ESP IJEMR
ESP International Journal of Emerging Multidisciplinary Research
Volume 1 Issue 1 October 2025 / Paper Id: IJEMR-V1I1P105/ Page No: 43-52

Original Article

Provocations from the Humanities for Generative AI Research

Devadharshini G[1], Kishalini C[2]
[1,2] *UG Scholar Holy Cross College, Tiruchirappalli, Tamil Nadu, India.*

Received Date: 28 August 2025 Revised Date: 27 September 2025 Accepted Date: 22 October 2025

Abstract: *The development of language, imagery, music and multimodal artefacts has already been disrupted by the swift progress of generative artificial intelligence (AI), leading to increasing confusion between algorithmic synthesis and human creativity. The interpretive, historical and cultural dimensions that the humanities have long studied are typically overlooked by engineering-centric research in generative models, which has focused heavily on optimisation, scale and benchmark performance. In arguing that interpretive modes such as literary theory, philosophy, archival intervention, and critical cultural analysis can help to illuminate significant blind spots in the practice of contemporary AI work, this paper advances several humanistic provocations for generative AI research. All of these provocations challenge the persistent notion that AI models are objective, context-free technologies. Instead, they highlight generative systems as interpretive agents enfolded in complex social networks that replicate biases, values and hierarchies which always need to be exposed to critical scrutiny. In so doing, this paper suggests that generative AIs should be approached as interlocutors of meaning rather than mere code-generating objects and that they could be tackled by hermeneutics, historiography and critical epistemology. At the same time, existing indicators do not account for provenance and authorship, chronology or rhetorical framing; in these humanities offer methodological tools that can be used to investigate such factors. For example, archival theory exposes data collection methodologies that produce structural silence and selective memory; close readings of models outputs reveal the narrative and ideologic factors enacted in representation. In this way the humanities ground critique in historically and materially particular ways, taking the conversation out of these kinds of abstract notions like bias or "ethics. The article reconceptualises generative AI as an interpretative space than mechanical output through nine provocations. These include calls to celebrate interpretative multiplicity, recognize labour and institutional power, remember historical contingency, privilege data provenance and treat models as interpretive objects among others. Further provocations draw attention to archival reflexivity, hermeneutic interpretability and ethically charged evaluative dimensions. Joined, they advocate for the integration of humanistic methodologies into AI research pipelines, prophylactically provenance-aware data infrastructures, participatory evaluation with and by affected communities, and "close reading" of models. The study concludes that, without a role for the humanities, there cannot be responsible development of generative AI. Ethical deliberation should form the basis, not as an afterthought to model generation, interpretation and implementation Humanistic inquiry should drive modelmaking. By defining generative systems as interpretive or cultural agents, and not only computational ones, researchers can create machine learning that respects historical context, multiple meanings and human creativity. Rather than stifling creativity, the provocations put forward work towards alternative and more reflexive, democratic, and culturally accented AI futures.*

Keywords: *Generative Artificial Intelligence, Humanities, Hermeneutics, Critical Theory, Data Provenance, Interpretive Plurality, Archival Studies, Rhetorical Analysis, Epistemic Justice, Cultural Context, Digital Humanities, AI Ethics, Sociotechnical Systems, Historical Contingency, Algorithmic Power.*

I. INTRODUCTION

Over the past ten years, generative AI has gone from a relatively unknown field of study to powering new and innovative advancements across business. > Algorithms that can generate code, art, music and natural language are already a feature of professional and personal life. Generative models have become human partners in meaning making, whether through automated journalism, academic help with understanding objects or phenomena, creative co-creation and policy draughting. The theoretical foundations of its design, however, continue to be dominated by the computer science and engineering models of safety, scalability and performance estimation (and optimisation), despite their widespread use. Even as these paradigms have produced stunning technical achievements, scant attention has been paid to the epistemic, cultural and historical implications of the generative systems they deploy. Rather than interpreting the models as participants in human-like conversation, the field often takes them to be passive tools.

Humanities, on the other hand, have decades of experience analysing text and meaning-making in pictures and objects. Fields that concentrate on contextual analysis, including by context the production / reading situation as well as the situation of address and power relations include such fields as literary studies, philosophy, history and cultural theory. Their methods — close reading, critique, historiography and hermeneutics — are exactly designed to investigate the same sort of

meaning-making procedures that underlie generative AI today. Language models and picture generators are culture-based as well as computation-powered acts of representation and narrative when they produce text or reconstruct images. They access shared linguistic, visual and acoustic repositories whose composition has been shaped by institutional biases, social hierarchies and historical exclusions. One needs interpretive literacy, not just algorithmic transparency, to make sense of these processes.

There are pros and cons in bridging generative AI and the humanities. On the other, AI opens up new horizons for research in digital humanities as it makes possible automatic annotation, wide-range text analysis, and original experimentation. But humanistic study of the humanities forces AI researchers to confront hard questions: Whose knowledge is included in those data? When culture becomes a set of vectors, who is left out? What sort of authority an authorship do machine-generated artifacts embody? How should culture evaluate or sustain these artificial artifacts? It's therefore important to connect such problems, which exceed the space of computational judgment, with the interpretative and ethical as well as political traditions of humanistic philosophy. To claim an authentic transdisciplinary epistemology, this paper posits that generative AI research needs to move beyond the binary of “technical performance versus ethical oversight”. The humanities provide rigorous analysis tools for understanding meaning-making, representation and interpretation as well as moral critique of technology. To think AI outside of this analytic model is not to treat machine-constructed conclusions as different things which need a different kind of judgment but as textual sources to be read. The colonial, racialized or gendered logics structuring the collection of data would be revealed if datasets were to be traced by their lineages from a historical perspective. How AI systems mediate authority—how some are given outputs of being neutral, authoritative, persuasive products—would be considered from a rhetorical approach. Together, these perspectives suggest that generative AI is a sociotechnical artifact and that cultural assumptions are woven together with technical design.

