



# Synthetic actors in digital research: a narrative review of authenticity, provenance, and evidentiary confidence

## Actores sintéticos en la investigación digital: una revisión narrativa de la autenticidad, la procedencia y la confianza en la evidencia

Rocío de la Cruz Hernández<sup>1</sup>  

Daniel Roman-Acosta<sup>2</sup>  

<sup>1</sup> Tecnológico Nacional de México/ITS de Centla. Tabasco, México.

<sup>2</sup> Plataforma de acción, gestión e investigación social, Colombia

Submitted: 30-05-2026 | Revised: 11-07-2026 | Accepted: 29-08-2026 | Published: 31-08-2026

**Corresponding Author:** Rocío de la Cruz Hernández 

### ABSTRACT

Bots, automated accounts, generative artificial intelligence, synthetic media, and AI-mediated interaction challenge a long-standing assumption of digital ethnographic research: that observable online traces originate from identifiable human participants. This narrative review examines how synthetic actors affect concepts of authorship, authenticity, cultural meaning, naturally occurring data, and evidentiary confidence in digital research. Relevant conceptual, methodological, and emerging empirical literature published mainly between 2021 and 2026 was identified across major academic databases and scholarly discovery platforms. Fifteen core studies were examined, with additional literature used for contextual interpretation. The synthesis focused on authenticity and provenance, cultural significance and human-machine hybridity, and methodological and ethical responses to uncertain or synthetic participation. Two findings recur across the literature. First, technical detection alone provides an insufficient basis for ethnographic judgement because machine-generated material cannot always be distinguished reliably from human-generated traces. Second, account authenticity and cultural significance are analytically distinct: synthetic actors can participate in communities, reproduce social categories, generate meaningful interaction, and contribute to harm or care even when authorship is uncertain or distributed. The review proposes a four-axis heuristic based on provenance, hybridity, interpretive validity, and evidentiary confidence. It is intended to make judgements about synthetic participation explicit and reportable, but it remains a conceptual synthesis that requires empirical testing in diverse digital fields.

Keywords: digital ethnography; generative artificial intelligence; synthetic actors; authenticity; provenance; evidentiary confidence

### RESUMEN

Los bots, las cuentas automatizadas, la inteligencia artificial generativa, los medios sintéticos y la interacción mediada por IA cuestionan una suposición histórica de la investigación etnográfica digital: que los rastros observables en línea proceden de participantes humanos identificables. Esta revisión narrativa examina cómo los actores sintéticos afectan los conceptos de autoría, autenticidad, significado cultural, datos digitales producidos de manera natural y confianza en la evidencia. Se identificó literatura conceptual, metodológica y empírica emergente publicada principalmente entre 2021 y 2026 en las principales bases de datos académicas y plataformas de descubrimiento científico. Se examinaron 15 estudios centrales y se utilizó literatura adicional para la interpretación contextual. La síntesis se centró en la autenticidad y la procedencia, el significado cultural y la hibridez humano-máquina, y las respuestas metodológicas y éticas ante formas de participación inciertas o sintéticas. Dos hallazgos se repiten en la literatura. En primer lugar, la detección técnica por sí sola constituye una base insuficiente para el juicio etnográfico, ya que el material generado por máquinas no siempre puede distinguirse de manera fiable de los rastros generados por personas. En segundo lugar, la autenticidad de una cuenta y su significado cultural son dimensiones analíticamente distintas: los actores sintéticos pueden participar en comunidades, reproducir categorías sociales, generar interacciones significativas y contribuir tanto al daño como al cuidado, incluso cuando la autoría es incierta o distribuida. La revisión propone una heurística de cuatro ejes basada en la procedencia, la hibridez, la validez interpretativa y la confianza en la evidencia. Su objetivo es hacer explícitos y reportables los juicios sobre la participación sintética, aunque continúa siendo una síntesis conceptual que requiere evaluación empírica en campos digitales diversos.

Palabras clave: etnografía digital; inteligencia artificial generativa; actores sintéticos; autenticidad; procedencia; confianza en la evidencia

## INTRODUCTION

Ethnographic research in digital settings has always been an exercise in mediated observation, and netnography and digital ethnography developed precisely as responses to that condition, treating online interaction as a site where cultural meaning is produced, not merely transmitted. Methodological reflection within this tradition has insisted that the rise of social media did not dissolve the basic principles of ethnography, notably immersion and reflexivity, and that day-to-day presence “on the sites” yields localized knowledge that big-data sketches cannot approximate.(1) Complementary work has shown that digital platforms are not neutral containers but interfaces whose affordances generate symbolic meanings, which is why participant observation in the places where platforms are accessed can be fruitfully paired with technical walkthroughs of those interfaces.(2)

At the same time, scepticism about the provenance of platform data is not new: the post-API turn made visible that researchers had long worked with proprietary black boxes whose data quality was rarely known with confidence,(3) and methodological comparisons between virtual and digital approaches have traced how the distinction between observing online worlds and observing with digital methods reshapes what counts as fieldwork.(4) The synthetic-actor problem therefore arrives on terrain that was already contested, but it alters the stakes.

What has changed is the composition of the field itself. Automated accounts have been shown to impersonate users at scale, with a small minority of highly active bot profiles accounting for a disproportionate share of political traffic during electoral contention.(5) Virtual influencers populated by human teams have been studied as a coherent class of non-human actors whose authenticity, scalability and controllability are actively managed.(6) Netnographic fieldwork inside an online AI community has documented how affordances of automated content creation, automated data analysis and AI-generated content dissemination progressively digitalize the practices of ordinary users.(7)

Bot-driven engagement farming and politically inflected synthetic imagery have been analysed as a genre of deception that targets the male gaze while generating interaction for its own sake.(8) Generative systems now produce human-like chatbot exchanges and manufactured text and images purporting to represent real people, places and events.(9) Beyond the digital-ethnography literature directly reviewed here, digital ethnography case studies have examined AI-generated “rage bait” whose creation and initial distribution involve no human actor at all,(10) and experimental work with a representative sample of social media users indicates that some AI assistance increases engagement while decreasing the perceived quality and authenticity of discussion, with negative spill-over into surrounding conversation.(11) In this literature, the challenge for netnographic practice is sustained saturation, not occasional contamination, though, as the methods and results below make explicit, the evidence supporting that framing in the present review is thinner and more narrowly distributed than that framing suggests.

