

Transparent Hearts: Balancing Privacy and Trust in AI-Generated Self-Presentation for Online Dating

Meryem Barkallah
University of Michigan-Flint
Flint, Michigan, USA
meryemba@umich.edu

Yosr Aissa
University of Michigan-Flint
Flint, Michigan, USA
yosr@umich.edu

Douglas Zytko
University of Michigan-Flint
Flint, Michigan, USA
dzytko@umich.edu

Abstract

Dating app users can now utilize AI features to streamline and improve their self-presentation to potential meeting partners, such as auto-generated bios or AI-modified photos. While these tools can aid impression management, they can also create doubts about authenticity and trust, raising new questions about if and how automated disclosure of AI-generated content should be implemented. To address these tensions, we designed a prototype called Transparent Hearts, informed by a self-reflective diary study and participatory design insights. By balancing needs for privacy and disclosure, Transparent Hearts seeks to empower daters to customize how much of their own usage of AI-modified content is disclosed to others, while also establishing settings for disclosure of AI in other users' content. In addition to user-controlled AI-disclosure toggles and explanations of AI-use, the prototype includes a conflict resolution mechanic for when two users have mismatched AI disclosure settings. This work aligns with SDGs 10 and 16, highlighting design considerations for fairness, inclusivity, and trust in AI-mediated social platforms.

Keywords

AI; online dating; dating apps; generative AI; self-presentation; transparency; authenticity; disclosure

ACM Reference Format:

Meryem Barkallah, Yosr Aissa, and Douglas Zytko. 2025. Transparent Hearts: Balancing Privacy and Trust in AI-Generated Self-Presentation for Online Dating. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25)*, April 26–May 01, 2025, Yokohama, Japan. ACM, New York, NY, USA, 8 pages. <https://doi.org/10.1145/3706599.3720311>

1 Introduction and Motivation

Online dating has revolutionized how individuals form romantic and social connections [10]. At the same time, generative AI tools have transformed self-presentation capabilities across a range of social computing contexts [9, 12, 19, 20]. In terms of online dating, this includes auto-generating profile bios, retouching profile photos, and suggesting chat responses to potential dating partners [8, 12, 19]. These capabilities have clear benefits for self-presentation. For example, a user who is not fluent in the platform's primary language

can utilize AI-generated bios to effectively communicate their interests and intentions. Likewise, a user could alter or generate profile pictures that best align with their personal identity.

AI-modified self-presentation in online dating also introduces critical challenges for users who are the recipients of such content [12]. If users suspect that profiles are heavily curated or “faked” through AI enhancements, trust can be eroded, undermining the authenticity of interactions. This skepticism leads to tensions around consent, authenticity, and disclosure, as individuals navigate the fine line between self-presentation and genuine connection.

Our research explores the emerging dual role of generative AI in online dating—enabling self-presentation while moderating disclosure of AI generated content. This investigation draws on insights from a participatory design study conducted by our lab, which examines how online daters anticipate and desire AI to shape contemporary platforms (including immersive VR contexts). Building on that work, we conducted a self-reflective diary study focusing on user-controlled AI disclosure and conflict resolution. By “moderating AI disclosure and conflicts” we refer to using AI capabilities in the app to detect whether content (bios, photos, messages) has been AI-generated or significantly modified, and then providing mechanics to resolve any disclosure conflicts that arise when users have differing preferences about revealing AI usage.

We propose Transparent Hearts, a design prototype that addresses these issues by allowing users to choose their preferred level of generative AI transparency. By doing so we aim to address United Nations Sustainable Development Goals (SDGs) 10 (Reduced Inequalities) and 16 (Peace, Justice and Strong Institutions), highlighting inclusivity and just moderation in online dating.

To ground our design, we address the following research questions in this initial exploration:

RQ1: *How can an online dating platform detect and label AI-generated or edited content while respecting user autonomy, privacy, and the potential desire for confidentiality?*

RQ2: *Which interface features might allow individuals to customize their level of AI disclosure, and how can we handle conflicts between users with different transparency preferences?*

RQ3: *In what ways can an AI intervene to moderate unwanted content, and how might this dynamic influence trust, transparency, or user satisfaction?*

Although the questions were not formally tested with a user population, they guided each iteration of our prototype design, illuminating potential trade-offs between transparency, privacy, moderation, and freedom of expression.

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 third-party components of this work must be honored. For all other uses, contact the owner/author(s).

CHI EA '25, Yokohama, Japan

© 2025 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-1395-8/25/04

<https://doi.org/10.1145/3706599.3720311>

2 Background and Related Work

2.1 AI in Online Dating

AI has long been integrated into online dating apps, primarily to enhance matchmaking and provide personalized recommendations by analyzing user data, such as interests and preferences, to suggest better matches. However, AI's role in dating apps has evolved significantly. Recent research highlights the need for AI to act as a risk detection tool, aiming to prevent interpersonal harm in dating apps [6, 7]. Moreover, many online dating platforms have invested in developing AI-driven safety tools [15, 16].

This expanding role of AI now encompasses addressing users' concerns about self-presentation. A significant number of online daters already use generative AI tools to navigate online dating and improve their chances of finding a match. A recent study reveals that 71% of people are interested in using AI to craft pick-up lines and conversation starters, 70% for developing dating app profiles, and 64% for enhancing photos [14].

