



Report: AI and Knowledge Management in Evaluation and Evidence Synthesis

Monday 24 – Wednesday 26 March 2025

In association with
Global Affairs Canada,
IOD Parc, Foreign,
Commonwealth &
Development Office



In association with



Foreign, Commonwealth
& Development Office



Global Affairs
Canada

Affaires mondiales
Canada

Report: AI and Knowledge Management in Evaluation and Evidence Synthesis

Monday 24 – Wednesday 26 March 2025

In association with

Global Affairs Canada, IOD Parc, Foreign, Commonwealth and Development Office

Executive summary

Effective use of existing global evaluation evidence is central to the adoption of successful foreign and development policies. Artificial intelligence (AI) technologies have the potential to significantly strengthen our ability to learn from existing evidence, distil key messages, and enhance evidence uptake. Meanwhile, combining AI with robust Knowledge Management (KM) practices and high-quality evidence synthesis can help decision makers and programme leads understand what works and what does not, enabling them to take an evidence-based approach to designing and implementing policy and programmes.

From 24-26 March 2025, experts and practitioners from around the world, including the global south, gathered at Wilton Park to discuss strategies for harnessing the potential of AI and KM in evaluation and evidence synthesis. Participants from a wide range of technical and policy perspectives shared their experiences, forged cross-sector partnerships, and committed to practical actions designed to shape the future use of AI and KM in advancing impactful and sustainable evidence-based interventions¹.

Throughout the discussions, several recurrent themes emerged. These included:

- The opportunities and challenges inherent in using AI for evaluation and evidence synthesis

¹ A glossary was issued to participants ahead of the event to ensure that discussions were based upon a common understanding of key terms and concepts. This glossary also informs the language used within this report, and can be found [here](#).

- The potential for AI to address or entrench existing inequalities across geographies and thematic areas
- The development of principles and standards around the use of AI
- The importance of human supervision and quality assurance
- The value of robust KM structures
- The role of AI and KM in supporting and enhancing evidence uptake
- The need for sustainability and appropriate funding models
- The benefits of greater cross-organisational collaboration

Participants agreed a series of action points and commitments to drive forward the use of AI and KM in evaluation and recognised the need to maintain momentum gathered during this conference via discussions at key upcoming events.

Contents

Executive summary	3
Context and opportunities	6
Knowledge management and its intersection with AI	7
Using AI to address data inequalities	12
Funding and sustainability	13
Co-operation and collaboration	14
Next steps	19

Context and opportunities

Recent decades have seen a significant uptick in the amount of high-quality evaluation evidence in the foreign policy and global development sphere. However, current efforts to synthesise this evidence are inconsistent and often take place in isolation, while the use of evidence to inform decision making remains low. AI technologies underpinned by effective KM approaches have the potential to facilitate a step-change by improving efficiency, enhancing the quality of evidence synthesis, and ensuring that decision makers can find, understand and use evidence in an effective and timely manner.

“How do we re-think development?”

Recognising the moment. In the context of reduced global development budgets, the use of AI and KM in evaluation and evidence synthesis – together with increasing cross-organisational co-operation – was viewed as essential to maximising the impact and value for money of interventions. Participants also recognised the critical function of AI and KM in guarding against evidence loss in times of geopolitical uncertainty.

Improving efficiency. Incorporating AI and KM into evaluation and evidence synthesis has the potential to lead to large-scale efficiency and productivity gains. Utilised effectively, complimentary AI tools and robust KM practices can bring together vast amounts of information quickly and cheaply, allowing researchers to plug gaps in the existing evidence synthesis landscape, answer emerging research questions responsively, and maximise the prospect of evidence being used by decision makers.

Doing things differently. AI was situated as having the capacity to engender a shift in the way development is approached. New technologies offer the opportunity to leverage knowledge of what works to alleviate poverty and improve lives at both a global and a local level, placing data and insights at the heart of global development. At the same time, AI tools have the power to put agency into the hands of local communities – particularly in the global south – by facilitating the inclusion of a broader range of evidence in synthesis products to ensure local relevance, widen access to evidence, and enable local policymakers to make evidence-informed decisions.