The conceptual difficulty is not simply that some accounts are false. It is that the vocabulary netnography inherited, including participant, informant, post, account and naturally occurring data, presumes a correspondence between an online trace and the social actor whose conduct it records. Synthetic actors sever that correspondence while remaining socially consequential. This is why the review treats the authenticity of an account and the cultural significance of synthetic content as distinct analytical objects. Authenticity itself has been reframed not as a stable property to be verified once, but as a problem experienced by ethnographers and participants alike, a matter of situated judgement, cross-contextual variation and lived uncertainty.(9)

Synthetic visual media intensify this reframing because generative imagery is often photorealistic yet non-indexical, unsettling long-standing conventions of co-presence, negotiated authorship and situated seeing.(12) Empirical digital ethnography of synthetic faces queried through reverse image search reaches a comparable conclusion from the opposite direction: a search service that owes its users a promise of indexicality does not discriminate between authentic portraits and realistic fakes, leaving the user to investigate further and generating ever more elaborate apparatuses of verification.(13) Where AI systems are treated as interlocutors and not only as artefacts, care-like meanings can emerge from sociomaterial interaction even though the systems possess no ethical consciousness,(14) and where writers work alongside large language models under platform pressure, authorship becomes distributed and ambivalent instead of simply disappearing.(15) Algorithmic mediation of memory, agency and identity has likewise been described as constitutive of contemporary self-formation, not merely a representation of a pre-existing self.(16) Cultural significance and verifiable humanness are therefore not the same question.

Netnography is unusually exposed to this problem because of how its evidence is constituted.

The tradition depends on immersion, sustained interaction and the co-construction of accounts with participants, and it treats archival traces as materials that acquire meaning only within an interpretive relationship. Where generative systems can produce plausible texture at volume, three of these dependencies come under strain at once: the participant may be an engineered persona or a hybrid configuration; the archive may be synthetic in whole or in part; and the ethnographer's own materials, including notes, transcripts, coding and even draft interpretations, may themselves be machine-assisted.

Recent methodological argument holds that artificial intelligence and chatbots make it tempting to interface with an entirely digital field site, yet immersive ethnographic practices and an ethnographic sensibility remain indispensable, and that digital research would benefit from explicit ontological and epistemological examination of what ethnography can and cannot deliver. (17) Proposals for AI-augmented netnography accept efficiency gains but insist that combining machine pattern recognition with human interpretive oversight is what preserves cultural nuance, while flagging privacy, transparency, authenticity and bias as unresolved concerns. (18) Parallel debates in adjacent fields press the same point: computational ethnography proposes that language models might assist interviewing and analysis while naming significant ethical, epistemic and resource challenges, (19) and multi-platform online ethnography has responded to dispersed and hidden populations by formalizing staged participation, participant verification, consent and triangulation. (20)

The central question is how bots, generative AI, synthetic text and images, deepfakes, automated accounts and AI-mediated interaction challenge netnography's assumptions about participants, authorship, authenticity, cultural meaning, naturally occurring digital data and evidentiary confidence, and what conceptual and methodological resources can guide practice in AI-saturated fields. Its objectives are to synthesize the relevant conceptual, methodological and emerging empirical literature; to sharpen the distinction between the authenticity of an account and the cultural significance of synthetic content; and to propose a framework for evaluating provenance, human-machine hybridity, interpretive validity and evidentiary confidence.

A narrative synthesis was used to integrate a small and methodologically diverse body of literature on synthetic participation in digital fields. The evidence is unevenly distributed across research domains and varies in source detail, so claims are kept at the level supported by the available literature. The analysis first establishes the methodological foundations of the problem, then compares findings on detection, provenance, cultural significance, and ethical practice before proposing an appraisal heuristic and a research agenda.

## METHODS

The study uses a structured narrative review of conceptual, methodological, and emerging empirical literature on synthetic participation in digital fields. Its analytical focus is how bots, automated accounts, generative AI, synthetic media, and AI-mediated interaction affect assumptions about participants, authorship, authenticity, cultural meaning, and evidence.

### Search and source selection

The source search covered multidisciplinary academic databases and discovery services, concentrating on literature published mainly between 2021 and 2026. Search concepts combined netnography and digital ethnography with bots, automated accounts, generative AI, deepfakes, synthetic media, authenticity, provenance, verification, reflexivity, research ethics, and evidentiary validity. Multilingual variants were used to broaden discovery, and no formal language exclusion criterion was applied.

Fifteen studies formed the primary analytical set, with additional sources used to contextualise methodological and ethical debates. Sources were included when they directly addressed qualitative digital fieldwork involving synthetic actors or contributed to questions of authenticity, provenance, verification, cultural significance, or researcher judgement. Purely technical detection studies without substantive relevance to qualitative digital research were not treated as core evidence.

### Analytical approach

Evidence was charted according to study context, type of synthetic or automated actor, methodological approach, claims about authenticity or provenance, cultural significance, ethical

implications, and reported limitations. The synthesis was thematic and organised around three domains: authenticity and provenance; cultural significance and human-machine hybridity; and methodological and ethical responses.

The depth of source access varied across the included material, so the review does not claim exhaustive coverage or uniform study-level verification. Full-text evidence was prioritised where available, while more limited records were used cautiously to identify conceptual patterns. The resulting framework is therefore interpretive and exploratory, not a formal risk-of-bias assessment or quantitative synthesis.

## RESULTS

The claims below are stated at the level of generality the reviewed evidence can bear, and no further. Because 11 of the 15 included core studies were synthesized from abstracts and bibliographic metadata instead of full text, and because the four full-text studies are not distributed evenly across the topics addressed, the synthesis should be read as the most defensible reading of a small and uneven evidence set, not as settled findings. Where a claim rests mainly on abstract-only evidence, or on single studies, this is noted at the point of the claim. Where the included evidence is silent, that silence is reported as a limit of the reviewed set and not as a property of the field.

