International Journal of Applied Social Science
(An International Double Blind Peer Reviewed / Refereed Research Journal of Social Science)
Volume 12 (11 & 12), November & December (2025) : 1025-1041

RESEARCH ARTICLE
ISSN : 2394-1405 (Print)

Received : 25.10.2025; Revised : 11.11.2025; Accepted : 25.11.2025 **DOI** : https://doi-ds.org/doilink/01.2026-58437543/IJASS/12.11&12/1025-1041

Artificial Intelligence and the Re-Emergence of Positivism in Social Science Research

BHAWNA SIKARWAR
Ph.D. Scholar
Jawaharlal Nehru University, New Delhi (India)

ABSTRACT

The increasing use of Artificial Intelligence (AI) in research workflows has brought forth the problematic aspect of its usage at the interpretative stage of social science research. On the one hand, AI-driven analyses with access to a larger database can swiftly recognise patterns and themes, pluralise interpretation, and simulate multi-framework readings. On the other hand, there is an imminent risk of reinforcing positivist assumptions in areas such as cultural hybridity, which thrive on ambiguity and context. Moreover, it can inadvertently simplify complex cultural phenomena, standardise meanings, and reinforce dominant discourses. In this manner, this paper argues that at the interpretative stage, AI functions not only as an efficiency tool but as a co-interpreter wherein it reshapes the very ontological ground of meaning. By proposing the AI-mediated research methodology, this paper calls for awareness of these epistemological tensions and advocates for the reconfiguration of scholarly agency and reflexivity in the age of AI.

Keywords: Artificial Intelligence (AI), Social science research, Positivism and Epistemic

INTRODUCTION

The world of academia has long been interspersed with Artificial Intelligence (AI), mostly at the margins, with assistance in data analysis pipelines and proofreading (Chaudhary and Alam, 2024; Hemachandran *et al.*, 2023). Initially, the role of AI was restricted to a comparatively few computationally trained scholars, with its higher threshold of technical expertise. However, the release of ChatGPT to the public in November 2022 'democratised' AI with the expansion of the scope and accessibility of these tools (Cappelli *et al.,* 2024). Thus, the lowering of technical barriers and no specific need for coding skills have made AI viable for Social Science scholars to integrate it at every stage of the research workflow (Bail, 2024; Davidson, 2024; Ivanov and Soliman, 2023). Tasks such as designing surveys, simulating respondents, analysing texts, and even drafting manuscript sections, which were once confined to computational experts, are now a few prompts away. However, this democratization, though productive, has started showing its effective cost. The same technologies that accelerate mundane scholarly labour are also facilitating unethical practices such as enabling convincing fabrication of texts, tables, figures and even references (Walters and Wilder, 2023), fueling paper mills and retractions (Chauhan and Currie, 2024; Nazarovets, 2024), and thus, raising urgent questions about integrity, transparency, and the epistemic foundations of the social sciences (Driessens and Pischetola, 2024). A cursory glance over the literature on this terrain showcases a double-edged consensus that agrees on the transformative influence of AI on every stage of the social science research workflow, and improving its efficiency (Anderson *et al.,* 2025; Tingelhoff *et al.,* 2024) and yet cannot forego the need for new forms of transparency and oversight (Farangi *et al.,* 2024; Ivanov and Soliman, 2023). More so, recent surveys have divided the researchers over the scope and ethics of AI in academic writing and review (Kwon, 2025). This has called for a proposal of a process-wide account of responsible and

How to cite this Article: Sikarwar, Bhawna (2025). Artificial Intelligence and the Re-Emergence of Positivism in Social Science Research. *Internat. J. Appl. Soc. Sci.,* **12** (11 & 12) : 1025-1041.

irresponsible uses of AI, along with pragmatic suggestions which are grounded in worked examples that track the standard sequence from ideation to manuscript evaluation (Ivanov, 2025). While these debates have focused on the role of AI for the Social Sciences as an external force, the internal workings have received the least attention. When ideation, design and data collection can be streamlined and augmented by computational systems, there is a high possibility of advances in AI reshaping how knowledge is produced, circulated and interpreted across disciplines.

The field of AI-mediated Communication (AI-MC) has already showcased the way AI acts 'as-if' an agent to intervene in human exchanges by modifying, augmenting and generating messages on behalf of communicators (Hancock *et al.,* 2020). This can be seen in a multitude of digital communication, spanning from Gmail's predictive replies to generative models that produce entire conversational threads. There are two critical aspects to be marked here: firstly, AI functions as a communicative mediator, as it is not merely transmitting information but actively shaping its form; secondly, AI does so here without intention or consciousness. Russel and Norvig refer to this act "as if" it were an agent, taking precepts as inputs and producing outputs in pursuit of the goal (2010, 2015). In this regard, AI has already assumed a role of functional agency within our everyday digital communication. When such an agency is applied to social science research, its implications deepen. The Research workflow requires AI not only in its functional roles, such as collecting, organising and classifying data, but it also intervenes in interpretation. Unlike survey sampling or database structuring, interpretive work thrives on ambiguity, contradiction, and negotiation. As Clifford Geertz (1973) argued, interpretation is "thick description," foregrounding small details and ambivalence. Similarly, Homi Bhabha's (1994) notion of cultural hybridity underscores that meaning often emerges from liminal spaces, where coherence is less critical than plurality and liminality. The risk, then, is that AI's interpretive interventions, which are shaped by assumptions of coherence, regularity, and efficiency, reinstate a positivist drift in a domain that depends on openness. Recent scholarship has highlighted the tendencies of AI to prioritise predictive performance over conceptual validity (Baden, 2022) and preference for coherence and uniformity, which risks collapsing novelty into "positivist drift" (Ivanov, 2025). Grossmann *et al.* (2023) identify the "illusion of coherence" in AI-generated interpretations, while Cappelli *et al.* (2024) point to the "implicit intelligence" of large language models, which align with human judgment in routine cases but falter under complexity, defaulting to moderation or what they call latent perplexity. These tendencies illustrate that AI is not a neutral assistant but an epistemic actor whose mediating role shape what counts as valid knowledge.

This paper situates itself at this critical junction. It is argued here that AI must be understood not simply as an automated tool but as a mediator and co-interpreter in social science research. This role has already been established in AI-MC, where communicative outcomes are co-produced by humans and machines. When transferred to the interpretive stage of social science research, however, AI's mediation carries epistemic consequences. By privileging coherence and efficiency, AI risks erasing the ambiguity and plurality on which interpretive traditions depend. To ground this argument, I use the case of Mongol cultural hybridity as a litmus test. The Mongol empire, often dismissed in Eurocentric historiography as an "anomaly" or mere borrower of sedentary cultures, in fact exemplifies hybridity through its multilingual decrees and fusion of Persian bureaucratic forms with Mongol yasa. Conventional interpretive methods highlight this liminality, foregrounding ambiguity as insight. By contrast, AI-mediated interpretations, whether symbolic, statistical, generative, or multi-agent, tend to collapse hybridity into neat categories. The result is not only methodological bias but also historiographical risk, that is, a re-inscription of Eurocentric narratives under the guise of algorithmic neutrality. Against this backdrop, our paper makes a different and complementary intervention. We focus on the interpretive stage of social science research and argue that AI's expanding role risks re-introducing positivist drift and epistemic bias precisely where interpretive pluralism, ambiguity, and reflexivity should be paramount. This paper advances three contributions. First, this study conceptualise AI-mediated interpretation as a distinct object of analysis, situations where computational systems directly intervene in the meaning-making moment (not merely in storage, indexing, or formatting). Second, we map how different AI paradigms of symbolic, statistical/ML, generative, and multi-agent, encode distinct assumptions that tilt interpretation toward coherence, regularity, and efficiency. Third, we stage a methodological litmus test using the case of cultural hybridity (with reference to the Mongol empire's linguistic and administrative assemblages) to show how AI-mediated pipelines can reproduce anomaly narratives or Eurocentric defaults unless governed by explicit reflexive safeguards. In doing so, we align with the emerging process-wide guidance (e.g., Ivanov, 2025) while shifting the spotlight to the interpretive juncture, where epistemic stakes are highest. Our argument is not anti-AI. Rather, it is a call to articulate the conditions under which AI can serve as a responsible co-interpreter and hence be transparent, auditable, multilingual, integrative, and subordinated to human reflexivity and theoretical aims. Henceforth, the literature review focuses on AI in the research process, research ethics, and responsible-use principles. We then analyse how specific AI paradigms mediate interpretation and detail the epistemic risks they introduce. Next, we develop our hybridity case to demonstrate the mechanics of flattening and the safeguards needed to resist it. We conclude with implications for authors, early-career researchers, editors, and funders, and outline a practical agenda for reflexive AI-mediated interpretation that preserves plurality rather than collapsing it into spurious coherence.