Developing the agenda. Although AI tools in combination with effective KM practices hold significant potential, their use in this field remains in its early stages. Participants highlighted the value of bringing together experts from evaluation, KM, and AI alongside representatives from a range of bilateral and multilateral organisations to discuss the possibilities and challenges of incorporating AI and KM into evidence

synthesis. They also underscored the importance of ongoing collaboration across the sector – through networks including the [Evidence Synthesis Infrastructure Collaborative](#), the [Multi-Donor Learning Partnership](#), and the [Global SDG Synthesis Coalition](#) – to ensure the fair and effective application of AI as the technology continues to develop.

Knowledge management and its intersection with AI

Effective KM is crucial to extracting maximum value from existing evaluation evidence. It is also a necessary precursor to the effective use of AI for evidence synthesis, with the ability of researchers and evaluators to realise the full potential of AI dependent on the quality of underlying systems of KM. Although a range of barriers were recognised, participants highlighted the criticality of good KM practices and offered a series of suggestions for overcoming these barriers.

“We can be doing so much more with the evidence that’s already out there”

Defining ‘knowledge management’. For some researchers effective KM means getting information to evidence users and decision makers at the right time, while for others it means developing an organisational culture around evidence sharing and reuse. KM is also seen as a process for managing risk, preserving knowledge, and creating space for learning and connections. There was broad agreement that KM represents a wide-ranging concept that intersects with evaluation and evidence synthesis to create an environment conducive to the production, curation and use of high-quality evidence. However, participants acknowledged the need for further work to clarify and contextualise the term to allow for a productive dialogue on the central role of KM.

The role of AI in evidence mapping and synthesis. Evidence mapping represents a foundational step in the development of an effective KM system. It enables an understanding of where gaps exist in a current body of evidence and provides an indicator of which geographies or methodologies may be underrepresented in an evidence base. Such exercises also identify areas where evidence is being over-produced, allowing researchers to avoid duplicating efforts and target increasingly limited resources more effectively. AI was recognised as having the capacity to perform mapping tasks to a high standard, and its use here was seen as both low-risk and scalable. While evidence and gap maps are increasingly utilised by research organisations, AI tools were also viewed as having the potential to support the future creation of ‘living’ synthesis products which are automatically updated to take account of emerging research.

AI and KM in maximising evidence use. Robust KM systems can also enhance the capacity of AI to contribute to evidence uptake. When used in combination with high-quality KM platforms, AI tools can improve the quality and applicability of insights derived from evidence synthesis products by allowing researchers to enter tailored prompts and refine results to focus on specific geographic or thematic areas. AI tools which interact with reliable KM systems were viewed as enabling policymakers to navigate complex knowledge bases, identify relevant findings in a timely manner, and more easily incorporate evidence into the decision-making process.

Promoting good knowledge management. Participants offered a series of suggestions for developing good KM practices. Drawing up a blueprint detailing internal processes, principles, systems, and partnerships was advocated, with the aim of codifying how KM is approached and communicating this widely. The creation of an organisation-wide knowledge agenda was also offered as a step towards creating a positive KM culture, with each team summarising their needs and priorities and setting out the actions and resources required to achieve these. More broadly, developing a cross-organisational maturity model for effective KM was recommended, focusing on detailing what success in this area would look like, setting achievement targets, and designing metrics to track progress.

Key challenges of using AI for evaluation and evidence synthesis

Despite its significant potential, participants identified several barriers to the adoption of AI in evidence synthesis. These included reservations around the transparency and quality of outputs, a shortfall in appropriate tools and the skills required to use them, concerns around data security and privacy, fears that AI may compound existing geographic and methodological biases, and concerns around the social and environmental impact of AI. More broadly participants recognised that the fundamental conceptualisation and framing of AI remains in flux, with a greater understanding required of its appropriate role and limitations. The production of a ‘dystopian’ theory of change to map out the full range of potential negative outcomes of using AI in evaluation and evidence synthesis was suggested, with participants identifying a series of actions designed to prevent the most severe of these from occurring.