This reframing is urgently needed. Talk of AI ethics is now often reduced to management checklists about AI’s impact on humanity, devalue human dignity and autonomy, avoid aligning with such far-reaching implications for the interpretation of precise questions that AI will respond to in our lives — like explainability dashboards, toxicity filters or fairness metrics. And, while such mechanisms are indispensable, they are insufficient for understanding how generative systems affect human meaning. For instance, it is not only about “bias,” but also about historical and rhetorical distortion when a model is depicted in a way as to create an inaccurate view of history or propagates (or reinforces) stereotypes through narrative constructions. The humanities have a unique capacity to site these aberrations and provide interpretive closure. Moreover, the core concepts of the humanities — authorship, originality, creativity, and authenticity — are placed under scrutiny by generative AI. As we star to create artificial text and images, what is the notion of generation/interpretation or semiosis need to be suited? Humanists can participate in working through the development of frameworks for human-machine co-authorship, citational ethics surrounding data usage, and new forms of cultural stewardship over objects created by AI. Once AI research acknowledges that (often) there are multiple context-dependent interpretations rather than one right response or output, it can incorporate interpretative pluralism in the evaluation of models that it makes use of using these frameworks.

There are also institutional and public implications for such interdisciplinary integration. Tech Follies Tech research and humanistic research are often divided into separate silos at universities, in funding organizations and corporate labs-and the latter is treated as an auxiliary facility of technology. But understanding the social and cultural significance and impact of generative AI is crucial for its acceptability and durability. Humanities skills can be used by AI research teams to enrich interpretation of datasets, user interface design, and framing of evaluation metrics. A result of such collaborations could be new ways to do research with both qualitative methods and quantitative approaches-"close reading prototype behavior, not just statistical performance." Therefore this paper aims to offer an organized overview of humanities-centred provocations for generative AI research. These provocations are methodological and ethical demands to rethink the very ways models are itself being conceptualized, evaluated, and as part of larger cultural systems rather than contrastive critiques. To demonstrate the fact that the humanities are indispensable to the critical development of generative technologies, we will place rhetorical analysis, historical contextualisation and hermeneutic sensitivity at the centre of AI research. Rethinking AI as interpretation makes it possible to build systems that honor the human complexity of meaning, memory and creativity even as they output material.

II. A SHORT LITERATURE FRAMING

Since both sides are struggling with the production, mediation and interpretation of meaning, the connection between artificial intelligence (AI) research and humanities scholarship has attracted an increasing attention in recent years. Frame-works for interpretation of how writings, images and artefacts construct and represent human knowledge have been long-established by humanities disciplines, including philosophy, literary studies history studies cultural theory media studied. Since generative-machine media (AI) today operates as vast sources of language and cultural meaning, their

tradition is an essential ground for analysis these machines themselves. The humanities provide indispensable tools for analysis that reveal how it is that AI systems embed assumptions about language, authorship, temporality and power—and they are not just there by accident. Drawing on rhetoric, history, archive theory, hermeneutics, and critical race; gender and empire studies this section looks at influential humanities ideas that inform our critical engagement with AI. It also synthesises relevant works from the fields of algorithmic responsibility, data ethics and sociotechnical critique in order to show how difference dimensions combine to challenge the technocratic supremacy that reigns over AI research.

A. The Multiple Meanings of Hermeneutics

Through its long association with text interpretation, hermeneutics offers one of the most direct conceptual connections between AI scholarship and humanities inquiry. Hermeneutics is based on the work of Friedrich Schleiermacher, Hans-Georg Ghadames and Paul Ricoeur and this theory argues that meaning is derived from interpretation rather than direct perception. Meaning occurs in the relation of an interpreter to a text, and it happens only from within a "horizon" constituted by history and culture, mediated through what Gadamer calls a fusion of horizons. This concept is against that models have a neutral output, like machine can stabilise meaning when it comes to generative AI. Instead, the model, its training data or human users enter into an interpretive dialogue with AI-generated artifacts. This idea has been extended to data visualisation and computational text by scholars like Johanna Drucker (2014) who argues that all digital representations are “capta”—modeled selections of data influenced by interpretive frames. This hermeneutics reframes AI outcomes as acts of writing that need to be read responsibly and critically rather than statistically determined estimates.