Literature Review:

This review synthesises three interrelated strands of scholarship that underpin the present study. Firstly, it delves into the functional role of AI as mediator, automator and co-interpreter in knowledge production. Secondly, it investigates the varied ways in which meaning is constructed and interpreted across the domain of social sciences. The emphasis here is on the intersubjective and constructivist outlook. Lastly, it explores the different types of AI technologies and how their integration may distort interpretative practices through mechanisms of norm reinforcement, coherence bias and epistemic flattening.

AI as Mediator, Automation, Co-Interpreter:

The concept of AI-Mediated Communication (AI-MC) has already established how AI acts as an intelligent agent when it intervenes in digital communications. This intervention occurs in the form of modifying, augmenting and generating messages on behalf of its human communicator (Hancock *et al.,* 2020). This functional agency of AI in this mediation is, however, without presuming consciousness or intent because AI is acting "as-if" it were an agent (Russell and Norvig, 2010). Even the theory of Chinese Room reminds us that machines manipulate symbols without understanding the meaning, reason and intent (Searle, 1980). Hence, AI is already mediating even if it's purely a simulation. When AI is ported to social science research with this logic, the implications deepen. As aptly posited, what distinguishes "an agent from a mere cause is that the agent acts for reasons" (Davidson, 1980). Thus, in social sciences, agency is closely tied to consciousness because interpretation requires self-reflexivity, intentionality and the ability of meaning-making (Taylor, 1971; Gadamer, 1960). What we can gather here is that in social sciences research workflow, the role of AI extends beyond the functional collection and organisation of data to interpretive mediation. AI already shapes what kind of data is accessed, which narratives are privileged, and how analytic categories are stabilised (Kaprouzi, 2025). Crucially, mediation is not neutral. As CTAM demonstrates via its different approaches of rule-based, supervised, and unsupervised, they all carry distinct assumptions about what counts as a valid signal (Boumans and Trilling, 2016; Baden *et al.,* 2022). Rule-based methods presume that categories are fully specifiable ex ante (e.g., dictionaries, formal parsers); supervised methods treat categories as learnable correlates of labeled data, and unsupervised methods induce structure from distributional regularities. This architectural diversity aligns with Baden's diagnosis of three persistent "gaps" (2022). Firstly, the validity gap pertains to predictive performance substituting for conceptual/operational validity. Secondly, an integration gap emphasises how AI tools focus on one kind of information even when constructs are multi-component, and lastly, a language gap which prioritises the English language and thus, disadvantages comparative and multilingual work (Bender, 2011). Thus, AI mediation functions with the essence of an agent but without reflexive intent, which is at the core of the interpretative stage of social science research.

When mediation scales into automation, the consequences become starker. The process by which tasks of decision-making, judgement or meaning-making making, which were traditionally undertaken by conscious, interpretative human agents, are delegated to systems that operate via formalised, rule-based or algorithmic procedure is referred to as Automation (Crawford, 2012, Vallor, Suchman, 2007). Ivanov (2025) cautions that such automation smooths complexity into coherence, generating an illusion of validity that can mask conceptual loss. Baden *et al.* (2022) likewise show how predictive success can ride on spurious correlates, famously, a classifier trained to detect political ideology that actually learned incumbency because it was easier to recognise (Hirst *et al.,* 2014). In practice, automation can thus reward correlations over constructs, encouraging "what works" statistically while sidelining "what means" conceptually. A third stance treats AI as a co-interpreter, useful if and only if researchers maintain reflexive oversight. Scholars have warned us of the delegation of interpretative tasks to unconscious agents such as AI. There is a risk of bypassing the ethical and reflective capacity that comes with human consciousness (Vallor, 2018). Cappelli *et al.* (2024) report that large language models (LLMs) often align surprisingly well with expert judgments in specialised surveys, a property they term implicit intelligence, even without fine-tuning. Yet they also identify latent perplexity, which is a tendency to hedge toward moderate, uniform outputs under uncertainty. Their findings underscore that effective scholarly use of GenAI requires a three-tiered practice of careful problem formulation, prompt engineering, and prompt interaction. In this manner, iteratively interrogating outputs for relevance, accuracy, and coherence. Without this reflexive loop, mediation collapses back into automation and inherits CTAM's structural risks (Baden *et al.,* 2022; Ivanov, 2025).

Meaning and Interpretation in Social Sciences:

When considering the role of AI in social science research at the Interpretative stage, the classic debates about truth, meaning-making making and context cannot be ignored. There is consensus among scholars regarding the AI being built on Cartesian and formalist assumptions that all ideas can be reduced to symbols and rules (Suchman, 2007; Dreyfus, 1992). While this semantic model of truth can work for logic, this Tarskian collapse fails in the Interpretative social sciences, which thrive on the very terrain of polysemy, irony and contested categories (Vallor, 2016; Crawford, 2021). Here, the meanings are not based on correspondence and prediction; rather, they are negotiated. Moving beyond correspondence, meaning-making is social and institutional, wherein rules of use and communal negotiation anchor reference and fix what counts as an "X" in context "C" (Searle, 1969, 1995; Putnam, 1975). These insights resonate with concerns about supervised and unsupervised CTAM. Unless labels inherit communal criteria or inductive patterns are re-anchored by theory, models drift toward proxies and plausibility rather than validity (Baden *et al.,* 2022; Hirst *et al.,* 2014). Similar assimilation about the aspect of interpretation and meaning-making is vouched for in the social sciences. Interpretation thrives on subtleties, nuances and contexts, often described as "thick description", which locates every action amongst webs of meanings (Geertz, 1973). Wendt (1999) argued that social reality is intersubjective, constituted by shared understandings rather than brute facts. Bhabha (1994) conceptualises hybridity as a "third space" of ambivalence and negotiation, while García Canclini (1995) theorises mestizaje as ongoing cultural mixing. In these frames, ambiguity is not noise to be removed but the signal itself. That, in turn, explains why tools optimised for regularity can sit uneasily with interpretive aims that require contradiction and multiplicity to remain visible. An information-philosophical lens clarifies the mechanism. Floridi (2011) argues that models operate at Levels of Abstraction (LoA) wherein selective reductions that privilege certain observations. Many AI tools fix relatively coarse LoAs (e.g., token windows; next-word probabilities), which are efficient for scale but prone to collapse relational or pragmatic dimensions central to interpretive constructs. The broader socio-technical picture matters as well. Crawford (2021) reminds us that AI is an infrastructure of power and standardisation; its data, benchmarks, and pipelines privilege what is abundant and digitised. This helps explain the English-first CTAM ecosystem that Baden *et al.* (2022) and Bender (2011) critique: multilingual, non-Western, or hybrid meaning systems are structurally harder to represent, evaluate, and compare.

Taken together, these perspectives converge on a tension central to this paper: interpretation in social science thrives on plurality, contradiction, and hybridity, whereas contemporary AI systems tend to optimise for coherence, regularity, and efficiency (Baden, 2022; Ivanov, 2025). The implication is not that AI is unusable for interpretation, but that reflexive design and oversight are required to prevent positivist drift at the very stage where interpretive openness is constitutive rather than incidental.

Types of AI and their Interpretative Risks:

A closer look at different AI paradigms helps to clarify how each carries particular interpretive affordances and liabilities when utilised in the research workflow.

Symbolic AI systems mainly rely on knowledge structures that humans explicitly define, including symbols and rules. Their strengths are transparency and concept-driven control. These features are mirrored in social-science practices as dictionary-based coding or rule-guided parsing (Baden *et al.,* 2022). They are limited in terms of rigidity and closure. With the categories fixed, ambiguity is lost. Kaplan and Haenlein (2019) emphasise that classical expert systems lack autonomous learning, which can lock researchers into initial assumptions and formalise a single interpretive lens. In interpretive research, this rigidity risks over simplifying constructs and suppressing alternative readings. Data-driven approaches of Machine learning infer patterns from examples or distributions. While they promise flexibility and scale, their liabilities include bias amplification, proxy substitution, and the replacement of validity with predictive performance (Baden *et al.,* 2022). The classic cautionary tale of how an ideology classifier that actually learned incumbency (Hirst *et al.,* 2014) illustrates how models can optimise on the wrong signal. Baden *et al.* (2022) further show that specialization, which requires measuring exactly one kind of pattern at a time, undermines analysis of multi-component constructs, while the English-first toolchain (Bender, 2011) hinders comparative and multilingual interpretation. In practice, the result can be automation drift. When the researchers accept machine outputs as "objective," though they rest on thin evaluation criteria that do not map onto thick, contextual constructs (Ivanov and Soliman, 2023). Large language models introduce a qualitatively new form of mediation by producing fluent, authoritative prose on demand. This is powerful for drafting and synthesis, but it risks an illusion of coherence. Here, the outputs might read as cogent yet may flatten contradiction or embed training-data biases (Grossmann *et al.,* 2023). Cappelli *et al.* (2024) demonstrate LLMs' implicit intelligence, which surprisingly aligns with expert surveys and yet carries latent perplexity, where models hedge to mid-scale or generic judgments in complex domains. Ivanov (2025) warns that, without human-in-the-loop reflexivity, GenAI can reintroduce positivist drift into interpretive stages with the smoothing of heterogeneity into stable narratives. In editorial and pedagogical contexts, this intersects with concerns about plagiarism and fabricated references (Walters and Wilder, 2023; Chauhan and Currie, 2024; Nazarovets, 2024), reinforcing the need for transparency and disclosure (Driessens and Pischetola, 2024). Although still nascent in this corpus, multi-agent configurations raise distinct interpretive questions. In principle, interacting agents could simulate multiple perspectives or interpretive traditions. However, the statistical design of AI tends to reduce interpretive multiplicity into a single "probable" output (Bommasani *et al.,* 2021). For interpretive social science, the challenge is to design interactions that preserve disagreement and plurality rather than driving toward least-resistance coherence. Across paradigms, the risks converge on a common thread already visible in CTAM practice. Symbolic AI tends toward rigidity (and epistemic closure). ML/NLP tend toward bias amplification and proxy learning, with predictive performance displacing validity (Baden *et al.,* 2022; Hirst *et al.,* 2014). LLMs tend toward coherence drift and latent perplexity, especially under uncertainty (Cappelli *et al.,* 2024; Grossmann *et al.,* 2023; Ivanov, 2025). And multi-agent approaches risk homogenization through convergence. These patterns explain why, as Baden *et al.* (2022) document, social scientists often forgo computational methods for more complex constructs and multilingual settings. It is not because of computational "illiteracy," but because current tools, as designed and deployed, can compromise operational validity, integrative measurement, and linguistic inclusivity.