### Detection is not a foundation: authenticity, provenance and the limits of indexicality

The first claim is that netnographic evidence presupposes a correspondence between trace and actor, and that automated and synthetic actors break this correspondence at a scale that affects the field, not only at its margins. The strongest support among the included studies is a full-text study of bot activity during the 2016 municipal elections in Rio de Janeiro, which identified 3,101 bot profiles responsible for 19,915 tweets, distinguishing light, medium and heavy bots; heavy bots represented 1.96 per cent of bot profiles but produced 20.12 per cent of the tweets in the sample, and 35 of the 61 heavy bots were media spambots that posted only links to media websites, with 21 of these tweeting links exclusively to one media group, accounting for 43.4 per cent of all tweets linking to the two dominant media groups.(5)

The study's classification of user-generated bots, media spambots and political bots, with campaigner and inciting-agent subtypes, shows that impersonation is not a single phenomenon but a differentiated repertoire serving different purposes, including the manufacture of apparent public opinion.(5) Analysis of bot-driven "I need husband" catfishing posts, available to this review only as an abstract, reaches a similar conclusion about purpose: these posts use synthetic images of women with impossible proportions ostensibly seeking partners, are designed to farm engagement, and reflect a mechanical rendering of the male gaze.(8) In both cases, the ethnographically salient fact is not only that the accounts are not what they appear to be, but that their apparent identity is instrumental to a communicative strategy.

The second claim is that detection cannot carry the epistemic weight that such findings invite it to carry. Work on generative AI and online ethnography, synthesized here from its abstract, argues explicitly against a generalized methodological exceptionalism, noting that efforts to estimate the extent of AI-generated online material are hampered by the considerable difficulty of reliably detecting machine-generated material.(9) Empirical digital ethnography of synthetic faces and reverse image search, one of the four full-text studies reviewed, supplies the sharpest available demonstration of why technical verification cannot settle the question: synthetic portraits from a "this person does not exist" generator, when used as queries against a facial reverse-search engine, returned results resembling a person who does not exist, and the engine did not discriminate between "authentic" portraits and "fake" images provided the latter were realistic enough.(13)

The authors conclude that the service falls short of the promise of indexicality on which its authority rests, that it falls to the user to investigate further, and that the consequential issue is not the capacity to produce realistic synthetic representations but the demand for indexicality that synthetic data triggers and the bureaucratic apparatuses of verification that emerge to contain it.(13) Conceptual work on generative imagery in ethnography, drawn from its abstract, converges: synthetic visuals are photorealistic but non-indexical, unsettling co-presence, negotiated authorship and situated seeing.(12) Where generative chatbots were examined systematically, contextual inaccuracies were observed even though no obvious fabricated claims were identified, which indicates that plausibility and factuality are separate properties and that an output can be generative, fluent and locally wrong at the same time.(14) This second claim rests on one full-text empirical study, two abstract-based conceptual arguments and one abstract-based empirical study; it is

convergent across those sources but is supported by a small number of cases.

The third, more constructive claim is that provenance is better treated as a situated and relational achievement than as a property that a detector can deliver. The clearest conceptual resource in the included literature is the contrast between representational and relational models of data, advanced in the full-text study of synthetic faces and reverse search: the analyst shifts from asking what an image represents to asking what relations, infrastructures and verification practices produce and sustain its evidentiary status.(13)The auto-netnographic study of generative chatbots, available as an abstract, operationalizes a related move by treating chatbot outputs as system-generated data and analysing them qualitatively alongside reflexive experiential engagement, attending to interactional framing, emotional attunement, specificity and performative features instead of focusing on truth value alone.(14)A netnographic and actor-network-informed reading of AI scribe failures similarly locates documentation problems in socio-technical interaction, not in model accuracy alone, which relocates provenance questions from the artefact to the assemblage that produces, reviews and corrects it.(21)

Counterevidence within the same evidence set complicates any shift toward provenance as a purely technical or artefact-centred matter. A longitudinal digital ethnography of Persian Twitter, read here from its abstract, found that malicious activity during the #MahsaAmini movement was not limited to false information and bot activism: authoritarian actors also deployed fabricated stories and human undercover agents, generating a condition in which users are bombarded with a mixture of true and false messages that undermines thinking altogether.(22)Provenance judgements therefore cannot be reduced to distinguishing machine from human, because some of the most consequential synthetic effects are produced by human actors using synthetic means.

Conceptual caution runs in the same direction: the call to avoid human-centric approaches founded on a principled separability of human and machine implies that verification regimes built on that separation will misdescribe the field they claim to secure.(9)Finally, the included evidence offers no validated verification procedure and no estimate of the error rates of netnographic provenance judgements; the studies that document inauthenticity do so through bespoke methodological work, not through transferable tools that other researchers can adopt without repeating that labour.(5,13)

### **Cultural significance without authentic subjects: synthetic actors in community life**

Among the included studies, the evidence most directly supporting the claim that synthetic and hybrid actors participate in, and sometimes constitute, communities whose cultural life is real comes from a full-text netnography of virtual influencers that monitored four high-profile accounts over ten months, analysing 497 posts, 66 videos and 387,107 comments belonging to profiles with roughly 3.6 million followers, alongside eight in-depth interviews with digital marketing specialists.(6)The study identified five categories for understanding these actors, namely anthropomorphism/humanization, attractiveness, authenticity, scalability and controllability, and found more convergences than divergences between virtual and real influencers and between humans and non-humans, with virtual influencers largely operated by human teams and dependent on them.(6)

The category of authenticity appears here not as a binary property that non-human actors lack, but as one dimension among several that practitioners manage, which supports the review's insistence that authenticity and cultural significance are separable questions. This is the strongest empirical support for that inference in the included evidence, and it concerns commercial actors on one platform ecology.

