelines or checklists, our plugin dynamically shows ACDPs within concrete interfaces, pairing them with design alternatives in a proactive way. Designers involved in our study (Section 6) highlighted the potentials of such an approach, but also suggested valuable design insights that could enhance the tool and inform future research on AI-powered design-support systems.

A first promising enhancement could involve integrating a real-time feature where the LLM continuously evaluates UIs against ACDPs while the designer works. Similar methods have been investigated in other domains. Lawton et al. [38], for example, introduced an interface enabling designers to collaborate with generative AI in real-time to create sketch drawings. Another valuable addition towards effective iteration would be a memory feature to retain information from previous scans. The importance of AI-powered design tools that contextualize current projects status based on prior versions has been emphasized in research, for instance by Vaithilingam et al. [70]. Incorporating such a memory feature would allow the LLM to reference designers' earlier efforts when addressing ACDPs, reducing redundant detections, offering personalized support, and mitigating frustration.

During the study, designers also expressed concerns about the plugin's overly assertive tone and its lack of transparency regarding its reliance on the LLM. This tension suggest that feedback should be sufficiently assertive to disrupt habitual practices, yet sufficiently tentative to preserve designer agency. Previous research by Li et al. [40], which explored UX practitioners' perceptions of generative AI, emphasized the importance of attributing ownership to AI-generated outputs. In this context, it would be beneficial for the plugin to explicitly communicate that ACDPs detections and alternative design suggestions are AI-generated, e.g., by displaying confidence levels, while noting that ACDPs definitions are derived from academic sources. Additionally, the plugin should clearly articulate its limitations, such as its inability to compare multiple screens analyzed simultaneously and its lack of the contextual understanding. These limitations align with findings by Duan et al. [18] regarding LLM-generated heuristic evaluations.

Designers also suggested that the plugin could support collaboration by enabling them to share their projects and ACDP detections with colleagues worldwide through a community section. This feature would allow designers to seek a "second human opinion" on ACDPs detection and alternative suggestions, engage with others' projects, and stay informed about emerging trends in attention-capturing design they need to be aware of. While the concept of sharing work-in-progress projects with a broader community is not novel, it has been demonstrated to be effective. For example, Kim et al. [31] developed an online platform for illustrators to share snapshots of their work, showing that such environments can help creators reflect on their creative processes. Their findings suggest that similar approaches could be applied across creative domains, including design.

### 7.2 The Designer's Role and Integration into Existing Workflows

Our findings highlight the broader implications of integrating such tools into design workflows. Our evaluations with experts in damaging designs and designers suggest that plugins like ours could be beneficial throughout the design process and beyond, from initial ideation to scenarios where plugin's outputs are used as educational resources, e.g., for junior designers. Interestingly, both experts and designers observed that the occurrence and severity of an ACDP in a given UI are context-specific, depending on factors like the platform type, target population, and other screens. We hypothesize that in real-world scenarios, designers—as the creators of the UIs under analysis—would possess the necessary contextual knowledge. Indeed, our findings suggest that our plugin may support designers in exercising informed judgment on ACDPs without requiring expert-level inspection. Thus, we consider the designer to be the *mediator* of the plugin's detections and suggestions, ultimately deciding whether and how to incorporate them into their work before the project is moved to other departments within the company.

This framing positions designers as ethical agents who interpret, translate, and negotiate values within organizational contexts, in line with prior work on ethical mediation in design practice [26]. Accordingly, we envision our plugin as a digital-wellbeing assistant that operates in the loop, raising awareness of ACDPs and supporting their detection and discussion among designers and within organizations, rather than enforcing specific design choices. This approach resonates with Peters' work on wellbeing-supportive design, which emphasizes that supporting users' wellbeing requires dialogue and shared consensus within design teams [59].

At the same time, promoting digital wellbeing through design remains structurally constrained by tensions between organizational business metrics and user wellbeing. Consistent with critiques of the attention economy, we treat ACDPs not as isolated design flaws but as structural outcomes of this business model. Designers in our study were aware that current engagement metrics embed normative assumptions about success, often leaving limited room to contest potentially harmful practices without institutional support. While our plugin surfaces alternative design choices that may mitigate the harms of ACDPs, their adoption therefore depends on how companies reconcile these alternatives with engagement-driven KPIs that frequently normalize attention-related harms as part of routine design decision-making.

