AI for Social Science: Reclaiming Research Sovereignty in the Age of Artificial Intelligence

The twenty-first century has witnessed an unprecedented explosion of data and an equally unprecedented concentration of the means to process it. While the digital revolution promised to democratise knowledge, the reality has been rather different: the infrastructures of artificial intelligence (AI) – the cloud servers, the training datasets, the algorithmic architectures – remain overwhelmingly concentrated in the hands of a few corporations headquartered in the Global North. For scholars in Africa, Asia, and Latin America, this asymmetry poses a fundamental challenge: how can we conduct rigorous social science research when the very tools we depend upon are designed elsewhere and for other purposes, embedding assumptions and priorities that may not align with our own?

This study emerges from the work of Global South Insights (GSI), a research initiative within Tricontinental: Institute for Social Research dedicated to developing analytical frameworks and technological capabilities that serve the needs of peoples in the periphery and semi-periphery of the world system. Over the past two years, our team has grappled with a deceptively simple question: can AI be harnessed to strengthen – rather than supplant – the intellectual labour of social scientists working on the problems that matter most to the Global South?

The answer we have arrived at is neither naïve techno-optimism nor reflexive technophobia. We have developed what we call AI4SS (AI for Social Science), a comprehensive framework that, grounded in trustworthy research infrastructure, operates across three interconnected layers: an ontological layer that reconceptualises the fundamental categories of research in the age of AI; a methodological layer that translates these concepts into concrete technical architectures; and a practical layer that implements these architectures in functioning systems. This framework is not merely theoretical. It has been tested in the production of research on digital sovereignty, technology competition, and development policy – research that has already begun to circulate among scholars and policymakers across the Global South.

What distinguishes AI4SS from the proliferation of AI tools now marketed to academics is its explicit commitment to what we call epistemic sovereignty: the principle that researchers in the Global South should control the means of knowledge production, rather than remaining dependent upon infrastructures and frameworks developed in and for the Global North. This does not mean rejecting Northern technologies – such a stance would be both impractical and counterproductive. It means building the capacity to adapt, extend, and redirect these technologies in accordance with our own research agendas and intellectual traditions.

The pages that follow present the AI4SS framework in its entirety: the problems it addresses, the ideas it embodies, the architecture it deploys, and the applications it has already enabled. Our hope is that this study will serve not only as an introduction to our work but as an invitation to collaboration. The challenges facing social science research in the Global

South cannot be solved by any single team or institution. They require collective effort, shared infrastructure, and solidarity across borders. We offer AI4SS as a contribution to that larger project.

1. The Crisis of Social Science Research in the Data Age

The contemporary social scientist confronts a paradox. Never before has so much information about human societies been available; never before have the tools for processing that information been so powerful. Yet the lived experience of researchers – particularly those working in under-resourced institutions across the Global South – is often one of frustration, fragmentation, and futility. The gap between what is theoretically possible and what is practically achievable grows wider with each passing year. To understand why this is so, we must examine the structural contradictions that define social science research in the current conjuncture.

1.1 The Bandwidth Paradox: Data Abundance Versus Cognitive Limits

The first contradiction arises from the collision between the explosive growth of available data and the biological constraints of human cognition. Consider the situation facing a researcher studying, say, the political economy of extractive industries in West Africa. Relevant materials exist in multiple languages (English, French, Portuguese, Arabic, and numerous African languages); in multiple formats (government reports, corporate filings, news articles, academic papers, social media posts, satellite imagery, audio recordings of community meetings); and across multiple jurisdictions (national, regional, international). The total volume of potentially relevant information far exceeds what any individual, or even any team, could possibly read, let alone analyse.

Faced with this deluge, researchers have traditionally resorted to two complementary strategies: sampling and abstraction. Sampling means selecting a manageable subset of the available data according to some criteria (random, purposive, convenience sampling). Abstraction means reducing the complexity of the data through categorisation, coding, and quantification. Both strategies are legitimate and often necessary. But both also entail significant losses. Sampling means that potentially crucial evidence may be excluded; abstraction means that the granular texture of social reality – the specific words people use, the particular contexts in which events unfold – is smoothed away in favour of generic categories.

The result is a kind of reductionist imperative built into the very foundations of social science methodology. Not because researchers prefer simplification, but because the cognitive bandwidth of human beings – what Herbert Simon termed 'bounded rationality', our limited capacity to read, remember, and reason – forces us to simplify. We lose breadth in order to gain depth, or sacrifice depth in pursuit of breadth. The possibility of achieving both simultaneously – of conducting research that is at once comprehensive and granular – has seemed foreclosed by the hard limits of human information processing.

1.2 The Rigour Gap: Generative Fuzziness Versus Computational Precision

The second contradiction emerges from the peculiar characteristics of the AI tools now most widely available. Large language models such as ChatGPT have captured the public imagination and rapidly penetrated academic practice. These systems excel at certain tasks: summarising texts, generating prose, and answering questions in natural language. But they suffer from well-documented limitations that render them problematic for serious scholarly work.

Chief among these is the phenomenon of hallucination: the tendency of language models to generate plausible-sounding but factually incorrect statements. Research has shown that over half of the academic citations generated by ChatGPT are fabricated, and even the more advanced GPT-4 still exhibits an 18% fabrication rate. A model asked to summarise a body of literature may invent citations that do not exist; asked to analyse a dataset, it may report statistical findings that bear no relation to the underlying numbers. These failures are not bugs to be fixed but features intrinsic to the probabilistic architecture of such systems. Language models are trained to predict the most likely next word in a sequence, not to ascertain truth or verify facts. Their outputs are fundamentally generative rather than analytical – they produce text that *sounds* right rather than text that *is* right.

This creates a profound disconnect between the capabilities of general-purpose AI tools and the requirements of rigorous social science. Academic research demands not only the generation of plausible narratives but their grounding in verifiable evidence; not only the identification of patterns but their validation through statistical inference; not only the articulation of arguments but their logical coherence and empirical adequacy. The language model that can write a serviceable first draft of a literature review cannot perform a regression analysis, verify that the sources it cites actually exist, or assess whether its conclusions follow from its premises.

We are thus confronted with a gap between what can be *calculated* and what can be *understood* – between the deterministic precision of statistical computation and the probabilistic fuzziness of language generation. Bridging this gap requires not merely better models but fundamentally different architectures: systems that can integrate the fluency of language models with the rigour of computational methods, the creativity of generative AI with the accountability of evidence-based reasoning.

