Qeios

Peer Review

Review of: "The Algorithmic Gaze: When Binary Code Meets Fluid Identities"

Antonio Pele[1]

1. LAW, Pontifícia Universidade Católica do Rio de Janeiro, Brazil

REFEREE REPORT

The Algorithmic Gaze: When Binary Code Meets Fluid Identities

Author: Hetarth Ajaybhai Parekh — Dept. of Computer Science and Engineering, Institute of Technology, Nirma University, India

Venue: Qeios preprint (Review Article), 22 June 2026 — DOI 10.32388/0H2BWW

Reviewer remit: Integrated assessment across three lenses — technical/empirical (algorithmic auditing and measurement), critical/trans-HCI (framing, epistemic stance, terminology), and methodology/governance (review design, legal reading, contribution).

Recommendation: Major revisions — a fundamentally sound and honest paper that should not be rejected, but whose framing, method reporting, and self-positioning need real rewriting before it earns its title.

1. Overview

This is a review article on how five distinct classes of digital system — automatic gender classification, biometric identity verification, administrative databases, automated content moderation, and generative-AI/retrieval systems — can disadvantage transgender, non-binary, and gender-diverse people. Its distinguishing move is a “source audit”: each factual claim inherited from an earlier draft is traced to a primary source, and several widely circulated figures are corrected, qualified, or removed. The manuscript is intellectually honest, unusually disciplined about causal language, and materially better than most of the advocacy-inflected literature it surveys. My reservations are not about integrity. They concern (a) a central metaphor that is invoked and then abandoned, (b) a method that brands itself a

"citation audit" but is not reported reproducibly, (c) a contribution that is never positioned against the obvious prior reviews, (d) an evidentiary hierarchy whose politics go unacknowledged, and (e) — ironically for a citation-audit paper — internal reference errors. Each is addressable, and the paper's own ethos points toward the fixes.

The three lenses converged on the same core issues. Where they diverged is instructive and is treated separately in §6.

2. Significance and contribution

The topic is timely, and the paper fills a real need: the trans-AI-harms discourse is saturated with figures that migrate between advocacy documents, press coverage, and academic work while losing their qualifications en route. A review that stops to ask what each number actually measures is valuable. On the evidence, the paper has two genuinely reusable contributions: the five-system taxonomy (§4), which directly attacks the conflation the paper rightly warns against ("Treating them as one technology can obscure where a failure occurs"), and the per-claim source-verification table (§3.1), which publishes what was wrong, why, and how it was corrected. Both deserve to be claimed explicitly — and, as §4 below argues, currently are not.

3. Strengths (to preserve in revision)

The statistical restraint is the paper's best feature. The reinterpretation of the Scheuerman et al. figure as "a 29.5% non-matching rate against the study's hashtag-based ground truth, rather than a universal claim that '38% of trans men are misidentified'" (§5.1) is exactly right, and the reading of the 0% result for #agender/#genderqueer/#nonbinary as "category mismatch" — a task-definition impossibility rather than a measured error rate — is the paper's sharpest analytic moment. It aligns with Keyes' argument about the misgendering of the task itself and is, if anything, more damning than the sloppy figure it replaces.

The causal discipline is exemplary throughout: the Uber lockouts are correctly held to "plausibly identify a mismatch among reference images, current appearance, account records, and review procedures" without being converted into evidence that the system "was attempting to classify the driver's gender" (§5.2); the Pravda/NewsGuard correction properly separates the 33% chatbot-repetition rate from anti-LGBTQ+ content specifically and identifies the mechanism error (retrieval/crawler-flooding was not shown to be training-data ingestion, §5.4). The refusal to launder the unsupported "70% of Republican advertisements" figure and the Moldova chatbot example is ethically as well as methodologically correct:

debunked numbers are weaponised against the community the paper studies, and removing them is protective, not timid.

