

# Design Implications for LLM-Supported Consent and Control Mechanisms

CATHERINE R K O'BRIEN, University College London, United Kingdom

AMID AYOBI, University College London, United Kingdom

MARK WARNER, University College London, United Kingdom

## ACM Reference Format:

Catherine R K O'Brien, Amid Ayobi, and Mark Warner. 2026. Design Implications for LLM-Supported Consent and Control Mechanisms. , (2026), 3 pages. <https://doi.org/XXXXXXXX.XXXXXXX>

## 1 Introduction

Drawing on our research into the analysis of platform policies and behaviour-related documents on online harms, Large Language Model (LLM)-supported user reporting, and emerging literature on the use of LLMs for privacy policy comprehension, this position paper explores the potential use of LLMs in supporting user consent and control. We argue that LLMs could help users make sense of privacy policies and link that understanding directly to consent and privacy settings. However, we emphasise that such systems must be governed by explicit safeguards and design principles that promote user sense-making without steering outcomes, obscuring accountability, or further displacing interpretive burdens onto users.

## 2 Drawing Parallels Between the Notice and Consent Model and User Reporting

The prevailing “notice and consent” framework has been widely criticised. Privacy policies are rarely read and frequently misunderstood due to legalistic language, high reading levels, and the impractical amount of time required to engage with policies across services. Research has found that even when read, policies often fail to clearly communicate how data will be used or may misrepresent data protection practices. Furthermore, beyond policy texts, consent interfaces have been critiqued for shaping and constraining user choice, rather than enabling meaningful control. Design choices such as “manipulative patterns” can push users towards data-sharing and require substantial motivation from users to re-configure. As a result, consent mechanisms often place unrealistic cognitive and interpretative demands on users through privacy self-management.

In a related vein, our work on characterisation, enforcement actions, and safeguarding against harms in platform policies and behaviour-related documents [4, 5] has surfaced similar issues. These documents rely on ambiguous language, contain internal inconsistencies, and shift responsibility for safety onto users. And then, user reporting requires users to interpret these policies and translate complex, relational, or evolving experiences into rigid, predefined categories. Across both user reporting and consent, we see a shared friction: policies remain

---

Authors' Contact Information: Catherine R K O'Brien, [catherine.obrien@ucl.ac.uk](mailto:catherine.obrien@ucl.ac.uk), University College London, London, United Kingdom; Amid Ayobi, [amid.ayobi@ucl.ac.uk](mailto:amid.ayobi@ucl.ac.uk), University College London, London, United Kingdom; Mark Warner, [mark.warner@ucl.ac.uk](mailto:mark.warner@ucl.ac.uk), University College London, London, United Kingdom.

---

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

© 2026 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM XXXX-XXXX/2026/-ART

<https://doi.org/XXXXXXXX.XXXXXXX>

ambiguous while interfaces demand decisive categorisation, at the expense of contextual sensitivity and reflection, placing the burden of interpretation on users while platforms retain discretion over enforcement and outcomes.

Recent work explores the use of LLMs to support interpretation of privacy policies, arguing they can extract, annotate, and surface relevant policy information [2], and reconcile policy comprehensiveness with user comprehensibility [6]. However, this work largely focuses on the “notice” component of the notice-and-consent model, stopping short of examining how LLMs might support users in acting on that understanding through control mechanisms.

Insights from our work on LLM-supported user reporting offer a productive extension. In this context, we have argued for reframing reporting tools as sense-making and support scaffolds rather than purely classificatory systems [1, 5]. Through co-design workshops with Instagram users reporting ambiguous harassment, we developed user-centred design principles for LLM-supported reporting that balance contextual sensitivity with user burden, validation with the risk of false reporting, and personalised guidance with user agency [1]. We argue that LLMs should scaffold reflection and surface relevant considerations without overriding user judgement, aligning with proposals for “thinking assistants” [7] that support deliberation rather than dictate outcomes. Crucially, this work goes beyond standalone guidance by embedding reflective support directly within reporting flows, allowing users to move fluidly between interpretation, reflection, and action while retaining control over what is ultimately reported. Applied to consent and control, this perspective suggests that LLMs could similarly be used to support users in making sense of privacy policies *in situ*, while directly linking that understanding to consent mechanisms and privacy settings. Rather than treating comprehension and consent as separate steps, LLM-supported systems could integrate sense-making and action within a single flow.

### 3 Design Implications

Drawing on prior work, alongside emerging research on LLMs for privacy policy comprehension, we propose a set of design implications for the use of LLMs in consent and control mechanisms.

***The LLM acting as an interpretive scaffold rather than offering categorical judgment:*** LLM-assisted consent and control mechanisms could function as interpretive scaffolds without leading users to consent through manipulative patterns. Our co-design workshops highlighted that participants valued systems that operated as “thinking assistants”, helping them reflect on their experiences and surface relevant considerations, without enforcing categorical judgements. Translating this insight for consent and control, LLMs might support users in understanding how particular data practices relate to their concerns and circumstances without presenting recommendations as definitive, or taking actions to control settings.

