Beyond Transparency: Using AI-Generated Narratives to Surface Algorithmic Identity

Meet Ahluwalia
Culture and Enterprise
Central Saint Martins, University of Arts London
London, United Kingdom
m.ahluwalia0620241@arts.ac.uk

ABSTRACT

Recommendation algorithms shape user identity through invisible profiling, yet users lack tools to understand how platforms see them. We present a Research-through-Design study using AI-generated narratives to surface algorithmic identity on Instagram. Through sessions with 25 participants (ages 23-41), we found that creative narration enables meaningful reflection on algorithmic selfhood. Our findings reveal three reflective mechanisms: conceptual gifting (AI provides language users lack to describe themselves), affective dissonance (cognitive tension between self-perception and algorithmic interpretation), and algorithmic authorship (desire to shape one's digital identity). We contribute design implications for reflective AI systems that support metacognition rather than technical transparency.

CCS CONCEPTS • *Human-centered computing → HCI theory, concepts and models*

KEYWORDS Algorithmic identity, reflection, AI, metacognition, Research- through-design, Explainable AI

1. INTRODUCTION

Recommendation algorithms on platforms like Instagram continuously profile users, constructing digital identities through behavioral inference. Every click, pause, and scroll trains the algorithm to recognize patterns in user interests, values, and emotional states. Yet this algorithmic identity remains invisible to users, creating a fundamental asymmetry: platforms know us in ways we cannot access.

While transparency initiatives reveal what data platforms collect, they fail to support reflection on who the algorithm thinks we are. Knowing that Instagram tracks our engagement tells us nothing about the identity it has constructed from that data. Technical explanations of algorithmic mechanics often increase cognitive load without enabling meaningful self-examination.

We introduce a reflective AI probe that generates narrative portraits from users' Instagram Explore feeds. Rather than explaining algorithmic mechanics, we use AI (GPT-4) to interpret algorithmic output as creative storytelling. Through 25 participant sessions, we examine how narrative-based reflection differs from information-based transparency, contributing to Theme 3 (AI Support for Reflection) and Theme 1 (Methods for Capturing Reflection) of this workshop.

2. REFLECTIVE AI vs TRANSPARENT AI

Algorithmic transparency research emphasizes explainability: showing users how systems work [3, 7]. However, transparency often increases cognitive load without supporting reflection [1]. Users presented with algorithmic explanations may understand the mechanism but fail to engage critically with how it shapes their experience.

Recent work on AI and identity suggests that generative AI can serve as a mirror for self-examination rather than just a production tool [2, 6]. Sterman et al. distinguish between productive AI (generating content) and reflective AI (supporting metacognition), finding that reflective integration preserves creative agency while productive use can diminish it.

We build on reflective design [4] and first-person methods [5] to explore how AI-generated narratives can scaffold metacognition about algorithmic identity. Our approach treats reflection as an emergent property of interpretation rather than information access. By reframing algorithmic output as narrative material, we shift from explaining systems to examining selves.

3. METHOD

Design Probe. We created a reflective probe using GPT-4 with a structured prompt designed to interpret Instagram Explore page screenshots as narrative portraits. The prompt instructs the AI to analyze three screenshots and generate a 200-250 word creative narrative describing 'who the algorithm thinks this person is' across dimensions of identity, values, lifestyle, and emotional landscape. We chose narrative over explanation because stories engage emotional and conceptual faculties that technical descriptions bypass.

The prompt structure balances open-ended interpretation with analytical focus. We specified four interpretive dimensions (identity markers, value systems, lifestyle patterns, emotional undertones) to ensure narratives addressed substantive aspects of

selfhood while allowing creative freedom in expression. This structured approach addresses the workshop question: Can LLM structure scaffold reflection without constraining surprise?

Sessions. Sessions lasted 30-60 minutes and followed a three-phase structure. First, participants completed a pre-reflection form asking: 'If Instagram's algorithm could describe you as a person, what do you think it would say?' This captured baseline self-awareness about algorithmic profiling. Second, participants took three screenshots of their Explore page (capturing the first visible content without scrolling) and submitted them through a Tally form. The AI generated a narrative interpretation within 30 seconds. Third, participants read the narrative and completed a post-reflection form capturing immediate reactions, surprises, and reflections on the gap between self-perception and algorithmic representation.