The same inference is available, though more weakly, from contexts where synthetic mediation is primarily infrastructural, not spectacular. A netnographic study combining more than nine months of naturalistic observation in an online AI community, generating 1,572 pages of data and known to this review through its abstract and bibliographic record, found that different types of affordances, including automated content creation, automated data analysis and AI-generated content dissemination, foster the gradual digitalization of individuals, and it identified constraints of generative AI together with potential interventions to prevent unintended consequences.(7)

In clinical documentation, a netnographic analysis of 162 Reddit threads across clinician-oriented communities, comprising 952 documents and 2,267 relevant data segments from 2023 to 2025, found that reported AI scribe problems such as misattribution, hallucinations and omissions, together with workflow disruptions including latency, crashes and copy-and-paste friction, were associated with interacting technological, organisational, environmental and individual conditions.(21)The study's central finding is that AI scribe safety is not solely a function of model accuracy, because documentation problems arise through socio-technical interactions that determine whether

errors are identified, corrected or carried forward, and that safe deployment requires strengthening integration, governance and verification processes alongside technical performance.(21) Here synthetic mediation is not an exotic intrusion but an ordinary condition of a professional community's discourse, and the community's own verification practices become part of the object of study.

Additional included evidence, again available only at abstract level, supports the stronger claim that meanings resembling care can emerge through sociomaterial interaction with systems that are not subjects. The auto-netnographic comparison of four generative chatbots, using two standardized prompts across all platforms within a single simulated patient role, mapped responses onto two intersecting continua, logical to empathetic orientation and general to specific framing, producing four interactional styles: structured overview, rational clarity, compassionate perspective and compassionate precision.(14)

The authors argue that care-like meanings may emerge through sociomaterial interactions between users and artificial intelligence systems, while noting explicitly that these systems lack ethical consciousness and cannot replace professional care, and that the study does not demonstrate generalized performance or clinical reliability.(14) Political communication supplies the inverse case, in which synthetic and fabricated material is absorbed into real contention and thereby acquires cultural force,(22) and the "I need husband" genre shows synthetic imagery circulating as a recognizable social type that reproduces a particular visual politics of gender.(8) Both of these last two cases rest on abstracts alone.

Counterevidence and limits qualify this claim in three ways. First, convergence between virtual and human actors is not equivalence: the finding that boundaries between real and virtual and between human and non-human are being reduced was established for a deliberately top-ranked set of commercial profiles, which the authors themselves identify as a possible source of selection bias and as a limit on representativeness.(6) Second, methodological argument warns against deriving from such convergence a licence to treat human and machine accounts as interchangeable evidence, calling instead for continued reflexivity about what is at stake in judging the ontological state of materials encountered online,(9) and insisting that immersive practices and ethnographic sensibility remain indispensable even when the temptation is to interface with an entirely digital field site.(17)

Third, the included evidence is unevenly distributed: its strongest evidence concerns marketing actors, political contention and clinical documentation, while the everyday meaning-making practices through which ordinary communities detect, ignore, joke about or incorporate synthetic presence are not documented in any included study; several studies also treat engagement metrics as the evidentiary surface, which measures attention, not meaning.(6,7)

### **Methodological resources: reflexivity, hybridity and evidentiary confidence**

Several methodological resources emerge from the literature; the fourth claim is that they should be treated as proposals, not validated procedures. The most explicit is a framework for AI-augmented netnography that addresses informed consent, data minimization, bias mitigation and accountability, and proposes combining AI-driven sentiment and pattern recognition with human interpretive oversight in order to capture the cultural nuances essential to understanding online social movements, illustrated through a case study of online activism across platforms such as Twitter and Instagram; the review has access to this framework through its abstract and bibliographic record.(18)

A second resource is a framework of contextuality, consent and criticality for engaging with generative synthetic visual media in ethnographic research, grounded in foundational ethnographic values and offered explicitly as a heuristic, not a prescriptive guide, as a starting point for ethical experimentation and positional reflection.(12) A third is the argument that existing approaches already contain the needed resources: multi-sited approaches allow experience of varying and cross-contextual understandings of authenticity, autoethnography directs attention to how researchers navigate lived uncertainty about the nature of online content, and more-than-human approaches position ethnography as an immersive means of embracing non-human actors, including AI-generated features, as an intrinsic part of online experience.(9)

Three further resources operate at the level of analytic framing instead of protocol. Interface ethnography pairs participant observation in the settings where platforms are accessed with technical walkthroughs, enabling researchers to trace the twofold meaning construction of communities whose work is oriented toward digital platforms.(2) Treating generative outputs as system-generated data analysed through actor-network sensibilities allows the analyst to keep both the system and the reflexive user within a single frame of sociomaterial interaction.(14) A so-

cio-technical lens organized across technology, organisation, person and environment domains allows documentation problems to be traced to the conditions under which errors are detected or carried forward, including individual verification practices.(21) Alongside these, the included literature contains a strong general claim, drawn from an abstract and a conceptual review, not from empirical demonstration: digital research with an ethnographic sensibility benefits from explicit examination of what ethnography can and cannot deliver, and researchers should practise participant observation, embrace reflexivity and attune to the body and sensations.(17)

Counterevidence and limits are substantial. The two frameworks are heuristic by design or by demonstration, and neither has been evaluated against conventional practice: one is illustrated through a single case study of online activism and is available here only as an abstract,(18) and the other states explicitly that it is heuristic, not prescriptive and offers only a starting point for ethical experimentation.(12) The argument that promising strategies lie in existing approaches is a claim about where to look, not a tested remedy, and it is accompanied by an explicit caution against wholesale methodological exceptionalism.(9)

Two methodological resources in the included literature predate generative systems entirely and address platform ethnography without synthetic actors, so their transfer to AI-saturated fields remains inferential and has not been demonstrated.(1,2) The included studies provide no comparative evaluation of verification procedures, no estimate of the frequency of erroneous provenance judgements, and no treatment of the false-positive problem, that is, the harm done to human participants who are wrongly suspected of being synthetic, even though this risk follows directly from the detection difficulties documented elsewhere in the same literature.(9,13) Finally, automating verification would import the same concerns the AI-augmented framework raises about privacy, transparency and bias, so automation cannot be assumed to resolve the epistemic problem it appears to address.(18)

The table 1 consolidates the principal controversies and the counterevidence identified across the included core studies. All sources cited in it are core studies.