In this context, several participants described the plugin less as a solution to this tension and more as a boundary object that enables designers to surface, articulate, and negotiate design decisions related to users' digital wellbeing within organizations. By making ACDPs explicit, naming their associated harms, and proposing functionally equivalent alternatives, the plugin provides concrete artifacts that designers can mobilize in discussions with product managers and other stakeholders, particularly at earlier stages of the workflow, when alternatives are still negotiable. In this sense, the tool may support designers who already seek to address social values in their work by enabling what Wong [73] describes as "tactics of soft resistance," rendering attention-related harms more visible and harder to dismiss during design deliberations. Similarly, in relation to the design tensions described by Watkins et al. [71], the plugin may function as a tangible instrument through which designers attempt to influence corporate culture, rather than having their ethical concerns routinely "nullified" or "suppressed."

Encouraging designers to critically reflect on attention-capturing elements within their interfaces may also prompt them to reconsider choices that are often introduced "out of habit." This is particularly important given emerging regulatory approaches, which increasingly focus on the effects and harms of dark patterns rather than designers' intent [67].

Overall, fostering discussion around ACDPs through our plugin can helps shifting the focus from short-term metrics toward longer-term considerations. For example, designers in our study noted that micro-friction interventions may initially frustrate users and reduce engagement, potentially lowering short-term economic performance. Yet such frictions can also prevent more severe or prolonged user dissatisfaction, leading to more positive and meaningful digital experiences over time. We hypothesize that users may be willing to accept these short-term trade-offs in favor of a more positive, meaningful digital experience. Indeed, research shows that users are increasingly aware of the negative effects of the Attention Economy and are actively seeking tools to improve their digital habits [44, 52, 69]. From an optimistic economic perspective, focusing on meaningful, wellbeing-oriented design may foster greater user loyalty over time and greater long-term user engagement, even if it leads to temporary declines in short-term engagement [17]. This approach suggests that prioritizing digital wellbeing could benefit both users and companies in the long run, aligning ethical design with sustained business success.

That said, we acknowledge that prioritizing digital wellbeing can be challenging. As such, the next section examines the importance of establishing new regulatory initiatives to reinforce our approach.

### 7.3 Towards New Regulatory and Legal Environments

Recently, Sánchez Chamorro et al. [65] highlighted that addressing the complex challenges of with dark patterns and their associated harms requires a collaborative, interdisciplinary approach. In line with recent work at the intersection of HCI and law [17, 67], we emphasize that our work should be complemented by collective action and stronger regulatory interventions. Indeed, while policymakers and regulators are increasingly recognizing the harms caused by certain design patterns, particularly in the areas of privacy [22] and data sharing [19, 20], how to regulate ACDPs remains an open question, especially when it comes to "seductive" patterns like infinite scroll and content autoplay [57]. Indeed, as recently reported by Santos et al. [67], the harms caused by these kinds of dark patterns have not been thoroughly examined from a legal perspective, and users are still *"exposed to dark patterns that cost their time, attention, cognitive capability, among others, without remedy"* (p. 2).

Recent ontologies and taxonomies [30, 57], along with new frameworks classifying the harms caused by these designs [67], are helping to bridge the gap between designers, technologists, regulators, and legal professionals [30]. We posit that ACDPs and their harms, like other types of dark patterns, could potentially be legally recognized and enforceable across existing legal initiatives, such as those currently in effect in Europe, including the GDPR [21], Digital Services Act [11], Artificial Intelligence Act [19], and Data Act [20].

Article 25 of the DSA [11], for instance, prohibits the design of online interfaces that deceive or manipulate users in ways that materially distort or impair their freedom of choice. This general mandate provides a plausible regulatory hook for several ACDPs, ranging from infinite scroll to neverending autoplay. Complementarily, Article 27 establishes transparency obligations for algorithmic recommender systems, requiring platforms to disclose information about the "main parameters" shaping content curation and to offer users meaningful options to modify or opt out of personalised recommendations.

In parallel, Article 5 of the AI Act [19] bans certain AI practices outright, including systems that deploy subliminal techniques or exploit users' vulnerabilities (e.g., related to age or disability) in ways that materially distort behaviour and cause harm. Because ACDPs often rely on exploiting psychological vulnerabilities and cognitive biases, some of these patterns could plausibly fall within the scope of this prohibition. More broadly, the AI Act articulates a principle that AI systems should not covertly coerce, exploit, or unfairly manipulate individuals, and introduces specific transparency obligations—for example, requiring AI systems that interact with humans (e.g., chatbots or virtual assistants) to disclose their AI nature—which resonates with ACDPs such as fake social notifications.