Finally, Recommendation 1 ("Retire unnecessary gender inference") and the interpretation rule that the problem "may call for retiring the inference task" articulate a near-abolitionist stance — a refusal of the premise that gender should be inferred at all — that sits in productive tension with the paper's otherwise cautious voice. It should be named and owned rather than left implicit.

4. Major points

M1 — Theorise the "algorithmic gaze", or drop it. The metaphor carries the title but appears in exactly one paragraph (§2: "a critical metaphor for the ways institutions translate complex identities into categories, scores, matches, and moderation decisions. The metaphor is useful only if it does not erase technical distinctions"), and then never recurs in §§4–7. The paper honours the second clause by effectively discarding the metaphor. Yet "the gaze" has a specific critical lineage — Foucault's disciplinary and clinical gaze, surveillance studies, Browne's work on racialised surveillance, and the trans-studies politics of legibility — none of which is engaged. This matters because the gaze names the one thing the paper otherwise insists the five systems do not share: the institutional will to render gender legible and sortable. Properly theorised, the metaphor could supply the connective tissue the paper says is missing; as written it is decorative. Either develop it into a working analytic frame or retitle and reframe around the systems taxonomy, which is what the paper actually delivers.

M2 — Report the review method reproducibly. §3 describes verification behaviour, not a search strategy: no named database or index, no Boolean strings (only "combinations of … terms"), no lower date bound (only the 16 June 2026 cut-off), and no record counts or search-flow account. A paper that demands provenance of others, brands itself a "citation audit", and stakes its contribution on traceability must meet a minimal provenance standard itself. Add a short reproducibility note: databases/registries queried, representative search strings, a date range, and an approximate count of claims audited versus revised. The genre labelling should also be settled — the document is variously a "white paper" (abstract), a "Review Article" (header), and a "targeted narrative review and citation audit" (§3); pick one, because it sets reader expectations about coverage.

M3 — Own the corpus-anchoring and the one-directional nature of the audit. §3 states the audit traced "each factual sentence and statistic in the earlier draft". The universe under review is therefore the claims already present in a prior version of this manuscript, not a systematic sweep of the field. Two consequences go unacknowledged. First, a pass that looks only for claims that are too strong will move

conclusions in one direction; it corrects false positives but is structurally blind to false negatives (under-cited or missed evidence). Second, if the "earlier draft" is the author's own, then the abstract's framing that "several widely repeated claims require correction or qualification" overstates the reach of what was done — the corrected claims are largely the paper's own prior claims. State explicitly whose claims were audited, and add a limitation acknowledging the asymmetry. If the contribution rests solely on self-correction of the author's own draft, the novelty claim needs rescoping accordingly.

M4 — Position the contribution against the obvious prior reviews. The paper cites the two most relevant syntheses — Keyes (2018), a review of automatic gender recognition, and Perilo, Valença & Telles (2024), an explicit catalogue of trans-inclusive AGR practices — yet offers no "related reviews" paragraph and no gap statement. A review article must situate itself. Add a short paragraph stating what this advances beyond Keyes and Perilo; on the evidence, the defensible novelties are the five-system taxonomy and the citation-audit discipline, and both should be claimed as such.

M5 — Complete the AI Act reading on the point that matters most for this paper. §7 characterises Article 5 correctly at a high level and rightly corrects the instrument ("Regulation (EU) 2024/1689, not a 2023 final act") and rightly flags that the recommendation "should not be presented as a verbatim requirement of the EU AI Act". But for a paper whose lead system is gender classification, the decisive detail is omitted: the Article 5(1)(g) prohibition enumerates race, political opinions, trade-union membership, religious or philosophical beliefs, sex life, and sexual orientation — gender and gender identity are not on that list (verified against the regulation and current legal commentary; see §7). Biometric inference of gender is therefore not caught by the Article 5 prohibition at all, and is addressed, if anywhere, through the Article 50 transparency duty and possible Annex III high-risk classification. This strengthens the paper's own thesis — it is precisely why "retire gender inference" must be a policy proposal rather than existing law — so making it explicit costs nothing and gains rigour. A governance-lens review of biometric gender inference should also at least flag GDPR Article 9 (special-category data), which is often the more directly operative constraint. The legal apparatus (a single top-level citation to the Regulation, no pinpoint to Art 5(1)(g) or recitals, no commentary) is thin for a paper that makes a legal correction.