**Table 1. Controversies and counterevidence in netnographic encounters with synthetic actors**

| Contested claim                                                                                                            | Contrary or complicating evidence                                                                                                                                                                                                                                                        |
|----------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Machine-generated and automated material can be reliably detected, and detection can anchor exclusion decisions. (9,12,13) | Detection of machine-generated material is difficult enough to hamper even estimation of its extent; reverse facial search fails to distinguish realistic synthetic portraits from authentic ones; generative visuals are photorealistic but non-indexical.                              |
| Authenticity is a property of an account that can be verified once and for all. (6,9)                                      | Authenticity is treated as a problem experienced by ethnographers and participants, varying across sites; it appears as one managed dimension among several in the operation of non-human influencers.                                                                                   |
| Synthetic or hybrid actors can be excluded from analysis without loss, because their content is culturally empty. (6,8,14) | Non-human influencers sustain large follower communities and converge with human influencers on multiple dimensions; care-like meanings emerge through sociomaterial interaction with systems lacking ethical consciousness; synthetic imagery circulates as a recognizable social type. |
| The decisive distinction for provenance is human versus machine. (9,22)                                                    | Malicious activity included fabricated stories and human undercover agents alongside bots; principled separability of human and machine is explicitly cautioned against.                                                                                                                 |
| Verification can be delegated to platforms, detectors or search infrastructure. (13,21)                                    | Reverse search shifts the burden of investigation back to the user and generates new verification apparatuses; safety and error propagation depend on integration, governance and verification capacity rather than on model accuracy.                                                   |
| Existing ethnographic resources suffice, or conversely that generative AI requires wholly new method. (1,2,9)              | Multi-sited, autoethnographic and more-than-human approaches are proposed as promising without being tested; methodological exceptionalism is rejected, yet no core study evaluates a verification procedure or reports error rates.                                                     |

## DISCUSSION

Read against the wider literature, the synthesis suggests that the synthetic-actor problem is better characterized as an acceleration of a long-standing evidentiary predicament, not a rupture. The critique of academic digital research in the post-API age, a supporting record outside the core evidence set, argued that platforms were always proprietary black boxes never intended for scholarly use, that researchers rarely knew what type or quality of data they held, and that the era of abundant access had not always been characterized by rigour.(3) That argument converges with the core findings on provenance (13,21) in locating evidentiary doubt in infrastructural opacity, not in researcher error, but it also bounds them: the post-API critique concerns access and coverage, whereas the core studies reviewed here extend the problem to authorship.

The two are convergent in diagnosis and divergent in locus, and that divergence is precisely the increment the synthetic-actor literature adds. It also reframes the contribution of frameworks such as the one proposed for AI-augmented netnography, which situates consent, data minimization, bias mitigation and accountability as ongoing obligations of the research process instead of gatekeeping rules applied only at entry.(18)

Experimental evidence from outside the core evidence set both corroborates and bounds the review's central conceptual distinction. A controlled study with 680 participants in a realistic social media environment, a supporting record, not a core study, found that some AI-assisted tools increased user engagement and the volume of generated content while decreasing the perceived quality and authenticity of discussion, and produced a negative spill-over effect on surrounding conversations.(11) This converges with the core claim that authenticity is a situated and contested achievement, not a directly verifiable property,(9,12) and it strengthens that claim by showing the decoupling of engagement from authenticity at the level of measured user perception, not through interpretive argument alone.

It bounds the core evidence in a different direction: the experimental design isolates tool effects under controlled conditions, whereas the core studies observe actors in situ, so the supporting study measures perceptions of authenticity while the core studies document its practical management. The design principles the authors derive, namely transparent disclosure of AI-generated content, user-focused personalization, context-sensitivity and intuitive interfaces,(11) operate at the platform level, whereas netnographic judgement operates at the level of the case; the gap between them is where the appraisal heuristic proposed below is intended to sit.

The supporting literature also complicates the assumption that synthetic content necessarily implies an intentional author, and here it extends beyond simply echoing the core evidence. Digital ethnography case studies of AI-generated rage bait describe a severe form of trolling that differs from other rage baiting because no human actor is involved in its creation or initial distribution, while nevertheless drawing on extant datasets, practices and norms to further embed rage in the digital ecology.(10) This converges with the core finding on bot-driven engagement farming (8) and with the core argument that detection cannot be foundational,(9) but it goes beyond both by supplying a case of cultural causality without intentionality, which the core evidence set asserts only indirectly through the circulation of synthetic imagery as a recognizable social type.

Work on generative AI in creative labour supplies a complementary account of hybrid authorship: a four-year multi-sited digital ethnography of 34 writers on a Chinese serial fiction platform, alongside sustained forum observation, found four coexisting modes of negotiation with large language models, namely strategic collaboration, cyborg authorship, conditional refusal and narrative incorporation, with writers preserving an unstable human zone sustained by reader vigilance, platform metrics and self-discipline.(15) This is convergent with the core evidence on hybrid configurations(6) but diverges from it in kind: the core findings concern commercial and political actors whose hybridity is constructed for audiences, whereas the supporting study documents hybridity as a working condition experienced from the inside. The observation that the preserved human zone is neither permanent nor guaranteed directly reinforces the core argument against treating hybridity as a stable category assignable to a participant once.(9)

Adjacent work on algorithmic mediation pushes the same conclusion further, describing platformized digital twins whose agency, memory and identity functions are algorithmically re-configured, not merely represented, and calling for human-centred frameworks to engage with the constitutive role of algorithms in self-formation.(16) Read against the core evidence set, this is convergent with the auto-netnographic finding that care-like meanings emerge through socio-material interaction(14) and with the relational, not representational, treatment of data developed in the study of synthetic faces,(13) but it makes a stronger constitutive claim than either: that algorithmic mediation forms the self instead of merely mediating a pre-existing one.