This regulatory recognition could encourage the adoption of established ethical design practices, providing our plugin with standardized "bright patterns [66]" as design alternatives that promote users' digital wellbeing.

### 7.4 Limitations and Future Works

While our findings are encouraging, it is important to acknowledge certain limitations that, although constraining the generalizability of our approach, also point to valuable directions for future research.

*7.4.1 Extending Conceptual and Methodological Horizons.* Our plugin aims to improve digital wellbeing by supporting designers in finding viable and healthier design alternatives to ACDPs. However, we recognize that digital wellbeing is a complex and socially constituted issue that extends beyond the scope of user attention. Consequently, while our approach provides a tangible starting point for promoting healthier design practices, it can be complemented by further research addressing the deeper socio-political and economic factors driving the commodification of attention and data.

Our work represents one of the first attempts in digital wellbeing research to shift the focus from an end-user-centric perspective—which has been found to be both ineffective in the long term [64] and overly simplistic [17]—to the design community, as we do believe that better design—one that supports users' sense of agency—has the potential to enhance digital well-being. Yet, we have not explored how following the plugin's suggestions might influence user behavior or digital wellbeing outcomes. Future work should address these gaps through rigorous user studies that empirically evaluates the impact of the suggested design alternatives on end users, providing insights into their efficacy in promoting meaningful technology uses.

*7.4.2 Future Directions in Empirical and Technical Development.* While we sought diversity in the areas of expertise and geographical backgrounds of the dark pattern experts involved in our study, their limited number inevitably constrains the breadth of perspectives represented. Future work should therefore extend this evaluation to a larger and more diverse panel of experts across industries and geographies to strengthen the generalizability of the findings.

Similarly, the study with UI/UX designers was primarily conducted with a relatively small group belonging to specific cultural context. Given that the participant pool was culturally homogeneous, the findings should be interpreted with caution regarding cultural perceptions of ACDPs. Future work should recruit a more globally diverse group of designers to better capture cross-cultural perspectives. Furthermore, although we encouraged designers to freely experiment with the plugin, we used pre-built prototypes from the UIs dataset for both studies to streamline the process. Consequently, aside from participants' qualitative feedback, we did not capture the impact of the plugin in a more realistic design scenario, where designers would start with a concept and create UI mockups from scratch. Future works should observe designers working on real-world projects to provide a more ecological assessment of the plugin's impact on actual design decisions and workflow integration.

These studies would be also helpful to further assess the approach over time. A potential contradiction emerging in our work, indeed, concerned the generally positive quantitative ratings of the plugin's suggestions and the more critical accounts articulated in designers' qualitative reflections, particularly in relation to habituation effects, business constraints, and longer-term adoption. We interpret this as an indication that designers valued the conceptual direction and reflective affordances of the tool, even when specific design alternatives were perceived as requiring further refinement or negotiation. However, we acknowledge that initial ratings may have been shaped by novelty and desirability effects, as is often the case with early-stage design tools.

From a technical perspective, our quantitative evaluation of the detection prompt is subject to important limitations. In particular, the use of a small dataset of mockups derived from canonical examples of ACDPs, combined with prompting the model using the same taxonomy, introduces a degree of alignment between the evaluation data and the model's instructions. While this setup enables controlled comparison across prompt configurations, it does not reflect the variability and ambiguity of real-world interfaces. Furthermore, we tested a single LLM (GPT-4o) as a tool to detect ACDPs within UIs and suggest design alternatives. As landscape of LLMs is continuously evolving, we cannot exclude the possibility that other models might perform better. Furthermore, current limitations in LLMs forced us to evaluate static screenshots (one at a time), preventing the plugin from assessing interactivity. We acknowledge that several ACDPs are inherently temporal, with their harmful qualities emerging through repeated interaction or transitions over time rather than from a single interface state. Recent work [29] shed light on temporally sensitive dark patterns. In our domain, patterns like Pull-to-refresh, Infinite Scroll, or Attentional Roach Motel rely on microinteractions, reward timing, and evolving user commitment, and therefore may not be fully captured in static representations. In contrast, other ACDPs such as Disguised Advertisements and Fake Social Notifications are often legible from a single screen. Our approach therefore prioritizes early-stage, static indicators of potential ACDPs, positioning the plugin as a reflective aid rather than a comprehensive detection system. We see this work as complementary to temporally informed analyses of dark patterns such as the one proposed by Gray et al. [29], suggesting future opportunities to integrate these two approaches, for example through enhanced prompting with interaction annotations or video-based analysis with human-in-the-loop involvement.

