Journal of Management Info

Journal Homepage: https://readersinsight.net/JMI

Research Article

POLAS: A Framework for responsible institutional governance of generative AI in social science research

Mohammad Rashed Hasan Polas[1]*,

Strategic Research Institute (SRI), Asia Pacific University of Technology and Innovation (APU), Technology Park Malaysia, 57000 Kuala Lumpur, Malaysia

***Corresponding Author email: rashedhasanpalash@gmail.com**

ABSTRACT

This paper examines the ethical challenges generative artificial intelligence (GenAI) introduces to social science research management, including algorithmic bias, data privacy risks, threats to research integrity, and diminished transparency, and proposes a governance framework to address them. The paper draws on a critical synthesis of recently published frameworks, including the Core Principles for Responsible AI Use in Research, the ETHICAL framework, the European Commission's Living Guidelines on AI in Research, and recommendations by Resnik and Hosseini, to develop an integrated governance model. The synthesis yields POLAS: Policies & Disclosure, Oversight & Quality Control, Literacy & Understanding, Accountability, and Stewardship & Social Impact. The framework argues that responsible GenAI adoption in research organizations requires more than policy compliance; it demands an active organizational culture of literacy, accountability, and stewardship. As a conceptual framework, POLAS remains empirically untested and is grounded primarily in qualitative research traditions. Future research should validate its effectiveness and extend it to quantitative and discipline-specific contexts. Unlike general-purpose AI ethics frameworks, POLAS is purpose-built for the social science research context, offering a memorable, citeable, and adaptable instrument to guide institutional governance of GenAI integration.

Keywords

Generative Artificial Intelligence;

Research Ethics;

Social Science;

POLAS Framework;

Research Integrity;

Responsible AI

Submitted:
11 August 2026

Revised:
15 November 2026

Accepted:
21 November 2026

1. INTRODUCTION

Artificial intelligence has not arrived suddenly into social science research, it has been present for decades, deployed in narrower and more technically demanding forms: automated text classification, sentiment analysis, computational content coding, and statistical modelling tasks that required researchers to possess substantial programming expertise before they could leverage the technology at all (Brown et al., 2020; Bommasani et al., 2021; Davidson & Karell, 2025; Qian & Xie, 2026). This technical barrier, for all its exclusivity, functioned as an inadvertent safeguard, limiting AI-assisted research to a relatively small community of computationally skilled scholars who had, by necessity, developed some working understanding of the tools they were using. The release of ChatGPT in November 2022 dismantled that barrier with startling abruptness (Davidson & Karell, 2025). By presenting the capabilities of large language models through a conversational interface requiring no programming knowledge, no specialist training, and no institutional infrastructure, generative AI effectively democratised access to a technology of considerable power, drawing into its orbit researchers at every career stage,

https://doi.org/10.31580/jmi.v13i4.3563
© 2026 The Authors. Published by Readers Insight Publisher. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/)

across every subdiscipline, and within institutions with vastly different levels of preparedness (Brown et al., 2020; Bommasani et al., 2021; Sebastian et al., 2025). The scope of tasks amenable to AI assistance expanded correspondingly, stretching across the full research lifecycle from initial literature mapping to final manuscript revision. Moreover, this democratisation carried a shadow: the same lowering of access costs that made GenAI available to conscientious researchers equally reduced the practical friction associated with unethical research conduct, placing tools capable of fabricating data, generating spurious citations, and ghost-writing undisclosed manuscripts within reach of anyone with an internet connection and the inclination to misuse them (Dergaa et al., 2023; Lund et al., 2023; Sebastian et al., 2025). Researchers and institutions, shaped by a slower-moving technological landscape, were structurally unprepared for the compound disruption that followed.

The transformative promise of GenAI for social science scholarship is both broad and profound. Beyond mere efficiency gains, these tools fundamentally reconfigure what is analytically possible, accelerating data processing pipelines, democratising access to sophisticated analytical capabilities, and meaningfully reducing the resource asymmetries that have long disadvantaged researchers operating in the Global South (Dwivedi et al., 2023). Large-scale qualitative synthesis, historically constrained by the irreducible limits of human bandwidth and institutional resourcing, becomes increasingly tractable when AI augments thematic analysis, pattern recognition, and interpretive coding across voluminous datasets. Equally significant are the implications for meta-scientific inquiry: bibliometric mapping, systematic reviewing, and content analyses that once demanded months of painstaking labour can now be completed within weeks, opening new frontiers for understanding how knowledge itself is produced, organised, and contested (Sebastian et al., 2025). In this sense, GenAI does not merely assist research, it expands the horizons of what social science can ask.

However, the adoption of GenAI also introduces a set of serious ethical risks that are amplified within the social sciences. Algorithmic bias embedded in training data may reproduce and amplify social inequalities, distorting the findings of research on marginalised communities (Bender et al., 2021; Blodgett et al., 2020; Jeon et al., 2025). The use of participant data in AI systems raises novel privacy and informed consent concerns that existing ethical review frameworks were not designed to address (Ansarullah et al., 2024; Qian & Xie, 2026). The near-undetectable fluency of AI-generated text creates new vectors for academic misconduct, threatening the integrity and credibility of published scholarship (Stokel-Walker & Van Noorden, 2023). And the opacity of GenAI systems, what scholars call the 'black box' problem, makes it difficult for researchers to account for or scrutinise the epistemic processes underlying AI-assisted conclusions (Rudin, 2019; Qian & Xie, 2026).