The scholarship surveyed here converges on a double imperative. First, GenAI and related tools undeniably extend capacity and efficiency across the research pipeline (Bail, 2024; Davidson, 2024; Dwivedi *et al.,* 2023; Andersen *et al.,* 2025; Tingelhoff *et al.,* 2024). Second, precisely because of that reach, their use in interpretive stages demands reflexive oversight, explicit articulation of constructs, and documentation of how ambiguity is preserved rather than smoothed (Baden *et al.,* 2022; Ivanov, 2025; Hancock *et al.,* 2020). It is at this disjunction that the present study enters, focusing on how AI's mediating tendencies, toward coherence, regularity, and efficiency, can reintroduce positivist drift and epistemic bias where interpretive pluralism, ambiguity, and contradiction are indispensable.

Case Study: Mongol Cultural Hybridity

Traditional historiography designates the Mongol Empire as a barbaric and destructive force (Grousset, 1970; Barthold, 1928). Eurocentric and Orientalist biases led the early accounts to focus on the Mongols' military ruthlessness and widespread destruction, downplaying cultural and administrative achievements (Dawson, 1955). In recent decades, revisionist historians have reframed the Mongol Empire by foregrounding cultural hybridity and cross-cultural exchange (Allsen, 2001; Rossabi, 1988). The Mongols are portrayed as "agents of cultural change" who catalysed unprecedented Eurasian connections, forging a shared, unique imperial culture (Biran, 2013). Understanding Mongol history demands a nuanced interpretation of sources. With few Mongolian records and most primary accounts by conquered peoples, interpreting the empire implies navigating multiple cultural perspectives, underscoring hybridity. Such complexity defines Mongol hybridity, and this is the ambiguity that AI might flatten. Floridi (2011) notes that any representation at a given Level of Abstraction highlights certain details and ignores others, and Wendt (1999) stresses that meaning is co-created through human negotiation. Without this human nuance (Hancock *et al.,* 2020), symbolic text analysis reduces hybridity to simple keyword counts (Baden, 2022). This danger is not only technical but also historiographical. Against this background, the Mongol case thereby serves as a litmus test: Can AI sustain the interpretive pluralism of hybridity, or does it collapse history back into the very stereotypes scholars have spent decades undoing?

Why Hybridity Matters:

The Mongol Empire has long occupied a paradoxical place in world history. Early Western historians emphasised the "barbarian" nature of the Mongols, stressing their conquests, massacres, and the apparent collapse of a unified polity after the succession conflict of 1260–1264 (Spuler, 1965, Barthold, 1956). In this framing, Mongol cultural activity was treated as derivative. The episodes of them patronising arts in Persia or bureaucracy in China were read as borrowing from sedentary civilisations, never by innovation of their own (Dawson, 1955). Hybridity in governance, law, or culture was either ignored or treated as a sign of Mongol dependence. Revisionist historians have since pushed back against this one-dimensional portrayal. Scholars have emphasised that the Mongols were not simply destroyers, but active integrators and adapters who forged unique and hybrid imperial institutions through cross-cultural interactions (Allsen, 2001, Rossabi, 1988). Rather than "dissolving" after 1260, the empire's partition into Yuan China, the Ilkhanate in Persia, the Chagatai Khanate in Central Asia, and the Golden Horde in Eastern Europe represented a transformation of governance (Weatherford, 2004). Despite their autonomy, these khanates remained bound by common Chinggisid traditions, continued commercial and diplomatic exchange, and fostered what has been called the Pax Mongolica (or Mongol Peace) (May, 2019). In this view, hybridity in the way of blending of nomadic traditions with Persian, Chinese, Islamic, and even European practices was not incidental but central to the Mongols' imperial project. This historiographical shift matters because it redefines what counts as Mongol legacy. Where the traditional view treated the Mongols as anomalous "borrowers," the revisionist approach sees them as agents of cultural change whose empire became a conduit of ideas, practices, and people across the East-West (Allsen, 2001). The very institutions that enabled Eurasia's unprecedented connectivity, such as the merchant associations (ortaq), the postal relay (yam), and multilingual legal decrees, were hybrid forms, created through negotiation between Mongol rulers and local traditions. Recognising hybridity therefore destabilises Eurocentric binaries (civilisation vs. barbarism, sedentary vs. nomadic) and shows how power operated in the liminal "third space" (Bhabha, 1994).

A striking example of hybridity lies in Mongol economic policy. Unlike Confucian Chinese traditions that often disparaged merchants, the Mongols elevated them to a privileged status (Weatherford, 2004). This reflects a nomadic pragmatism in which steppe societies, lacking their own industries, relied on exchange and valued mobility as a form of wealth (Schorkowitz, 2020). Under imperial rule, this pragmatism fused with Persian and Chinese commercial practices to generate hybrid institutions like the merchant (ortoq) partnerships (Endicott-West, 1989). These state-sponsored merchant associations pooled resources across ethnic and regional lines, with Mongol nobles often supplying capital and foreign merchants providing networks. The empire even extended low-interest loans to merchants, effectively serving as venture capital for long-distance trade (Vaissiere, 2016). Such arrangements were inconceivable in earlier sedentary empires, yet under the Mongols, they became the backbone of Eurasian commerce. Traditional historians often acknowledged Mongol patronage of merchants but interpreted it as opportunistic, an external borrowing from the cultures they conquered. Revisionist scholars, by contrast, highlight it as evidence of active hybridisation that the Mongols did not simply adopt Persian commercial models but re-engineered them to fit a nomadic ethos of risk-sharing and mobility (Allsen, 2001). The result was an integrated economic network where East and West met in markets from Khanbaliq to Tabriz. The fourteenth-century merchant Pegolotti famously noted how travel from the Black Sea to China had become faster and safer than in centuries past, a testament to the hybrid commercial structures of Pax Mongolica (trans. Yule, 1914). The postal relay system further illustrates this negotiation. While courier systems had precedents in the steppe and in Chinese dynasties, the Mongols expanded and fused them into an empire-wide communications grid. Relay stations spaced across vast distances provided horses, provisions, and shelter for envoys, couriers, and merchants (Rossabi, 1988). At the centre of this system was the paiza (gerege), a tablet of authority that entitled its bearer to imperial privileges. Historians note that the paiza itself was a hybrid artefact. It drew on Chinese and Khitan precedents of ranked tablets but was redesigned under Mongol auspices to integrate Persian, Uyghur, and Mongolian scripts. One paiza discovered near the Dnieper River bore trilingual inscriptions, which is a literal embodiment of Mongol hybridity in governance (Atwood, 2004). Traditional historiography saw the Yam and paiza as pragmatic borrowings; revisionists insist they demonstrate a sophisticated capacity to integrate steppe and sedentary technologies into new imperial forms. For scholars of hybridity, the Yam exemplifies liminality: it was both nomadic and bureaucratic, both local and universal. Its role extended beyond administration to enable cross-cultural contact, allowing figures like Marco Polo, William of Rubruck, and Ibn Battuta to traverse Eurasia (Dawson, 1955). In this sense, the Yam was not merely logistical infrastructure but a cultural artery, shaping how people encountered one another and how knowledge moved across continents. Another domain where Mongol hybridity is evident is religion and intellectual life. The khans patronized Buddhist lamas, Nestorian monks, Islamic jurists, and Daoist sages, often simultaneously. Karakorum and Khanbaliq became cosmopolitan centers where multiple religious traditions coexisted under imperial protection (Blair, 2005).