1.3 The Disciplinary Archipelago: Methodological Isolation Versus Interdisciplinary Imperatives

The third contradiction concerns the organisation of academic knowledge itself. The social sciences have long been divided into distinct disciplines – economics, sociology, political science, anthropology, history – each with its own theoretical traditions, methodological conventions, and institutional structures. This division of labour has been productive in many ways, enabling the development of specialised expertise and the accumulation of disciplinary knowledge. But it has also created what we might call a Methodological Archipelago: a landscape of isolated islands, each speaking its own language, each employing its own techniques, with limited communication between them.

Consider the challenge of assessing the development strategy of a Global South country. An economist might analyse the health of the economic structure through macroeconomic indicators – GDP growth rates, capital formation ratios, trade dependency. A political scientist might examine governance capacity, the formulation and implementation of industrial policy, and the constraints that international dependency relations impose on policy space. A sociologist might focus on structures of inequality, the stratification of labour markets, and the distribution of development outcomes across different social groups. A historian might trace how colonial legacies have shaped the present economic structure and institutional arrangements. Any single dimension yields only a partial picture; the complete judgement – whether a country's development path is sustainable, whom it benefits, what structural forces constrain it – can only emerge through integration. But integration is precisely what disciplinary structures impede.

The problem is not merely intellectual but practical. Different disciplines use different software, data formats, and analytical protocols. The qualitative researcher working with interview transcripts in NVivo (qualitative data analysis software) cannot easily combine their analysis with the quantitative researcher's Stata (statistical software) output. The historian reading archival documents has no standardised way to link their findings to the economist's time-series data. Cross-disciplinary collaboration, when it occurs, typically requires laborious manual translation between methodological frameworks – a process so costly in time and effort that it happens far less often than it should.

What is needed is not the abolition of disciplinary specialisation but its transcendence through methodological interoperability: the capacity for tools and techniques developed in one field to be readily employed in another, for findings generated through one method to be systematically integrated with findings from other methods. Such interoperability cannot be achieved through appeals alone; it requires technical infrastructure that enables different methodological traditions to communicate.

1.4 The Labour Inversion: Mechanical Toil Versus Intellectual Creation

The fourth contradiction is perhaps the most insidious because it concerns not the nature of research outputs but the allocation of research labour. Ask any working social scientist how they spend their time, and you will likely hear a litany of tasks that have little to do with the creative intellectual work that drew them to scholarship in the first place: formatting citations, cleaning datasets, searching for literature, converting files between formats, navigating bureaucratic systems. These tasks are necessary but not sufficient; they must be done, but doing them does not advance knowledge.

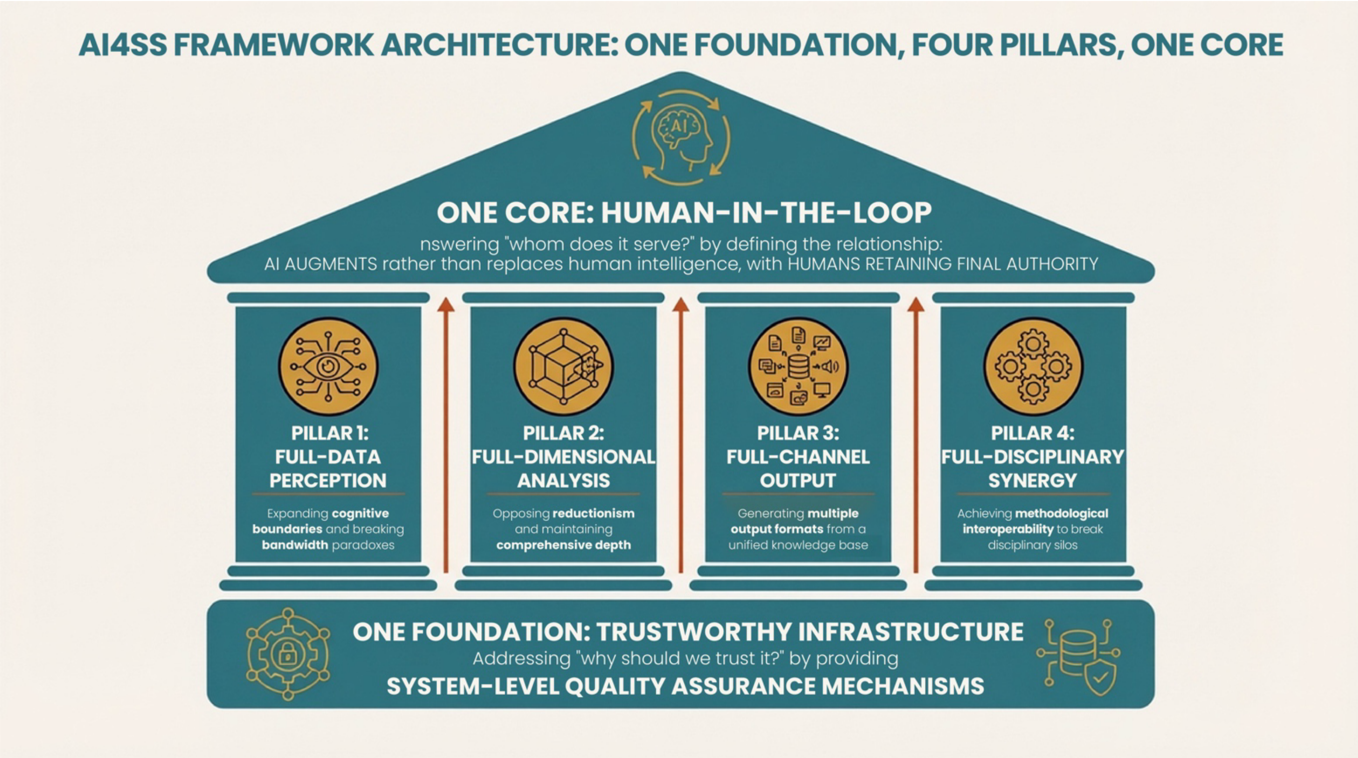
The result is a perverse inversion of value: researchers spend the bulk of their time on low-value mechanical labour while the high-value intellectual labour – theoretical construction, interpretive insight, the formulation of genuinely new ideas – is squeezed into whatever hours remain. This inversion is especially damaging in the Global South, where researchers often lack the institutional support (research assistants, administrative staff, technical infrastructure) that their Northern counterparts take for granted. A professor in Lagos or La Paz may spend more time on bureaucratic and technical tasks than a professor in London or Los Angeles, leaving less time for the research that could establish their reputation and advance their career.