Existing ethical frameworks have begun to address these concerns at a general level. The Springer Nature Core Principles for Responsible AI Use in Research enumerate eight principles including transparency, bias mitigation, and broader impact (Knöchel et al., 2025; Qian & Xie, 2026). The ETHICAL framework offers a seven-step practitioner checklist for managing AI use in research settings (Eacersall et al., 2025). The European Commission's

Living Guidelines on Responsible AI in Research provide broad normative guidance to researchers, institutions, and funders (European Commission, 2024). Resnik and Hosseini (2025) offer specific recommendations on disclosure, authorship, and stakeholder engagement. These contributions are important, but they share a limitation: they are largely domain-agnostic, and do not adequately attend to the distinctive methodological traditions, ethical commitments, and social stakes of the social sciences.

This paper addresses that gap. Drawing on a critical synthesis of existing frameworks and the broader literature on AI ethics, research methodology, and science studies, it introduces the POLAS framework, an acronym standing for Policies & Disclosure, Oversight & Quality Control, Literacy & Understanding, Accountability, and Stewardship & Social Impact. POLAS is designed as a structured, memorable, and practically implementable guide for social science researchers, enabling ethical and responsible GenAI integration across the full research lifecycle. The paper provides a detailed exposition of each POLAS pillar, maps it to existing principles, and offers a phase-specific implementation checklist. It concludes with a discussion of challenges and directions for future inquiry.

2. LITERATURE REVIEW

2.1. Generative AI in Research: Emerging Landscape

The landscape of GenAI adoption in academic research has expanded rapidly since 2022. Studies across disciplines document increasing use of LLMs for tasks including systematic review screening, qualitative thematic analysis, research question formulation, statistical code generation, and manuscript drafting (Lund et al., 2023; Noy & Zhang, 2023; Qian & Xie, 2026). Within the social sciences specifically, researchers have begun using GenAI to assist with interview transcript analysis (Tiwari et al., 2023), survey pre-testing (Bail, 2024), literature mapping, and even participant simulation for pilot hypothesis testing. The velocity of adoption has outpaced the development of governance structures, creating a gap between practice and ethical guidance.

A distinctive feature of the social science context is the field's epistemological diversity. Positivist, interpretivist, critical, and post-structuralist traditions each bring different assumptions about knowledge, objectivity, and the researcher's role (Davidson & Karell, 2025; Qian & Xie, 2026). These traditions diverge significantly in how they would evaluate GenAI's appropriateness: while positivist-oriented quantitative research may find GenAI useful for data processing with manageable risks, interpretivist qualitative research raises deeper concerns about how AI mediation reshapes the meaning-making process (Gioia et al., 2013; Davidson & Karell, 2025; Qian & Xie, 2026). Any ethics framework for GenAI in social science must be sensitive to this epistemological pluralism.

2.2. Ethical Dimensions of GenAI in Social Science

The ethics of GenAI use in social science research spans several intersecting dimensions, each of which has received attention in the recent literature. These dimensions are not independent: they interact in ways that require holistic, systemic thinking.

2.2.1. Bias and Representation:

GenAI systems are trained on large corpora of text that reflect historical patterns of representation and exclusion. When deployed in social research, these biases can reproduce and amplify inequalities (Bender et al., 2021; Davidson & Karell, 2025; Qian & Xie, 2026). For instance, LLMs have been shown to exhibit gender and racial biases in text generation tasks (Blodgett et al., 2020; Qian & Xie, 2026), which may distort findings when used for qualitative coding of sensitive social phenomena. Social scientists studying marginalised or underrepresented communities are particularly vulnerable to these distortions.

2.2.2. Privacy and Informed Consent:

Social science research frequently involves sensitive personal data, interview transcripts, clinical narratives, social media records. The use of GenAI platforms to process such data raises novel informed consent challenges. Participants who consented to have their data used for research purposes may not have anticipated secondary processing by commercial AI systems. Moreover, the potential for GenAI systems to retain or reproduce training data introduces privacy risks that IRB frameworks were not designed to address (Ansarullah et al., 2024; Dellafiore et al., 2026).

2.2.3. Research Integrity and Authorship:

The ease with which GenAI produces fluent, plausible academic prose creates risks of fabrication, hallucination, and plagiarism that are difficult to detect with standard tools (Stokel-Walker & Van Noorden, 2023). The question of authorship whether AI should receive credit, and what it means to 'author' research with AI assistance has prompted significant debate and inconsistent policy responses across journals and institutions (Else, 2023). Most leading journals now prohibit listing AI as an author, but policies on disclosed AI assistance vary considerably.

2.2.4. Transparency and Reproducibility:

Scientific norms of transparency and reproducibility require that research methods be described in sufficient detail to allow independent replication. GenAI introduces new reproducibility challenges: LLM outputs are non-deterministic, vary across model versions, and depend on prompts that researchers may not fully document. The result is a potential 'reproducibility gap' in AI-assisted research (Kapoor & Narayanan, 2023; Davidson & Karell, 2025; Qian & Xie, 2026). This gap is particularly acute in qualitative social science, where the researcher's role in interpretation is already a site of reflexive scrutiny.