B. The Provenance Question and the Theory of Archives

A similarly valuable lens through which to study the data infrastructures that support generative AI is one grounded in archival theory. No, archives are curated systems that reflect social, institutional and political authority — there is no such thing as the “neutral” archive. Leading theorists, for example Verne Harris (2002) and Jacques Derrida in Archive Fever (1995), have shown how archiving processes silence some voices while amplifying others. Provenance and gaps and silences are the things that contemporary archive studies scholars—and, of course, this is not unrelated to Butcher’s worries about AI training datasets—emphasise. Generative models are trained on large digital text and image archives, but these archives reflect and reproduce existing exclusions and disparities by often favoring dominant languages, regions, perspectives. As noted by Terry Cook (2013) and Michelle Caswell (2016), ethics of archive should take into account representational justice and recognition of absence. Using archive theory for AI encouragingly involves recognizing that data curation, labeling and selection are profoundly moral and historical, rather than technical preparation tasks. It makes clear that provenance is essential when trying to evaluate whether the generated processes can be trusted, or are valid.

C. Persuasion, Rhetoric, and Machine Authority

One further approach for understanding AI talk comes from rhetorical theory – since ancient times, it has been concerned with how to persuade someone and how to establish authority. Classical rhetoric from Aristotle through to Quintilian was heavily oriented toward a typology of expressions: pathos, which Rosenstone describes as "the emotional power" WC possesses, ethos – "character," while the last component is logos -"a statement about the way things were". That has been extrapolated to include the rhetorical situation and genre as a template for social action, in modern rhetorical theory by scholars like Carolyn Miller (1984) and Kenneth Burke (1969). Thus in the case of AI, model outputs simulate neutrality, coherence, and confidence to exercise persuasive authority. Even if we don’t understand the underlying epistemic processes, rhetorical flourishes of language models like GPT and picture systems like DALL·E or Midjourney make it seem as if synthetic outputs are trustworthy — that they’re true. Such is what Alex Campolo and Kate Crawford (2020) refer to as the “rhetoric of neutrality” in AI systems, a veneer of impartiality that conceals how data and design decisions are socially constructed. By viewing AI as a rhetorical agent, scholars have been urged to "read closely" model outputs, examining how narrative form, tone and language style are mobilised to authorise themselves and make an impact on human perception.

D. Historical Analysis and Temporal Context

From historiography, the domain that studies how history is documented and shaped, AI critique acquires a temporal layer. It seems to highlight that the knowledge is always situated in a specific time, place and paradigm of interpretation. Reinhart Koselleck’s work in the history of semantics, and Michel Foucault’s concept of the “archive” as an epistemic development both show that categories of thought shift over time. Generative AI that brings to life words that were never written, what will it do to synthesis of historical difference when trained on huge sweeps of text from the contemporary media to millennia old writing? This confusion produces anachronisms, supports ancient social codes and makes the evolution of concepts (e.g. citizenship/gender/race) murky. This tendency of AI for dehistoricization is challenged by humanistic modalities of temporality. The demand for considering temporal situatedness is also noted by scholars such as Benjamin Peters (2016) and N. Katherine Hayles (2017), among others, highlighting that to prevent the framing of AI

systems as timeless or universal knowledge tools it is crucial to take into account temporal mediations: If taken heed, another momentous spatial element arises, inviting pondering design in relation to postcolonialist critique.

E. Critical Empire, Gender and Race Relations of Power

Critical theory, as it developed in feminist, postcolonial and critical race studies, has demonstrated the interpenetration of power and knowledge production. Thinkers such as Bell Hooks (1994), Donna Haraway (1988) and Gayatri Spivak (1988) have emphasised "situated knowledges" and the significance of recognising partial perspectives. understand the societal implications of generative AI need to understand this. Both Ruha Benjamin's Race After Technology (2019) and Safiya Umoja Noble's Algorithms of Oppression (2018), for example, heavily detail how bias in data and design assumptions across algorithmic systems conspire to uphold racial hierarchies. Likewise, "data feminism," which transforms computational methods toward equity and responsibility, is advanced by feminist data scholars such as Catherine D'Ignazio and Lauren Klein (2020). These critical traditions force AI researchers working on generative systems to think about the epistemic inequality generated in the design of these systems and what aspect of which others' cultural narratives get foregrounded or backgrounded when models are activated. These findings challenge AI researchers to think of inclusion in terms of epistemic diversity and not demographic representation.

F. Critical Data Studies and the Humanities. Open Connection Paul S. awaii reception Paul S.

These conceptual novelties have been recently implemented in empirical research by a number of scholars working at the intersection between data studies and humanities. Infrastructure of data and model design are interwoven with labor practices and institutional politics, as seen in Rob Kitchin's (2014), The Data Revolution and Nick Seaver's work on algorithmic culture (2022). This point is elaborated in Kate Crawford's Atlas of AI (2021), which connects labour and environmental exploitation to extractive data practices, theorising the planetary materiality of AI. These vignettes illustrate how AI systems are sociotechnical conglomerations entangled in ecologic, economic and political networks rather than merely computer artifacts. The powerhouse paper On the Dangers of Stochastic Parrots (2021) by Emily Bender and co-authors crystallizes a core humanistic preoccupation in this heterogeneous domain: at what point has size and speed outstripped interpretability, provenance, and accountability? They argue that context, intention and effect should not be sidelined in our rush towards giant killers, a critique of the process of making meaning perhaps as much as it is of the humanities.