The irony is that many of these mechanical tasks are precisely what machines ought to be good at. Computers excel at repetitive operations, pattern matching, format conversion, information retrieval. Humans excel at judgement, interpretation, creativity, and the kind of contextual understanding that comes from lived experience. As Norbert Wiener observed in *The Human Use of Human Beings* (1950), a rational division of labour would assign to machines what machines do best and reserve for humans what humans do best. Instead, we have the opposite: machines (which often hallucinate) are used for thinking, while humans perform the drudgery that machines could handle.

2. Towards a New Paradigm: The Ontological Foundation of AI4SS

The contradictions outlined above are not inevitable features of the human condition; they are historical products of specific technological and institutional configurations. As those configurations change, new possibilities emerge. The AI4SS framework represents an attempt to seize those possibilities – to reconceptualise the fundamental categories of social science research in light of what artificial intelligence now makes possible. This reconceptualisation operates at the ontological level: it concerns not merely the tools researchers use but the very nature of research itself – the relationships among researcher, data, method, and output.

We organise the ontological foundation of AI4SS as a three-tiered structure we call ‘One Foundation, Four Pillars, One Core’. The **Foundation** is Trustworthy Infrastructure, answering the question ‘why should we trust it?’ – it provides system-level quality assurance mechanisms that are the prerequisite for all capabilities to operate reliably. The **Four Pillars** are Full-Data Perception, Full-Dimensional Analysis, Full-Channel Output, and Full-Disciplinary Synergy, answering the question ‘what can it do?’ – four mutually independent yet mutually reinforcing functional capabilities, each responding to the structural contradictions identified above. The **Core** is Human-in-the-Loop, answering the question ‘whom does it serve?’ – it defines the relationship between AI and human researchers, and constitutes the value anchor of the entire system.



The ontological foundation of AI4SS: One Foundation, Four Pillars, One Core

2.1 The Foundation: Trustworthy Infrastructure

To apply artificial intelligence to serious academic research, we must first confront a fundamental problem: the hallucination phenomenon in general-purpose AI tools is not a software defect awaiting a fix but an architectural feature intrinsic to probabilistic language generation. The 'rigour gap' identified in the previous section will not disappear simply because models become larger or more capable. Addressing it requires not better models but system-level constraint architectures.

This is precisely what distinguishes AI4SS from the proliferation of general-purpose AI tools now flooding the market. We are not offering researchers a more intelligent chatbot; we are building a trustworthy research infrastructure – a system designed from the outset with academic standards as its binding constraints. Trustworthiness unfolds along several interconnected dimensions. Execution paths must be predictable and auditable: the system operates along predetermined tracks, AI is not permitted to skip steps or take shortcuts, and every intermediate step can be inspected and traced. Every assertion must be traceable to its original source: the evidential chain of academic research cannot be broken, and the path from final output to underlying data must be preserved in its entirety. Content production and quality verification must be separated: as in the peer review system of academic publishing, the agent that generates content cannot review its own output – independent verification agents must perform the checking. Critical conclusions require multi-source cross-validation: judgements that rely on a single source carry an inherent risk of systematic bias, and robust conclusions can only emerge from the convergence of multiple independent streams of evidence.

Without this foundation, the more powerful the four capabilities above it become, the greater the risk of producing unreliable content at scale. A system capable of perceiving vast quantities of data, conducting multi-dimensional analysis, and generating multi-channel outputs would, in the absence of quality assurance mechanisms, simply produce errors with greater efficiency and at greater scale. Capability and trustworthiness must be built in tandem – this is the fundamental premise of the entire AI4SS architecture.

2.2 Full-Data Perception: Expanding the Cognitive Horizon

The first pillar responds to the bandwidth paradox by fundamentally expanding the range of data that can be brought within the researcher's purview. Full-Data Perception means not merely accessing more data but transforming the relationship between researcher and data such that comprehensiveness and granularity become simultaneously achievable.

This transformation operates along three axes. The first concerns data ownership and access. Traditional social science research has relied heavily on publicly available datasets – government statistics, survey repositories, published texts. But some of the most valuable information for understanding social phenomena exists behind various barriers: proprietary corporate data, confidential government records, paywalled academic literature, and the tacit knowledge embedded in field notes and interview recordings that researchers accumulate but rarely systematise. Full-Data Perception involves the construction of what we call Institutional Cognitive Moats: secure repositories of high-value data assets that become the foundation for sustained research programmes. Through appropriate protocols for access control, encryption, and local storage, institutions can build cumulative knowledge bases that compound in value over time.

The second axis concerns format integration. Social reality does not present itself in neatly structured spreadsheets. It manifests in texts and images, in audio recordings and video footage, in maps and diagrams and handwritten notes. Traditional research methods have struggled to integrate these diverse modalities, forcing researchers to analyse each format separately and synthesise the results manually. Advances in multimodal AI now make it possible to process text, image, audio, and video within unified analytical frameworks – to extract information from a photograph and relate it to information from a document, to transcribe an interview and link its contents to quantitative datasets.

The third axis concerns language. The majority of human knowledge is not in English, yet English dominates the infrastructure of global scholarship. For researchers in the Global South, this creates a double bind: either restrict investigation to English-language sources and miss crucial local materials, or invest enormous effort in translation and lose productivity on other fronts. Semantic alignment technologies based on vector embeddings now enable a different approach: the mapping of texts in different languages into a common semantic space where meaning can be compared and retrieved regardless of the original language. A researcher need not be fluent in Portuguese to search a Brazilian archive, nor in Arabic to analyse Egyptian policy documents. The language barrier does not disappear, but it becomes permeable.

Full-Data Perception, then, is not merely about having more data. It is about establishing a new relationship between the researcher and the totality of relevant information – a relationship in which the limitations of individual cognitive bandwidth are overcome through technological augmentation, while the granularity and contextual richness of original sources are preserved rather than abstracted away.

2.3 Full-Dimensional Analysis: Against Reductionism

The second pillar addresses the reductionist imperative that has historically constrained social science methodology. Full-Dimensional Analysis means conducting research that is comprehensive in scope without sacrificing depth of engagement – that analyses the full dataset rather than a sample, that preserves the micro-texture of social reality rather than smoothing it into aggregate categories.