2.2.5. Power, Equity, and Access:

The governance of AI reflects existing power asymmetries: leading GenAI systems are developed by a small number of primarily Anglophone, Western technology companies whose design choices embed cultural and commercial assumptions. Researchers in the Global South may face additional barriers to access, and their communities may be inadequately represented in training data (Mohamed et al., 2020; Davidson & Karell, 2025;

Qian & Xie, 2026). A framework for ethical AI use in social science must attend to these structural dimensions of AI governance.

2.3. Existing Frameworks: Strengths and Limitations

Several frameworks have emerged to guide ethical AI use in research. Each represents a meaningful contribution but also exhibits limitations that motivate the present proposal. The Core Principles for Responsible AI Use in Research, proposed by Hosseini et al. (2025) in AI and Ethics (Springer), enumerate eight principles: regulatory compliance, data security, quality control, originality, bias mitigation, accountability, transparency, and broader impact. This is one of the most comprehensive domain-general frameworks available and provides a useful taxonomy of concerns. However, the principles are articulated at a high level of abstraction and do not provide specific operational guidance for the social science context.

The ETHICAL framework (Eacersall et al., 2025) is structured as a seven-step practitioner checklist: Examine policies; Think about social impacts; Harness technical understanding; Indicate use; Critically engage with outputs; Access secure versions; Look at user agreements. This framework is commendably practical and mnemonic, but its social impact component is relatively brief, and it does not foreground the epistemological or methodological specificities of social science research. The European Commission's Living Guidelines on Responsible AI Use in Research (2024) offer normative guidance across three audiences: researchers, institutions, and funders and address important structural dimensions of AI governance including data governance, transparency, and the protection of fundamental rights. The guidelines are authoritative but necessarily general, intended to span all scientific disciplines.

Resnik and Hosseini (2025), writing in Research Ethics, offer perhaps the most practically grounded recommendations for individual researchers. They advise on bias reduction through prompt engineering and output auditing; the importance of disclosure in methods sections; stakeholder engagement with communities affected by research; transparency regarding synthetic data; the prohibition of AI authorship; and the need for ongoing education. These recommendations are valuable and influential, but they are synthesised from existing principles rather than constituting a standalone framework.

A qualitative study by Jeon et al. (2025) exploring narratives among early-career social scientists identified three dominant orientations toward GenAI: the Equalizer narrative (AI democratises research capacity), the Meritocracy narrative (AI rewards those who master it), and the Community narrative (AI reshapes scholarly relationships and norms). These narratives reveal that ethics-in-practice is not reducible to policy compliance: it involves ongoing negotiation of identity, community, and values. A comprehensive framework must speak to this lived dimension of research ethics. Taken together, existing frameworks provide essential conceptual resources but leave a gap at the level of social science–specific, operationalised guidance. The POLAS framework, developed in the following section, is designed to address this gap.

3. THE POLAS FRAMEWORK: ETHICAL GENERATIVE AI USE IN SOCIAL SCIENCE RESEARCH

The POLAS framework is a structured model for the ethical and responsible use of GenAI across the social science research lifecycle. It comprises five pillars: Policies & Disclosure (P), Oversight & Quality Control (O), Literacy & Understanding (L), Accountability (A), and Stewardship & Social Impact (S). The acronym is designed to be memorable and easily operationalised, facilitating adoption by researchers, research teams, and institutions. POLAS is grounded in three foundational commitments. First, it is social science–specific: each pillar is elaborated with reference to the distinctive methodological traditions, ethical norms, and social stakes of social science research. Second, it is lifecycle-oriented: the pillars apply across the planning, data collection, analysis, writing, and dissemination phases of research. Third, it is integrative: POLAS synthesises and extends existing frameworks rather than replacing them, providing a coherent and mappable structure.

Table 1. The POLAS Framework

Element	Pillar	Description	Linked Principles
P	Policies & Disclosure	Align with institutional and global guidelines; transparently report GenAI tools, prompts, and the extent of AI involvement in each phase of the research process.	Regulations, Transparency (Core); Examine, Indicate (ETHICAL)
O	Oversight & Quality Control	Implement peer-review of AI-generated outputs; conduct bias checks; ensure data privacy and governance compliance; maintain audit trails.	Quality Control, Accountability (Core); Critically Engage (ETHICAL)
L	Literacy & Understanding	Educate researchers on GenAI mechanisms, limitations, and biases; promote critical engagement with AI outputs and foster informed interpretation.	Harness Understanding (ETHICAL); Education (Resnik & Hosseini)
A	Accountability	Define responsible parties; clarify the human oversight role; ensure all AI outputs are verified, validated, and attributable to identifiable researchers.	Accountability (Core); No AI Authorship (Resnik)
S	Stewardship & Social Impact	Assess societal implications of GenAI use; engage communities and stakeholders; address equity, long-term impacts, and public trust dimensions.	Broader Impact (Core); Think Social Impacts (ETHICAL)

3.1. P — POLICIES & DISCLOSURE

The first pillar of POLAS concerns the policy environment within which GenAI is used and the obligations of researchers to disclose their use transparently. Social scientists work within overlapping policy frameworks, institutional research policies, journal editorial policies, national research council guidelines, and international frameworks such as the EU AI Act, that are evolving rapidly in response to GenAI. Researchers bear responsibility for monitoring and complying with applicable policies, and for engaging actively with policy development processes.