G. Toward the Humanistic Foundation for AI Study

Taken together, these arcs of scholarship indicate that ways of knowing in the humanities are constitutive (not corrective) dimensions for theory building in AI. Theories of power reveal the politics of representation inscribed into technological systems; rhetoric envelops performative impact and writing model; historiography returns sense to temporality and contextual contingency protested against by archival theory, which annexes data from the histories of preservation and deletion in doing justice to them, while hermeneutics highlights the dynamism and multiplicity of meanings that no one will rule or compartmentalise. By expressing a demand for reflexivity, accountability, and interpretative profundity, these traditions together resist the shrinking of AI to efficiency and optimisation. Thus the following provocations in this essay are informed by the literature, both theoretically and empirically. These provocations make the case that generative AI is a cultural and epistemic phenomenon, something in which humanists must participate rather than just observe.

III. PROVOCATIONS — WHAT THE HUMANITIES DEMAND OF GENERATIVE AI RESEARCH

Conceptual provocations that challenge first principles of generative AI research are provided by the humanities studies. The humanities invite scholars to think about such systems as interpretative, historical and cultural forces woven into sophisticated meaning making practices, not just as technical aids. Humanistic conceptions of interpretation, provenance, plurality and power are leveraged in each provocation to trouble entrenched engineering assumptions about neutrality, data, temporal Taken together, these provocations argue that our focus on generative AI should be revalued, advocating for a consideration and development of generative AI through modes of interpretive literacy, ethical awareness and historical reflexivity as well as metrics of optimisation.

The first provocation, compared with agnostic tools, holds that generative models are interpretive artefacts. From a humanities perspective, it has been shown that all produced artefacts are acts of interpretation while engineers often understand the output of models as probabilistic reconstructions of the training distribution. AI-produced language, art or music is not just "representing" meaning (or an image in paint or sound), but constituting the shape of meaning, enacting power over rhetoric and framing worlds of reference rather than merely copying patterns. From a humanities point of view, model outputs are texts, to be submitted to detailed analysis in order to uncover the metaphors, rhetorical topoi and ideological assumptions they contain. Not only how models perform, but also what and whom they perform for may be open to inquiry by developing qualitative framework for interpretive analysis in addition to quantitative stocktaking measures.

Provenance is a second provocation. In engineering research, training data is generally considered as a large anonymous pool; and once the volume reaches a critical value, individual sources are no longer important. In the humanities (notably in historiography and in archives studies), the emphasis is placed on origin, that is, who created an artefact, when, where and for whom. Provenance-less datasets reinforce systems of erasure and hiatus, in particular for marginal voices, by hiding the editorial, cultural and political conditions of their production. You'd use detailed model cards, dataset ledgers, and lineage metadata to carefully capture the data histories in a way that would treat provenance tracking as a first class research priority. We should move away from the blind scraping approach and curation must become provenance aware for high-stakes use-cases.

The third provocation is temporality. In the generative AI literature, data are frequently treated as if content of static snapshots is exchangeable and time-invariant. But the humanities remind us that meanings and ideas are also historical. English period are squished flat and the sense must be carried by the model trained on some mix of temporal corpora, with either the resulting anachronisms or retained old-fashioned forms. There is a need for historical sensitivity to stop models from overgeneralising across contradictory eras, or from assuming presentist assumptions into the past. Models can also have better respect for how language and cultural meaning evolve over time by being trained with a temporal awareness built into the model as well as its evaluation. This can be achieved using timestamped corpora, diachronic baselines or user interfaces that provide information of the ap- plication time

The forth provocation challenges the notion of a single, neutral ground truth. AI evaluation often relies on reference-based measures that assume one right answer per prompt. Instead, humanistic traditions stress the legitimacy of multiple readings and interpretive plurality. The written word, the painted image, and the performed gestures are all of their nature polysemous; that is what gives them life. Thus, novel or contextually appropriate deviations are punished in single-label evaluation tasks. Several annotators, evaluation criteria and interpreta- tive guidelines representing a diversity of plausible interpretations should better capture the complexity found in human understanding. Evaluation would become not a measure of rationality but an account of expressive variety if we accepted pluralism in readings. The realization of institutional labour and power in AI research is the sixth provocation. Although the human labour that makes this possible– content moderation, data annotation and documentation and platform maintenance–is often buried from view, models are instead frequently figured as autonomous technical achievements_HERSHEYo. The humanities, with the guidance of political economy and critical labour studies that embolden their keepers to confront these injustices head on, show such invisibility to be neither neutral, nor convenient but as a means of re-affirming hierarchies of value and recognition. To make sense of generative AI as a sociotechnical system, it is crucial to consider conditions of production including the global circulation of cognitive and emotional labour. Hence, labor impact assessment and sociotechnical analysis that demonstrate how power and value flow within the model's ecosystem are required during research pipeline.

The sixth provocation asks us to notice local knowledge systems and non-central epistemologies. Engineering paradigms often assume that model architectures and universalist datasets can generalize relatively easily across domains. On the other hand, humanities departments focus attention on the multiple epistemic traditions — feminist, postcolonial, indigenous and community-based — that have their own definitions of what counts as knowledge. These epistemologies are commonly invisible or misrepresented by dominant frameworks in big datasets. By including cultural safeguards and consent practices, co-designing data sets and assessment standards with local communities would ensure that generative AI is developed in a way that respects epistemic diversity and minimises representational harm.