This is not a rejection of sampling or abstraction as such. There are contexts in which sampling remains necessary (when the population is genuinely open-ended or inaccessible) and abstraction remains valuable (when the research question specifically concerns aggregate patterns). But these should be methodological choices, not technological necessities. The researcher should be able to choose between sampling and exhaustive analysis based on the demands of the inquiry, not the limitations of available tools.

Full-Dimensional Analysis involves several interconnected capabilities. Exhaustive Computation means processing the entirety of a dataset rather than a subset, identifying patterns that sampling might miss, detecting outliers and edge cases that would be excluded from statistical averages. Holographic Analysis means maintaining access to the original data even as higher-level patterns are identified – the ability to drill down from an aggregate finding to the specific instances that compose it, to move fluidly between the macro and micro levels of analysis. Iterative Feedback means establishing dynamic loops between empirical material and theoretical construction, allowing theory to guide data exploration while remaining open to revision in light of what the data reveal.

Perhaps most importantly, Full-Dimensional Analysis seeks to integrate what have traditionally been separate methodological traditions: the quantitative and the qualitative, the statistical and the interpretive. This integration is not a matter of simply combining numbers and narratives; it requires systems capable of mutual validation, where statistical findings can be checked against qualitative evidence and qualitative interpretations can be tested against quantitative patterns. The goal is not to dissolve the distinction between methods but to enable their productive interaction.

2.4 Full-Channel Output: Multiple Outputs from Unified Knowledge

The third pillar concerns the relationship between research and its audiences. Traditionally, the primary output of social science research has been the academic article or monograph – a format optimised for peer review and disciplinary credibility but often ill-suited to reaching wider audiences or influencing policy. Researchers who wish to communicate beyond the academy must engage in a separate process of translation, rewriting their findings in accessible language, reformatting their arguments for different media, adapting their presentations for different contexts. This translation work is time-consuming and often undervalued by academic institutions.

Full-Channel Output inverts this relationship between research and output. Rather than producing a single canonical text that must then be translated into other formats, the researcher works with a unified knowledge base from which multiple outputs can be generated as needed. The same underlying analysis can manifest as an academic paper for a peer-reviewed journal, a policy brief for government officials, a long-form article for an educated general audience, a presentation for a public lecture, even an audio or video production for broader dissemination. Each output is adapted to its intended audience and medium, but all draw from the same evidentiary foundation and maintain the same argumentative integrity.

This capability depends on several technical features. Adaptive Content Generation means the ability to adjust register, vocabulary, and rhetorical structure according to the target audience. Stylistic Emulation means the capacity to match specific conventions – the format of a particular journal, the house style of a particular publication, the presentational norms of a particular discipline. Multimodal Output means generating not only text but also visualisations, graphics, and potentially audio and video. And Traceable Provenance means maintaining what we call the 'digital umbilical cord' that links every output back to its evidentiary sources – the ability to verify any claim by tracing it to the underlying data.

Full-Channel Output thus serves both efficiency and accountability. It allows researchers to maximise the impact of their work by reaching diverse audiences through appropriate channels, while ensuring that this diversification does not compromise scholarly rigour or evidential grounding.

2.5 Full-Disciplinary Synergy: Methodological Interoperability

The fourth pillar addresses the methodological archipelago by establishing infrastructure for cross-disciplinary integration. Full-Disciplinary Synergy means not merely encouraging collaboration between disciplines but building the technical systems that make such collaboration practically feasible.

The key mechanism is Capability Encapsulation: the packaging of disciplinary methods as modular units that can be used by researchers from other fields without requiring mastery of the underlying technical details. An economist's regression toolkit, a sociologist's qualitative coding system, a historian's archival search protocol – each can be wrapped in a standardised interface that exposes its functionality while hiding its implementation complexity. A researcher who needs to conduct a statistical analysis need not become a statistician; they need only know how to invoke the relevant capability and interpret its outputs.

This encapsulation unfolds at two levels. At the tool level, standard protocols such as the Model Context Protocol (MCP) enable heterogeneous systems to exchange data and commands in consistent formats – tools built by different teams, in different programming languages, for different purposes, can nonetheless work together within integrated workflows. At the knowledge level, the emerging skills ecosystem (such as the open agent skills platform Skills.sh) enables methodological knowledge to be packaged as one-click installable, cross-platform reusable capability units – an analytical workflow developed by one researcher can be directly invoked by a researcher on the other side of the world, much like installing a software plugin. MCP addresses how tools connect; the skills ecosystem addresses how knowledge circulates. Together, the researcher becomes not a specialist in any single method but an orchestrator – capable of assembling multi-method workflows appropriate to the question at hand.

Full-Disciplinary Synergy also involves Fractal Collaboration: the ability to scale cooperative research from individual projects to large multi-institutional programmes. The same infrastructure that enables a single researcher to invoke multiple methodological tools enables teams to coordinate their work, institutions to share their capabilities, and networks of researchers across the Global South to pool their resources. The vision is one of cumulative, collective capacity-building rather than fragmented, duplicative effort.

2.6 The Core: Human-in-the-Loop – Augmentation, Not Replacement

As the core of the entire framework, Human-in-the-Loop is in some ways the most fundamental idea, for it concerns the relationship between human researchers and artificial intelligence itself. Human-in-the-Loop is both a technical architecture and a philosophical commitment: the principle that AI systems should augment human intelligence rather than substitute for it, should serve human purposes rather than impose their own.