Disclosure is a cornerstone of research integrity and scientific transparency. When GenAI is used in any phase of the research process, literature synthesis, data analysis, writing, translation, this use must be disclosed explicitly in the methods section. Disclosure should specify which tools were used; the model version; the tasks for which AI was employed; the prompts or prompt strategies used; and the extent of AI contribution relative to human

authorship. This level of disclosure enables readers to assess AI's potential influence on findings and supports reproducibility.

The social science literature on positionality and reflexivity provides additional justification for the disclosure imperative. Just as social scientists are expected to reflect on how their social location shapes their interpretive choices, they are equally obligated to reflect on how AI mediation, with its embedded assumptions, biases, and limitations, shapes their research process and findings (Bourdieu, 1990; Alvesson & Skoldberg, 2009). GenAI disclosure is not merely a compliance requirement; it is an expression of the epistemological humility central to rigorous social inquiry. Recommended practices under this pillar include: maintaining an AI usage log throughout the research process; including a standardised GenAI disclosure statement in manuscript methods sections; engaging with institutional AI policies and contributing to their development; and consulting current journal-specific policies before submission.

3.2. O — Oversight & Quality Control

The second pillar addresses the human oversight and quality control mechanisms necessary to ensure that AI-assisted research produces valid, reliable, and unbiased findings. GenAI systems are prone to a range of failure modes with relevance for social science: hallucination (generating plausible sounding but factually incorrect content), reproducing biases present in training data, misrepresenting statistical relationships, and producing culturally inappropriate or contextually insensitive outputs (Ji et al., 2023; Weidinger et al., 2021).

Human oversight is not optional: it is the primary safeguard against these failure modes. Oversight mechanisms should operate at multiple levels. At the individual researcher level, this means critically evaluating all AI-generated outputs rather than treating them as authoritative. At the team level, it means implementing peer verification procedures in which team members independently review and validate AI-assisted analyses. At the institutional level, it means establishing data governance frameworks that specify permissible uses of GenAI with participant data and ensure compliance with data protection legislation.

Bias auditing deserves particular attention. Social scientists researching marginalised communities, sensitive social phenomena, or cross-cultural contexts should conduct systematic bias audits of GenAI outputs before incorporating them into research. These audits may involve comparing AI-generated categorisations with human expert judgements, testing for differential performance across demographic groups, and examining how AI framings may implicitly reproduce dominant cultural assumptions. Methods for bias auditing are an active area of development in the AI accountability literature (Raji et al., 2020).

Data privacy and ethical governance are further dimensions of oversight with particular urgency in social science. Researchers processing sensitive interview data, clinical narratives, or other personal information through GenAI platforms must attend carefully to the data governance implications, including whether platform terms of service permit

commercial use of submitted data. Where GenAI is used in data collection, for example, as an interviewing assistant—this must be disclosed to participants and addressed in IRB protocols.

3.3. L — Literacy & Understanding

The third pillar addresses the need for researchers to develop substantive AI literacy: a deep and critical understanding of how GenAI systems work, what they can and cannot do, and what risks they introduce. AI literacy, in this context, goes beyond operational familiarity with specific tools. It encompasses understanding of transformer architectures and the implications of training data composition; awareness of the empirical literature on LLM failure modes; the capacity to critically appraise AI-generated outputs; and reflexive awareness of how AI tools reshape the research process.

AI literacy has important equity dimensions. There is currently significant variation in GenAI literacy across the social science research community, correlating with institution type, career stage, disciplinary background, and geographic location. Junior researchers and those at under-resourced institutions may be particularly vulnerable to uncritical AI adoption without adequate technical background. Research institutions and professional associations bear responsibility for investing in AI literacy education and ensuring equitable access to training resources.

Literacy also requires critical engagement with AI outputs. Social scientists are well-positioned to apply their disciplinary skills—critical reading, source evaluation, reflexive analysis—to GenAI outputs, but this requires conscious effort and methodological design. Recommended practices include triangulating AI-generated analyses with human expert review; maintaining detailed audit trails of AI interactions; integrating AI literacy into research methods training curricula; developing field-specific guidelines for critically appraising AI assistance; and fostering communities of practice that share best practices and failure cases.

The ETHICAL framework's injunction to 'Harness Understanding', meaning not just operational competence but informed critical engagement, resonates closely with this pillar. POLAS extends this by foregrounding the equity and institutional dimensions of literacy development.

3.4. A — Accountability

The fourth pillar concerns the clear assignment of responsibility for AI-assisted research decisions and outputs. The fluency and accessibility of GenAI create a risk of accountability diffusion: when errors, biases, or misconduct arise in AI-assisted research, it may be unclear who is responsible. POLAS holds that human researchers bear ultimate and undivided responsibility for all aspects of their research, regardless of the extent to which AI tools were employed in its production.