An eighth provocation is to re-frame interpretability as a hermeneutic. Interpretability, as it is commonly used in current AI debate, often means mathematical transparency via saliency visualization or feature attribution. This notion is further explored within the humanities more broadly, which understand interpretation as contextual and dialogic rather than reducible to visualisation. Narrative and rhetorical framing that make sense to human users are necessary when explaining the behaviour of a model. Instead of constituting mere parameters that are uncovered, interpretability becomes a hermeneutical method of giving meaning to, or interpreting, data. The question of interpretability could be reframed from a technical problem to a communicative, ethical and aesthetic practice if provenance tales, narrative-based explanatory objects and user-centred interpretative research were developed. The seventh provocation focuses on archival places to think methodologies. Data sources are usually processed as raw material for calculation, yet the humanities insist that archives are constructed and contested. What is part of the record and what isn't are decisions made in processes of collection, digitization, and preservation. As such, models trained on digitised archives learn their silences as well as their wealth of content. The digitisation of politics and bringing the archivist into view as a collaborator might shine a light on biases in what AI learns. As such the archival theory contributes to principles of accountability, transparency and inclusivity in curatorial choices as a critique and methodology towards ethical data building.

Political and ethical practice is deterritorialized in the ninth provocation to reconceptualize appraisal as a political and ethical process. Technical measures such as accuracy and perplexity are often assumed to be objective performance criteria, but the metrics themselves embody humanistic values. What counts as effective, and whose views are validated, depends on what is measured. Accordingly, assessment is normative in the sense that it privileges certain epistemic outcomes. Complementing quantitative analyses with context-sensitive qualitative work, and the making of metric choices more open and participatory, would draw attention to the ethical and political dimensions of these technical practices. Thus, evaluation should be a design principle, not an afterthought.

Together, these nine provocations drive a reframing of the scope and approach to generative AI research. Implicitly, they show that generating and evaluating generative models is a matter of cultural production informed by interpretive, historical, and ethical considerations rather than only computation or optimisation. Meaningfully bridling AI to human complexity, and thus incorporating such provocations as humanistic ones, enlarges rather than narrows its epistemic reach without impeding scientific progress. The very critical frameworks that would make technology reflexive, accountable, and humane are in fact enunciated by the humanities themselves; they are not external to technology. Looked at from this multiplicity of perspectives, generative AI can evolve or be framed out of its status as a machine that creates reproductions and serve to become a tool for conversation which properly respects diversity, history, and the rich depth of human interpretation.

IV. METHODOLOGICAL BRIDGES: COMBINING HUMANITIES METHODS WITH AI RESEARCH

Beyond ethical reflection, operationalising the aforementioned provocations requires methodological experiments which marry the empirical precision of computer research and the hermeneutic depth made by humans. Methods that allow the ways of thinking industrious within humanistic disciplines – rhetorical, historical, ethical – to direct work on generative models (a la development over testing over explanation) is required for bridging different epistemic cultures. In this section we put forward six interlocking methodological bridges—close-reading ensembles, provenance pipelines, temporal slicing and counterfactual modelling, participatory evaluation, narrative explainability and archive impact assessment—that aim to transform humanistic critique into research design. Individually, each bridge produces a more critical and ethically minded science, by anchoring the interpretive rigour of the humanities in the experimental framework of AI.

The first bridge, close-reading ensembles, marries the textual strategies of literary and cultural criticism with computational experimentalism. C]-Findings: Our approach is to assemble an interdisciplinary team from which both ReadMe-measurements and model-explicating reads are generated as modes of performance assessment rather than using only numeric metrics. To understand how a model is interpreting, this approach involves analyzing the utterances output by the generator for tone (ideological and metaphorical) and rhetorical strategy—factors which are inscrutable to ecologically irrelevant accuracy metrics. In doing so, close-reading ensembles turn qualitative phenomena into analytical categories that organize potential new criteria. For example, it may be the case that a language model's repeated use of deficit metaphors to describe oppressed populations gives rise to stylistic diversity indices or rhetorical-bias metrics. It is this sort of integration which allows scholars to see models as participants in rhetorical traditions and not only statistical entities. A second bridge between archival theory to data engineering practice is provided by provenance pipelines. Humanities scholarship has long emphasized how the provenance of knowledge, its place of origin, authorship and circulation influences it. To adapt this intuition to generative AI, metadata collection needs to be enforced across the complete data pipeline. Each data point going into a corpus should be accompanied by contextual information such as source, publication date, author's demographics (to the extent ethically permissible), editorial line, and genre. Making it easy for models to condition on source characteristics, provenance metadata facilitates the study of how training environment influences output. Provenance pipelines provide an appealing model for ethical auditing as well as model conditioning. With explicit tracing of data lineage, researchers can detect over-coverage or systematic under-representation and correlate dataset content with past accountability and fairness norms. This activity is in line with archival ethics that understand the role of documentation for maintaining epistemic integrity rather than as a bureaucratic requirement.