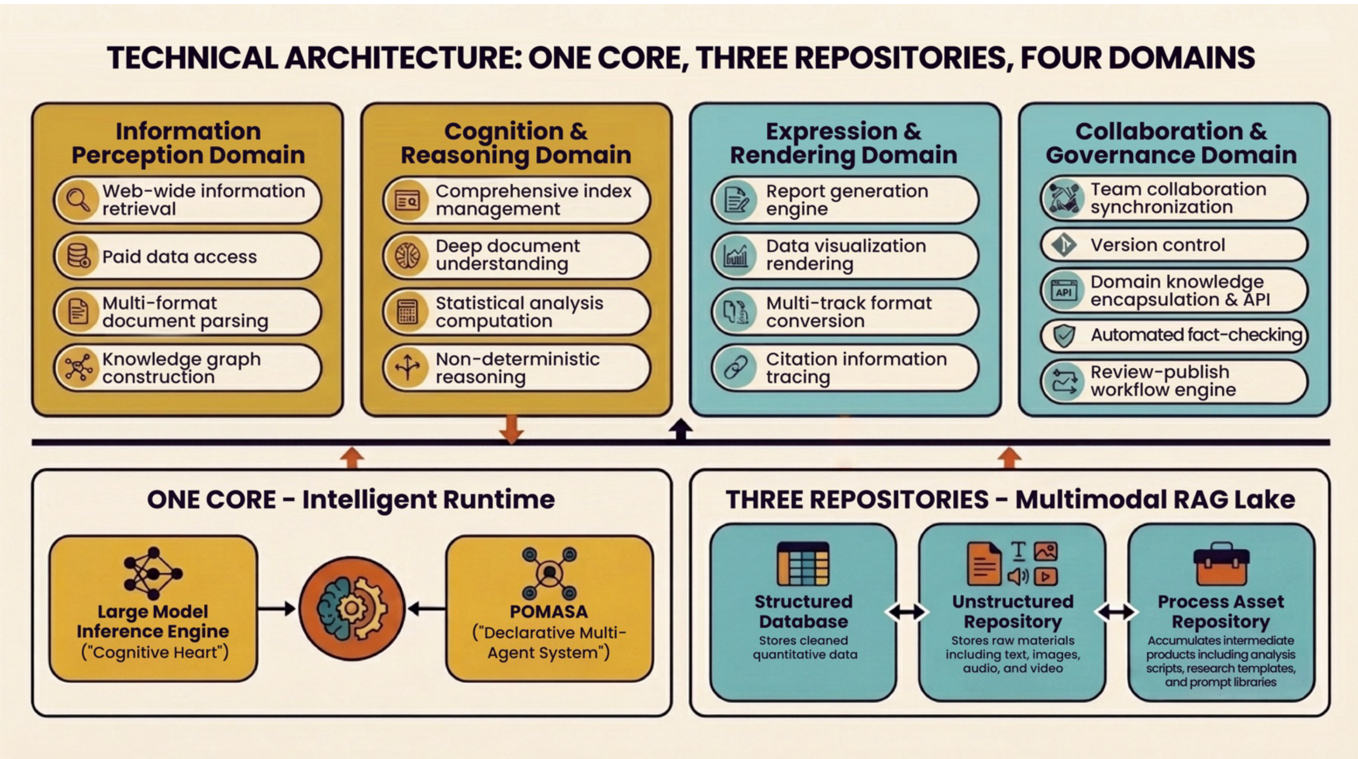
This principle has several dimensions. Value Anchoring means that human researchers retain authority over the ethical frameworks, aesthetic standards, and social concerns that guide inquiry. The AI system does not determine what questions are worth asking or what answers are desirable; it operates within parameters set by humans according to human values. Cognitive Sovereignty means that humans retain ultimate veto power over the outputs of AI systems, particularly in domains requiring theoretical construction, meaning interpretation, and normative judgement. Every output generated by AI is a proposal subject to human review, not a final product; the system must make its reasoning processes visible and its evidential chains accessible, enabling researchers to verify and, where necessary, overturn algorithmic conclusions. The machine may propose; the human disposes.

At the same time, Human-in-the-Loop mandates the automation of mechanical labour – the delegation to machines of precisely those tasks that machines can perform reliably and that consume human time without engaging distinctively human capacities. Literature search, citation formatting, data cleaning, file conversion: these are activities that AI systems can handle more efficiently than humans and whose automation frees researchers to focus on work that genuinely requires human intelligence. But here a critical distinction must be drawn: the goal of automation is not an end-to-end unmanned pipeline but an iterative cycle of human-machine collaboration. AI4SS favours a Plan-Do-Check-Adjust iterative process in which human researchers maintain the capacity to intervene at every stage – to reassess direction, modify parameters, and pursue unexpected leads that the system alone would not have identified.

The ultimate goal is what we call Re-Embedding Experience: liberating the cognitive bandwidth of researchers so that they can return to the field, to the archive, to direct engagement with the social realities they study. The irony of the current situation is that researchers increasingly experience their subjects only through the mediation of datasets and documents, while spending their time on administrative and technical tasks that could be automated. Human-in-the-Loop reverses this pattern: machines handle the mediated, repetitive work while humans reclaim the capacity for direct encounter – for listening to the 'silent voices' that no dataset can fully capture. The time and energy saved through automation should be reinvested in activities that require human presence and perception: fieldwork, interviews, archival immersion, theoretical reflection.

3. The Technical Architecture: One Core, Three Repositories, Four Domains

The ontological concepts described above must be realised in concrete technical systems. The key to implementation is modularity: different capabilities are realised as independent components connected through standard interfaces, rather than bundled in a single monolithic system. This enables flexibility (components can be upgraded, replaced, or extended without affecting the whole), resilience (the failure of one component does not crash the system), and collaborative development (institutions can contribute specialised components and share them with the broader community). This section presents the methodological layer of AI4SS: the architecture that translates conceptual framework into operational infrastructure. We describe this architecture using a schematic formula – 'one core, three repositories, four domains' – that captures its essential structure.



The AI4SS technical architecture: One Core, Three Repositories, Four Domains

3.1 The Intelligent Runtime ('One Core')

At the heart of AI4SS lies what we call the Intelligent Runtime: the computational engine that drives all system operations. This runtime has two primary components.

The first is a Large Model Inference Engine providing capabilities for natural language understanding, semantic analysis, and content generation. This component leverages the remarkable advances in language modelling over the past several years, enabling systems to process unstructured text at scale, to understand the meaning of documents rather than merely matching keywords, and to generate coherent prose expressing analytical findings.

The second component is a Declarative Multi-Agent System Architecture, which we have formalised under the name POMASA (Pattern-Oriented Multi-Agent System Architecture). Rather than writing procedural code that specifies exactly how tasks should be performed, POMASA enables the definition of agent blueprints – declarative specifications of what agents should accomplish, what resources they can access, and how they should interact. These blueprints are interpreted by the runtime, which instantiates and orchestrates the specified agents to accomplish complex research tasks.

The combination of language model capabilities with multi-agent orchestration enables a qualitative leap in what AI systems can accomplish. Individual language models, for all their power, operate in isolation: they process inputs and generate outputs but cannot coordinate with other systems, cannot maintain state across interactions, cannot implement complex workflows involving multiple steps and multiple specialised functions. Multi-agent architecture overcomes these limitations by decomposing complex tasks into subtasks handled by specialised agents, coordinating the flow of information between agents, and managing the overall process from initiation to completion.