Operationalising accountability in AI-assisted research requires deliberate structural design. Research teams should designate a lead AI-responsible researcher who maintains

an overview of AI use throughout the project; ensures compliance with ethical guidelines and institutional policies; oversees the integrity of AI-assisted analyses; and takes responsibility for disclosure and reporting. This is analogous to the established role of principal investigator in research ethics governance.

Accountability also requires robust documentation practices. Version control of prompts, model specifications, and AI-generated outputs enables researchers to trace the provenance of specific findings, facilitates audit in cases of disputed results, and supports reproducibility. The scientific community's emerging norm that AI should not be listed as an author, adopted by most leading journals, reflects a fundamental accountability principle: authorship entails responsibility, and AI systems cannot bear responsibility for their outputs (Else, 2023; Knöchel et al., 2025).

A further accountability dimension concerns the relationship between researchers and the communities whose data they analyse. Social science research frequently involves power asymmetries between researchers and participants, and these are not dissolved by AI mediation. On the contrary, GenAI tools may introduce additional layers of opacity that further distance participants from the knowledge-production process. Accountability in this broader sense requires ongoing community engagement, transparent communication about research methods and findings, and responsiveness to community concerns about AI use.

3.5. S — Stewardship & Social Impact

The fifth and final pillar of POLAS is in many ways the most distinctively social science oriented. Stewardship & Social Impact concerns the researcher's obligations to consider, monitor, and mitigate the broader social consequences of GenAI use in their research—and to actively contribute to the responsible governance of AI in the research ecosystem.

Social scientists have long recognised that research is never socially neutral: it produces knowledge that can be used to inform policy, shape public discourse, and affect the lives of communities (Merton, 1973; Harding, 1986). GenAI extends and intensifies this social embeddedness. Research findings produced with GenAI assistance may carry implicit assumptions derived from AI training data; they may circulate more rapidly and widely than traditionally produced research; and they may be adopted in high-stakes policy contexts without adequate scrutiny of their AI-assisted provenance.

The stewardship pillar calls on social scientists to conduct impact assessments prior to deploying GenAI in research—considering questions such as: Who may be harmed if AI-generated analyses are inaccurate or biased? What communities are represented (and misrepresented) in the training data of the AI tools being used? How might AI-assisted findings be used or misused by policymakers, media, or other actors? These questions are analogous to the social impact considerations already embedded in IRB review processes, and POLAS argues that they should be extended to encompass AI-related risks.

Stewardship also encompasses a broader responsibility to the research community and public. Researchers who develop expertise in ethical GenAI use are positioned to contribute

to the policy conversations, journal guidelines, and professional standards that shape the field's collective response to AI. Sharing lessons learned—including failures—contributes to the collective intelligence of the research community. Publishing AI disclosure statements openly, participating in cross-institutional working groups, and contributing to the development of social science–specific AI ethics guidance are all expressions of this stewardship role.

The equity dimension of stewardship is particularly salient. Social science has a special obligation to attend to who benefits from and who is harmed by new research technologies. If GenAI systematically advantages well-resourced researchers and institutions at the expense of others—deepening existing inequalities in academic knowledge production—then social scientists have a responsibility to name and resist this dynamic. POLAS calls for active engagement with the political economy of AI in research, not merely passive compliance with technical guidelines.

4. IMPLEMENTATION GUIDELINES: THE POLAS CHECKLIST

The practical value of any ethical framework lies in its implementability. POLAS is designed not only as a conceptual model but as a working instrument that research teams can apply throughout the research lifecycle. The following checklist translates each POLAS pillar into concrete, phase-specific actions. The checklist is intended to be adapted to disciplinary contexts and institutional settings, and researchers are encouraged to supplement it with field-specific guidance.

Table 2. POLAS Implementation Checklist

POLAS Pillar	Actionable Steps	Phase
P — Policies & Disclosure	Reference global/institutional AI guidelines; include a GenAI usage statement in the methods section; disclose all tools, model versions, and prompts used; report extent of AI-generated content.	Planning, Writing, Submission
O — Oversight & Quality Control	Document peer verification of AI outputs; maintain data governance logs; conduct automated and manual bias audits; ensure IRB/ethics compliance when AI is used in data collection.	Data Collection, Analysis, Review
L — Literacy & Understanding	Organize team AI literacy training; integrate AI limitations into research design; critically appraise all AI-assisted interpretations; use multiple verification passes.	Pre-study, Ongoing Training
A — Accountability	Appoint a lead AI-responsible researcher; maintain model/prompt version control; validate AI outputs against primary sources; do not list AI as author; retain human final authority.	Throughout All Phases
S — Stewardship & Social Impact	Conduct societal impact assessments; involve stakeholder communities; address equity and representation gaps; publish AI disclosures openly; consider long-term implications of AI-derived findings.	Publication, Dissemination, Follow-up

4.1. PLANNING PHASE

The planning phase is the critical window for establishing the ethical architecture of AI-assisted research. Under the Policies pillar, researchers should identify applicable institutional, journal, and regulatory policies governing GenAI use before beginning the project. A written AI usage plan, specifying which tools will be used, for what purposes, and subject to what safeguards, should be developed and, where appropriate, submitted to the

IRB alongside standard ethics applications. Under the Literacy pillar, the planning phase should include an AI literacy assessment for all team members, with targeted training for those with limited experience. Under the Stewardship pillar, researchers should conduct a preliminary social impact assessment, identifying communities who may be affected by the research and considering how AI use may introduce or amplify risks to those communities.