The third methodological horizon, counterfactual modelling and time slicing is deployed to solve the problem of temporality. Anachronisms and outdated norms emerge when generative models conditioned on chronologically unmarked data conflate epochs. The humanities departments, by contrast, see time as a constitutive element of meaning. On the division of training corpora by era: temporal slicing is a mixed bag Itt 2019 Temporal slicing helps understand relative change in a model's representations over time From monolingual to multilingual \cite{AM_etal} Moreover, counterfactual prompting encourages models to "give voice end by the models whenev no matterwhatus To 'speak from period X,' permitting them to emulate the logic or language of a particular moment in history. This approach highlights temporal drift in generative outputs, and illustrates how well/ poorly the language model is retaining history. Temporal modelling thus serves as a methodological stepping stone to temporally responsible AI and an exploratory device to diagnose the flattening

of history in machine learning. Participatory evaluation, a fourth bridge, bring democratic and community approaches to the evaluation of AI. Traditional evaluation systems are less attentive to the lived experiences people in datasets represent, and prefer benchmarks authored by researchers. Feminist and post-colonial research traditions offer methods that upend traditional practices, including those in place in medicine and population health science, which include the voices of people most affected. The method of participatory assessment can be applied to generative AI by collaborating with community partners to co-identify categories of harm, codevelop evaluation tasks and design what is fair or culturally respectful. In this way the participatory co-production ensures that evaluation measures will reflect multiple priorities instead of only those of the program logic developers. It also introduces accountability mechanisms for aligning AI research with the principles of reciprocity, transparency and consent.

Narrative explainability, the fifth methodological bridge, looks at interpretability and rethinks it in rhetorical and hermeneutic terms. With this focus on which inputs impact which outputs, the XAI methodologies in current use are plagued by a propensity to conflate interpretation with feature attribution. The humanities view is that explanation is narrative, not just causal: events and processes are rendered in relation to intention, purpose and the way things stand at a particular time. (Narrative explainability features textual or multimodal narratives that recover and describe the interpretive conditions of generation— what sources the model was exposed to, what rhetorical frameworks it enacted and whose "voice" it employed in relation to training data are shown side-by-side with data visualizations.) By facilitating a dialogic communication with explanations, these narrative artefacts build confidence through storytelling and knowledge rather than mere openness. AI systems can render interpretability in forms interpretable to non-specialist audiences while preserving analytical standards by aligning explanation with narrative coherence. Finally, AA effectively bridges critical archival theory and data governance. Each data set is a growing archive, and embodies institutions' decisions about what to preserve, leave out or make invisible. In turning to archive theory as a method for producing datasets, it is important that we hold on to both inclusion criteria and "negative space" (Kaplan 2019) i.e., the silences and absences for our models. Therefore, archival impact studies examine the cultural and epistemic effects of dataset curation that are akin to those produced by environmental impact reports. This prompts teams to disclose who collected the data, why certain categories were prioritised, and what biases could be accentuated in digital form. As this method supports researchers locating the genealogy of data and to anticipate interpretive counter-effects when combined with provenance processes.

Together, these six methodological bridges constitute what might be called an integrated humanistic engineering, a mode of inquiry in which technical design is informed by interpretive understanding and cultural critique is checked against empirical evidence. Their aim is to embed reflexivity in its cognitive architecture, not humanize it by adding a false skin. They computationally operationalize humanities methods into heuristics: archival reflection into dataset governance, provenance in metadata ontology, temporality in corpus segmentation, participation in evaluative design and close reading in feature analysis. AI is operationalised by the humanities; and AI operationalises the humanities, which see and improve AI.

Table 1: Methodological Bridges between Humanities and Generative AI Research

Methodological Bridge	Humanities Foundation	Operational Mechanism in AI Research	Research Implications
Close-Reading Ensembles	Hermeneutics, literary analysis	Pair qualitative close reading with computational metrics	Reveals ideological framing, informs new evaluative metrics
Provenance Pipelines	Archival theory, historiography	Embed metadata (source, author, context) during ingestion	Enables ethical auditing and provenance-conditioned modeling
Temporal Slicing & Counterfactuals	Historiography, cultural temporality	Train and test on temporally partitioned corpora; simulate historical perspectives	Detects anachronisms and temporal bias
Participatory Evaluation	Feminist and postcolonial methodologies	Co-create evaluation tasks with communities represented in data	Aligns fairness and safety with community values
Narrative Explainability	Rhetoric and hermeneutics	Combine quantitative feature attribution with narrative explanation	Improves interpretability and user trust
Archival Impact Assessment	Critical archival studies	Document inclusion/exclusion and dataset "negative space"	Identifies epistemic gaps and biases in data construction

By formalising these methodological bridges, research in generative AI can develop from restricted optimisation paradigms into a genuinely interdisciplinary science of meaning. The subsequent practices prompt thought about what models mean, whose histories they epitomize, and how their machinations change the cultural fabric as much as what models can do. So it is that instead of being the handmaiden to AI, humanities are its epistemic mate, enriching what can be known (through a coupling to more or less all there is), understood (in terms of depth and perspective on questions) and imagined morally in AI research.

V. TWO ILLUSTRATIVE VIGNETTES

Two case histories of how humanistic critique reshapes technical practice demonstrate the practical fallout of including humanities methodologies in generative AI research. Through their case studies, these short stories reveal how small acts of generation can give rise to ethical and interpretive challenges and show how collaboration across the disciplines opens possibilities for research results that are historically and culturally sound.