3.2 The Multimodal RAG Lake ('Three Repositories')

The second architectural element concerns data storage and retrieval. AI4SS operates on a multimodal Retrieval-Augmented Generation (RAG) lake: a comprehensive data infrastructure that stores research materials in forms optimised for AI processing and retrieval.

This infrastructure comprises three distinct repositories. The Structured Data Repository contains cleaned, validated quantitative datasets: statistical tables, survey results, economic indicators, and other numerical information organised in formats amenable to computational analysis. The Unstructured Materials Repository stores original texts, images, audio recordings, video footage, and other primary materials in their native formats, indexed for semantic retrieval but preserving full fidelity to the originals. The Process Assets Repository accumulates the intermediate products of research – working notes, preliminary analyses, methodological templates, best practices – enabling institutional learning and knowledge transfer.

The RAG architecture ensures that AI-generated outputs remain grounded in actual evidence. When the system formulates a response, it does not rely solely on patterns learned during training (which is the source of hallucination in standard language models); it retrieves relevant materials from the repositories and grounds its generation in this retrieved context. The result is outputs that are not only fluent but verifiable – claims that can be traced back to specific sources in the underlying knowledge base.

3.3 The Technical Capability Ecosystem ('Four Domains')

The third architectural element organises the diverse technical capabilities required for AI-assisted research into four functional domains, each corresponding to a distinct phase of the research process.

3.3.1 The Information Perception Domain

The Information Perception Domain handles the ingestion and preliminary processing of data. Capabilities include: web scraping and API access to open sources; integration of commercial databases and paywalled materials; multi-format document parsing (PDF, Word, Excel, PowerPoint); image understanding through multimodal models; audio and video transcription; knowledge graph construction from unstructured text; and data archiving with classification and access control. This domain corresponds to the 'sensing' function of the research system – its interface with the external world of information.

3.3.2 The Cognitive Reasoning Domain

The Cognitive Reasoning Domain performs deep analytical processing. Capabilities include: full-text indexing at multiple granularities (sentence, paragraph, document, corpus); deep document comprehension that grasps meaning and argument structure; research planning and pathway determination; multi-dimensional data exploration across different levels of abstraction; dynamic generation of specialised agents for specific tasks; statistical analysis and hypothesis testing; econometric methods including panel data analysis, instrumental variables, and difference-in-differences; and probabilistic reasoning for pattern recognition, semantic induction, and causal inference. This domain corresponds to the 'thinking' function – the analytical core of the research system.

3.3.3 The Expression Rendering Domain

The Expression Rendering Domain handles the generation and presentation of outputs. Capabilities include: report generation with logical structuring; data visualisation through charts, dashboards, and infographics; natural-language-driven visualisation generation – researchers describe what they need in a single sentence and receive a finished chart; content management with version control; multi-format typesetting for academic papers, policy briefs, presentations, articles, and videos; multi-channel distribution adaptation; data lineage tracking and citation management; and audit logging for compliance and academic integrity. This domain corresponds to the 'speaking' function – the system's interface with human audiences.

3.3.4 The Collaborative Governance Domain

The Collaborative Governance Domain manages coordination and quality control. Capabilities include: team collaboration and real-time synchronisation; Git-style version control for data and documents; domain knowledge encapsulation with API exposure; cross-system protocol adaptation; citation format management (Chicago, APA, and other standards); automated fact-checking; bias detection and flagging; and workflow engines for review and publication processes. This domain corresponds to the 'regulating' function – the system's mechanisms for ensuring quality, consistency, and accountability.

Together, these four domains provide a comprehensive capability architecture that supports the full lifecycle of research from data acquisition through analysis to publication and dissemination.

4. From Theory to Praxis: Applications of AI4SS

A framework is only as good as its applications. The concepts and architectures described above would remain merely theoretical were they not embodied in functioning systems that produce actual research outputs. Global South Insights has developed over forty applications within the AI4SS framework, spanning domains from political economy to agricultural research, from regional and country studies to crisis analysis. These applications run on the GSI Unified Data Platform – according to GSI internal documentation, a comprehensive data infrastructure integrating 96 datasets from institutions including the United Nations, the World Bank, and the International Monetary Fund, encompassing 41,100 indicators and 3.45 billion rows of data. This section presents four representative examples that demonstrate the viability and utility of the framework.

4.1 Meta-RAG: The Knowledge Infrastructure

Meta-RAG is the flagship implementation of AI4SS's multimodal RAG lake concept. It is not merely a tool but an infrastructure: a comprehensive system for transforming implicit research methodologies into configurable software workflows. We describe this approach as 'Research Methodology as Configuration' – the idea that the procedures researchers follow can be explicitly specified, systematically implemented, and reliably repeated.

The system comprises over thirty atomic agents, including Evaluators, Orchestrators, and Validators, that can be combined into custom workflows. Users can design research processes incorporating planning, execution, and evaluation loops with human oversight at specified checkpoints. The underlying data architecture uses PostgreSQL vector databases to store multimodal knowledge bases, each containing tens of thousands of vector embeddings that enable semantic retrieval across text, images, and other materials. For any research topic, a researcher can build a specialised knowledge base containing thousands of high-quality documents in approximately one week at a cost of roughly $200 – work that would traditionally require months of literature accumulation.

Meta-RAG has been deployed to support research in political economy, agricultural technology, and Marxist studies, among other fields. A representative case demonstrates its capabilities: an African national development strategy project used Meta-RAG's 'New Structural Economics Lens' workflow to produce structured policy recommendations. The system autonomously integrated macroeconomic frameworks with sector-specific industrial policies, generating a comprehensive multi-chapter report with standardised structure. Total processing time was twenty-seven minutes; total cost in computational resources was $2.91.

Such efficiency gains are significant, but they are not the primary point. What matters is that the research process remains transparent, verifiable, and reproducible. Every output can be traced to its sources; every analytical step can be examined and, if necessary, contested. The speed and economy of AI-assisted research do not come at the cost of scholarly rigour.

4.2 The Digital Sovereignty Index: Assessing Global South Autonomy

The Digital Sovereignty Index (DSI) project applies AI4SS to a substantive research question of direct relevance to the Global South: the degree to which nations control their own digital destinies. Drawing on a theoretical framework developed by our team, the DSI evaluates countries across four dimensions – data ownership, digital infrastructure, cyberspace governance, and digital capability – to produce comprehensive assessments of digital sovereignty.