4.2. Data Collection and Analysis Phase

During data collection and analysis, the phases most directly affected by GenAI capabilities, oversight mechanisms are paramount. All AI-assisted analyses should be subjected to human verification before being accepted as findings. Bias audits should be conducted, particularly when AI is used to categorise or interpret data relating to social groups. Data governance procedures should be followed rigorously: sensitive participant data should not be submitted to commercial AI platforms without explicit consent and institutional authorisation. Under the Accountability pillar, the lead AI-responsible researcher should maintain a detailed log of all AI interactions during this phase, including prompts, model versions, and outputs. This log constitutes both an audit trail and the primary documentation source for the methods section.

4.3. Writing and Submission Phase

In the writing and submission phase, the Policies pillar comes to the fore. Researchers should include a standardised GenAI disclosure statement in the methods section, specifying tools, tasks, and extent of AI involvement. Journal-specific AI policies should be consulted, and submission cover letters should address GenAI use where required. Under the Accountability pillar, final review should ensure that all AI-generated content has been verified and that no AI system is listed as an author.

4.4. Dissemination and Follow-up Phase

The dissemination phase is where the Stewardship pillar is most salient. Researchers should actively consider how AI-assisted findings may be received and used by different audiences and take steps to communicate the role of AI in the research clearly and accessibly in all dissemination contexts—not only in the academic paper but in press releases, policy briefs, and public communications. Post-publication, researchers should monitor for misuse or misrepresentation of their findings and be prepared to respond publicly where AI-related limitations are relevant.

5. DISCUSSION

5.1. POLAS in Comparative Perspective

Table 3 below presents a systematic comparison of POLAS with the Core Principles framework and the ETHICAL framework, the two most directly relevant existing

frameworks. This comparison situates POLAS within the existing landscape and clarifies its distinctive contributions.

Table 3. Comparative Analysis of Ethical AI Frameworks

Dimension	Core Principles (Springer Nature)	ETHICAL Framework	POLAS Framework
Scope	General AI/research ethics	Practitioner-focused guidance	Social science–specific
Format	Eight principles	Seven-step checklist	Five-pillar acronym
Disclosure	Transparency principle	Indicate & Examine steps	P — explicit disclosure
Bias/Quality	Quality Control & Bias Mitigation	Critically Engage	O — systematic oversight
Education	Implied	Harness Understanding	L — formal literacy
Human Responsibility	Accountability principle	Partial — user responsibility	A — explicit assignment
Social Context	Broader Impact	Think social impacts	S — community-centred

The comparison reveals that POLAS is distinguished from existing frameworks in three principal respects. First, it is explicitly social science–specific in its scope: each pillar is elaborated with reference to the methodological traditions, ethical norms, and social stakes particular to the social sciences. The Core Principles and ETHICAL frameworks, while valuable, are designed for general research contexts and do not address the distinctive epistemological and methodological concerns of interpretive, qualitative, or critical social research.

Second, POLAS foregrounds the equity and structural dimensions of ethical GenAI use—particularly in the Stewardship and Literacy pillars, in a way that existing frameworks do not. The attention to AI literacy equity, the political economy of AI development, and the obligations of researchers to communities affected by AI-mediated research reflects the social science tradition's concern with power, positionality, and social impact.

Third, POLAS is designed as a lifecycle-oriented framework, with each pillar explicitly linked to research phases and operationalised in a practical checklist. This lifecycle orientation enables researchers to apply POLAS systematically rather than retrospectively, embedding ethical consideration throughout the research process rather than treating it as a compliance exercise at the point of submission.

5.2. Challenges and Limitations

The implementation of POLAS faces several challenges that must be acknowledged. The pace of GenAI development means that any framework risks obsolescence: model capabilities, failure modes, and policy contexts are evolving continuously. POLAS is designed to be adaptable, its principles are not tool-specific—but the checklist and guidelines will require periodic revision as the landscape changes. We recommend that professional associations in the social sciences establish standing committees charged with updating POLAS guidance in response to technological and regulatory developments.

Resource constraints pose a significant implementation challenge, particularly for researchers at under-resourced institutions, those working independently, or early-career scholars without institutional support. The AI literacy training, bias auditing, and impact assessment activities implied by POLAS require time, expertise, and sometimes financial

resources that are not equitably distributed across the research community. Addressing these constraints requires coordinated action at the institutional and funding level, not merely individual researcher effort.

Disciplinary diversity within the social sciences means that POLAS cannot be applied in a uniform way across all fields and methodological traditions. The framework's elaboration in this paper focuses primarily on issues salient to qualitative and interpretive research, but the quantitative social sciences, computational social science, and mixed-methods research each present distinct AI ethics challenges that deserve field-specific guidance. Future work should elaborate POLAS for specific disciplinary and methodological contexts.