M6 — Acknowledge the politics of the evidentiary hierarchy. The method prioritises "peer-reviewed studies and official legal or administrative sources", uses investigative reporting for incidents, and treats advocacy submissions as "not … independent confirmation". GATE — a trans-led organisation — is explicitly demoted to "advocacy and policy analysis, not … primary empirical verification". Three points

deserve a response in the text rather than silent operation. First, the scepticism is unidirectional: the sharpest scalpel falls on community-derived claims, while "official records" (the Indian portal counts; the 2011 census "transgender/other" category) sit near the top of the hierarchy even though §5.2 shows state record-keeping is itself a site of harm. Second, the situated-knowledge tradition (standpoint epistemology; D'Ignazio & Klein's "counterdata") is absent; demanding "a traceable underlying dataset" as the price of admission structurally silences harms that were never datafied. Third, this undercuts Recommendation 8: the paper asks trans people to be substantive knowers in governance while treating trans-community-produced knowledge as insufficient evidence in the review. The fix is not to abandon evidentiary discipline — that discipline is a strength — but to make it reflexive: name the trade-off, apply the scepticism symmetrically to official records, and let the discipline coexist with an acknowledgement of what it cannot see.

5. Minor points

1. Internal citation error, in a citation-audit paper. §5.3 attributes the ctrl+alt+lgbt survey ("50% of surveyed LGBTQ+ adults reported anti-LGBTQ+ bullying or harassment online") to reference [2], but that survey is reference [7]; [2] is the Indian Transgender Persons Rules, 2020. Correct this and re-check the full callout set — this is exactly the failure mode the paper polices.
2. Reference title error. Reference 1 renders the Scheuerman et al. title as "Commercial Racial Analysis and Image Labeling Services." The published title is "Commercial Facial Analysis and Image Labeling Services" (verified). Confirm whether this is an OCR artefact or a manuscript typo and correct.
3. Report between-service dispersion for the Scheuerman figures. §5.1 gives four single averages (98.3 / 97.6 / 87.3 / 70.5%) but a minimum only for #transman (61.7%). A paper that elsewhere demands model versions and coding rules should report the range across the four services (and ideally n per hashtag cell) for the other categories too; four bare averages invite the over-reading the paper is trying to prevent. Verify the ten/five/four service cascade verbatim against the source.
4. Close the "38%" provenance loop. §5.1 rejects "38% of trans men are misidentified" but does not say where the figure comes from. Since 100 - 61.7 ≈ 38, noting that the circulating number traces to the single lowest-performing service — not the mean — would make the flagship correction airtight.
5. Verify the Hartmann (2025) quantifiers. §5.3's two strong claims — that "all providers under-moderated implicit hate" and did so "particularly" for LGBTQIA+-directed hate — should be

confirmed source-exact, since this is one of only two current technical audits carrying the moderation conclusion. Soften if the study is more qualified.