That claim exceeds what the core evidence can support, and it is therefore treated here as a bounding hypothesis, not a finding. A netnographic study of how a developer community adapted to an image-generation tool similarly found that participants remained in a state of permanent experimentation, deploying collective bricolage to navigate a tool whose outputs could not be explained, which converges with the core account of how communities absorb generative systems into ordinary practice(7) and with the AI-augmented framework's insistence on iterative human oversight,(18) while suggesting a research posture, provisional, iterative and openly experimental, that the core frameworks recommend but do not model.(23)

Two further strands of supporting literature correct for possible over-attribution to machine agency, and both diverge from the core evidence set in instructive ways. An analysis of conservative appropriation of algorithmic-bias discourse on X during the months following the public

release of a major language model showed how the concept of bias, originally grounded in critical accounts of structural power, was repurposed to advance an ideological agenda through a mixture of thematic strategies and coordinated social action.(24) Where the core study of Persian Twitter locates manipulation in authoritarian state actors deploying fabricated stories and undercover agents,(22) this supporting case locates it in partisan user discourse within a democratic setting; the two are convergent on the existence of coordinated manipulation but divergent on its source, which bounds any reading of synthetic mediation as the master variable in digital conflict.

Likewise, click-farm research in Brazil and Colombia shows that accounts, photographs and bots circulate as resalable commodities within historically informal labour markets whose boundaries with illegality are actively negotiated.(25) This converges with the core finding that impersonation is instrumental and differentiated(5) but relocates its explanation from communicative strategy to political economy: authenticity in this account is not a metaphysical property but a traded resource embedded in regional digital culture, which grounds the provenance problem in labour and commercial arrangements, not in engineering alone.

The supporting literature also points toward operational responses, each of which can be compared against a specific core finding. A staged, ethics-attentive, multi-platform online ethnographic protocol for hidden and stigmatized populations moves from low-intrusion public mapping to ethically managed engagement in semi-private spaces and, where appropriate, to direct interaction, participant verification, informed consent, offline interviewing and triangulation, treating hashtags, recommendations, symbolic codes and cross-platform pathways as field signs, not merely recruitment tools.(20)

Participant verification and staged escalation function as a provenance procedure developed for reasons unrelated to synthetic actors; relative to the core evidence set, it supplies a procedure where the core AI-augmented framework supplies only principles,(18) but it diverges by assuming verifiable human participants, so it sets limits on the synthetic-actor problem without resolving it. A proposal for computational ethnography in public health similarly positions language models as complements to ethnographic workflows, through assisted interviewing and computational support for inductive, deductive and abductive analysis, while naming the ethical, epistemic and resource challenges that must be addressed before such complementarity is realized(19); this is convergent with the core argument that automation should supplement, not replace, interpretive oversight,(18) and it is bounded by the same untested status that limits the core frameworks.

In clinical documentation, a parallel netnographically informed analysis of clinician discourse demanded enforceable mechanisms for provenance tracking, error contestation and vendor accountability, on the grounds that a signature plus passive visibility is a fragile guarantee of accountability(26); this converges closely with the core socio-technical finding that error propagation depends more on review capacity than on model accuracy(21) and translates that finding into an institutional demand, which is the form the evidentiary-confidence axis proposed below takes.

Finally, the longer arc of digital ethnography shows that hybrid online-offline reflexivity is not new, which bounds the review's claims about novelty. A reflexive account of a religious festival performed mainly through social media during the 2020 global health crisis documented the inherent complexity of interpreting ethnographic data when body, emotion and participation are displaced,(27) and smartphone ethnography of a Muslim dating application described users being shaped by the religious framework the app provides as much as shaping it.(28) Both converge with the core argument that interface and infrastructure shape what participants can be and do,(2) while diverging from the synthetic-actor literature by presupposing human participants throughout, which is exactly the presupposition the core evidence set puts in question.

Affect-inflected participant observation of livestreaming channels made comparable claims about repetitive, incoherent and asymmetrical interaction as ethnographic material and not as noise,(29) and methodological comparison of virtual and digital approaches has long insisted that the choice of method changes what is seen,(4) which is convergent with the core insistence that what ethnography can and cannot deliver be examined explicitly.(17)

The framework proposed here is a heuristic, in the same spirit as the contextuality-consent-criticality framework for generative visual media and the AI-augmented netnography framework for responsible practice, and it is intended to be used iteratively and revised, not applied mechanically.(12,18) Its rationale is that the included evidence contains strong findings about what synthetic actors do and strong warnings about what detection cannot deliver, but no procedure for connecting the two at the point of fieldwork. The heuristic has four axes, each of which is a question the researcher answers and reports.

The first axis is provenance: how did this material come to exist, and through what infrastructure did it reach the field? Instead of asking only whether an artefact is synthetic, the axis asks what

relations, platforms, accounts, operators and verification practices sustain its presence, following the relational, not representational, treatment of data developed in the study of synthetic faces and reverse search.(13) Reported provenance should include how the researcher encountered the material, what was known and unknown about its production, and which platform affordances or governance arrangements mediated it, since safety and error propagation depend on integration, governance and verification capacity and not on model behaviour alone.(21)

The second axis is hybridity: what configuration of human, machine and institutional agency produced this trace, and how is that configuration sustained over time? The reviewed evidence indicates that the relevant configurations are mixtures, not alternatives, including non-human influencers operated by human teams,(6) fabricated stories and undercover human agents operating alongside automated activity,(22) and writers curating machine-generated prose under platform pressure.(15) The axis therefore explicitly rejects the assumption of principled separability of human and machine that underpins many verification regimes,(9) while retaining hybridity, not undecidability, as the reported object.