Traditional historians interpreted this pluralism as political expedience. The Mongols, having no religious orthodoxy of their own, tolerated all faiths out of indifference. Revisionist accounts see it differently. For Allsen (2001) and Rossabi (1988), this pluralism was an intentional policy that leveraged the spiritual authority of many traditions to legitimate Mongol rule. It was not absence but hybridity: the khans fused steppe traditions of shamanic openness with institutionalised religious patronage from conquered lands. This hybridity extended to science and medicine. The Ilkhan Hülegü founded the Maragha Observatory in Persia in 1259, staffed by astronomers from China, the Islamic world, and beyond. Knowledge flowed in both directions: Chinese astronomy informed Persian calculations, while Islamic planetary models circulated eastward. Rashid al-Din's encyclopedic histories incorporated Chinese agricultural and medical practices, while Kublai Khan established institutes for Islamic medicine in Yuan China (Rossabi, 1988). Revisionist historians emphasise these as deliberate acts of knowledge hybridisation, producing cosmopolitan centres of learning unmatched in their time.

Henceforth, Mongol hybridity matters because it forces us to confront the tension between coherence and contradiction in both history and methodology. For the historian, it is a reminder that cultural encounters are messy, negotiated, and productive. For the social scientist, it is a warning that AI-mediated interpretation, unless carefully managed, may smooth away precisely the contradictions on which hybridity depends.

Conventional Interpretation vs. AI-Mediated Interpretation:

Interpretation of Mongol cultural hybridity has long depended on the human scholar's agency engaged in the framing of questions, the careful curation of sources, and the contextual close reading of texts. Revisionist historians stress that multilingual decrees and administrative borrowings were not mere "copying" but negotiations that produced hybrid forms (Biran, 2013). For example, Persian bureaucratic terms embedded in Mongol legal decrees signal a deliberate fusion of nomadic law and sedentary governance. In the Conventional research workflow, this ambiguity is not dismissed as noise but treated as the core of interpretation, with hybridity serving as precisely the tension between Mongol steppe traditions and Persian or Chinese institutional logics. A human historian's close reading foregrounds this liminality. A bilingual decree, issued in Persian and Mongolian, would be read for its polyvocality as to how Mongol titles and Persian divan terms sit uneasily together, reflecting the delicate balance of power. Conventional historians debate whether such a decree reflected Mongol appropriation of Persian forms or, conversely, Persian adaptation to Mongol frameworks. Importantly, these debates preserve the contradictions, and they resist the urge to resolve hybridity into a singular meaning. By contrast, AI-mediated interpretation privileges coherence. Symbolic AI, in the style of computational text analysis (CTAM), would reduce hybridity to keyword counts such as "X% Mongol terms, Y% Persian terms." As Baden (2022) warns, such rule-based approaches replace operational validity with plausible measurement. The result is a checklist wherein hybridity is an additive presence rather than a negotiated meaning. The relational dimension that Persian words in a legal decree might mean something different than in a Persian-only context is lost.

The risk deepens with machine learning approaches. A supervised classifier trained on Persian-dominant corpora could easily conflate hybridity with Persian bureaucratic dominance, because Persian sources were often the best preserved and most widely transmitted. Unsupervised models, such as topic clustering, would likely partition texts into "Mongol legal" vs. "Persian administrative," thereby dissolving hybridity into separable categories. Both reproduce what Ivanov (2025) calls predictive fit, replacing construct validity. Revisionist historians insist that the khanates' continuation after 1260 was not dissolution but transformation, with hybridity binding disparate regions together under a Chinggisid framework. Yet AI mediation risks reinscribing the old "anomaly" thesis. If clustering or classification pushes Persian bureaucratic elements into one group and Mongol legal terms into another, the hybrid experiment appears to vanish; the Mongols become once more "borrowers" rather than originators. This is precisely the danger Bhabha (1994) warns against, wherein hybridity is flattened into binary opposites rather than preserved as a "third space" of negotiation. Generative AI (LLMs) introduces subtler distortions. A decree brimming with contradiction such as Persian fiscal terminology alongside Mongol invocations of Eternal Heaven might be paraphrased by an LLM into a fluent, coherent summary. Grossmann *et al.* (2023) call this the illusion of coherence wherein a narrative that sounds authoritative but strips away tension. Cappelli *et al.* (2024) note that LLMs exhibit implicit intelligence in routine contexts but default to latent perplexity with its averaged interpretations when faced with complexity. In the Mongol case, this could mean erasing ambiguity and delivering a smoothed version of hybridity as "Persian-inspired Mongol law," a phrase that misses the liminal negotiations that revisionists foreground. Even multi-agent AI, imagined as simulating different perspectives (a Mongol chronicler, a Persian vizier, a European traveler), tends toward premature consensus. As Cappelli et al. show, simulated agents converge on moderation, ironing out disagreements for efficiency (2024). For hybridity, this is catastrophic because it's the clash between Rashid al-Din's Persian universalism and Marco Polo's Venetian cosmopolitanism that is precisely what makes Mongol sources rich. A system that harmonises them risks erasing the contestation that revisionist scholars work hard to sustain.

Underlying these AI-mediated distortions are pre-analytic interventions that already bias interpretation. If the curated corpus privileges Persian chronicles over Mongol decrees, the system "sees" hybridity as Persianization. If the prompt asks for "administrative influence," the AI is steered toward attributing agency to Persian bureaucrats rather than Mongol innovations. As Baden (2022) emphasises, computational tools rarely align with social-scientific concerns for construct validity, leaving meaning hostage to design choices. Moreover, AI introduces feedback loops into the interpretive process. A researcher dissatisfied with an initial AI output ("Persian dominance") might adjust prompts to emphasise Mongol agency. Yet each iteration risks reinforcing the model's bias toward coherence. Ivanov (2025) warns of this automation drift because iterative querying nudges interpretation toward the already-known rather than the ambiguous or marginal. In effect, the interpretive locus expands into a human–AI–text interaction, but one where the machine's structural bias toward regularity exerts pressure. Seen through the lens of agency–structure interaction, AI becomes a co-constructor of meaning. Human historians bring theoretical frameworks of hybridity (Bhabha, 1994; Canclini, 1995), but AI systems bring structural constraints wherein corpora are shaped by archival dominance, algorithms tuned for coherence, and metrics optimised for predictive fit. Meaning emerges from their interaction, but unless reflexivity is sustained, the structure overwhelms agency. The result is an epistemic regression, and hybridity is reinterpreted as an anomaly with contradiction smoothed into coherence.

Thus, in comparing conventional and AI-mediated interpretations of Mongol hybridity, the stakes become clear. Human close reading foregrounds ambiguity as substance; AI mediation flattens it into coherence. Revisionist historiography insists on hybridity as negotiation; AI risks returning us to positivist binaries of "borrowers" and "originators."

Historiographical Stakes:

The study of the Mongol Empire has always been entangled with larger historiographical debates about civilisation, empire, and cultural legitimacy. In early Western scholarship, the Mongols were seen as a violent eruption and "barbaric destroyers" whose empire was "a tempest rather than a tradition" (Grousset, 1973; Dawson, 1955). This conventional reading emphasised death tolls and devastation, frequently citing Rashîd al-Dîn or Juvaynî for accounts of slaughter and ruin, while overlooking the rich institutional and cultural syntheses that emerged in the wake of conquest. Berthold Spuler and V.V. Barthold similarly stressed dissolution after 1260, interpreting the fragmentation into khanates as the collapse of a failed, derivative polity. Such narratives, produced within a Eurocentric frame, reinforced a civilizational binary between nomadic destructiveness and sedentary creativity. Revisionist historians have worked to undo this interpretive bias by foregrounding hybridity as the key to Mongol rule. Thomas Allsen, Michal Biran, Marie Favereau, Nicola di Cosmo, and Morris Rossabi argue that the khanates remained bound together by shared Chinggisid institutions and continued cross-cultural exchange, even after political partition. They emphasise hybrid institutions such as the state-sponsored merchant partnerships, the postal relay system, and multilingual chancery practices as evidence that Mongol governance was not mere appropriation but a deliberate synthesis of steppe and sedentary traditions. Rashîd al-Dîn's Compendium of Chronicles, written in Persian yet incorporating Chinese historiographical techniques, exemplifies how hybrid forms of knowledge were consciously crafted under Mongol patronage. The very presence of Chinese motifs in Persian miniature painting, or Persian terminology embedded in Mongol yasa decrees, testifies to a creative imperial culture that thrived on negotiation and liminality rather than on uniform imposition. It is precisely here that the stakes of AI-mediated interpretation emerge. As Baden (2022) shows, computational text analysis methods (CTAM) often prioritise predictive performance over conceptual validity, thereby privileging what is measurable at the expense of what is meaningful. A symbolic AI approach, for instance, might parse Rashîd al-Dîn's Persian text as evidence of Persian dominance in the Ilkhanate, reinforcing the conventional claim that Mongols were cultural borrowers. An unsupervised clustering model could split Mongol decrees into "Mongolian" and "Persian" categories, dissolving hybridity into neat partitions and erasing the liminal space that Bhabha (1994) identifies as the "third space" of cultural negotiation. Generative AI compounds this risk: by producing fluent and coherent summaries, it generates what Grossmann *et al.* (2023) call the "illusion of coherence," masking the very contradictions that revisionist historians like Biran and Allsen treat as analytically central.