6. Align summary to body on Haimson. The Executive Summary's "disproportionate … removals" is stronger than §5.3's correctly hedged "often described … falling into moderation gray areas"; "disproportionate" implies a rate comparison the qualitative design cannot firmly establish. Match the summary to the body.
7. Reconcile title and body vocabulary. "Fluid identities" (title) never recurs in the body, which uses "transgender, non-binary, and gender-diverse" throughout. Beyond the inconsistency, "fluid" risks conflating non-binary with genderfluid and subtly reinstating the trope of the unstable trans subject against the solid machine; and "binary code" puns on a harm that actually lives in binary categories/schemas, not binary code. Reconsider the title's antithesis so it names the analysis underneath it.
8. Reconcile the non-AI layer with the frame. The administrative-database row is flagged as possibly involving "no AI at all" (Limitation 5). This honesty is welcome, but one reconciling sentence in §4 — that the taxonomy maps the digital-identity pipeline, not only ML components — would resolve the tension with the "algorithmic gaze" framing. Relatedly, note in the table that AGR-from-a-face is itself a form of biometric categorisation, so Systems 1 and 2 are distinguished by function, not by input modality.
9. Add a one-line appraisal of study quality. Several load-bearing conclusions rest on a single study each (Scheuerman for AGR; Hartmann for moderation). A brief risk-of-bias note would reinforce the evidence-standard ethos the paper champions.
10. Trim Executive Summary / Introduction overlap. Both open on the same ctrl+alt+lgbt datapoint; in a 14-page review the restatement is avoidable. Also check the boxed callouts ("Bottom line", "Interpretation rule", "Evidence standard") render with the label above rather than below the box.

6. A note on the productive disagreement between lenses

The panel split on one axis is worth flagging for the author. The technical and methodological readings treat the source audit as the paper's principal strength; the critical reading treats the same audit as an unexamined epistemic politics that discounts trans situated knowledge. Both are correct, and the paper is strongest if it holds them together rather than choosing. The evidentiary discipline should stay — it is what makes the paper trustworthy and protects the community from backlash-inviting overstatement. What is missing is one or two paragraphs of reflexivity: an admission that prioritising datafied,

institution-held evidence has a cost, that the cost falls hardest on harms that were never recorded, and that the paper's own governance recommendation (community members as substantive knowers) implies those forms of knowledge must eventually count. Adding that reflexivity would resolve M6, shore up Recommendation 8, and — not incidentally — give the "algorithmic gaze" of M1 something real to do, since the gaze is exactly the apparatus that decides what gets recorded and what does not.

7. Verification performed by this reviewer

Two load-bearing claims were checked directly rather than taken on trust. First, the Scheuerman et al. (2019) title is "How Computers See Gender: An Evaluation of Gender Classification in Commercial Facial Analysis and Image Labeling Services" (ACM CSCW; DOI 10.1145/3359246) — the manuscript's "Racial" is an error. Second, EU AI Act Article 5(1)(g) enumerates the prohibited biometric-categorisation inference targets as race, political opinions, trade-union membership, religious or philosophical beliefs, sex life, and sexual orientation, with carve-outs for labelling/filtering of lawfully acquired datasets and law-enforcement uses; gender and gender identity are absent from that list. This confirms and sharpens the manuscript's own (correct) instinct that its gender-inference recommendation is a policy proposal, not an existing legal requirement.

8. Prioritised revision checklist

1. Decide the fate of the "algorithmic gaze": theorise it into a working frame or retitle around the systems taxonomy (M1).
2. Add a reproducible method note — databases, search strings, date range, claim counts — and settle the genre label (M2).
3. State whose claims were audited and add the one-directional-audit limitation (M3).
4. Insert a short novelty/positioning paragraph against Keyes and Perilo (M4).
5. Complete the Article 5 reading (gender absent from 5(1)(g)); flag GDPR Art 9; add pinpoint legal citations (M5).
6. Add reflexivity on the evidentiary hierarchy and strengthen Recommendation 8's foundation (M6, §6).
7. Fix the internal citation error ([2]→[7]) and the "Racial"→"Facial" title error, and re-audit the reference callouts (minor 1–2).
8. Report figure dispersion, close the 38% loop, verify the Hartmann and service-cascade claims (minor 3–5).

The manuscript is worth the investment these revisions require. Its restraint is a model for a field that badly needs one; what it lacks is the reflexive step of turning that same rigour on its own frame, method, and evidentiary choices.

Declarations

Potential competing interests: No potential competing interests to declare.