Some studies have explored the impact of AI-driven enhancements on users' perceptions of attractiveness and trust in online dating [12, 19]. Wu et al [19] conducted a study with 48 participants who were presented with the text of 10 dating profiles and were told that the profiles had been written by humans or with the help of AI. Findings of this study indicate that the perceived involvement of AI did not significantly affect the profile's attractiveness, but did result in a notable decrease in the perceived trustworthiness of the profile author.

Recent proposals recommend labeling AI-generated content [18], yet few systematically investigate how much control users want over such labeling.

2.2 Content Moderation in Digital Platforms

Content moderation in social and dating apps has typically relied on a combination of human oversight and algorithmic filtering to detect unwanted behaviors [2, 11]. With the overwhelming surge of harmful content on social media platforms, the manual methods for identifying such content are no longer sufficient. Major social media platforms like Facebook, Twitter, and YouTube have made huge investments to fully automate this process of detection and moderation and also framed policies to detect and moderate such detrimental content [2, 11].

This shift brings significant benefits, such as minimizing human exposure to harmful content. Additionally, AI surpasses human capabilities by analyzing and detecting vast amounts of user-generated content at an unparalleled scale and speed [5]. However, some scholars warn against heavy-handed filtering that can suppress legitimate expression [1].

Content moderation in dating apps has a significant impact on users' ratings of trust, as it can raise questions around deception, consent and authenticity of interactions. Moderation of such content is a necessity, but balancing user autonomy with protective guardrails remains a pressing challenge in HCI.

3 Participatory Design as Impetus for Transparent Hearts

Our self-reflective diary study is situated within a broader, multi-year research initiative focused on understanding and designing computer-mediated consent as a lens for mitigating unwanted behavior, particularly in sexual experiences.

Before conducting our self-reflective diary study, our lab undertook a participatory design (PD) approach with 17 women and LGBTQIA+ individuals about contemporary online dating technologies and their role both in perpetuating harm and maintaining user safety, particularly AI and AR/VR environments (e.g., [3, 17]). Most crucial to this paper's focus on AI-generated self-presentation and automated disclosure of AI-generated content, participants talked at length about anticipated interactions with AI to communicate and clarify information between dating partners. Specifically, they imagined AI as a vehicle for communicating consent to particular interactions across fully virtual and physical interactions. Likewise, they also expressed concern about new capabilities for augmenting—and distorting—self-presentation online that departs from one's physical-world self. While this was mostly discussed in regard to virtual avatar construction, we reflected on these two findings collectively to motivate a subsequent research focus on the implications of generative AI on informed consent to interaction in online dating contexts.

4 Methods

4.1 Self-Reflective Diary Study: Rationale and Limitations

Given the early, speculative nature of our work, we adopted a self-reflective diary study [4, 13] to rapidly iterate on design ideas without the logistical constraints of recruiting external participants. Between December 20, 2024, and January 17, 2025, members of the research team sketched and refined prototypes daily, keeping detailed diaries of our evolving ideas and critiques. We deliberately avoided becoming overly attached to any single prototype to prioritize a human-centered focus: how each design could empower individual users to define their own privacy, articulate personal authenticity, control self-presentation, establish trust, and resolve conflicts in AI-mediated dating scenarios. We continually documented how well each prototype addressed these goals and used daily reflections to critique our previous iterations through a fresh, neutral lens.

Although this self-reflective approach was internal, it was informed by insights from a prior VR participatory design (PD) study in our lab, which identified core concerns regarding AI-based consent, authenticity, and self-presentation in online dating contexts. Their input also shaped our focus on conflict resolution, guiding us to consider how AI might mediate or help resolve disagreements between users who have differing opinions or comfort levels. In practice, we integrated these PD findings into our diary reflections to ensure our speculative ideas resonated with real-world user perspectives rather than existing in isolation.

We acknowledge that having the researchers serve as both designers and participants limits the generalizability of our findings.

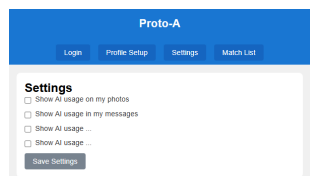
Additionally, relying on daily self-critiques means that our iterative process may have been influenced by our team’s own biases and personal experiences. To mitigate these risks, we consistently reviewed each day’s prototype with a fresh lens and challenged assumptions rooted in our prior ideas. Looking ahead, we plan follow-up user tests to gather broader, more diverse feedback and further refine our designs.

Despite these limitations, the self-reflective diary study proved an efficient first step in articulating and exploring key design features—especially around balancing privacy, disclosure, trust, and conflict resolution—in AI-mediated dating platforms.

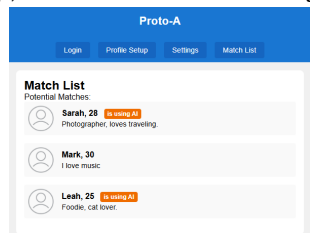
4.2 Prototyping Transparent Hearts

Our design process spanned four weeks (December 20, 2024–January 17, 2025) of near-daily reflections, mockups, and iterative testing. The resulting prototypes, labeled Proto-A, Proto-B, Proto-C, and Proto-Final, evolved to address user autonomy, conflict resolution, and color-coded AI labeling. Below, we recount pivotal design shifts in chronological order.