***Personalisation without overriding user judgement:*** Aligning with critiques of opaque personalisation, our participants required such adaptability be designed with transparency and consent in mind. Similarly, in consent systems, personalised explanations or prompts should adapt to user needs without overriding user judgement or narrowing perceived options. Thus, personalisation should reduce interpretive and cognitive labour, without pre-empting decision-making. Furthermore, users may differ in their privacy expectations and vulnerabilities, for example, between adults and minors, and may require different protections and tailoring.

***Shared accountability through human–AI collaboration:*** Our prior work discussed how platforms often shift responsibility onto individuals in behaviour-related documents [4]. To avoid reproducing this dynamic, LLM-supported tools should be embedded within governance structures that maintain visible human accountability, with handover to human support when required, acting in human-AI collaboration.

***Seamless but transparent links between sense-making and action:*** In our reporting work, participants sought transparency and seamless integration between reporting, outcomes, and support. In consent context, LLM-supported sense-making should be directly, and transparently, linked with control mechanisms, allowing users to move between learning, reflection, and action. Thus retaining user control, and empowering them to

configure their own privacy controls with the appropriate knowledge. This supports well-reported work that argues for the integration of “notice” with usable tools for “consent”.

**Integrating aftercare, feedback, and ongoing support:** Finally, our co-design findings emphasise the importance of aftercare and ongoing support following reporting of harms. This is relevant to consent in two respects. Firstly, privacy concerns are often surfaced when harms occur and users seeking to re-configure their consent and understand the use of their data may be doing so in the context of the emotional distress caused by privacy violations. Thus, emotional support may also be relevant in this context. Secondly, understanding privacy as “situational”, with context-dependent views of privacy emerging in relation to experience, we might consider how what one consents to may change over time. Therefore, LLM-supported systems could provide follow-up information, reminders, or opportunities to revisit decisions – without engaging in manipulative patterns that cause consent fatigue. This would position consent as an ongoing and supported process, rather than a one-off decision, reflective of arguments for “dynamic consent” [3].

#### 4 Conclusion and Future Directions

This position paper has argued that LLMs have the potential to reshape engagement with consent and control through transforming static notice and consent interfaces into situated sense-making activities embedded within the context of use. Instead of asking users to read and interpret dense policies and translate their concerns into predefined settings, LLM-supported systems can help users reflect, surface relevant policy implications and relate specific data practices to their lived realities – all in language they can comprehend. Furthermore, integrating interpretation of policy with action within a single flow lowers cognitive and interpretive burdens, as users can informedly move directly from sense-making to consent. However, it is important to acknowledge that the same affordances that enable such adaptive support can also introduce risks around nudging or displaced accountability. To ensure that such LLM-supported consent and control mechanisms empower, rather than manipulate users, we argue for safeguards that ensure LLMs scaffold user sense-making without steering outcomes, preserve user agency, share accountability between humans and AI, and link interpretation, action, and ongoing support in transparent and revisable ways. Of course, the discussion presented here is based predominantly on our empirical findings from co-design workshops with users in the context of reporting. Consequently, future work should look to understand user needs in the consent and control context specifically.

#### References

- [1] Izzy Ferreira-Alturas, Catherine RK O’Brien, Amid Ayobi, and Mark Warner. Under Review. Co-Designing an LLM-Based User Reporting Tool in the Context of Ambiguous Targeted Harassment with UK Instagram Users. 0, 0 (Under Review), 0.
- [2] Arda Goknil, Femke B Gelderblom, Simeon Tverdal, Shukun Tokas, and Hui Song. 2024. Privacy policy analysis through prompt engineering for llms. *arXiv preprint arXiv:2409.14879* (2024).
- [3] Hyunsoo Lee and Uichin Lee. 2022. Toward dynamic consent for privacy-aware pervasive health and well-being: A scoping review and research directions. *IEEE Pervasive Computing* 21, 4 (2022), 25–32.
- [4] Catherine RK O’Brien, Nuur Alifah Roslan, Steven J Murdoch, Ruba Abu-Salma, Douglas Zytko, and Mark Warner. 2025. Online Dating Platform Safeguards and Self-Protection: How Dating Platforms Characterise, Respond to, and Safeguard Against Harms. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–8.
- [5] Catherine RK O’Brien, Douglas Zytko, Ruba Abu-Salma, Leonie Marie Tanczer, Nuur Alifah Roslan, Steven J Murdoch, and Mark Warner. In Preparation. Harassing Harms: A Framework for Understanding How Online Social Platforms Characterise Harassment. 0, 0 (In Preparation), 0.
- [6] Przemysław Pałka, Francesca Lagioia, Rūta Liepina, Marco Lippi, and Giovanni Sartor. 2025. Make privacy policies longer and appoint LLM readers. *Artificial Intelligence and Law* (2025), 1–33.
- [7] Soya Park, Hari Subramonyam, and Chinmay Kulkarni. 2023. Thinking assistants: Llm-based conversational assistants that help users think by asking rather than answering. *arXiv preprint arXiv:2312.06024* (2023).

Received ; revised ; accepted