“AI is not a silver bullet”

Transparency. The importance of ensuring that the outputs of AI are traceable for audiences of evaluation and synthesis products was emphasised. Participants warned that too often AI tools represent a ‘black box’, obscuring our understanding of their

mechanics, preventing replicability, and making effective human oversight challenging.

Developing principles around transparency was seen as critical to ensure that AI tools are developed and utilised in a responsible way with a focus on public good.

Quality. Although AI tools are capable of achieving a high degree of accuracy when undertaking specific tasks, the potential for AI to hallucinate and falsify information was seen as undermining trust in AI technologies to produce valid and reliable results. Additional concerns were expressed that AI software may miss important outliers during analysis, filtering out complexity to present an artificially clean summary of data.

Participants recommended setting out a clear definition of what is meant by ‘quality’ in the context of AI, creating robust measures against which AI tools can be assessed, and maintaining scientific rigor in AI-assisted synthesis via quality assurance and clear communication to users of uncertainties and limitations.

Tool applicability. Although open-source AI can perform a small number of discrete tasks to a high standard, the capability of off-the-shelf tools to undertake synthesis-specific functions is often limited. Meanwhile, the pace at which technology is developing means that it can often be unclear which tools have been rigorously tested. Although the development of in-house AI has the potential to result in more applicable tools this can often be resource-intensive, while working on tool development in silos risks significant duplication of efforts.

Participants issued a call for researchers to consider what collective action might be taken to shape the market in this area, and for organisations to work more closely together to make the process of developing bespoke AI tools more efficient.

Skills and capability. Uneven levels of awareness and engagement with existing AI tools have led to a high degree of variance in the skills and capabilities of researchers within and across organisations. Meanwhile, the pace at which the field is advancing has resulted in challenges not only in training researchers to use AI but in keeping their skills up to date. Further, placing AI tools in the hands of evidence users without equipping them with an understanding of their limitations risks amplifying unsound or incorrect information.

Ensuring the systematic rollout of AI tools and complementing this with a focus on training and capacity-building among both researchers and evidence users was seen as key to addressing these concerns.

Data security. Concerns around the use of AI to analyse sensitive data has led to researchers typically limiting its use to open-access documentation, while fears around intellectual property infringement have further narrowed the scope of information researchers are willing to expose to AI tools.

A drive to encourage governments and development organisations to broaden access to the evaluation data that they hold was recommended, while participants suggested funding future research only on the condition that data and results are publicly accessible, discoverable, and machine-readable.

Bias. Participants emphasised that any form of bias in the underlying data analysed by AI tools will be reproduced in their outputs, with the results of AI-assisted evaluation and evidence synthesis replicating – or even compounding – the uneven distribution of current evidence across geographies, themes, and methodologies. Meanwhile, as AI models are typically trained using data derived from randomised control trials largely emanating from health literature, such models may struggle to properly capture social science and evaluation publications that may rely more heavily on qualitative or mixed methods.

The importance of developing tools that can accurately interpret a broad range of evaluation literature was underscored, as was the

foundational importance of expanding the current evidence base to include evidence produced in the global south and addressing a wider span of thematic issues.

Resources and funding. The time and money required to design, set up, and roll out high-quality AI tools was flagged by participants as a key constraint, with processes for trialling and capacity building requiring additional resource.

Developing sustainable ways of financing innovation in this area was identified as crucial, while communicating the value of AI in evaluation and evidence synthesis to funders was positioned as central to securing future investment.