We conducted informal discussions after participants completed the forms, asking them to elaborate on their reactions and explore what the experience revealed about their relationship with Instagram. These discussions lasted 10-20 minutes and provided rich contextual data about how participants made sense of their algorithmic identities.

Participants. Twenty-five Instagram users (ages 23-41, mean age 29, convenience sample) participated. All were active users familiar with their Explore feed, using Instagram at least weekly for the past year. Participants represented diverse professional backgrounds (artists, academics, designers, students, working professionals) and varied Instagram usage patterns (casual browsers to content creators). We prioritized interpretive depth over statistical representativeness, consistent with Research-through-Design methodology that values rich qualitative insights over generalizable patterns.

Analysis. We conducted reflexive thematic analysis [8] on pre-reflection forms, post-reflection forms, and discussion transcripts. Two researchers independently coded the data, generating initial codes that captured specific reactions (e.g., 'surprise at accuracy,' 'language for feelings,' 'desire to manipulate feed,' 'validation of self-image,' 'discomfort with representation'). We then met to discuss emerging patterns, grouping initial codes into provisional themes through iterative refinement. For example, codes around 'language for feelings' and 'words I didn't have' clustered into the theme of conceptual gifting. We validated themes by checking their presence across multiple participants and their alignment with our research questions about reflection mechanisms. The analysis process itself modeled the interpretive approach we advocate: moving from data points to meaningful patterns through creative synthesis.

4. FINDINGS: REFLECTION THROUGH NARRATIVE

Finding 1: Conceptual Gifting. The AI narrative provided language participants lacked to articulate their identity. One participant noted: 'I never had words for this feeling of being caught between cultures, but seeing it written out made it real.' Another described the narrative as 'giving me permission to claim parts of myself I thought were contradictory.' The narrative acted as a conceptual gift, offering vocabulary that enabled deeper self-reflection than raw data could provide.

This finding addresses the workshop question: Can LLM structure scaffold creative reflection? Our probe demonstrates that structured prompts can generate interpretive language that supports metacognition. Participants valued not just what the AI said, but that it said anything at all. The mere act of receiving a coherent narrative about their algorithmic self made that self feel real and worthy of examination. Several participants saved the narrative text to revisit later, treating it as a artifact for ongoing reflection.

Finding 2: Affective Dissonance. Participants experienced cognitive tension between self-perception and algorithmic interpretation. Some felt validated ('This is exactly me, it's scary how accurate this is'), while others felt misrepresented ('This makes me sound shallow' or 'I'm more than just these interests'). This dissonance sparked reflection on the gap between private identity and public algorithmic selfhood.

The emotional response became a site for critical thinking about platform representation. Participants who felt accurately described began questioning whether the algorithm knew them too well. Those who felt misrepresented examined why the algorithm might construct that particular identity. Both responses prompted metacognition: thinking about how one thinks, or in this case, how one is thought about by algorithmic systems. This addresses the workshop concern about preserving opportunities to reflect on surprises within AI constraints.

Finding 3: Algorithmic Authorship. Seeing their algorithmic identity externalized prompted participants to consider shaping it intentionally. Several described wanting to 'curate' their Explore feed to reflect their values rather than habits: 'I realize I engage with a lot of anxiety-inducing content. I want to train the algorithm to show me things that make me feel good instead.' Others considered performing different identities: 'I might start liking more art posts to shift what Instagram thinks I care about.'

This relates to the workshop question about creative agency: our findings suggest that reflective AI preserves and even strengthens user agency by making algorithmic identity visible and mutable. Rather than feeling controlled by the algorithm, participants felt empowered to author their digital selves strategically. The narrative format made this authorship feel creative rather than manipulative, like writing a character rather than gaming a system.

5. DISCUSSION: DESIGNING FOR REFLECTION

Our findings contribute three design implications for reflective AI systems that support metacognition rather than mere information

access. These principles apply beyond algorithmic transparency to any context where AI might scaffold human reflection.

Interpretive over informational. Design AI systems that interpret rather than explain. Narrative reflection proved more generative than technical transparency because it engaged users emotionally and conceptually rather than cognitively overloading them with mechanism explanations. Information answers 'how does this work?' while interpretation addresses 'what does this mean?' The latter question better supports reflection on identity, values, and lived experience.