Finally, enforcement and incentives remain an unresolved challenge. Ethical frameworks are only effective insofar as researchers are motivated to implement them. At present, the incentive structures of academic publishing and research evaluation do not consistently reward rigorous AI ethics disclosure; in some contexts, they may inadvertently penalise transparency by drawing attention to methodological limitations. Aligning institutional incentives with ethical AI practice—through journal policies, grant requirements, and research assessment criteria—is a systemic challenge that extends beyond the scope of any single framework.

5.3. Implications for Research Practice and Education

The implications of POLAS for research practice are wide-ranging. At the level of the individual researcher, POLAS calls for a shift in orientation: from treating GenAI as a transparent tool to recognising it as an active participant in knowledge production that must be understood, managed, and accounted for. This shift requires investment in AI literacy and critical reflection on the epistemological implications of AI mediation.

At the team and institutional level, POLAS implies new organisational roles and processes, AI usage planning, bias auditing, impact assessment, that require resourcing and coordination. Research teams that integrate POLAS into their standard operating procedures will be better positioned to produce rigorous, ethical, and reproducible AI-assisted research.

For research methods education, POLAS provides a curriculum framework for teaching ethical AI use. Methods courses in the social sciences at both undergraduate and postgraduate levels, should incorporate AI literacy modules covering the principles of POLAS, practical experience with GenAI tools and their limitations, and critical engagement with the social and ethical dimensions of AI in research.

5.4. Future Research Directions

The present paper is intended as a contribution to a developing conversation, and several important questions remain for future research. Longitudinal studies tracking the impact of AI ethics frameworks on research quality and integrity outcomes would provide valuable evidence for the effectiveness of different governance approaches. Comparative studies examining how POLAS and analogous frameworks are implemented across different

disciplinary, institutional, and national contexts would illuminate the practical challenges and facilitating conditions for ethical AI adoption.

The development of validated instruments for assessing AI literacy in the social science research community would support both empirical research and educational programme evaluation. Field-specific elaborations of POLAS, for example, for social work research, public health research, or critical race studies, would enhance the framework's practical utility. And engagement with the perspectives of research participants and community stakeholders on GenAI use in social research, voices largely absent from the current literature, would enrich the Stewardship pillar considerably.

6. CONCLUSION

The integration of generative artificial intelligence into social science research represents one of the most significant methodological developments in the field's recent history. It brings genuine opportunities, for analytical power, efficiency, and democratisation, alongside serious ethical risks that the social sciences are uniquely positioned to analyse and address. The urgency of developing robust, field-specific ethical guidance has never been greater.

This paper has introduced the POLAS framework, Policies & Disclosure, Oversight & Quality Control, Literacy & Understanding, Accountability, and Stewardship & Social Impact—as a structured, practical, and social science–specific model for ethical GenAI use in research. Building on and extending existing frameworks from the Core Principles, ETHICAL, European Commission guidelines, and Resnik and Hosseini, POLAS synthesises the most important ethical insights of the field into a coherent, memorable, and operationalizable structure.

POLAS is offered not as a compliance checklist but as an orientation: a set of commitments that express what it means to be a responsible social science researcher in the age of generative AI. These commitments, to transparency, to human oversight, to critical knowledge, to clear accountability, and to social stewardship, are continuous with the deepest traditions of the social sciences. They reflect the field's long-standing recognition that research is a moral practice embedded in social relationships and power structures, and that researchers bear responsibility not only for the technical quality of their work but for its social consequences.

The adoption of POLAS will require effort at multiple levels: individual researchers developing AI literacy and disclosure practices; research teams building oversight structures; institutions investing in training and governance; journals refining submission policies; and funders incorporating ethical AI requirements into grant conditions. No framework can substitute for this collective work. But POLAS can provide a common language and a structured reference point for the conversation that the social science community needs to have, and to keep having, as the technology and its ethical landscape continue to evolve.

The study invites engagement, critique, and collaboration from social science researchers, ethicists, and AI developers in refining and extending POLAS. The challenges of ethical GenAI use in social science are too important and too complex to be addressed by any single framework or any single community. They require the kind of sustained, reflexive, community-driven inquiry that the social sciences are, at their best, uniquely equipped to provide.

Author Contributions:

Author alone contributed to the study conception and design, data collection, and initial thematic analysis.

Funding:

The authors received no specific funding for this work.

Institutional Review Board Statement:

N/A

Informed Consent Statement:

N/A

Data Availability Statement:

N/A

Acknowledgments:

The author would like to thank Sanzida Esha (Author's wife) who was engaged in early discussions about the lack of humanity in AI and specifically Large Language Models.

Conflicts of Interest:

The author declares no competing interests.

Declaration on the Use of Generative Artificial Intelligence

Generative artificial intelligence tools (M365 Copilot, Claude) were used to assist with language editing and stylistic refinement of this manuscript. All scientific content, analyses, and interpretations were developed by the authors, who assume full responsibility for the accuracy and integrity of the work.

References

Alvesson, M., & Sköldberg, K. (2009). Reflexive methodology: New vistas for qualitative research (2nd ed.). SAGE Publications.