Environmental and societal impact. Broader issues were also raised about the wider negative impacts of AI technologies. These included fears over the intensive use of resources including water and energy, as well as concerns over the working conditions of those tasked with tagging and annotating large volumes of data. Meanwhile, reservations were aired over the prospect of governments and development organisations designing AI tools for evaluation and evidence synthesis alongside large technology companies who may have vested economic and political interests.

Options proposed for mitigating these concerns included prioritising smaller-scale AI models and ‘green AI’, while a suggestion was offered that these preferences might be enshrined in industry-wide guidelines.

Conceptualisation of AI. Participants also discussed the wider challenges involved in appropriately framing AI in the context of evaluation and evidence synthesis. While acknowledging the gravity of potential issues with hallucination, bias, and reliability, the considerable advantages of using AI in a targeted and pragmatic manner were emphasised. This was encapsulated by the proposition that AI tools should be thought of in a similar way to a human possessing significant aptitude in some areas but lacking skills in others.

Participants advocated for the adoption of an informed and nuanced view of the advantages and risks involved in AI utilisation, noting the importance of putting AI to good use while being alive to its shortcomings.

Using AI to address data inequalities

The development of effective AI tools for evidence synthesis offers an opportunity to broaden the evaluative evidence base, reduce inequalities, and facilitate more equitable access to data. However, if improperly managed, technological advances have the potential to enhance disparities in evaluation evidence between the global north and south, between information in English and in other languages, and between quantitative data derived from randomised control trials or quasi-experimental studies and more traditional or qualitative information. Participants discussed a range of strategies to ensure that the integration of AI in evaluation and evidence synthesis has a positive impact upon data equity.

"It's about creating meaningful and relevant insights from a diverse body of material"

Global north and south partnership and engagement. Participants highlighted a fundamental gap in understanding of AI capabilities outside of the global north. Although significant technical knowledge and skills exist outside of the global north there is limited awareness of and engagement with these. Increased collaboration with partners in the global south is required to develop paths to cooperation, investment, and interoperability across different contexts. Meanwhile, models built to suit the requirements of researchers in the global north are typically trained using data and documentation derived from traditionally dominant geographies. This leads to a marginalisation of evaluative data gathered elsewhere and a reduction in their wider accuracy and validity. Participants emphasised the importance of making use of a diversity of training data to minimise the risk of biased outputs, and of developing tools that are better able to integrate local and indigenous knowledge into synthesis platforms and repositories.

Language capabilities. The majority of AI tools typically perform poorly when reading and interpreting languages other than English, resulting in the over-representation of English-language data sources in their outputs. Although the challenges involved in addressing this issue are substantial, participants noted ongoing work aimed at

developing large language models (LLMs) capable of understanding languages other than English together with an increased focus on the production of minority language datasets which can be used to train AI models. The necessity to bring this work together under a common governance structure was emphasised, as was the importance of establishing minimum standards for training AI tools using non-English language data.

Machine readability. Due to their limitations in interpreting qualitative data and information written in languages other than English, current AI tools typically struggle to account for informal, indigenous, or community-generated knowledge that forms a key part of the evidence base in many geographic contexts – particularly in the global south. As a result, policy dialogues and exchanges that are often unique to specific settings risk being overlooked despite their centrality to decision making. Active methods for ensuring that AI models can understand and learn from such evidence were called for, with participants noting that evaluators in the north are currently missing out on important innovations emerging from these contexts.

Systems and incentives. Many AI models draw heavily on data published in academic journals. However, systems for publishing articles via these channels can marginalise evidence emerging from the global south, while leading journals often require readers to pay for access. This represents an important barrier to the ability of AI tools to reflect vital and innovative bodies of knowledge emerging from contexts outside of the global north, and creates inequalities in data access. Participants advocated for open-access publication as an effective route to ensuring greater equality in the data that AI models can draw on, while highlighting its role in widening access to evaluation data more broadly. Participants also suggested that current systems which incentivise researchers to publish data in paid-for journals should be revised, with researchers instead encouraged to pursue the more impactful and equitable dissemination of evidence via other means.