For the first, a team of researchers have developed a generative model that can produce letters with ancient sounds in them for the museum display. The goal is to encourage visitors to participate by creating the illusion of authentic letters through time. On the face of it, the technique works: he writes in prose that is emotionally resonant and linguistically flowing, transmitting a sense of period realism. Yet upon further scrutiny, there are several problematic implications which threaten both ethical coherence and historical verisimilitude. Unwittingly echoing colonial themes taken from subordinated literary sources, the model— trained on heterogeneous and temporally undifferentiated material— amalgamates idioms across centuries. Its produced letters, where characters tend to have the agency and ambition that would enable them to carry out their original project — imagining themselves into history so they can free or be kind to those damaged by historical oppression — generally end up romanticising as tales of discovery or kindness what were actually difficult histories of domination, creating a false alibi for its own emotional myopia toward enslaved black people or colonised brown ones.

These delusions are exposed by a humanities-informed assessment as more than mere technical errors, but interpretive failures stemming from erasure, anachronism, and unreflecting cultural conservatism. The model's behavior underscores what happens when the training data are not suited to temporal and provenance considerations when looked at through the lenses of historiography and postcolonial critique. One solution is to cross-disciplinary methodological bridges from the humanities. Provenance-sensitive prompts would ensure that generated text is written with a knowledge of the author’s context and power relations, whereas temporally partitioned training data would prevent linguistic and ideological slippage across different eras. The inclusion of curatorial review panels comprised of archivists and historians provides another level of interpretive oversight, localizing that the materials being generated are not just historically accurate but ethically true to the people and times they represent. This short story drives home how historical simulation court disaster in rendering archives as instruments not of history’s more precise recording but as devices by which history, it’s better laying out into an actionable argument for acting and interacting with institutions, becomes instrumentally defibrillated.

In the second scenario, a generative AI "assistant" is deployed in the classroom to support learners in literary interpretation and summarisation of poetry. That device, designed to democratize access to complex texts, does its work cleanly and gracefully. Repetition veers in one direction with consistent reply: formalist, historicist, critical race and gender analyses are marginalized; universalist readings of canonical male poets continue to have pride of place. Subtly its products mirror and relay the ideological hierarchies that construct traditional canons of literature, affirming and disseminating the received interpretive grids.

That the bias is an error in structure which limits interpretive possibilities, rather than a mere personal or anti-human attitude, is a central feature of its construction as problem and subject. Plurality of readin gs is here turned into algorithmic consensus by privileging some readings as more "natural". The absence of multi-perspective training objectives and reliance on single-reference evaluation benchmarks are the reasons for this loss in meaning diversity. Solution: Multi-perspective summarisation frameworks, encouraging incorporation of alternative critical lenses—feminist, postcolonial, formalist, or Marxist etc.—into generative outputs. Therefore, rather than punishing deviation from common readings, evaluation standards should encourage hermeneutic plurality. The discursive character of these objectives is made certain, rather than the homogenisation that it implies, through involving teachers and literary scholars in co-designing the tool.

Together these examples illustrate why interpretive literacy is necessary for the ethical and epistemic integrity of generative AI. Without humanist intervention, algorithms risk substituting knowledge for fluency, aestheticizing prejudice as style and re-inscribing the epistemic injustices of their training data. Rather than an unwitting pawn in the persistence of inequity by default through inheritance, generative AI might act as a fellow cultural-heuristic and employ humanities mode of operation—its temporal sensitivity, its provenance awareness, its participatory evaluation, and its interpretive plurarity— in design and deployment. These examples show that decent generative methodology requires a control of the technical

apparatus as well as an ability to interpretively recognise and transform the stories technology tells about art, history and humanity at large.

VI. CONCLUSION

The humanities, according to this study, present generative AI research with crucial provocations — conceptual, methodical, and ethical issues that bend the field's assumptions regarding neutrality as well as what is meant by meaning and value – rather than mere food for thought. The humanities reconstitute generative systems as cultural artefacts and not only as computational tools, by exploring the interpretive, historical-political dimensions of model generation. This shift requires that interpretive multiplicity, provenance awareness, and ethical reflexivity be built into all stages of model development, evaluation, and application in AI research methods. In this way, automated replication systems that make up generative AI can become dialogic interfaces with the Panoply of human knowledge. Much work on AI is based on a technocentric perspective, and this is challenged by the provocations of this paper. The rhetorical and ideological work which generative systems accomplish becomes more evident in this kind of approach, where models are treated as interpretive artefacts. If provenance and temporality are key aspects of what brings a data set into being, we can learn how meaning making is shaped by historical and cultural context. By addressing labour, institutional power and marginalised epistemologies the AI as a sociotechnical ecology becomes rephrased rather than an isolated technology; rejecting the pretence of a single ground truth highlights diversity in human interpretation. Taken together, these interventions blur the divide between technê and humanistia to demonstrate that both are needed for epistemic consistency.

Methodologically, this article proposed six bridges—close-reading ensembles, provenance pipelines, temporal slicing and counterfactuals, participatory evaluation, narrative explainability and archival impact assessment—that translate humanistic insight into computational scholarship. Each is an interdisciplinary approach in which knowing is co-constructed through interpretive and quantitative components. Together, they articulate an approach to interdisciplinary rigour that does not romanticize the humanities critic and devalue him or her relative to the engineer. Instead, they offer a genuine hybrid of research in which measurement and substance mutually support each other. The literature assistant and historical letter generator are two examples that illustrate the impact of not having such an integration. In both cases, generative modes of understanding become the grounds for interpretive exclusion and historical distortion in ways that mobilize fluency as ideological reproduction. Conversely, these very systems also serve as instruments of ethical innovation, education and preservation when human-centered approaches drive them. The intersection of the humanities and AI has implications that reach well beyond academic research. The interpretive frameworks that characterize generative models have political implications insofar as they increasingly mediate cultural utterance, such as writing and art and education –– history and governance. Recognizing these systems as agents of discourse compels makers, curators, and lawmakers to acknowledge how they shape public memory and imagination. An engagement of the humanities means that our AI system will act on its promise to be responsible in the crafting shared futures, being at least not just a mode of critique but also a methodological mode of stewardship.