The political consequences of such flattening are profound. As Karpouzis (2025) argues in the context of digital humanities, AI tools often "reproduce colonial legacies embedded in archives." Pavlidis (2022) adds that digital methodologies tend to privilege well-preserved, often Eurocentric sources, sidelining voices from non-Western contexts. Catelan (2025) shows that AI-driven heritage projects in the Arabian Gulf can structurally determine "whose voices are visible." Münster (2024) similarly warns that AI applications in digital heritage risk "smoothing over contested legacies under the guise of innovation." If these warnings are transposed onto Mongol studies, they suggest that AI could inadvertently amplify older Eurocentric narratives, re-centring Chinese or Persian voices while minimising Mongol agency. This convergence of empirical bias and technological mediation highlights the danger of what Ivanov (2025) calls "overreliance on AI," which risks steering interpretation toward the already known and the statistically dominant. In the case of the Mongols, this means sliding back into Grousset's or Spuler's categories of barbarism and dissolution, not because historians choose them, but because AI's training data and optimization for coherence nudge outputs toward such frames. As Floridi (2011) reminds us, every representation of reality is constructed at a Level of Abstraction (LoA). AI models, operating at coarse LoAs that privilege frequency and regularity, may erase the subtle markers of hybridity in Mongol decrees, art, or diplomacy, treating them instead as anomalies to be explained away. Hopf's (2013) notion of "common-sense hegemony" underscores the risk. Hegemonic discourses are powerful not because they are true, but because they are easy to believe and reproduce. If AI systems are trained on corpora saturated with Eurocentric assumptions, they will embed this "common sense" into their outputs. In practice, an AI summarizer might confidently state that "the Mongols relied on Persian administrators because they lacked their own institutions," echoing Spuler's anomaly thesis, while ignoring the evidence of Mongol innovation in the *ortoq* or yam systems. Such outcomes illustrate how AI's mediation is never neutral but epistemically loaded, channeling interpretation along dominant grooves. The historiographical stakes are thus double. First, there is the danger of regression: decades of revisionist scholarship that painstakingly recovered Mongol hybridity could be undermined by AI systems that naturalize earlier stereotypes. Second, there is the broader risk of epistemic homogenization. As Bommasani *et al.* (2021) caution, foundation models trained across domains can act as "epistemically and culturally homogenising" forces. In historical research, this means not just losing Mongol voices but systematically erasing the polyvocality that interpretive scholarship depends on. The Mongol case, with its reliance on translated sources (Persian chronicles, Chinese annals, European travelogues), is especially vulnerable to the "English-before-everything" gap (Baden, 2022; Bender, 2011; Crawford, 2021), since AI systems overwhelmingly prioritize Anglophone materials. For revisionist historians, the challenge is clear. The Mongol Empire is not only an object of study but also a methodological test case. If AI-mediated interpretation can sustain hybridity—preserving the contradictions between Persian and Mongol terms in decrees, or the coexistence of Buddhist, Christian, and Islamic practices at the Yuan court—then it may offer useful support. But if, as current evidence suggests, LLMs display "latent perplexity" (Cappelli *et al.*, 2024) and converge toward moderate, coherent answers in the face of complexity, then AI risks failing precisely where interpretive openness is most needed. The empire that once connected Eurasia in unprecedented ways may, in AI-generated histories, be remembered once more as a mere "barbarian tempest"—a flattening that reveals not the past itself but the biases of our computational mediators.

Discussion: Bias, Validity and Reflexivity

The use of AI at the Interpretative stage of Social Science Research highlights critical issues of bias, validity and the need for reflexivity. The prospect of a nuanced interpretation that is immensely relevant to the study of social phenomena doesn't go along with the algorithmic logic of AI. What AI brings to the Qualitative research workflow in terms of high efficiency and scope can result in the loss of subtleties, ambivalence, and contradictions. Henceforth, there is a need for human researchers to remain at the centre stage of the Interpretative paradigm of Qualitative Research workflow. This calls for vigilance on the part of human researchers to keep intact the contextual understanding and final interpretation and not accept the AI output uncritically. This section outlines the Reflexive strategies for AI-assisted research, and keep in mind design implications and ethical imperatives while using AI such that it keeps the values of Interpretative Research intact.

Reflexivity strategies

The practice of systematically and critically analysing one's assumptions, positionality and methodological choices is described as reflexive practice. It has long been at the core of interpretative social sciences (Geertz, 1973; Bhabha, 1994). The advent of AI into the research workflow necessitates the extension of reflexivity onto the technology itself, requiring the human researcher to reflect upon how AI mediates data, reshapes analysis and introduces its own epistemic biases. As Grossmann *et al.* (2023) rightly point out, such systems often create "an illusion of coherence" which smoothens over ambiguity in ways that conflict with interpretative traditions that thrive on contradiction and plurality. For mitigating such risks, researchers must incorporate reflexive practices into their integration of AI in research through strategies of triangulation with non-AI models, bias Auditing of AI outputs, and openness to contradictory Readings.

The technique of triangulation requires the simultaneous use of a multitude of data sources, methods, and researchers to come to a comprehensive understanding of a phenomenon. Denzin's typology distinguishes between method, investigator, theory and data triangulation (1978). In AI-mediated interpretation, this principle takes an even more urgent outlook. This process can play out as drawing parallels between AI-assisted codding with manual coding or using different AI models to determine if they come to similar thematic analyses and even pairing AI-assisted topic modelling with thick description (Geertz, 1973) or validating machine-coded survey simulations against archival sources. In this regard, Ivanov (2025) correctly points out how overreliance on AI for data analysis may give misleading results and irrelevant conclusions. Thus, the issue of transparency, traceability and explainability that come with AI analyses can be averted with such cross-verification. By treating the AI as one of the interpretive lenses, rather than the sole analyst, researchers can test the consistency of results across approaches. This convergence of evidence strengthens the credibility of qualitative insights. A more practical extension of triangulation can be the identification of the three-tiered practices of "problem formulation, prompt engineering and prompt interaction" (Cappelli *et al.*, 2024). Triangulation can be enacted across these tiers by comparing AI outputs with human-coded interpretations, manual close readings, or alternative computational models. This is particularly relevant for constructs like cultural hybridity, where reductive operationalization, such as loanword counts or sentiment scores, risks erasing liminality. As Baden (2022) posits, computational text analysis methods (CTAM) tend to privilege predictive accuracy over conceptual validity, thereby flattening interpretive nuance. Triangulation also means refusing to cede epistemic authority to black-box systems. Instead, as Mokander and Schroeder (2021) argue, AI must be situated as part of broader socio-technical systems where its knowledge claims are checked against existing theoretical traditions.

There is a tendency of AI to replace construct operationalisation with a powerful algorithm that is trained to identify any patterns and indicators that draw upon correlation with provided annotations, thus supplanting construct validity with predictive performance (Baden, 2022). As a result, AI systems can inadvertently reproduce societal biases present in training data. In an interpretive research context, bias auditing refers to rigorously investigating AI-generated results for unfair patterns, such as consistently negative interpretations of texts from a certain group and examining whether the model might be reflecting dominant cultural assumptions. An influential audit by Buolamwini and Gebru (2018) revealed that commercial facial recognition AI had near-perfect accuracy for white male faces but erred up to 35% on darker-skinned women. This stark disparity, uncovered through an algorithmic audit, illustrates how AI can encode racial and gender biases if left unchecked. Researchers like Noble (2018) have shown that ostensibly neutral algorithms (e.g. search engines) can reinforce racial or gender stereotypes. Therefore, before concluding, one should perform "bias tests" on AI outputs and examine results for any systematic exclusion or misrepresentation of marginalised perspectives. In case of any issues, researchers can adjust the AI (through re-training on more diverse data or prompt engineering) and document these mitigations. Without deliberate checks, algorithms may misclassify constructs, privileging what is easily correlated rather than conceptually valid. Hirst *et al.* (2014), for instance, found that a machine classifier trained to recognise political ideology ended up classifying incumbency, a proxy easier to detect but conceptually distinct. Ivanov (2025) links this issue to ethical standards of transparency and fairness, noting that black-box AI "violates methodological transparency" and that English-first CTAM structurally "privileges certain populations and disadvantages others". In cultural hybridity research, bias auditing requires checking how training data shape interpretive categories. If Persian administrative sources dominate a corpus on the Mongol empire, AI may systematically skew interpretation toward "Persianization," erasing the liminality of Mongol governance. Bias auditing can help identify and correct such distortions before they calcify into scholarship. Bias auditing, therefore, serves both epistemic and ethical functions, ensuring that marginalised voices are not erased in the pursuit of computational efficiency. Thus, proactive auditing and *debiasing* AI outputs can help ensure that the conclusions reflect a balanced interpretation rather than amplifying existing inequities.