The technical implementation uses the POMASA architecture with six specialised agents coordinated by an orchestrator. Agent blueprints total approximately 21,000 words; reference materials (including scoring rubrics, decision matrices, and background documentation) total approximately 188,000 words. The evaluation methodology specifies over 5,400 individual assessment criteria organised under sixteen indicators, with each indicator accompanied by multi-level scoring standards.

To date, the system has completed assessments of all eleven BRICS nations. For each country, the research process collects over 700 pieces of primary evidence, which are then deduplicated and quality-checked to yield 400–500 high-quality evidence items (a deduplication rate of 35–42%). The final output for each country is a structured assessment report of 11,000–13,000 words. Total processing time per country is approximately six hours.

External validation has been encouraging. Pablo Martinez of the National University of General Sarmiento (UNGS) in Argentina, responding to the DSI assessment of Argentina in 2024, observed: 'Really insightful work. The capability paradox and the 2.0/5.0 score are critical findings for us.' Such feedback suggests that AI-assisted research can produce outputs that specialists find genuinely useful – not merely plausible but substantively valuable.

4.3 The Regional and Country Studies System

The Regional and Country Studies System represents the most ambitious implementation of AI4SS to date: a comprehensive infrastructure for end-to-end academic research production. It is not an assistant that helps researchers write; it is a production system that can autonomously generate publication-ready research outputs from specified inputs and parameters.

The scale of the system reflects its ambition. It incorporates 317 declarative agents organised into 64 pipeline orchestrations. It connects to six MCP servers (providing a total of 123 tools) with access to paywalled journals, optical character recognition for scanned documents, and various analytical services. The quality assurance architecture includes 294 'Quality Gates' – automated checkpoints distributed across workflow layers including literature extension, research observation, and deep analysis – that verify outputs against specified standards and flag potential hallucinations.

The efficiency gains are dramatic. Research projects that would traditionally require six to twelve months of full-time work can be completed in 40–60 hours of system runtime – an efficiency improvement of 100–200 times. Final outputs are substantial: reports of 40–80 pages (60,000–100,000 Chinese characters or equivalent), incorporating over 3,000 precisely formatted citations in Chicago or APA style.

The system has been used to produce country studies covering topics such as social governance, capital accumulation, and international dependency relations. These are not superficial overviews but deep analytical treatments that engage seriously with primary sources and secondary literature, that construct arguments and evaluate evidence, that meet the standards expected of professional research institutions.

4.4 Rapid Exploration: From Question to Insight Without Gatekeepers

The final example is not a major research programme but an impromptu exploration that illustrates a different dimension of AI4SS's potential: the capacity for rapid, self-directed inquiry.

An Indian industry expert approached us with a question: what are the strategic contradictions in US semiconductor export controls, and what opportunities might they create for the Global South? In the traditional knowledge economy, such a question would initiate a familiar pattern of dependency: the expert would need to find someone with relevant expertise, request access to materials, wait for a literature review to be compiled, and ultimately depend on others to provide the analysis. Knowledge flows through gatekeepers; those without access to the right networks remain excluded.

Instead, we simply fed his question directly into POMASA's generator tool. No additional information was provided beyond his original problem framing. Within thirty minutes, the system had created a complete multi-agent research apparatus – seven specialised agents implementing fourteen architectural patterns, configured to explore precisely the question he had posed. The system then ran autonomously for approximately two hours, curating 106,000 words of source material, verifying evidence, conducting analysis, and producing a twelve-chapter report of 7,600 words with 122 references.

The critical realisation is this: he could have done this himself. The entire process required no privileged access, no special expertise, no gatekeeper's permission. POMASA is an open-source framework; the intelligent runtime environments that execute it are commercially available; the methodological patterns are documented and learnable. A researcher anywhere in the world, armed with a question and basic familiarity with the tools, can conduct the same kind of rapid exploration.

This represents a profound democratisation of exploratory research. Traditionally, the capacity to quickly survey a field, identify key debates, and generate preliminary findings has been concentrated in well-resourced institutions with extensive libraries, research assistants, and established networks. Researchers in the periphery have had to wait – for access, for translations, for someone else's synthesis. AI4SS collapses this dependency. The cost of intellectual experimentation drops to near zero. A hypothesis can be probed in hours rather than months. A question can be explored before committing to a full research programme.

This agility does not replace rigorous, sustained inquiry; it enables it by allowing researchers to identify which questions are worth deeper attention. And crucially, it shifts the locus of control. The researcher with the question becomes the researcher with the answer – or at least with a substantive first approximation that they themselves have produced.

5. Implications for the Global South

The applications described above demonstrate that AI4SS is technically viable and practically productive. But what does this mean for the broader project of knowledge production in and for the Global South? This section considers the implications of AI-assisted social science research for scholars, institutions, and movements in the periphery and semi-periphery of the world system.

5.1 Epistemic Sovereignty

The most fundamental implication concerns what we have called epistemic sovereignty: the capacity of Global South researchers to determine their own research agendas, develop their own analytical frameworks, and control the infrastructures through which knowledge is produced and disseminated.

The current global knowledge economy is characterised by profound asymmetries. The major academic publishers are headquartered in the North; the dominant citation indices privilege Northern journals; the most powerful AI systems are developed by Northern corporations and trained on predominantly Northern data. Researchers in the Global South operate within this system but have limited influence over it. Their work is evaluated by Northern standards, published in Northern venues, and often framed in terms of Northern theoretical concerns.

AI4SS offers a path – not to autarky, which would be neither possible nor desirable, but to greater autonomy within interdependence. By building institutional knowledge bases from local materials, researchers create cognitive moats that cannot be simply appropriated or overridden. By developing analytical frameworks suited to Southern concerns – AI4SS's theory-agnostic architecture enables the same system to operate freely across Marxist, liberal, realist, and other frameworks, so that researchers are not held captive by ideological assumptions embedded in their tools – they ensure that AI systems serve Southern agendas rather than imposing Northern categories. By maintaining human control over research direction and interpretation, they preserve the space for intellectual traditions that differ from the globalised mainstream.

This is not a matter of rejecting Northern knowledge or technology. Much of what the North has produced is genuinely valuable and should be engaged seriously. But engagement should be critical and selective, not passive and wholesale. Epistemic sovereignty means the capacity to learn from others without being subordinated to them.

5.2 Capacity Multiplication

The second implication concerns research capacity. Global South institutions typically operate with fewer resources than their Northern counterparts: smaller budgets, heavier teaching loads, less administrative support, weaker infrastructure. These constraints limit what researchers can accomplish regardless of their intellectual abilities.