## 8 CONCLUSIONS

This paper presented an LLM-powered Figma plugin that helps designers detect ACDPs and discover alternative designs that may safeguard users' time and attention. We refined and evaluated our approach through a series of studies, namely a performance analysis of our multi-modal prompting technique, an expert assessment of the LLM's outputs, and a user evaluation of the plugin with 15 professional UI/UX designers. Findings suggest that the plugin can inspire designers to reconsider their choices early in the design process and highlight the tool's potential to support the negotiation of alternatives to ACDPs within companies. Through the mediation of the designer and, ideally, new regulatory and legal initiatives around ACDPs, we envision our plugin being integrated into current design workflows, thereby making digital wellbeing a critical design dimension.

## REFERENCES

[1] Reem S. Al-Mansoori, Dena Al-Thani, Raian Ali, and Alhasan Alsammarraie. 2023. Designing for User's Digital Wellbeing: Co-Creating Nudges with Designers: Designing for User's Digital WellbeingCo-Creating Nudges with Designers. In *Proceedings of the 2023 ACM Conference on Information Technology for Social Good* (Lisbon, Portugal) *(GoodIT '23)*. Association for Computing Machinery, New York, NY, USA, 156–164. https://doi.org/10.1145/3582515.3609530

[2] Adam L. Alter, Chadly Stern, Yael Granot, and Emily Balcetis. 2016. The "Bad Is Black" Effect: Why People Believe Evildoers Have Darker Skin Than Do-Gooders. *Personality and Social Psychology Bulletin* 42, 12 (2016), 1653–1665. https://doi.org/10.1177/0146167216669123

[3] Eric P.S. Baumer and M. Six Silberman. 2011. When the implication is not to design (technology). In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. Association for Computing Machinery, New York, NY, USA, 2271–2274. https://doi.org/10.1145/1978942.1979275

[4] Mohammed Bedjaoui, Nadia Elouali, and Sidi Mohamed Benslimane. 2018. User Time Spent Between Persuasiveness and Usability of Social Networking Mobile Applications: A Case Study of Facebook and YouTube. In *Proceedings of the 16th International Conference on Advances in Mobile Computing and Multimedia* (Yogyakarta, Indonesia) *(MoMM2018)*. Association for Computing Machinery, New York, NY, USA, 15–24. https://doi.org/10.1145/3282353.3282362

[5] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101. https://doi.org/10.1191/1478088706qp063oa

[6] Christopher Burr, Nello Cristianini, and James Ladyman. 2018. An Analysis of the Interaction Between Intelligent Software Agents and Human Users. *Minds Mach.* 28, 4 (dec 2018), 735–774. https://doi.org/10.1007/s11023-018-9479-0

[7] Christopher Burr, Mariarosaria Taddeo, and Luciano Floridi. 2020. The Ethics of Digital Well-Being: A Thematic Review. *Science and Engineering Ethics* 26 (2020), 2313–2343. https://doi.org/10.1007/s11948-020-00175-8

[8] Canva. 2024. Visualize your design ideas with Magic Design. https://www.canva.com/magic-design/ Accessed: 2024-07-11.

[9] Jieshan Chen, Jiamou Sun, Sidong Feng, Zhenchang Xing, Qinghua Lu, Xiwei Xu, and Chunyang Chen. 2023. Unveiling the Tricks: Automated Detection of Dark Patterns in Mobile Applications. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (San Francisco, CA, USA) *(UIST '23)*. Association for Computing

Machinery, New York, NY, USA, Article 114, 20 pages. https://doi.org/10.1145/3586183.3606783

[10] Hyunsung Cho, DaEun Choi, Donghwi Kim, Wan Ju Kang, Eun Kyoung Choe, and Sung-Ju Lee. 2021. Reflect, not Regret: Understanding Regretful Smartphone Use with App Feature-Level Analysis. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2, Article 456 (oct 2021), 36 pages. https://doi.org/10.1145/3479600

[11] European Commission. 2022. The Digital Services Act package. https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package Accessed: 2024-07-11.