Grossmann *et al.* (2023) warns about that the creation of an "illusion of coherence," where outputs appear plausible but suppress tension and ambiguity. Cappelli *et al.* (2024) document a related phenomenon of "latent perplexity," where large language models (LLMs) default to moderate, midpoint answers when uncertain models tended to gravitate towards the midpoints of the scale. While this tendency can be mistaken for balanced judgment, it represents a flattening of interpretive possibility. For interpretive social science, where contradiction and plurality are not noise but substance, such smoothing is deeply problematic. As Geertz (1973) argued, interpretation requires "thick description," which foregrounds the irreducible complexity of cultural meaning. Bhabha's (1994) notion of hybridity likewise stresses the productive tension of the "third space," where identities remain in negotiation. To accept AI's coherent answers uncritically is to foreclose this space. Therefore, it won't be wrong to say that perhaps the most important reflexive strategy is maintaining openness to contradictory readings. Maintaining openness means deliberately resisting AI's gravitational pull toward uniformity. In qualitative analysis, this aligns with seeking negative cases and alternative explanations as part of analytic rigour. Rather than accepting the themes an AI finds as final, an analyst should ask: What might I or the AI be missing? It is advisable to actively look for evidence in the data that challenges the AI-suggested patterns. This might involve double-checking transcripts or field notes for different interpretations or inviting other researchers (or community members) to offer alternate readings of the same material. By embracing contradictory readings, researchers practice reflexivity wherein they acknowledge that one's (or the AI's) interpretation is not the only possible one. This strategy echoes the idea of *theoretical sensitivity* in grounded theory (Glaser and Strauss) and the practice of constantly comparing data against emerging interpretations. Maintaining this openness guards against the complacency that AI's seeming objectivity can induce. It reminds us that AI outputs are 'constructed' interpretations, subject to error or bias, and thus must be weighed against context and human insight. Wendt's (1999) constructivist emphasis on intersubjectivity underscores that meaning is never singular but always co-constructed, a principle directly at odds with AI's tendency to deliver one "best" output. In essence, researchers should cultivate what some have termed interpretative humility, which pertains to the understanding that any result is provisional, and other viewpoints may reveal different truths. By being willing to revise or even contradict AI-derived findings, scholars ensure a more reflexive and robust interpretive process. In this light, a reflexive approach of the researcher entails critically interrogating the utility of the AI tool by asking who built and trained the model, whose voices and values are embedded in it, and what blind spots that entails. With such questions in focus, scholars ensure that the use of AI remains a subject of analysis in its own right, rather than an invisible given. This meta-reflexivity, wherein one reflects on how AI influences the research, needs to be considered a vital part of methodological rigour when technology is involved.

Design Implication:

The way we design and deploy AI in research can either support the nuanced understanding of cultures or inadvertently encode cultural biases and hybridities as fixed entities. Cultural hybridity refers to the mixing and blending of cultural identities and meanings. This concept was famously articulated by Homi K. Bhabha. Bhabha describes culture as emerging in a "Third Space of enunciation," an ambivalent in-between area where new, hybrid meanings form beyond strict cultural binaries. Rather than treating cultures as pure or static, this perspective highlights that all cultural expressions are influenced by multiple contexts and are constantly evolving (1994). Recognising this, AI tools used in social research should be designed to handle and even embrace such fluidity. The design of AI systems for social science should reflect reflexive priorities. Rather than encoding hybridity into fixed categories, tools must support interpretive practices that sustain ambiguity. This requires Interfaces that foreground multiple, even contradictory outputs, Options for researchers to document and annotate uncertainty in AI results and Corpus design that incorporates multilingual and marginal sources, countering English-dominant bias (Baden *et al.*, 2022). As Cappelli *et al.* (2024) emphasise, careful prompt engineering and ongoing interaction are key to steering AI toward meaningful outputs. But design must also resist automation drift, where human oversight gradually cedes to machine authority. Embedding reflexive checkpoints into AI pipelines can help maintain interpretive pluralism.

Currently, one danger is that AI systems trained on large datasets might flatten cultural nuances. If AI is fed predominantly Western-text data, for example, it may interpret inputs through a Western-centric lens, thereby "encoding" a particular cultural standpoint as the default. To avoid this, AI design must prioritise cultural sensitivity and adaptability. Recent work advocates for inclusive design strategies in AI development, emphasising the need to involve diverse stakeholders and represent multiple cultural contexts in the training and tuning of AI models. By incorporating voices from different cultural backgrounds (through participatory design or community consultation), developers can identify where an AI might misunderstand local idioms, values, or historical contexts. This can lead to features that allow context-switching or culturally contextualized responses, rather than one-size-fits-all outputs. Another design implication is the importance of continuous adaptation. Culture is not static, and neither should AI models be. Nelson *et al.* (2025) highlight that successful culturally aware AI systems often incorporate mechanisms for feedback and adaptation, learning from user interactions in different cultural settings. For example, an AI text analysis tool could allow researchers to input cultural context notes or select interpretive frameworks (e.g., feminist, indigenous, postcolonial perspectives) that guide its analysis. Such features would enable the AI to support cultural hybridity by adjusting to different interpretive lenses, rather than encoding only the dominant cultural narrative. In practical terms, to support cultural hybridity means the AI should help reveal the intersections and blending of cultures in the data. For instance, detecting multiple vernaculars or reference frames in participant interviews, instead of forcing data into a single cultural frame. It is equally crucial to avoid designing AI in ways that reify or exoticise culture. Bhabha warned against the "exoticism of multiculturalism" that treats cultures as monolithic others, advocating instead for recognising the empowering nature of hybridity (1994). An AI that supports cultural hybridity would, for example, allow a researcher to see how participants navigate mixed cultural influences in their narratives, rather than pigeonholing participants into preset cultural categories. Technical implementations might include algorithms that can handle code-switching in language, or image recognition systems attuned to diverse aesthetic norms, all developed with input from culturally varied data. By contrast, an AI that simply encodes hybridity might, say, label an interviewee's perspective as "hybrid" and then treat it as a fixed attribute, losing the dynamic process aspect. Therefore, the design goal is to keep AI flexible and context-aware. Therefore, adapting AI tools for interpretative research in a culturally complex world involves embracing polycultural approaches. This entails training AI on diverse datasets, using inclusive design (co-design with communities, interdisciplinary teams), and building in options for contextual calibration (so the AI can adjust its analysis based on cultural context input by the researcher). By doing so, AI can become a vehicle for exploring cultural hybridity, illuminating how global and local influences intertwine in the data and, rather than a blunt instrument that unknowingly reinforces cultural hierarchies or stereotypes. As one report puts it, a holistic, culturally sensitive AI design recognises and respects cultural differences, ultimately ensuring more equitable and meaningful insights for all users.

Ethical Imperatives:

Finally, the integration of AI into interpretative social science research brings ethical imperatives that scholars must heed. Three interrelated principles stand out, including responsible use of AI, transparency about its role, and interpretative humility in handling its outputs. Responsible use of AI means employing these tools in ways that do not cause harm and that uphold the integrity of research. Researchers have a duty to ensure that AI assistance does not lead to ethical violations such as privacy breaches (e.g., if using AI on sensitive interview transcripts) or the marginalisation of certain voices. This may involve obtaining informed consent for any AI analysis of participant data and ensuring data are handled securely. It also means actively working to mitigate biases such as auditing the AI and correcting bias is part of being responsible. In line with emerging AI ethics guidelines, scholars should evaluate the fairness of AI models and refrain from using those known to produce discriminatory outcomes. Responsible use might also include setting clear boundaries on what tasks are delegated to AI. For example, using AI for draft analysis is fine, but perhaps not letting AI solely decide which quotes from an interview are "important," since that value judgment might hide bias. In essence, the researcher remains the moral and analytical compass of the project, using AI to augment, not replace, careful human interpretation. Transparency is a fundamental ethical and scholarly requirement when AI is involved. Transparency entails openly documenting and communicating how AI was used in the research process. This includes reporting in publications in which the analyses were AI-assisted, what software or model was used, and how its outputs were verified or adjusted. Christou (2023) has argued that researchers need to explicitly acknowledge the use of AI in qualitative studies, rather than leaving it unmentioned, because this disclosure is crucial for evaluating the study's trustworthiness. The rationale is that readers and other researchers should be able to understand the role AI played in shaping findings. Moreover, transparency is tied to accountability. If an AI contributed to an insight, disclosing that invites scrutiny of whether the insight might reflect the AI's limitations. Best practices now encourage maintaining an "audit trail" of AI interactions, including logging the prompts given, the versions of models, and any parameters, as part of the project's documentation. By being transparent, researchers also contribute to a learning culture where the community can better assess and improve AI methods for qualitative research. On the flip side, a lack of transparency (treating AI as a black box assistant) could erode trust in the research, as stakeholders might wonder what biases or errors lurk behind undisclosed AI contributions. Thus, transparency serves both ethical integrity and the reflexive validity of the research.