4.2.1 December 20–24, 2024: Initial Conceptualization. On December 20, the first sketches revolved around a simple “AI-Enhanced?” toggle next to each photo (see Figure 1b (Proto-A2)). Diary notes suggested an “Explainable AI” feature that might quantify how much an image was AI-edited (e.g., “This image is 40% AI-edited”). Within days, the single toggle felt too coarse; by December 23, Proto-A introduced a basic settings screen (see Figure 1a (Proto-A1)) allowing more nuanced controls over AI usage in bios, photos, or messages. A diary entry read, “It’s too minimal... Next iteration needs more nuance.”



(a) Proto-A1: AI disclosure settings



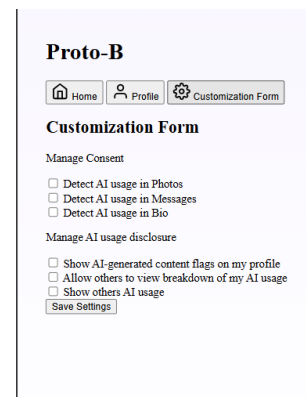
(b) Proto-A2: AI disclosure in Match List.

Figure 1: Proto-A: Early AI-Disclosure Features. (a) AI disclosure settings screen with toggle options. (b) Match List interface labeling profiles that use AI-generated content.

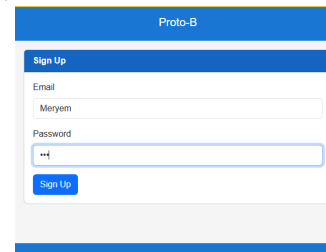
4.2.2 December 25–30, 2024: First Iterative Refinement. By December 26, a **Customization Form** (see Figure 2a (Proto-B1)) was redesigned into a collapsible tab or wizard to improve usability and allow users to configure AI settings more intuitively. **Proto-B**

introduced toggles for *privacy, analytics, and custom AI thresholds*, providing users with greater control over AI-generated content. However, initial testing revealed that placing too many options on a single screen was overwhelming. On December 28, a diary note reflected this: “*It took me far too long to scroll... imagine how new users might feel!*”

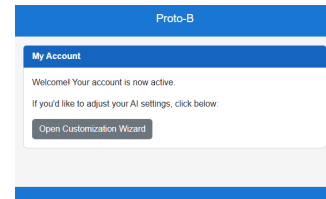
To address this, we implemented a **step-by-step wizard** (see Figure 2b (Proto-B2)) during sign-up, ensuring users could set preferences early without information overload. The wizard segmented settings into structured sections like “*Personal Info*,” “*AI Detection*,” and “*Ethical Settings*,” making navigation more intuitive. After account activation, users were prompted to *open the customization wizard* (see Figure 2c (Proto-B3)) to choose and define their AI preferences.



(a) Proto-B1: Initial customization form.



(b) Proto-B2: AI settings wizard during sign-up.



(c) Proto-B3: Account activation and customization setup

Figure 2: Proto-B: Early Steps in the AI Customization Wizard Flow. (a) Initial customization form with privacy settings. (b) AI settings wizard introduced during sign-up to guide user preferences. (c) Account activation screen prompting users to open the customization wizard for additional AI settings.

Further refinements improved onboarding flow. Figure 3a (Proto-B4) shows the wizard’s introduction to AI settings, providing an overview of personal preferences. Once users proceed, they adjust AI detection and transparency settings (see Figure 3b (Proto-B5)) before finalizing choices. These refinements reinforced a **progressive disclosure approach**, where essential settings were introduced early while allowing adjustments later.

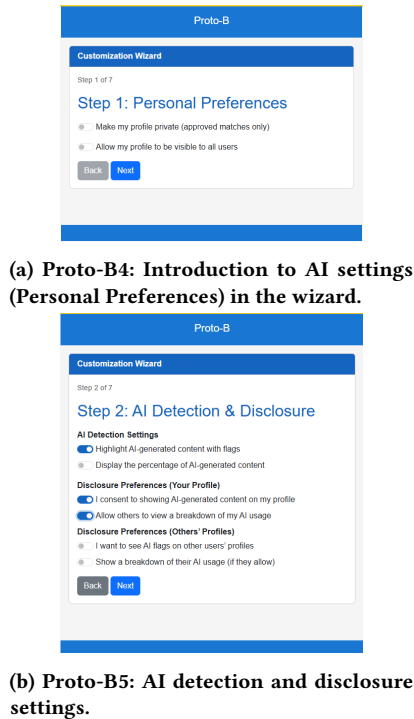


Figure 3: Proto-B: Advanced Steps in the AI Customization Wizard Flow. (a) The introduction screen guiding users through AI customization. (b) AI detection and disclosure settings for user-controlled transparency.

4.2.3 *December 31, 2024–January 10, 2025: Addressing “Disclosure Conflict” and Refining AI Disclosure Indicators.* By December 31, it became evident that mismatched user preferences—where one user prefers full transparency while another opts to hide AI usage—posed a challenge to user trust and communication. In internal discussions, we explored possible solutions to this issue, considering whether AI disclosure should be enforced, made fully optional, or mediated in a way that respects both perspectives. After weighing the trade-offs, we determined that a “conflict resolution mechanic” was necessary to allow users to negotiate disclosure preferences dynamically.