Accordingly, these modulations that cross boundaries among disciplines should be institutionalised for counts or corpuses in subsequent research: by means of shared provenance tracking infrastructures, efforts toward a collaboratively developed evaluation framework (one with interpretive transparency) and training initiatives that blend hermeneutic competence and computational literacy. AI epistemology needs the humanities critical vocabulary—interpretation, context, power, and temporality—as actively working categories. Only through such an integration can the discipline achieve to move beyond its image of objectivity and refexive self-critical science of meaning that is aware of its own cultural embeddedness. And finally, the humanities are not on the edge of generative AI; they are its interpretive horizon and its conscience. Through their questioning, they suggest to scholars that models can be considered fellow travelers in the ongoing human project of understanding and representing the world rather than neutral predictors. And so, perhaps, generative AI informed by humanist insight can be a technology for machines to learn how to read, remember and respond with moral imagination rather than one more fancy way to close the books on our creativity as humans. The problem is, of course not to humanize the epistemologies thanks to which one has these robots but to re-humanize them.

VII. REFERENCES

[1] Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?* Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency.

[2] Birhane, A. (2021). *Algorithmic injustice: A relational ethics approach.* Patterns, 2(2), 100205.

[3] Bolukbasi, T., Chang, K.-W., Zou, J. Y., Saligrama, V., & Kalai, A. T. (2016). *Man is to computer programmer as woman is to homemaker? Debiasing word embeddings.* NeurIPS.

[4] Borgman, C. L. (2015). *Big Data, Little Data, No Data: Scholarship in the Networked World.* MIT Press.

[5] Bourdieu, P. (1991). *Language and Symbolic Power*. Harvard University Press.

[6] boyd, d., & Crawford, K. (2012). *Critical questions for Big Data.* Information, Communication & Society, 15(5), 662–679.

[7] Chun, W. H. K. (2016). *Updating to Remain the Same: Habitual New Media.* MIT Press.
[8] Crawford, K. (2021). *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence.* Yale University Press.
[9] Crenshaw, K. (1991). *Mapping the margins: Intersectionality, identity politics, and violence against women of color*. Stanford Law Review, 43(6), 1241–1299.
[10] Daston, L., & Galison, P. (2007). *Objectivity*. Zone Books.
[11] Foucault, M. (1972). *The Archaeology of Knowledge.* Pantheon Books.
[12] Gebru, T. (2020). *Datasheets for datasets*. Communications of the ACM, 64(12), 86–92.
[13] Green, B. (2021). *The flaws of policies requiring human oversight of government algorithms.* Computer Law & Security Review, 41, 105528.
[14] Haraway, D. (1988). *Situated knowledges: The science question in feminism and the privilege of partial perspective.* Feminist Studies, 14(3), 575–599.
[15] Heidegger, M. (1977). *The Question Concerning Technology*. Harper & Row.
[16] Hutchinson, B., Prabhakaran, V., Denton, E., Webster, K., Zhong, Y., & Denuyl, S. (2021). *Towards accountability for machine learning datasets: Practices from software engineering and infrastructure.* NeurIPS.
[17] Iliadis, A., & Russo, F. (2016). *Critical data studies: An introduction.* Big Data & Society, 3(2).
[18] Kitchin, R. (2014). *The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences.* Sage.
[19] Latour, B. (2005). *Reassembling the Social: An Introduction to Actor-Network-Theory*. Oxford University Press.
[20] Marcus, G. (2022). *Deep Learning Is Hitting a Wall.* The Gradient.
[21] Noble, S. U. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism.* NYU Press.
[22] O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown.
[23] Parisi, L. (2019). *Critical computation: Digital automata and general artificial thinking.* Theory, Culture & Society, 36(2), 89–121.
[24] Pasquale, F. (2015). *The Black Box Society: The Secret Algorithms That Control Money and Information.* Harvard University Press.
[25] Raji, I. D., & Buolamwini, J. (2019). *Actionable auditing: Investigating the impact of publicly naming biased performance results of commercial AI products*. AIES.
[26] Ricoeur, P. (1981). *Hermeneutics and the Human Sciences*. Cambridge University Press.
[27] Suchman, L. (2007). *Human-Machine Reconfigurations: Plans and Situated Actions.* Cambridge University Press.
[28] Suresh, H., & Guttag, J. V. (2021). *A framework for understanding sources of harm throughout the machine learning life cycle.* Equity and Access in Algorithms, Mechanisms, and Optimization.
[29] Tufekci, Z. (2015). *Algorithmic harms beyond Facebook and Google: Emergent challenges of computational agency*. Colorado Technology Law Journal, 13, 203–218.
[30] Winner, L. (1986). *The Whale and the Reactor: A Search for Limits in an Age of High Technology*. University of Chicago Press.