Evidence Synthesis Infrastructure Collaborative (ESIC) planning process:
Summary Report - WG3 Safe & Responsible use of AI

Global SDG Synthesis Coalition (GSDGSC)
Building a Global Evidence Synthesis Community (BGESC)
Pan-African Collective for Evidence (PACE)
Center for Rapid Evidence Synthesis (ACRES)

Advancing safe and responsible use of artificial intelligence

Artificial Intelligence (AI) is rapidly reshaping the landscape of evidence synthesis. From semi-automated screening to large language models (LLMs) generating outputs, new AI tools have the potential to enable faster and more scalable synthesis. Yet this transformation is outpacing governance, raising concerns about trust, bias, tool performance, transparency, and global equity. Working Group 3 (WG3) proposes a governance-led approach to ensure that AI is used safely, responsibly, and in service of a more inclusive, effective and efficient evidence ecosystem.

The goal is not to slow innovation, but to guide it, enabling the confident adoption of AI tools across regions and contexts, while safeguarding against exclusion, misuse, and low-quality outputs. WG3's solutions are built around transparency, co-governance, and structured guidance, reflecting a global commitment to equity, trust, and the public good.

The landscape: who needs artificial intelligence and why?

Evidence producers, intermediaries, and decision-makers are increasingly relying on AI to manage the growing volumes of research and accelerate the delivery of timely syntheses. AI tools promise significant efficiency gains, especially in under-resourced contexts where human capacity is limited. Early adoption is increasing in both Global South and Global North, particularly for rapid reviews and horizon scanning. However, uptake remains uneven, and trust in AI remains low. Key actors, including ministries of health, regional evidence networks, and humanitarian agencies, are calling for clear standards and guidance to determine whether, when and how to adopt AI-driven methods in their workflows.

Capability gaps and maturity: where are we now?

While innovation in AI for synthesis is accelerating, infrastructure and governance systems lag behind. Few tools are validated, and even fewer provide transparency about training data, performance, or any embedded ethical safeguards. There is no global standard for evaluating AI-enabled synthesis, and no mechanism for developers to exercise accountability to users or affected communities.

Capability disparities are stark. Users in the Global South often lack access to the necessary infrastructure, licensing, or institutional support to adopt AI responsibly. This creates a growing risk that without protective measures AI will widen, rather than narrow, global equity gaps.

Key issues: what's holding us back?

Several issues inhibit the safe and responsible use of AI in evidence synthesis. These include a lack of transparency in proprietary tools, weak validation mechanisms, and the absence of governance frameworks that reflect the interests of all stakeholders. Many current AI systems embed biases that reinforce existing power asymmetries, particularly between the Global North and the Global South. Trust is further eroded by inconsistent disclosure of AI use, poor-quality AI outputs, and limited opportunities for end-users to provide feedback to tool developers. Without shared guidance, users face uncertainty about when AI can be relied upon and when it should be avoided entirely.

Solutions for progress: what can we do next?

WG3 identifies seven solutions to ensure the safe, responsible, and equitable use of AI in evidence synthesis. These solutions are interdependent, through phased implementation with an emphasis on inclusive governance, transparency, and capacity-building.

The ESIC planning process is funded by the Wellcome Trust.

3.1 AI-assisted software for all stages of the evidence synthesis process:
Develop a modular, AI-enabled platform that allows users to assemble and customise synthesis workflows. The Evidence Synthesis Studio (ESS) will include support for mixed-methods synthesis, multilingual outputs, and explainable AI with human-in-the-loop validation.

3.2 Inventory of AI tools for evidence synthesis (CESPIA):
Establish and maintain a live, validated repository of AI digital evidence synthesis tools (DEST) with peer-reviewed benchmarks. The Comprehensive-Evidence Synthesis Plug-in Architecture (CESPIA) will ensure transparency, usability, and ethical alignment through continuous community-driven oversight.

3.3 Federated repository of living evidence data:
Build a centralized archive to standardise and systematically store structured synthesis outputs with unique identifiers and secure metadata. Enable interoperability across platforms to improve transparency, traceability, and reproducibility. Aligns with federated repository of synthesis data (WG2 2.1).

3.4 Crowdsourcing training platform to support training and adoption of AI models:
Deliver inclusive training and mentorship programs on AI in synthesis, with multilingual and accessible formats. Address digital divides and build global capacity for ethical AI engagement.

3.5 Framework for validation of technology performance:
Co-create a validation framework to define performance benchmarks, validation protocols, and model transparency standards. It will integrate citizen input and interdisciplinary review, with documentation standards such as “model cards” and “data statements” to ensure reproducibility and social accountability.

3.6 Implementation of best practices and governance of synthesis technologies:
Consolidate ethical governance frameworks into a participatory framework that ensures accountability, fairness, and transparency in AI-DEST development. Define clear roles, data protection measures, and oversight mechanisms that are rooted in community and citizen engagement.

3.7 Research into error assessment and reliability of AI-assisted synthesis:
Suitable topics might include determining acceptable tool error thresholds, bias propagation, and optimising user interfaces across AI-enabled evidence synthesis. This will inform the ESS (3.1) and CESPIA (3.2) by producing real-world performance insights and methodological improvements, ensuring system reliability and relevance for policy use.

Outcomes: what is likely to change?

WG3’s proposed solutions will accelerate, standardise, and improve the legitimacy of AI use in evidence synthesis. The ESS will reduce review time, increase flexibility, enable transparent replication of reviews, and support a range of synthesis types through modular, customisable workflows. CESPIA and the DEST validation framework will provide a trusted global inventory of validated AI tools, embedding community-led governance and ethical benchmarks. The ES Data Store will enable structured data sharing, reproducibility, and seamless integration, making AI-generated outputs more accessible, auditable, and contextually relevant.

Expanded training and capacity-building infrastructure will address digital divides and enable more equitable participation in AI-DEST development, particularly in the Global South. Ethical guidance and ongoing meta-research will enhance accountability, minimise misuse, and facilitate continuous learning about tool performance and user needs. Together, these solutions will foster trust, reduce risk, and ensure that AI tools support safe, inclusive, impactful syntheses.

The ESIC planning process is funded by the Wellcome Trust.