[12] Thomas H. Davenport and John C. Beck. 2001. *Attention Economy: Understanding the New Currency of Business.* Harvard Business School Press.

[13] Pieter Desmet and Anna Pohlmeyer. 2013. Positive Design An Introduction to Design for Subjective Well-Being. *International Journal of Design* 7 (12 2013).

[14] Linda Di Geronimo, Larissa Braz, Enrico Fregnan, Fabio Palomba, and Alberto Bacchelli. 2020. UI Dark Patterns and Where to Find Them: A Study on Mobile Applications and User Perception. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376600

[15] Inc. Diagram Technologies. 2024. Design tools from the future. https://diagram.com/. Accessed: June 12, 2024.

[16] Niall Docherty. 2021. Digital Self-Control and the Neoliberalization of Social Media Well-Being. *International Journal of Communication* 15, 0 (2021). https://ijoc.org/index.php/ijoc/article/view/17721

[17] Niall Docherty and Asia J. Biega. 2022. (Re)Politicizing Digital Well-Being: Beyond User Engagements. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 573, 13 pages. https://doi.org/10.1145/3491102.3501857

[18] Peitong Duan, Jeremy Warner, Yang Li, and Bjoern Hartmann. 2024. Generating Automatic Feedbacksantogae-gaegeageagaegaegeageaggeageagea on UI Mockups with Large Language Models. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 6, 20 pages. https://doi.org/10.1145/3613904.3642782

[19] European Commission. 2024. Artificial Intelligence Act. https://artificialintelligenceact.eu/ Accessed: 2024-09-11.

[20] European Commission. 2024. The Data Act. https://digital-strategy.ec.europa.eu/en/policies/data-act Accessed: 2024-09-11.

[21] European Commission. 2024. Data Protection in the EU. https://commission.europa.eu/law/law-topic/data-protection/data-protection-eu_en Accessed: 2024-09-11.

[22] Federal Trade Commission. 2024. FTC, ICPEN, and GPEN Announce Results of Review of Use of Dark Patterns Affecting Subscription Services and Privacy. https://www.ftc.gov/news-events/news/press-releases/2024/07/ftc-icpen-gpen-announce-results-review-use-dark-patterns-affecting-subscription-services-privacy Accessed: 2024-09-11.

[23] Figma. 2024. Redesign the way you jam with FigJam AI. https://www.figma.com/figjam/ai/. Accessed: June 12, 2024.

[24] Batya Friedman and David G. Hendry. 2019. *Value Sensitive Design: Shaping Technology with Moral Imagination.* The MIT Press.

[25] Google. 2024. Digital Wellbeing Initiative. https://wellbeing.google/ Accessed: 2024-11-25.

[26] Colin M. Gray and Shruthi Sai Chivukula. 2019. Ethical Mediation in UX Practice. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–11. https://doi.org/10.1145/3290605.3300408

[27] Colin M. Gray, Shruthi Sai Chivukula, and Ahreum Lee. 2020. What Kind of Work Do "Asshole Designers" Create? Describing Properties of Ethical Concern on Reddit. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference* (Eindhoven, Netherlands) *(DIS '20)*. Association for Computing Machinery, New York, NY, USA, 61–73. https://doi.org/10.1145/3357236.3395486

[28] 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* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3173574.3174108

[29] Colin M. Gray, Thomas Mildner, and Ritika Gairola. 2025. Getting Trapped in Amazon's "Iliad Flow": A Foundation for the Temporal Analysis of Dark Patterns. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 225, 10 pages. https://doi.org/10.1145/3706598.3713828

[30] 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* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 289, 22 pages. https://doi.org/10.1145/3613904.3642436

[31] Joy Kim, Maneesh Agrawala, and Michael S. Bernstein. 2017. Mosaic: Designing Online Creative Communities for Sharing Works-in-Progress. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (Portland, Oregon, USA) *(CSCW '17)*. Association for Computing Machinery, New York, NY, USA,