Ansarullah, S. I., Kirmani, M. M., Alshmrany, S., & Firdous, A. (2024). Ethical issues around artificial intelligence. In A biologist's guide to artificial intelligence (pp. 301–314). Academic Press.

Bail, C. A. (2024). Can generative AI improve social science? Proceedings of the National Academy of Sciences, 121(21), Article e2314021121.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (pp. 610–623). Association for Computing Machinery.

Blodgett, S. L., Barocas, S., Daumé, H., III, & Wallach, H. (2020). Language (technology) is power: A critical survey of "bias" in NLP. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (pp. 5454–5476). Association for Computational Linguistics.

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., Artetxe, M., et al. (2021). On the opportunities and risks of foundation models (arXiv:2108.07258). arXiv. https://doi.org/10.48550/arXiv.2108.07258

Bourdieu, P. (1990). The logic of practice (R. Nice, Trans.). Stanford University Press.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., et al. (2020). Language models are few-shot learners. Advances in Neural Information Processing Systems, 33, 1877–1901.

Davidson, T., & Karell, D. (2025). Integrating generative artificial intelligence into social science research: Measurement, prompting, and simulation. Sociological Methods & Research, 54(3), 775–793.

Dellafiore, F., Saba, A., Collaro, C., & Artioli, G. (2026). Artificial intelligence in qualitative research: Insights from experts via reflexive thematic analysis. Qualitative Health Research, 36(2–3), 145–165.

Dergaa, I., Chamari, K., Zmijewski, P., & Ben Saad, H. (2023). From human writing to artificial intelligence generated text: Examining the prospects and potential threats of ChatGPT in academic writing. Biology of Sport, 40(2), 615–622.

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. 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, Article 102642.

Eacersall, D., Pretorius, L., Smirnov, I., Spray, E., Illingworth, S., Chugh, R., Strydom, S., Stratton-Maher, D., Simmons, J., Jennings, I., Roux, R., Kamrowski, R., Downie, A., Thong, C. L., & Howell, K. A. (2025). Navigating ethical challenges in generative AI-enhanced research: The ETHICAL framework for responsible generative AI use. Journal of Applied Learning & Teaching, 8(2), 1–14.

Else, H. (2023). Abstracts written by ChatGPT fool scientists. Nature, 613, 423.

European Commission. (2024). European living guidelines on the responsible use of generative AI in research. Directorate-General for Research and Innovation. https://research-and-innovation.ec.europa.eu

Gioia, D. A., Corley, K. G., & Hamilton, A. L. (2013). Seeking qualitative rigor in inductive research: Notes on the Gioia methodology. Organizational Research Methods, 16(1), 15–31.

Harding, S. (1986). The science question in feminism. Cornell University Press.

Jeon, J., Kim, L., & Park, J. (2025). The ethics of generative AI in social science research: A qualitative approach for institutionally grounded AI research ethics. Technology in Society, 81, Article 102836.

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., et al. (2023). Survey of hallucination in natural language generation. ACM Computing Surveys, 55(12), 1–38.

Kapoor, S., & Narayanan, A. (2023). Leakage and the reproducibility crisis in machine-learning-based science. Patterns, 4(9), Article 100804.

Knöchel, T.-D., Schweizer, K. J., Acar, O. A., Akil, A. M., Al-Hoorie, A. H., Buehler, F., et al. (2025). Core principles of responsible generative AI usage in research. AI and Ethics, 5(6), 6371–6377.

Lund, B. D., Wang, T., Mannuru, N. R., Nie, B., Shimray, S., & Wang, Z. (2023). ChatGPT and a new academic reality: AI-written research papers and the ethics of the large language model publishing landscape. Journal of the Association for Information Science and Technology, 74(5), 570–581.

Merton, R. K. (1973). The normative structure of science. In N. W. Storer (Ed.), The sociology of science: Theoretical and empirical investigations. University of Chicago Press.

Mohamed, S., Png, M.-T., & Isaac, W. (2020). Decolonial AI: Decolonial theory as sociotechnical foresight in artificial intelligence. Philosophy & Technology, 33(4), 659–684.

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. Science, 381(6654), 187–192.

Qian, P., & Xie, X. (2026). Task–technology fit leads to conflict: The double-edged-sword effect of generative artificial intelligence on scientific creative performance in humanities and social sciences research. International Journal of Human–Computer Interaction, 42(3), 1867–1891.

Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., et al. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (pp. 33–44). Association for Computing Machinery.

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence, 1(5), 206–215.

Sebastian, R., Kottekkadan, N. N., Thomas, T. K., & Niyas KK, M. (2025). Generative AI tools (ChatGPT*) in social science research. Journal of Information, Communication and Ethics in Society, 23(2), 284–290.

Stokel-Walker, C., & Van Noorden, R. (2023). What ChatGPT and generative AI mean for science. Nature, 614, 214–216.

Tiwari, A., Chugh, A., & Sharma, A. (2023). Uses of artificial intelligence with human-computer interaction in psychology. In Innovations in artificial intelligence and human-computer interaction in the digital era (pp. 173–205). Academic Press.

Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P.-S., et al. (2021). Ethical and social risks of harm from language models (arXiv:2112.04359). arXiv. https://doi.org/10.48550/arXiv.2112.04359