This principle extends beyond our specific probe. Any reflective AI system should prioritize meaning-making over mechanism-explaining. For creative practitioners, this might mean AI that interprets creative work rather than optimizing it, or AI that narrates design rationales rather than generating design solutions.

Structured prompts as scaffolds. LLMs can scaffold reflection through carefully designed prompt structures that balance open-ended interpretation with focused analysis. Our prompt's dimensional framework (identity, values, lifestyle, emotions) provided just enough structure to generate coherent narratives without constraining surprise. Too much structure produces formulaic outputs; too little produces incoherent ones. The sweet spot enables creative interpretation within meaningful boundaries.

This addresses the workshop's central tension: AI can support reflection if designed as interpretive scaffolding rather than prescriptive guidance. The key is giving AI enough structure to be coherent but enough freedom to surprise. Designers should treat prompts as architectural constraints that shape possibility space without determining outcomes.

Dissonance as design resource. Rather than seeking accuracy or validation, design for productive tension. The gap between self and algorithmic identity became the most reflective moment in our study. Participants who felt accurately described questioned algorithmic surveillance; those who felt misrepresented examined their self-presentation strategies. Both responses prompted valuable metacognition.

Systems should surface this dissonance as an opportunity for metacognition rather than smoothing it away. Perfect alignment between AI interpretation and self-perception would eliminate the reflective space where critical thinking emerges. Designers should embrace productive friction, creating moments where users must reconcile competing representations of self.

Limitations. Our study has several limitations. The convenience sample limits generalizability to broader populations. We captured reflection at a single moment; longitudinal study would reveal how reflective insights develop over time. The probe relies on Instagram's Explore page, which may not represent other algorithmic contexts. Finally, we cannot isolate whether effects stem from narrative specifically or from any externalized representation of algorithmic identity.

6. CONCLUSION

This work demonstrates how AI can support reflection on algorithmic identity through narrative interpretation rather than technical transparency. Our probe reveals that generative AI's strength lies not in explaining systems but in providing conceptual resources for metacognition. The three mechanisms we identified - conceptual gifting, affective dissonance, and algorithmic authorship - suggest pathways for designing reflective AI that enhances rather than diminishes human agency.

As AI becomes embedded in creative practice, designing for reflective engagement over productive efficiency offers a path toward preserving human agency and critical thinking. The workshop's questions about LLM scaffolding, surprise preservation, and creative agency find preliminary answers in our study: structured prompts can support reflection without constraining it, and interpretive AI can strengthen rather than weaken user agency.

Future work should explore how reflective AI can scaffold ongoing metacognition rather than one-time interventions, how to design for collective reflection on shared algorithmic experiences, and whether narrative approaches transfer to other domains beyond social media. The broader challenge is creating AI systems that help us think about thinking, rather than thinking for us.

REFERENCES

[1] Eslami, M., Vaccaro, K., Karahalios, K., & Hamilton, K. (2017). *Be careful; things can be worse than they appear: Understanding biased algorithms and users' behavior around them in rating platforms.* ICWSM.

[2] Jakesch, M., Buçinca, Z., Amershi, S., & Olteanu, A. (2022). *How different groups prioritize ethical values for responsible AI.* ACM FAccT.

[3] Karahalios, K., & Motahhare, E. (2015). *Reasoning about invisible algorithms in news feeds.* CHI.

[4] Sengers, P., Boehner, K., David, S., & Kaye, J. (2005). *Reflective design.* ACM Critical Computing.

[5] Lucero, A., & Arrasvuori, J. (2010). *PLEX Cards: A source of inspiration when designing for playfulness.* Fun and Games.

[6] Sterman, S., Kreminski, M., & Mateas, M. (2024). *Productive vs. Reflective: How different ways of integrating AI into design workflows affect cognition.* CHI.

[7] Vaccaro, K., Sandvig, C., & Karahalios, K. (2020). *At the end of the day Facebook does what it wants: How users experience contesting algorithmic content moderation.* CSCW.

[8] Braun, V., & Clarke, V. (2021). *One size fits all? What counts as quality practice in reflexive thematic analysis.* Qualitative Research in Psychology, 18(3), 328-352.