246–258. https://doi.org/10.1145/2998181.2998195

[32] Emre Kocyigit, Arianna Rossi, Anastasia Sergeeva, Claudia Negri Ribalta, Ali Farjami, and Gabriele Lenzini. 2025. DeceptiLens: an Approach supporting Transparency in Deceptive Pattern Detection based on a Multimodal Large Language Model. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency (FAccT '25)*. Association for Computing Machinery, New York, NY, USA, 1942–1959. https://doi.org/10.1145/3715275.3732129

[33] Geza Kovacs, Zhengxuan Wu, and Michael S. Bernstein. 2018. Rotating Online Behavior Change Interventions Increases Effectiveness But Also Increases Attrition. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 95 (Nov. 2018), 25 pages. https://doi.org/10.1145/3274364

[34] Emily Kuang, Minghao Li, Mingming Fan, and Kristen Shinohara. 2024. Enhancing UX Evaluation Through Collaboration with Conversational AI Assistants: Effects of Proactive Dialogue and Timing. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 3, 16 pages. https://doi.org/10.1145/3613904.3642168

[35] Klodiana Lanaj, Russell E. Johnson, and Christopher M. Barnes. 2014. Beginning the workday yet already depleted? Consequences of late-night smartphone use and sleep. *Organizational Behavior and Human Decision Processes* 124, 1 (2014), 11 – 23. https://doi.org/10.1016/j.obhdp.2014.01.001

[36] Simone Lanette, Phoebe K. Chua, Gillian Hayes, and Melissa Mazmanian. 2018. How Much is 'Too Much'?: The Role of a Smartphone Addiction Narrative in Individuals' Experience of Use. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW, Article 101 (Nov. 2018), 22 pages. https://doi.org/10.1145/3274370

[37] Tomas Lawton, Kazjon Grace, and Francisco J Ibarrola. 2023. When is a Tool a Tool? User Perceptions of System Agency in Human–AI Co-Creative Drawing. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference* (Pittsburgh, PA, USA) *(DIS '23)*. Association for Computing Machinery, New York, NY, USA, 1978–1996. https://doi.org/10.1145/3563657.3595977

[38] Tomas Lawton, Francisco J Ibarrola, Dan Ventura, and Kazjon Grace. 2023. Drawing with Reframer: Emergence and Control in Co-Creative AI. In *Proceedings of the 28th International Conference on Intelligent User Interfaces* (Sydney, NSW, Australia) *(IUI '23)*. Association for Computing Machinery, New York, NY, USA, 264–277. https://doi.org/10.1145/3581641.3584095

[39] Noah Levin. 2024. AI: The next chapter in design. https://www.figma.com/blog/ai-the-next-chapter-in-design/. Accessed: June 12, 2024.

[40] Jie Li, Hancheng Cao, Laura Lin, Youyang Hou, Ruihao Zhu, and Abdallah El Ali. 2024. User Experience Design Professionals' Perceptions of Generative Artificial Intelligence. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 381, 18 pages. https://doi.org/10.1145/3613904.3642114

[41] Kai Lukoff. 2022. *Designing to Support Sense of Agency for Time Spent on Digital Interfaces.* Ph. D. Dissertation. University of Washington. http://hdl.handle.net/1773/49196

[42] Kai Lukoff, Ulrik Lyngs, Karina Shirokova, Raveena Rao, Larry Tian, Himanshu Zade, Sean A. Munson, and Alexis Hiniker. 2023. SwitchTube: A Proof-of-Concept System Introducing "Adaptable Commitment Interfaces" as a Tool for Digital Wellbeing. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 197, 22 pages. https://doi.org/10.1145/3544548.3580703

[43] Kai Lukoff, Ulrik Lyngs, Himanshu Zade, J. Vera Liao, James Choi, Kaiyue Fan, Sean A. Munson, and Alexis Hiniker. 2021. How the Design of YouTube Influences User Sense of Agency. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 368, 17 pages. https://doi.org/10.1145/3411764.3445467

[44] Kai Lukoff, Cissy Yu, Julie Kientz, and Alexis Hiniker. 2018. What Makes Smartphone Use Meaningful or Meaningless? *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 1, Article 22 (mar 2018), 26 pages. https://doi.org/10.1145/3191754

[45] Ulrik Lyngs, Kai Lukoff, Petr Slovak, Reuben Binns, Adam Slack, Michael Inzlicht, Max Van Kleek, and Nigel Shadbolt. 2019. 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* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–18. https://doi.org/10.1145/3290605.3300361

[46] Ulrik Lyngs, Kai Lukoff, Petr Slovak, William Seymour, Helena Webb, Marina Jirotka, Jun Zhao, Max Van Kleek, and Nigel Shadbolt. 2020. 'I Just Want to Hack Myself to Not Get Distracted': Evaluating Design Interventions for Self-Control on Facebook. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–15.