To illustrate this challenge, Figure 4a (Proto-C1) shows a scenario where “User A” requests AI transparency, but “User B” has disabled AI disclosure. To address this, we introduced a pop-up conflict resolution mechanic (see Figure 4b (Proto-C2)) on January 1. This prompt informs User B: “User A requests AI transparency, but you have it disabled. Do you want to proceed or revise your settings?” This approach was inspired by VR-based discussions on AI moderation, where users expressed a desire for an AI-driven

“consent agent” to help mediate disputes rather than enforcing rigid rules.

The final refinement of this feature (see Figure 4c (Proto-C3)) shows the outcome when User B chooses to disclose their AI usage. By allowing users to make real-time decisions rather than imposing static rules, this iterative process aimed to balance user autonomy with mutual transparency expectations, ensuring that disclosure conflicts are resolved in a way that respects both parties’ preferences.

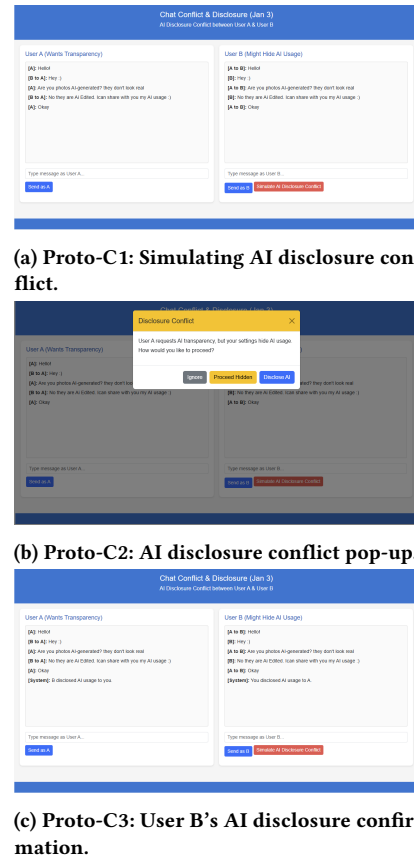
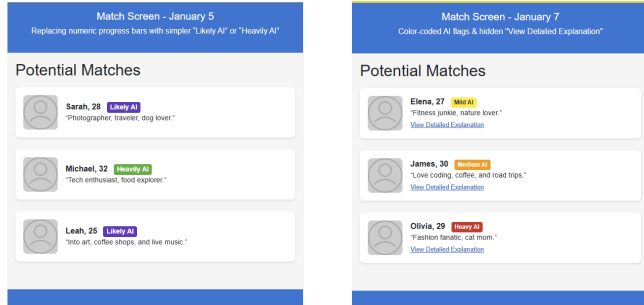


Figure 4: Proto-C: AI Disclosure Conflict Resolution. (a) A feature to simulate AI disclosure mismatches before interactions. (b) A system prompt requesting User B to adjust AI transparency settings. (c) User B confirming AI disclosure, resolving the conflict dynamically.

On January 3, Proto-C introduced “color-coded progress bars” to indicate the extent of AI-generated modifications in profile photos, displaying percentages such as “30%, 60%, or 90% AI” (see Figure 5c (Proto-C6)). However, our feedback revealed that numerical values alone were not intuitive enough and contributed to cognitive overload.

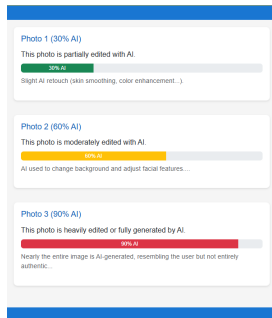
To simplify interpretation, these bars were replaced on January 5 with “text-based labels” such as “Likely AI” or “Heavily AI” (see Figure 5a (Proto-C4)), providing a more user-friendly way to communicate AI involvement without overwhelming users with

numbers. Further refinements followed on January 7, introducing a "color palette system" to enhance visual clarity. "Yellow" was assigned for "mild AI edits," "orange" for "moderate AI use," and "red" for "fully AI-generated images" (see Figure 5b (Proto-C5)). This revision aimed to strike a balance between "awareness and usability," ensuring that users were informed about AI modifications without experiencing unnecessary alert fatigue.



(a) Proto-C4: Text-based AI disclosure labels.

(b) Proto-C5: Color-coded AI modification levels.



(c) Proto-C6: AI detection with percentage indicators.

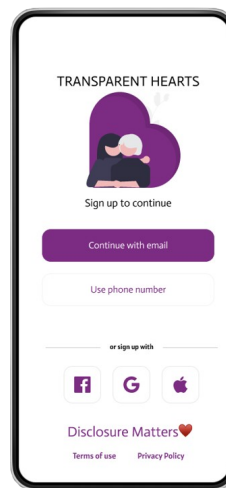
Figure 5: Proto-C: Refinements in AI Disclosure. (a) Simplified text labels replace numerical AI usage indicators. (b) Color-coded AI modification levels enhance clarity. (c) AI detection interface displaying AI-generated content as percentage values.