AI4SS offers a form of capacity multiplication that partially compensates for resource disparities. The mechanical labour that consumes so much researcher time – literature review, data processing, formatting, quality control – can be largely automated, freeing researchers to focus on activities that require human intelligence and cannot be delegated to machines. Small teams can produce outputs that previously would have required large institutions; individual researchers can tackle questions that previously would have demanded collaborative programmes.

This multiplication is not unlimited. AI systems require infrastructure (computing resources, stable electricity, reliable internet), expertise (to design, deploy, and maintain them), and judgement (to evaluate and verify their outputs). These requirements themselves impose constraints, particularly for the most under-resourced institutions. But within those constraints, significant gains are possible. The cases presented above already demonstrate that this promise is not hollow: $200 and one week to build a specialised knowledge base, six hours to complete a comprehensive country assessment, research projects that traditionally take six to twelve months completed in 40–60 hours.

5.3 Cross-Border Collaboration

The third implication concerns collaboration across the Global South. Southern researchers have long recognised the value of South-South cooperation: the exchange of experiences, the sharing of resources, the collective development of perspectives that differ from Northern hegemonic frameworks. But such cooperation has often been hampered by practical obstacles – distance, language barriers, incompatible systems, limited funding for travel and exchange.

AI4SS provides infrastructure that can facilitate South-South collaboration at scale. Standardised protocols enable researchers using different tools and speaking different languages to work together within integrated workflows. Shared knowledge repositories allow institutions to pool their resources, avoiding duplicative effort and building cumulative knowledge bases. The GSI Unified Data Platform already offers an example: according to GSI internal documentation, the platform's 96 datasets include 274 universal aggregation categories – BRICS nations, the African Union, groupings by income level, and more – enabling researchers from different countries to collaborate on a common data foundation. Modular architectures mean that capabilities developed in one context can be readily adapted for use in others.

The vision is one of distributed, cooperative capacity-building across the Global South. Rather than each institution developing its own tools in isolation, researchers can contribute to and draw upon a common infrastructure. Rather than competing for scarce resources, they can collaborate in mutual development. The result would be greater than the sum of its parts – a network of Southern research capacity that collectively matches or exceeds what any Northern institution can achieve alone.

5.4 Critical Vigilance

The final implication concerns the dangers as well as the opportunities of AI-assisted research. Artificial intelligence is not a neutral technology; it embeds the assumptions, biases, and priorities of those who develop and deploy it. Language models trained predominantly on English-language materials from Northern sources may reproduce Northern perspectives even when applied to Southern questions. Algorithmic systems optimised for efficiency may sacrifice nuance and context. The very power of AI tools creates risks of over-reliance – of substituting machine outputs for human judgement, of mistaking fluency for accuracy.

AI4SS responds to these dangers through the Human-in-the-Loop principle and the safeguard mechanisms that have already been engineered into practice – 294 Quality Gates, the architectural separation of content production from quality verification, end-to-end evidential traceability – these are not merely stated principles but the actual operating reality of the system. Yet mechanisms are only effective when actively wielded by critically vigilant researchers. AI systems should be treated as tools, not oracles – useful instruments that must be supervised, verified, and where necessary overridden.

Human judgement must remain paramount, especially on matters involving interpretation, valuation, and normative assessment.

Critical vigilance also means attention to the political economy of AI itself. Who benefits from the development and deployment of AI systems? Whose interests do they serve? How do they redistribute power and resources? These questions should inform not only how we use AI but how we participate in its development and governance. The Global South should not be merely a consumer of technologies developed elsewhere but a participant in shaping the technological future.

6. Conclusion: Towards a Decolonised Research Paradigm

This study has presented AI4SS as a comprehensive framework for artificial intelligence in social science research, developed by and for researchers in the Global South. We have argued that contemporary social science faces structural contradictions – the bandwidth paradox, the rigour gap, the disciplinary archipelago, the labour inversion – that limit what researchers can accomplish and disadvantage those working in under-resourced institutions. We have proposed an ontological reconceptualisation through the structure of 'One Foundation, Four Pillars, One Core' – Trustworthy Infrastructure as the foundation, Full-Data Perception, Full-Dimensional Analysis, Full-Channel Output, and Full-Disciplinary Synergy as the four pillars, and Human-in-the-Loop as the core – that together define a new paradigm for AI-assisted research. We have described a technical architecture that realises these concepts in practice and demonstrated viability through concrete applications.

The path forward requires simultaneous development across all three layers of the framework. At the ontological level, we must continue to refine our understanding of what AI makes possible and what it threatens, how it can augment human intelligence and where it must remain subordinate to human judgement. At the methodological level, we must build and improve the technical systems that translate concepts into capabilities, ensuring that these systems embody our principles and serve our purposes. At the practical level, we must apply these systems to research questions that matter, producing knowledge that advances the interests of the peoples of the Global South.

This is not work that any single team can accomplish alone. AI4SS is offered not as a finished product but as a contribution to a collective endeavour. We invite researchers across the Global South to engage with this framework – to use it, critique it, adapt it, extend it. The tools we have developed are available; the architectures we have designed can be implemented elsewhere; the principles we have articulated can guide work in contexts quite different from our own. What we have begun, others can continue and improve.

The broader stakes are significant. The age of artificial intelligence is upon us, and its contours are still being shaped. The Global South need not be a passive recipient of technologies designed by and for the North; it can be an active participant in determining how AI develops and what purposes it serves. A decolonised research paradigm – one that harnesses technological power while preserving human agency, that learns from global knowledge while serving local needs, that builds Southern capacity while fostering global solidarity – is within reach. The task is to grasp it.

We conclude with an invitation and a challenge. To researchers in the Global South: engage with AI not as a threat to be feared or a fashion to be followed but as a capability to be mastered and directed. To institutions: invest in the infrastructure and expertise that AI-assisted research requires, and build the collaborative networks that multiply its impact. To movements: insist that technology serve the people, that the benefits of AI be broadly shared, that the age of artificial intelligence become an age of genuine human flourishing.

The future is not determined. It is made – by choices, by struggles, by collective action. AI4SS is one contribution to making a future in which the peoples of the Global South are not objects of research but its subjects, not consumers of knowledge but its producers, not followers in technological change but its pioneers. That future will not arrive on its own. It must be built. The building begins now.