Interpretative humility is perhaps a newer ethical ideal in the context of AI, but it is deeply resonant with qualitative traditions. It means approaching AI-generated analyses with caution, modesty, and openness to correction. The researcher should avoid overstating what an AI's output signifies. For example, if an AI finds a pattern in interview data, interpretative humility urges the researcher to present it as one possible interpretation, not as an objective truth or a conclusive result. This stance recognises the fallibility of AI. As Conitzer, Hadfield and Vallor (2023) note, algorithms themselves are not "biased" in a moral sense, but they reflect the biases of data and design and can amplify human prejudices. An ethically humble interpreter of AI results will acknowledge these limitations. They will also remain aware that social phenomena are complex and often can't be fully captured by patterns an AI detects. In practice, interpretative humility can manifest as researchers double-checking AI findings with participants or community interpretations and being willing to discard an AI-suggested theme if it doesn't hold up upon closer human-centric examination. It also involves being cautious in claims, for instance, instead of saying "The AI proved X," a report might state, "An AI-assisted analysis suggested X, but further examination revealed Y, indicating the need for nuanced interpretation." Embracing this humility aligns with the qualitative ethos of treating data and their meanings as co-constructed and context dependent. Finally, reflexivity carries ethical imperatives. International frameworks emphasize transparency, auditability, fairness, and accountability as principles of responsible AI use (OECD, 2024; European Commission, 2024). Ivanov (2025) translates these into the research process when he posits that "Responsible AI in research must be situated within the established research process, which runs from ideation through interpretation and manuscript evaluation". For interpretive social science, this means disclosing not only which AI tools were used but also how their outputs were challenged, audited, and contextualised. Transparency without reflexivity risks tokenism; reflexivity without transparency risks invisibility. Humility is equally critical. Cappelli *et al.* (2024) stress that LLMs exhibit "implicit intelligence" but also "latent perplexity". To treat them as co-interpreters rather than replacements requires humility about their limits and about researchers' own complicity in shaping their outputs. As Grossmann *et al.* (2023) note, the danger is not that AI will interpret too little but that it will interpret too smoothly, masking the complexity that is the lifeblood of interpretive scholarship. Therefore, the ethical imperatives demand that researchers use AI thoughtfully, tell others how they used it, and remain humble about what its outputs mean. By clearly communicating the involvement of AI and critically evaluating its contributions, scholars demonstrate an ethical commitment to honesty and rigor. And by staying humble and acknowledging the AI's suggestions without idolising them, they ensure that the interpretative authority ultimately rests on reasoned, reflexive understanding, not on the allure of machine "objectivity."

Taken together, bias mitigation, reflexive practice, culturally-aware design, and ethical transparency form a comprehensive framework for integrating AI into interpretative social science. These measures help transform AI from a potential threat to validity into a helpful partner that, when guided by conscientious scholars, can deepen and enrich qualitative inquiry without sacrificing its core humanistic values. The promise of this partnership is substantial where AI can prompt new ways of seeing and questioning data but realising it requires that we adapt the tool to our critical frameworks, rather than adapting our frameworks uncritically to the tool. By doing so, researchers uphold the spirit of interpretative inquiry in the age of AI by remaining ever-attuned to context, complexity, and the co-creation of meaning, even as we leverage new technologies to explore them.

Conclusion:

This paper argues that AI must be understood not only as a computational tool but as a mediator and, at times, a co-interpreter in the social sciences. The precedence is already set by AI-mediated communication, in which machines intervene in the construction of messages. This paper extends the concept to the interpretative stage of social science research, wherein the process is deeply involved in meaning-making (Hancock *et al.*, 2020). At this stage, the stakes at the level of knowledge production are significant as interpretation thrives on ambiguity, ambivalence, and plurality, and reduction into coherent, uniform categories is highly condemnable. Through a conceptual simulation of Mongol cultural hybridity, this study has demonstrated how different AI paradigms intervene in distinct yet convergent ways. Across these paradigms, on the one hand, AI-mediated research, even in social science, follows a pattern marked by coherence, regularity, and efficiency; on the other hand, interpretive research demands openness, contradiction, and plurality. This tension connects to broader theoretical traditions. For constructivism, meaning is intersubjectively constituted (Wendt, 1999); for cultural hybridity, meaning emerges in liminal "third spaces" (Bhabha, 1994; Canclini, 1995); for interpretive anthropology, meaning resides in "thick description" (Geertz, 1973). These traditions underscore that interpretation is dialogic, negotiated, and plural. By contrast, AI systems, shaped by their training corpora and optimisation goals, risk reinscribing positivist epistemologies, privileging what is legible to machines over what is meaningful to humans. As Crawford (2021) reminds us, AI infrastructure is never neutral; it encodes histories of power, exclusion, and bias. Thus, the challenge is not whether AI can assist interpretation but how it mediates, and whether researchers remain reflexively aware of its assumptions and limitations. Without reflexive safeguards, AI's mediation can ossify into automation, privileging dominant narratives and flattening ambiguity. With reflexive oversight, however, AI can serve as a co-interpreter. A partner that amplifies scale and iteration, while humans preserve the interpretive openness needed for constructs like hybridity. The Mongol case demonstrates that what is at stake is not merely methodological but historiographical. Thus, AI-mediated interpretation risks reinforcing Eurocentric narratives that dismiss the Mongols as anomalies or borrowers. However, the utilisation of reflexivity in the use of AI-mediation makes way for pluralising archives, scaling comparative inquiries, and modelling the dynamics of hybrid governance without erasing its ambiguities. In this sense, the contribution of this paper is twofold. Firstly, it extends the framework of AI-mediated communication into AI-mediated interpretation. Secondly, the case of Mongol cultural hybridity highlights the essential role of reflexivity in AI mediation, without which there lies the risk of reinscribing positivist drift and epistemic bias into interpretive traditions. The immense importance of plurality, contradiction, and negotiation in social science research, particularly at the level of interpretation, cannot be denied. And this is even more problematic with the use of AI at the stage of research methodology. The resolution to this issue is not to refuse AI, but to ensure that its role as mediator and co-interpreter is situated within reflexive, plural, and ethically responsible practice.

REFERENCES

Awasthi, A.and Awasthi, K. (2022), Homi K. Bhabha's third space theory and cultural identity today, *Prithvi Academic Journal*, **5** : 171–181.

Atwood, Christopher (2004). *Encyclopedia of Mongolia and the Mongol Empire*. Facts On File.

Allsen, T.T. (2001). *Culture and Conquest in Mongol Eurasia*. Cambridge University Press.

Amitai, R. (1995). *Mongols and Mamluks: The Mamluk-Ilkhanid War, 1260–1281*. Cambridge University Press.

Baden, C., Pipal, C., Schoonvelde, M. and van der Velden, M. A. C. G. (2022). Three gaps in computational text analysis methods for social sciences: A research agenda. *Communication Methods & Measures*, **16**(1): 1–18. URL: https://doi.org/10.1080/19312458.2021.2013665

Bail, C. A. (2024). Can generative AI improve social science? *Proceedings of the National Academy of Sciences of the United States of America*, **121**(21): 231421121. URL:https://doi.org/10.1073/pnas.2314021121

Barfield, Thomas (1989). *The Perilous Frontier: Nomadic Empires and China*. Basil Blackwell.

Barthold, V.V. (1956). *Turkestan Down to the Mongol Invasion*, Trans. by T. Minorsky, London: Luzac

Biran, M. (2013). The Mongol Empire in world history: The state of the research. *History Compass*, **11**(11) : 1021–1033. Accessed online on 4 September 2023. URL: https://doi.org/10.1111/hic3.12074

Bhabha, H. K. (1994). *The Location of Culture*. Routledge.

Bommasani, R., Hudson, D. A., et al. (2021). On the opportunities and risks of foundation models. *Stanford University, Center for Research on Foundation Models (CRFM) Report*. URL:https://doi.org/10.48550/arXiv.2108.07258

Boumans, J. W. and Trilling, D. (2016). Taking stock of the toolkit: An overview of relevant automated content analysis approaches and techniques for digital journalism scholars. *Digital Journalism*, **4**(1) : 8–23. URL:https://doi.org/10.1080/21670811.2015.1096598

Buolamwini, J. and Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research*, **81** : 1–15.