The third axis is interpretive validity: what claims about meaning, community or culture can this material support, and at what level of generality? This axis follows the distinction that organizes the review. Where an actor is not authentic in the conventional sense but is culturally consequential, the defensible claim concerns the meanings, norms and relations the material participates in, not the interior life or intention of a subject.(8,14) The axis also requires attention to the possibility that the researcher's own interpretive materials are machine-assisted, since AI integration in netnography raises the same questions of authenticity and transparency it raises for participants.(18)

The fourth axis is evidentiary confidence: how much weight can this judgement bear, and what would change it? Here the included evidence is instructive in its limits, since it documents detection difficulty without supplying error rates.(9,13) Confidence should be reported as a bounded judgement with named conditions of revision, in the same spirit as ethnographic approaches that treat authenticity as a problem of situated judgement requiring multi-contextual experience and reflexive attention to uncertainty,(9) and in line with the insistence that digital research examine explicitly what ethnography can and cannot deliver.(17) Two safeguards belong specifically to this axis: false-positive caution, so that human participants are not reclassified as synthetic on the basis of stylistic suspicion, and the avoidance of automation as a substitute for judgement, given that automated verification would import the privacy, transparency and bias concerns that the reviewed frameworks identify.(18)

The table 2 sets out the principal gaps this framework exposes. Rows drawing on supporting contextual records, not core studies are the regional and infrastructural inequality row and the institutional provenance row; the remaining rows rest on core evidence.(25,26)

The heuristic proposed here should be read as an argument for transparent decision rules, not as a validity claim. Its axes are derived from frameworks that their own authors describe as heuristic, from argument against methodological exceptionalism, and from ethnographic sensibilities that emphasize reflexivity and immersion, so the framework inherits their limits along with their rationale.(9,12,17) It also inherits an unresolved tension that the evidence does not settle: if hybridity is ordinary, not exceptional, the reporting burden of full provenance documentation may exceed what any single study can bear, and over-documentation could itself distort the field by encouraging researchers to substitute suspicion for immersion. The research gaps set out above are stated in that spirit, as conditions under which the framework could be tested and, if necessary, abandoned.

## Limitations

The evidence base is small and uneven in depth. Full-text access varied across the included literature, which limited verification of methodological details in some studies. The available evidence also clusters in areas such as marketing, political communication, and clinical documentation, while everyday community settings and several geographic and linguistic contexts remain less represented.

The synthesis is narrative and interpretive, and the proposed four-axis heuristic has not been validated against independent datasets or established assessment instruments. In addition, the literature documents the difficulty of detecting synthetic content more clearly than it quantifies detection error. The conclusions should therefore be read as methodological propositions for further testing, not as settled rules for digital fieldwork.

**Table 2. Research gaps and corresponding elements of a methodological agenda**

| Gap                                                                                                                                          | Why it matters for netnographic evidence                                                                                                                                                                                             | Indicative next step                                                                                                                                                       |
|----------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Almost no sustained netnographic accounts of how ordinary communities themselves detect, ignore, joke about or normalize synthetic presence. | The documentation focuses more heavily on producers, platforms, and professional communities than on everyday sense-making, meaning claims about cultural significance rely on inferences drawn from commercial and political cases. | Multi-sited, long-duration fieldwork in communities where synthetic participation is routine, including attention to how members themselves establish authenticity. (9,15) |
| Frameworks for responsible AI-augmented and visually oriented ethnography are proposed but not evaluated.                                    | Without evaluation, the reporting standards they imply cannot be shown to improve interpretive validity or to reduce ethical risk.                                                                                                   | Comparative testing of the appraisal axes against conventional practice, including documentation of how provenance judgements are revised over time. (12,18)               |
| No reported error rates for provenance judgements, and no treatment of false-positive classification of human participants as synthetic.     | Evidentiary confidence cannot be calibrated, and the ethical costs of suspicion fall on participants rather than on researchers.                                                                                                     | Prospective studies reporting the frequency and consequences of mistaken attributions, with explicit harm analysis. (9,13)                                                 |
| Consent and privacy provisions are framed for human participants and for data minimization, not for hybrid or synthetic interlocutors.       | Hybrid configurations blur who can consent and what obligations attach to system-generated material retained as data.                                                                                                                | Explicit consent models for hybrid scenes and clear rules on retaining and reporting system-generated material. (14,18)                                                    |
| Evidentiary status of engagement metrics is not distinguished from evidentiary status of meaning.                                            | Attention and cultural significance are conflated, so synthetic activity can appear culturally weighty when it is merely voluminous.                                                                                                 | Report engagement data and interpretive material as separate evidentiary tiers with distinct confidence statements. (6,7)                                                  |
| Regional and infrastructural inequality in the evidence base, including dependence on platform access conditions.                            | Provenance practices are shaped by platform governance and by regional labour and commercial arrangements, not only by model capability.                                                                                             | Comparative research across regions and platform types, including the political economy of account and image circulation. (5,25)                                           |
| Verification is treated as an individual researcher's burden after infrastructure fails.                                                     | This redistributes risk to researchers and participants and obscures institutional accountability for verifiable provenance.                                                                                                         | Institutional provenance requirements and enforceable error-contestation mechanisms, analogous to demands emerging in clinical documentation. (21,26)                      |

## CONCLUSIONS

Synthetic participation changes the evidentiary conditions of digital research because authorship, agency, and authenticity can no longer be assumed from the appearance of an online trace. Technical detection may inform investigation, but the reviewed literature does not support using it as the sole basis for deciding whether material is ethnographically meaningful or trustworthy.

The cultural significance of synthetic content can persist even when its authorship is uncertain. For this reason, digital researchers need to distinguish questions about who or what produced a trace from questions about the meanings, relationships, harms, and practices that develop around it.

The proposed framework of provenance, hybridity, interpretive validity, and evidentiary confidence provides a structured way to make these judgements transparent. Its value lies in clarifying how claims are bounded and revised, but further empirical work is needed before it can be treated as a validated methodological instrument.