4.2.4 January 11–17, 2025: Final Integration and Polishing. The final prototype, Proto-Final, was assembled in mid-January. This high-fidelity mockup introduced toggles like “Share My AI Usage Stats” and user-specific settings to see or hide others’ AI usage in photos, bios, or messages. A concluding diary entry on January 17 read: “now have a comprehensive system that addresses user autonomy, color-coded disclosures, and an AI ‘consent agent.’ But it’s still a personal reflection with no direct user input on generative AI in dating.” These prototypes set the stage for future research and user testing.

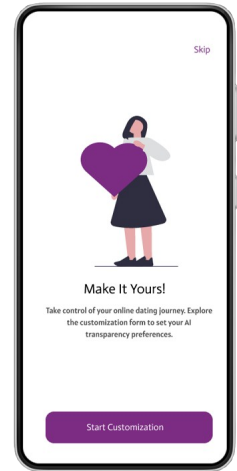
The onboarding process begins with a login screen (see Figure 6a (Final1)), followed by a prompt to start AI customization (see Figure 6b (Final2)). Users navigate AI settings through a collapsible menu (see Figure 7a (Final3)), where they can manage personal preferences (see Figure 7b (Final4)), AI detection settings (see Figure 8a (Final5)), and profile insights (see Figure 8b (Final6)). Additional

options include ethical AI feedback (see Figure 9a (Final7)) and AI detection threshold customization (see Figure 9b (Final8)).

AI usage indicators are visible in the match list (see Figure 10a (Final9)) and chat (see Figure 10b (Final10)). If a user questions another’s AI usage, a conflict resolution prompt appears (see Figure 11a (Final11)), followed by a system pop-up (see Figure 11b (Final12)). Users then receive confirmation messages indicating whether AI usage was disclosed (see Figure 12a (Final13)) or kept private (see Figure 12b (Final14)).

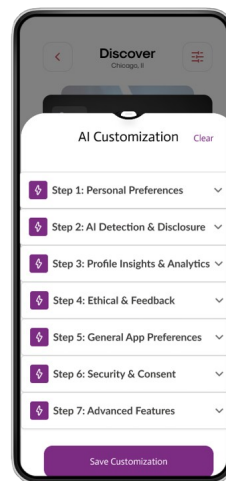


(a) Final1: Login Page.

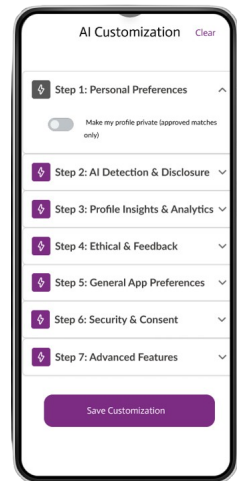


(b) Final2: Start AI Customization.

Figure 6: Proto-Final: Sign-up and Initial AI Customization Prompt.

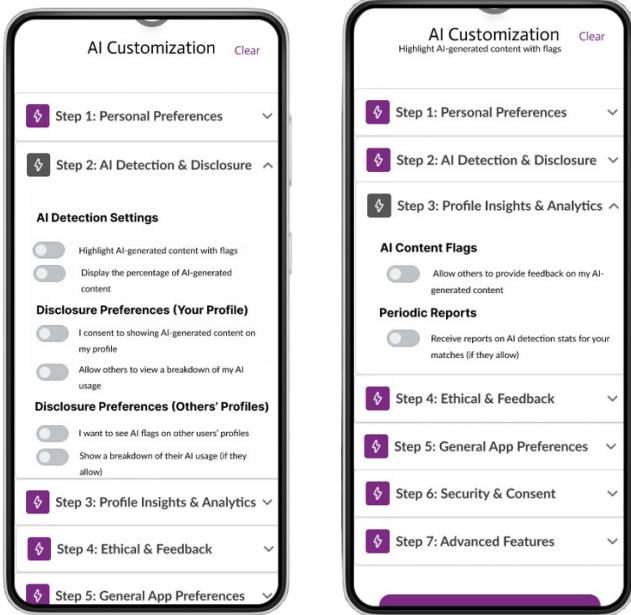


(a) Final3: AI Customization collapsible menu.



(b) Final4: Personal Preferences.

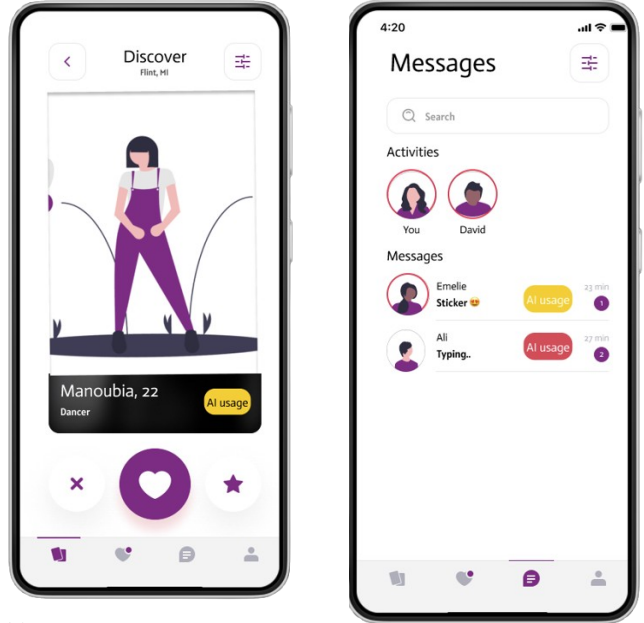
Figure 7: Proto-Final: AI Customization collapsible menu and Personal Settings.