Cappelli, O., Aliberti, M. and Praino, R. (2024). The "implicit intelligence" of artificial intelligence: Investigating the potential of large language models in social science research. *Political Research Exchange*, **6**(1): 1-32. URL: https://doi.org/10.1080/2474736X.2024.2351794

Carter, N., Bryant-Lukosius, D., DiCenso, A., Blythe, J. and Neville, A. J. (2014). The use of triangulation in qualitative research. *Oncology Nursing Forum*, **41**(5) : 545–547. URL:https://doi.org/10.1188/14.ONF.545-547

Chauhan, C. and Currie, G. (2024). The impact of generative artificial intelligence on research integrity in scholarly publishing. *American J. Pathology*, **194**(10) : 2234–2238. URL:https://doi.org/10.1016/j.ajpath.2024.10.001

Conitzer, V., Hadfield, G. K. and Vallor, S. (2023). Technical perspective: The impact of auditing for algorithmic bias. *Communications of the ACM*, **66**(1) : 80–82. URL:https://doi.org/10.1145/3578114

Crawford, K. (2021). *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press.

Dawson, C. (Ed.). (1955). *Mission to Asia*, Harper & Brothers.

Denzin, N. (1978). *The Research Act: A Theoretical Introduction to Sociological Methods*. McGraw-Hill.

Driessens, O. and Pischetola, M. (2024). Danish university policies on generative AI: Problems, assumptions and sustainability blind spots. *MedieKultur: Journal of Media and Communication Research*, **40**(76) : 31–52. URL:https://doi.org/10.7146/mk.v40i76.143595

Dwivedi, Y. K., *et al.* (2023). "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, **71**: 102-42. URL: https://doi.org/10.1016/j.ijinfomgt.2023.102642

Floridi, L. (2008). The method of levels of abstraction. *Minds & Machines*, **18**(3) : 303–329. URL:https://doi.org/10.1007/s11023-008-9111-7

Floridi, L. (2011). *The Philosophy of Information*, Oxford University Press.

Franklin, G., Stephens, R., Piracha, M., Tiosano, S., Lehouillier, F., Koppel, R. and Elkin, P.L. (2024). The sociodemographic biases in machine learning algorithms: A biomedical informatics perspective. *Life*, **14**(6) : 600-652. URL:https://doi.org/10.3390/life14060652

García Canclini, N. (1995). *Hybrid Cultures*, University of Minnesota Press.

Geertz, C. (1973). *The Interpretation of Cultures*, Basic Books.

Grossmann, I., et al (2023). AI and the transformation of social science research. *Science*, **380**(6650): 1108–1109. URL: https://doi.org/10.1126/science.adi1778

Gurung, T. W. (2025). Does AI make qualitative research more inclusive or less human? *Qualz.ai Blog*. Retrieved from https://www.qualz.ai/what-is-quality/

Hancock, J. T., Naaman, M. and Levy, K. (2020). AI-mediated communication. *Journal of Computer-Mediated Communication*, **25**(1) : 89–100. URL:https://doi.org/10.1093/jcmc/zmz036

Hodous, F. (2015). Clash or compromise? Mongol and Muslim law in the Ilkhanate (1258–1335). In A. Krasnowolska & R. Rusek-Kowalska (Eds.), *Studies on the Iranian World*, Jagiellonian University Press.

Hopf, T.G. (2013). Common-sense constructivism and hegemony in world politics. *International Organization*, **67**(2) : 317–354. Accessed online on 18 July 2023. URL:https://doi.org/10.1017/S0020818313000205

Howell, N., Hartsoe, W. F., Amin, J. and Namani, V. (2024). Reflective design for informal participatory algorithm auditing: A case study with Emotion AI, *Proceedings of Nordi CHI.* URL:https://doi.org/10.1145/3679318.3685411

Huang, L. T.-L. and Huang, T.-R. (2025). Generative bias: Widespread, unexpected, and uninterpretable biases in generative models and their implications. *AI and Society*. Advance online publication. URL:https://doi.org/10.1007/s00146-025-02533-1

Hughes-Warrington, M. (2025). *Artificial Historians*, Routledge.

Ivanov, S. (2025). Responsible use of AI in social science research. *Service Industries Journal*. Advance online publication.URL:https://doi.org/10.1080/02642069.2025.2537115

Ivanov, S. and Soliman, M. (2023). Game of algorithms: ChatGPT implications for the future of tourism education and research. *Journal of Tourism Futures*, **9**(2) : 214–221. Accessed online on 22 May 2025. URL:https://doi.org/10.1108/JTF-02-2023-0038

Juvaini, Ata-Malik (1958). *Genghis Khan: The History of the World-Conqueror*, Trans J. A. Boyle, Harvard University Press.

Kaplan, A. and Haenlein, M. (2019). Siri, Siri, in my hand: Who's the fairest in the land? *Business Horizons*, **62**(1) : 15–25. Accessed online on 22 October 2023. URL: https://doi.org/10.1016/j.bushor.2018.08.004

Karpouzis, S. (2025). Algorithmic consensus and the loss of polyvocality in historical simulations. *Journal of Computational History*, **3**(1) : 50–67. URL:https://doi.org/10.4995/jch.2025.19894

Kuhn, A. (2024). The impact of Eurocentric bias in AI-driven historical research, *Historica*: Digital Media and Culture Blog.

Lu, J. G., Song, L. L. and Zhang, L. D. (2025). Cultural tendencies in generative AI. *Nature Human Behaviour*. URL: https://doi.org/10.1038/s41562-025-02242-1

Mokander, J. and Schroeder, R. (2021). Artificial social intelligence and social theory. *AI & Society*, **37**(3) : 677–695. Accessed online on 22 October 2023. URL:https://doi.org/10.1007/s00146-021-01222-z

Münster, S., et al. (2024). Artificial intelligence for digital heritage innovation: Setting up a research agenda for Europe. *Heritage AI Journal*, **1**(1) : 10–24.

Naeem, M., Smith, T. and Thomas, L. (2025). Thematic analysis and artificial intelligence: A step-by-step process for using ChatGPT in thematic analysis. *International Journal of Qualitative Methods*, **24**(1) : 1–14. URL:https://doi.org/10.1177/16094069231182243

Nah, F. F.-H., Zheng, R., Cai, J., Siau, K. and Chen, L. (2023). Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration. *Journal of Information Technology Case and Application Research*, **25**(3) : 277–304. https://doi.org/10.1080/15228053.2023.2233814

Noble, S. U. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism*. NYU Press.

Paschalidis, A. (2025). AI and the great linguistic flattening. UNESCO. Retrieved from https://www.unesco.org

Pavlidis, G. (2022). AI trends in digital humanities research. *Trends in Computer Science and Information Technology*, **7**(2) : 26–34.

Putnam, H. (1975). *Mind, Language, and Reality*. Cambridge University Press.

Qadri, R., Diaz, M., Wang, D. and Madaio, M. (2025). The case for 'thick evaluations' of cultural representation in AI. *arXiv Preprint* arXiv:2503.19075.

Rossabi, M. (1988). *Khubilai Khan: His Life and Times*. University of California Press.

Russell, S. and Norvig, P. (2010). *Artificial Intelligence: A Modern Approach* (3rd ed.). Pearson.

Searle, J. R. (1969). *Speech Acts*. Cambridge University Press.

Searle, J. R. (1995). *The Construction of Social Reality*. Free Press.

Schorkowitz, Dittmar (2020). Mobility and Immobility in the Mongol Empire, *Mongolian Studies,* **12**(3): 430-445

Shani, C., Jurafsky, D., LeCun, Y. and Shwartz-Ziv, R. (2025). From tokens to thoughts: How LLMs and humans trade compression for meaning. *arXiv Preprint*.

Spuler, B. (1965). *Die Mongolen im Iran (The Mongols in Iran)*. Brill.

Tarski, A. (1944). The semantic conception of truth and the foundations of semantics. *Philosophy and Phenomenological Research*, **4**(3) : 341–375. https://doi.org/10.2307/2102968

Vaissiere, Etienne De La (2016). "Trans-Asian trade, or the Silk Road deconstructed (antiquity, middle ages)" in Larry Neal and Jeffrey G. Williamson (eds.), *The Cambridge History of Capitalism,* Cambridge: Cambridge University Press.

Waldron, P. (2025). AI suggestions make writing more generic and Western. *Cornell Chronicle*. Retrieved from https://news.cornell.edu/stories/2025/04/ai-suggestions-make-writing-more-generic-western

Walters, W. H. and Wilder, E. I. (2023). Fabrication and errors in the bibliographic citations generated by ChatGPT. *Scientific Reports*, **13**, 14045. https://doi.org/10.1038/s41598-023-41032-5

Wendt, A. (1999). *Social Theory of International Politics*. Cambridge University Press.

Williams, R. T. (2024). Paradigm shifts: Exploring AI's influence on qualitative inquiry and analysis. *Frontiers in Research Metrics and Analytics*, 9, Article 133-158. Accessed online on 8 February 2025. URL:https://doi.org/10.3389/frma.2024.1331589

****