## REFERENCES

1. Wang D, Liu S. Doing Ethnography on Social Media: A Methodological Reflection on the Study of Online Groups in China. *Qualitative Inquiry*. 2021;27(8-9):977-987. doi: 10.1177/10778004211014610.
2. Ritter CS. Rethinking digital ethnography: A qualitative approach to understanding interfaces. *Qualitative Research*. 2021;22(6):916-932. doi: 10.1177/1468794211000540.
3. Tromble R. Where Have All the Data Gone? A Critical Reflection on Academic Digital Research in the Post-API Age. *Social Media + Society*. 2021;7(1). doi: 10.1177/2056305121988929.
4. Padricelli GM, Punziano G, Saracino B. Virtual vs Digital: Examples of Netnography and Digital Ethnography in Tourism Studies for a Comparison between Methods. *ATHENS JOURNAL OF SOCIAL SCIENCES*. 2021;8(4):259-278. doi: 10.30958/ajss.8-4-4.
5. Santini RM, Salles D, Estrella CE, Barros CE, Orofino D. Bots as online impersonators. *The International Review of Information Ethics*. 2021;30(1). doi: 10.29173/irie402.
6. Batista da Silva Oliveira A, Chimenti P. "Humanized Robots": A Proposition of Categories to Understand Virtual Influencers. *Australasian Journal of Information Systems*. 2021;25. doi: 10.3127/ajis.v25i0.3223.
7. Park HEG. The double-edged sword of generative artificial intelligence in digitalization: An affordances and constraints perspective. *Psychology & Marketing*. 2024;41(11):2924-2941. doi: 10.1002/mar.22094.
8. Rosenbaum J. I need husband: AI beauty standards, fascism and the proliferation of bot driven content. *AI & SOCIETY*. 2025;41(4):2749-2760. doi: 10.1007/s00146-025-02491-8.
9. Hine C. Does generative AI mean the end for online ethnography?. *Methodological Innovations*. 2026;19(3):137-147. doi: 10.1177/20597991261443793.
10. Cover R. AI generation of rage bait: Implications for digital harms. *New Media & Society*. 2025. doi: 10.1177/14614448251400675.
11. Møller AG, Romero DM, Jurgens D, Aiello LM. The impact of generative AI on social media: an experimental study. *Scientific Reports*. 2026;16(1). doi: 10.1038/s41598-026-40110-8.
12. Chen J. AI-generated imagery in ethnography: A reflective framework. *Big Data & Society*. 2025;12(4). doi: 10.1177/20539517251403957.
13. Ridgway R, Malevé N. Synthetic Data and Reverse Image Search: Constructing New Surveillant Indexicalities. *Surveillance & Society*. 2024;22(4). doi: 10.24908/ss.v22i4.18332.
14. Christiansen M, Eriksson H, Fagerström L. Generative AI Chatbots as Digital Adjuncts for Sexual Health Information After Prostate Cancer in Men Who Have Sex With Men: Auto-Netnographic Study. *JMIR Cancer*. 2026;12:e81745-e81745. doi: 10.2196/81745.
15. Ge L. Sympoietic creativity and the boundaries of qing: digital romance writers negotiating generative AI. *AI & SOCIETY*. 2026. doi: 10.1007/s00146-026-03253-w.
16. Bar-Gil O. The Google self as digital human twin: implications for agency, memory, and identity. *AI & SOCIETY*. 2025;41(4):3027-3042. doi: 10.1007/s00146-025-02692-1.
17. Fu D, Nielsen RA, Schatz E. Ethnography and Ethnographic Sensibility in Political Science. *Annual Review of Political Science*. 2026;29(1):371-390. doi: 10.1146/annurev-polisci-041224-084600.
18. Cheah CW. AI-Augmented Netnography: Ethical and Methodological Frameworks for Responsible Digital Research. *International Journal of Qualitative Methods*. 2025;24. doi: 10.1177/16094069251338910.
19. McKay F. Computational ethnography and public health: Scaling and deepening lived experience research on social determinants of health with large language models. *Public Health*. 2025;248:105973. doi: 10.1016/j.puhe.2025.105973.
20. Chan GH, Lee GKW, Lo TW. Online Ethnography: A Research Methodology for Reaching Hard-to-Reach Groups in the Contemporary Era. *International Journal of Qualitative Methods*. 2026;25. doi: 10.1177/16094069261473438.
21. Atiku S, Owolanke K, Olakotan O. Evaluation of Socio-Technical Mechanisms Shaping AI Scribe Documentation Failures: A Netnographic Study. *Journal of Evaluation in Clinical Practice*. 2026;32(5). doi: 10.1111/jep.70554.
22. Kermani H. The art of delirium: social media suppression in authoritarian regimes. *Communication Theory*. 2025;35(4):197-213. doi: 10.1093/ct/qtaf006.
23. da Silva RC, Rüling CC, Duymedjian R, Zejnilovic L. Learning about unknowable tools: Bricolage in communities of practice for generative AI. *Technological Forecasting and Social Change*. 2026;230:124738. doi: 10.1016/j.techfore.2026.124738.
24. Cohen M, Toch E. Hijacking algorithmic bias: analyzing the political discourse around ChatGPT on social media. *Information, Communication & Society*. 2025;29(4):1171-1196. doi: 10.1080/1369118x.2025.2561046.
25. Grohmann R, Clara Aquino M, Rodrigues A, Matos É, Govari C, Amaral A. Click farm platforms. *Work Organisation, Labour & Globalisation*. 2022;16(2). doi: 10.13169/workorglaboglob.16.2.0007.
26. Atiku S, Olakotan O. An Evaluation of AI-Generated Clinical Notes in the OpenNotes Era: A Thematic Analysis of Clinician Discourse. *Journal of Evaluation in Clinical Practice*. 2026;32(5). doi: 10.1111/jep.70516.
27. Cocco C, Bertran A. Rethinking Religious Festivals in the Era of Digital Ethnography. *Social Analysis*. 2021;65(1):113-122. doi: 10.3167/sa.2021.650107.
28. Hasan F. Keep It Halal! A Smartphone Ethnography of Muslim Dating. *Journal of Religion, Media and Digital Culture*. 2021;10(1):135-154. doi: 10.1163/21659214-bja10042.
29. Zhang G. Richang: An Affect-Inflected Ethnography of Chinese Livestreams. *Asiascape: Digital Asia*. 2021;8(1-2):15-42. doi: 10.1163/22142312-12340130.

## **FUNDING**

No financing

## **CONFLICT OF INTEREST**

None

## **AUTHORSHIP CONTRIBUTIONS**

Drafting – original draft: Rocío de la Cruz Hernández, Daniel Roman-Acosta.  
Writing–review and editing: Rocío de la Cruz Hernández, Daniel Roman-Acosta.