(a) Final5: AI Detection & Disclosure.

(b) Final6: Profile Insights & Analytics.

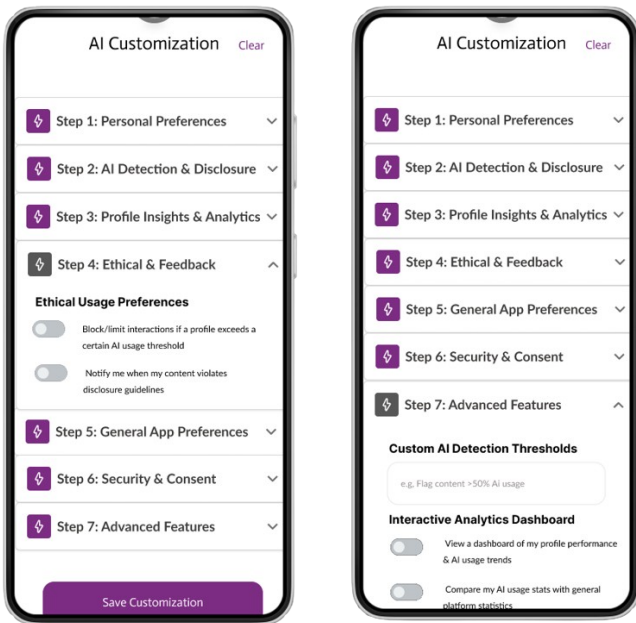
Figure 8: Proto-Final: AI Detection and Profile Insights Settings.



(a) Final9: AI Usage in Match List.

(b) Final10: AI Usage in Chat.

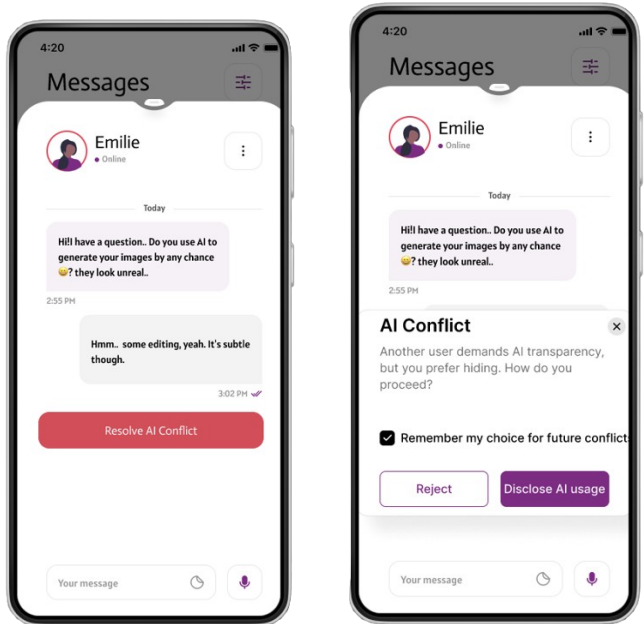
Figure 10: Proto-Final: AI Usage Indicators in Match List and Chat Interface.



(a) Final7: Ethical AI & Feedback.

(b) Final8: AI Detection Threshold Customization.

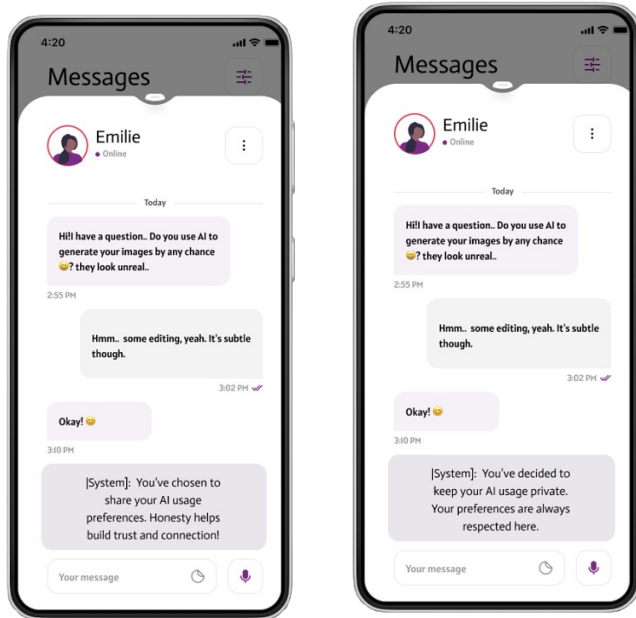
Figure 9: Proto-Final: Ethical AI Controls and Custom AI Detection Settings.



(a) Final11: AI Usage Conflict Triggered.

(b) Final12: AI Conflict Resolution Pop-up.

Figure 11: Proto-Final: AI Disclosure Conflict Detection and Resolution Prompt.



(a) Final13: AI Usage Disclosed Confirmation.

(b) Final14: AI Usage Kept Private.