[47] S M Hasan Mansur, Sabiha Salma, Damilola Awofisayo, and Kevin Moran. 2023. AidUI: Toward Automated Recognition of Dark Patterns in User Interfaces. In *Proceedings of the 45th International Conference on Software Engineering* (Melbourne, Victoria, Australia) *(ICSE '23)*. IEEE Press, 1958–1970. https://doi.org/10.1109/ICSE48619.2023.00166

[48] Microsoft. 2024. Microsoft Designer - Stunning designs in a flash. https://designer.microsoft.com/ Accessed: 2024-07-11.

[49] 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* (Pittsburgh, PA, USA) *(DIS '23)*. Association for Computing Machinery, New York, NY, USA, 2362–2374. https://doi.org/10.1145/3563657.3595964

[50] Thomas Mildner and Gian-Luca Savino. 2021. 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*. Association for Computing Machinery, New York, NY, USA, Article 464, 7 pages.

[51] Alberto Monge Roffarello and Luigi De Russis. 2019. The Race Towards Digital Wellbeing: Issues and Opportunities. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3290605.3300616

[52] Alberto Monge Roffarello and Luigi De Russis. 2021. Coping with Digital Wellbeing in a Multi-Device World. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 538, 14 pages. https://doi.org/10.1145/3411764.3445076

[53] Alberto Monge Roffarello and Luigi De Russis. 2023. Nudging Users Towards Conscious Social Media Use. In *Proceedings of the 25th International Conference on Mobile Human-Computer Interaction* (Athens, Greece) *(MobileHCI '23 Companion)*. Association for Computing Machinery, New York, NY, USA, Article 14, 7 pages. https://doi.org/10.1145/3565066.3608703

[54] Alberto Monge Roffarello, Luigi De Russis, and Kai Lukoff. 2025. The Digital Attention Heuristics: Supporting the User's Attention by Design. *ACM Trans. Comput.-Hum. Interact.* 32, 4, Article 38 (Aug. 2025), 41 pages. https://doi.org/10.1145/3725215

[55] Alberto Monge Roffarello, Luigi De Russis, and Massimiliano Pellegrino. 2024. Digital Wellbeing Lens: Design Interfaces That Respect User Attention. In *Proceedings of the 2024 International Conference on Advanced Visual Interfaces* (Arenzano, Genoa, Italy) *(AVI '24)*. Association for Computing Machinery, New York, NY, USA, Article 51, 5 pages. https://doi.org/10.1145/3656650.3656674

[56] Alberto Monge Roffarello, Luigi De Russis, R. X. Schwartz, and Panagiotis Apostolellis. 2022. Designing for Meaningful Interactions and Digital Wellbeing. In *Proceedings of the 2022 International Conference on Advanced Visual Interfaces* (Frascati, Rome, Italy) *(AVI 2022)*. Association for Computing Machinery, New York, NY, USA, Article 85, 2 pages. https://doi.org/10.1145/3531073.3535255

[57] 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* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 194, 19 pages. https://doi.org/10.1145/3544548.3580729

[58] OpenAI. 2024. Hello GPT-4o. https://openai.com/index/hello-gpt-4o/. Accessed: 2024-07-26.

[59] Dorian Peters. 2022. Wellbeing Supportive Design – Research-Based Guidelines for Supporting Psychological Wellbeing in User Experience. *International Journal of Human–Computer Interaction* 0, 0 (2022), 1–13. https://doi.org/10.1080/10447318.2022.2089812

[60] Savvas Petridis, Michael Terry, and Carrie J Cai. 2024. PromptInfuser: How Tightly Coupling AI and UI Design Impacts Designers' Workflows. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference* (IT University of Copenhagen, Denmark) *(DIS '24)*. Association for Computing Machinery, New York, NY, USA, 743–756. https://doi.org/10.1145/3643834.3661613

[61] Aditya Kumar Purohit, Louis Barclay, and Adrian Holzer. 2020. Designing for Digital Detox: Making Social Media Less Addictive with Digital Nudges *(CHI EA '20)*. Association for Computing Machinery, New York, NY, USA, 1–9. https://doi.org/10.1145/3334480.3382810

[62] Kris Rasmussen. 2024. Meet Figma AI: Empowering designers with intelligent tools. https://www.figma.com/blog/introducing-figma-ai/. Accessed: June 28, 2024.

[63] Jan Ole Rixen, Luca-Maxim Meinhardt, Michael Glöckler, Marius-Lukas Ziegenbein, Anna Schlothauer, Mark Colley, Enrico Rukzio, and Jan Gugenheimer. 2023. The Loop and Reasons to Break It: Investigating Infinite Scrolling Behaviour in Social Media Applications and Reasons to Stop. *Proc. ACM Hum.-Comput. Interact.* 7, MHCI, Article 228 (sep 2023), 22 pages. https://doi.org/10.1145/3604275

[64] Alberto Monge Roffarello and Luigi De Russis. 2023. Achieving Digital Wellbeing Through Digital Self-Control Tools: A Systematic Review and Meta-Analysis. *ACM Trans. Comput.-Hum. Interact.* 30, 4, Article 53 (sep 2023), 66 pages. https://doi.org/10.1145/3571810

[65] Lorena Sánchez Chamorro, Romain Toebosch, and Carine Lallemand. 2024. Manipulative Design and Older Adults: Co-Creating Magic Machines to Understand Experiences of Online Manipulation. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference* (Copenhagen, Denmark) *(DIS '24)*. Association for Computing Machinery, New York, NY, USA, 668–684. https://doi.org/10.1145/3643834.3661513

[66] Hauke Sandhaus. 2023. Promoting Bright Patterns. CHI '23 Workshop: Designing Technology and Policy Simultaneously. arXiv:2304.01157 [cs.HC]

[67] Cristiana Santos, Viktorija Morozovaite, and Silvia De Conca. 2024. No harm no foul: how harms caused by dark patterns are conceptualised and tackled under EU data protection, consumer and competition laws. (2024). June 26, 2024. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4877439.

[68] Nada Terzimehic, Florian Bemmann, Miriam Halsner, and Sven Mayer. 2023. A Mixed-Method Exploration into the Mobile Phone Rabbit Hole. *Proc. ACM Hum.-Comput. Interact.* 7, MHCI, Article 194 (sep 2023), 29 pages. https://doi.org/10.1145/3604241

[69] Jonathan A. Tran, Katie S. Yang, Katie Davis, and Alexis Hiniker. 2019. Modeling the Engagement-Disengagement Cycle of Compulsive Phone Use. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3290605.3300542

[70] Priyan Vaithilingam, Ian Arawjo, and Elena L. Glassman. 2024. Imagining a Future of Designing with AI: Dynamic Grounding, Constructive Negotiation, and Sustainable Motivation. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference* (IT University of Copenhagen, Denmark) *(DIS '24)*. Association for Computing Machinery, New York, NY, USA, 289–300. https://doi.org/10.1145/3643834.3661525

[71] Christopher Rhys Watkins, Colin M. Gray, Austin L. Toombs, and Paul Parsons. 2020. Tensions in Enacting a Design Philosophy in UX Practice. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference* (Eindhoven, Netherlands) *(DIS '20)*. Association for Computing Machinery, New York, NY, USA, 2107–2118. https://doi.org/10.1145/3357236.3395505

[72] WillowTree. 2024. Contrast. https://www.figma.com/community/plugin/748533339900865323/contrast Accessed: 2024-09-05.

[73] Richmond Y. Wong. 2021. Tactics of Soft Resistance in User Experience Professionals' Values Work. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2, Article 355 (Oct. 2021), 28 pages. https://doi.org/10.1145/3479499

[74] Wei Xiang, Hanfei Zhu, Suqi Lou, Xinli Chen, Zhenghua Pan, Yuping Jin, Shi Chen, and Lingyun Sun. 2024. SimUser: Generating Usability Feedback by Simulating Various Users Interacting with Mobile Applications. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 9, 17 pages. https://doi.org/10.1145/3613904.3642481

[75] 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 *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 60, 19 pages. https://doi.org/10.1145/3491102.3517722

[76] Shoshana Zuboff. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power.* PublicAffairs, New York.