Figure 12: Proto-Final: AI Usage Disclosure and System Confirmation Messages.

5 Design Requirements

We synthesized four main requirements during the iterative design process. Table 1 summarizes each requirement, its description, and rationale.

Table 1: Design Requirements for Transparent Hearts

Requirement	Description and Rationale
AI-Disclosure Toggles	Let users choose how to display AI usage (e.g., in bios, images). This addresses the tension of authenticity vs. privacy.
Conflict Resolution Mechanic	Provide an overlay for mismatched transparency preferences (e.g., pop-up if users disagree). Minimizes friction and fosters disclosure.
Color-Coded AI Indicators	Use a simple color scheme (yellow, orange, red) for mild, moderate, or fully generated content, providing quick interpretability for trust and transparency.
Flexible Moderation Levels	Let users select Light, Moderate, or Strict content filtering. Avoids paternalistic oversight; acknowledges varying comfort levels. Transparency is relative.

6 Linking to VR-Based Insights on Consent and AI

Throughout the diary process, we drew on the participatory design work in VR dating contexts, where participants imagined an AI “monitor” or “consent agent.” For instance, one VR participant (P5) described wanting “a third individual... maybe an AI bot” to intervene when things go wrong, while another called for “consent of both parties” before system-level monitoring was triggered. These insights shaped the “AI consent agent” in Transparent Hearts—a feature that warns or blocks unwanted things but remains configurable.

Users in VR contexts worried about overreach—“flagging certain phrases” could be helpful but also intrusive. Similarly, we enabled toggles for AI disclosure (rather than forcing universal labeling) and built a “Disclosure Conflict” pop-up for mismatched preferences. Although we have not yet tested these features with actual mobile daters who regularly use generative AI, the VR themes around AI-based monitoring and user discomfort with paternalistic oversight provided invaluable guidance.

7 Preliminary Observations and Discussion

By the conclusion of the diary study, the Transparent Hearts prototype offered a flexible approach to AI-influenced profile creation and moderation. Our observations indicate that user-controlled toggles help mitigate stigmatization for those relying on AI (e.g., for language translation), while color-coded indicators increase transparency for users concerned about authenticity. The “Disclosure Conflict” overlay addresses a real tension: some users want more honesty, others want privacy.

However, without direct user studies, these features remain speculative. Next steps will include user testing—through interviews, surveys, or pilot deployments—to gauge whether the conflict resolution mechanic truly resolves transparency disputes and how comfortable users are with color-coded AI labels. A deeper investigation of pop-up fatigue or perceived intrusiveness is also needed. Understanding diverse user populations and cultural norms will be crucial for refining these design elements.

8 Alignment with SDGs

Our work aligns with SDG 10 (Reduced Inequalities) by offering users who depend on AI—for instance, non-native language speakers—a way to communicate effectively without being unfairly labeled as “fake.” Simultaneously, SDG 16 (Peace, Justice, and Strong Institutions) is supported through the optional AI “consent agent” that flags or filters unwanted content while remaining configurable to avoid oppressive overreach. This approach to equitable and just moderation underscores the potential for AI-driven content to foster social good in digital dating platforms.

9 Conclusion and Future Work

Transparent Hearts demonstrates a design approach that seeks to balance AI-enhanced self-presentation with user consent and authenticity in online dating. Developed through a self-reflective diary study and informed by VR-based participatory design insights, the prototype introduces four key features: AI-disclosure toggles, a

conflict resolution mechanic, color-coded AI indicators, and flexible moderation levels. These features collectively address tensions between privacy vs. transparency and freedom vs. protection from unwanted content.

Moving forward, we plan to conduct user interviews, large-scale surveys, and potentially pilot tests of Transparent Hearts to validate its real-world effectiveness. Questions remain regarding the usability of multi-level toggles, the acceptability of color-coded AI labels, and the viability of conflict resolution prompts across diverse cultures and dating norms. Ultimately, we aim to refine Transparent Hearts into a widely acceptable tool for ethical AI-driven matchmaking, contributing to fair, inclusive, and secure interactions online.

Acknowledgments

This material is based upon work partially supported by the U.S. National Science Foundation under Grant No. 2211896, 2401775, 2339431. We would like to thank our lab members Braeden Burger, Devin Tebbe, and Emma Walquist for their invaluable and insightful feedback. We also extend our gratitude to the University of Michigan–Flint for supporting this research endeavor.

References

- [1] Oghenefejiro Isaacs Anigboro, Charlie M. Crawford, Danaë Metaxa, and Sorelle A. Friedler. 2024. Identity-related Speech Suppression in Generative AI Content Moderation. arXiv:2409.13725 [cs.CL] <https://arxiv.org/abs/2409.13725>
- [2] Oversight Board. 2024. *Content Moderation in a New Era for AI and Automation*. haha. <https://www.oversightboard.com/news/content-moderation-in-a-new-era-for-ai-and-automation/>
- [3] Braeden Burger, Devin Tebbe, Emma Walquist, Toby Kind, and Douglas Zytko. 2025. Saying No to “Yes Means Yes”: Limitations of Affirmative Consent for Mitigating Unwanted Behavior Online According to Women and LGBTQIA+ Stakeholders. In *CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) (CHI '25). ACM, New York, NY, USA, 17 pages. doi:10.1145/3706598.3713236
- [4] Maral Babapour Chafi, Ulrike Rahe, and Owain Pedgley. 2012. The Influence of Self-reflective Diaries on Students’ Design Processes. <https://api.semanticscholar.org/CorpusID:107173882>
- [5] Rem Darbinyan. 2022. Council Post: The Growing Role Of AI In Content Moderation. <https://www.forbes.com/councils/forbestechcouncil/2022/06/14/the-growing-role-of-ai-in-content-moderation/>
- [6] Isha Datey, Hanan Khalid Aljasim, and Douglas Zytko. 2022. Repurposing AI in Dating Apps to Augment Women’s Strategies for Assessing Risk of Harm. In *Companion Publication of the 2022 Conference on Computer Supported Cooperative Work and Social Computing*. ACM, Virtual Event, Taiwan, 150–154. doi:10.1145/3500868.3559472
- [7] Isha Datey, Hanan Khalid Aljasim, and Douglas Zytko. 2022. Repurposing AI in Dating Apps to Augment Women’s Strategies for Assessing Risk of Harm. In *Companion Publication of the 2022 Conference on Computer Supported Cooperative Work and Social Computing (CSCW’22 Companion)*. Association for Computing Machinery, New York, NY, USA, 150–154. doi:10.1145/3500868.3559472
- [8] Jaleesa Elliott. 2024. ‘Singles in America’ study: Daters breaking the ice with AI. <https://news.iu.edu/live/news/34137-singles-in-america-study-daters-breaking-the-ice>
- [9] Nicole Ellison, Rebecca Heino, and Jennifer Gibbs. 2006. Managing Impressions Online: Self-Presentation Processes in the Online Dating Environment. *Journal of Computer-Mediated Communication* 11, 2 (Jan. 2006), 415–441. doi:10.1111/j.1083-6101.2006.00020.x
- [10] Eli J. Finkel, Paul W. Eastwick, Benjamin R. Karney, Harry T. Reis, and Susan Sprecher. 2012. Online Dating: A Critical Analysis From the Perspective of Psychological Science. *Psychological Science in the Public Interest* 13, 1 (Jan. 2012), 3–66. doi:10.1177/1529100612436522
- [11] Vaishali U. Gongane, Mousami V. Munot, and Alwin D. Anuse. 2022. Detection and moderation of detrimental content on social media platforms: current status and future directions. *Social Network Analysis and Mining* 12, 1 (Sept. 2022), 129. doi:10.1007/s13278-022-00951-3
- [12] Jeffrey T Hancock, Mor Naaman, and Karen Levy. 2020. AI-Mediated Communication: Definition, Research Agenda, and Ethical Considerations. *Journal of Computer-Mediated Communication* 25, 1 (March 2020), 89–100. doi:10.1093/jcmc/zmz022
- [13] Pelin Karaturhan, Ecem Arıkan, Pelin Durak, Asim Evren Yantac, and Kemal Kuscı. 2022. Combining Momentary and Retrospective Self-Reflection in a Mobile Photo-Based Journaling Application. In *Nordic Human-Computer Interaction Conference (NordiCHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–12. doi:10.1145/3546155.3546676
- [14] Malea Lamb-Hall and Courtney Rowles. 2024. Swipe Right on Safety: New Study from Norton Finds 64% of Online Daters are Interested in Using AI as a Dating Coach. <https://investor.gendigital.com/news/news-details/2024/Swipe-Right-on-Safety-New-Study-from-Norton-Finds-64-of-Online-Daters-are-Interested-in-Using-AI-as-a-Dating-Coach/default.aspx>
- [15] Tinder Newsroom. 2020. Tinder Introduces Safety-Focused Updates. <https://au.tinderpressroom.com/tinder-introduces-safety-updates>
- [16] Tinder Newsroom. 2021. Tinder Introduces Are You Sure?, an Industry-First Feature That is Stopping Harassment Before It Starts. <https://au.tinderpressroom.com/news?item=122506>
- [17] Devin Tebbe, Meryem Barkallah, Braeden Burger, and Douglas Zytko. 2024. Risk of Harm in VR Dating from the Perspective of Women and LGBTQIA+ Stakeholders. doi:10.48550/arXiv.2405.05914 arXiv:2405.05914 [cs.HC] arXiv:2405.05914 [cs].
- [18] anna washenko. 2024. Bumble wants users to report AI-generated images. <https://www.engadget.com/bumble-wants-users-to-report-ai-generated-images-203627777.html>
- [19] Yihan Wu and Ryan M. Kelly. 2021. Online Dating Meets Artificial Intelligence: How the Perception of Algorithmically Generated Profile Text Impacts Attractiveness and Trust. In *Proceedings of the 32nd Australian Conference on Human-Computer Interaction (OzCHI '20)*. Association for Computing Machinery, New York, NY, USA, 444–453. doi:10.1145/3441000.3441074
- [20] Douglas Zytko, Sukeshini A. Grandhi, and Quentin Jones. 2014. Impression Management Struggles in Online Dating. In *Proceedings of the 2014 ACM International Conference on Supporting Group Work (GROUP '14)*. Association for Computing Machinery, New York, NY, USA, 53–62. doi:10.1145/2660398.2660410