Funding and sustainability

Throughout discussions, participants underscored the importance of establishing stable and appropriate systems of funding to ensure the development of tailored and accessible AI tools that meet the needs of researchers and evaluators globally. Participants also emphasised the value of creating an institutional culture that enables organisations to encourage innovation and embrace new technology, allowing progress in this area to continue despite changes in leadership and shifting priorities.

"Let's make it very clear what our vision of success is, because we have the wind in our sails"

Promoting the value of AI. Although external interest in robust evaluation data and evidence synthesis runs in peaks and troughs, participants expressed a belief that demand for evidence and synthesis products and innovative tools to exploit these is currently high. To take advantage of this momentum and persuade funders of the importance of continued technological advancements, the value that AI tools and good KM approaches can bring to global development and foreign policy must be clearly demonstrated. Creating demand systems was positioned as central to achieving ongoing financial investment in the development of AI tools, with researchers and evaluators encouraged to agree a unified and positive narrative around the potential of AI and KM in evaluation and identify areas where successes can already be shown.

Developing appropriate funding pathways. While substantial funding commitments from UK Research and Innovation and Wellcome Trust were welcomed by participants, it was acknowledged that sustainable systems of funding are required to support continued technological developments in this area in the long-term. Participants highlighted the importance of funding mechanisms that are agile, responsive, and encourage innovation, and that offer support to researchers and evaluators in the global south to develop context-specific tools that fit their requirements.

A sustainable ecosystem. Participants acknowledged the fragility of institutional frameworks and the potential impact of political decisions upon organisational priorities; while individuals in positions of influence may be supportive of developing this agenda, momentum can be lost in the event of a change in personnel or a shift in the political environment. The importance of creating a sustainable ecosystem that incubates developments in AI and KM was underlined, with participants advocating for the distribution of responsibilities and capacities both within and across organisations and a focus on building resilience to protect the sector's investment in AI and evaluation synthesis tools.

Co-operation and collaboration

In any area of rapid growth, the risk of fragmentation and duplication of efforts is high. For the sector to maximise the potential advantages of incorporating AI into evaluation and evidence synthesis, participants acknowledged a duty to foster links between institutions, to pool resources, and to formalise collaboration to advance the responsible and efficient use of AI and KM.

"How can we collectively use the power of AI to do something that couldn't be done before?"

Collaboration between institutions. Although organisations are working to develop tools that suit their specific needs, much of this work is taking place in isolation. Participants called for the dismantling of institutional silos and the creation of an ecosystem that spans the organisational level to ensure that the work of individual institutions complements rather than duplicates the efforts of others. Suggestions for collaboration included sharing the results of experiments with existing off-the-shelf AI tools, discussing the strengths and weaknesses of different approaches to developing and using bespoke software, and uniting to build skills and capability across the sector.

Connecting evidence synthesis efforts. While the volume of evidence synthesis work being undertaken has increased significantly in recent years, a co-ordinated system linking these individual efforts together is lacking. As a result, there is currently a patchwork of evidence synthesis products and repositories that provide only a partial picture of the evidence base. Meanwhile, these products are often static, quickly becoming outdated upon the production of new evidence. Participants suggested pooling evidence that currently sits on institutional platforms, creating links between existing evidence repositories to form a ‘repository of repositories’, and working collaboratively towards the development of open and accessible living synthesis products.

Creating multi-disciplinary alliances. Participants also underlined the importance of collaborating with experts from other disciplines with alternative perspectives and complimentary skill sets. This included working with technical specialists to tailor emerging AI products to suit researcher needs, with governments to co-generate evidence and promote the use of AI in synthesis and dissemination, and with policymakers to ensure that their priorities are understood and catered for by emerging technologies. Several examples of successful collaboration were spotlighted including the [World Bank’s Impact AI initiative](#) and the work of the [Deep Learning Indaba](#), while participants suggested establishing regular hackathon-style events incorporating technical experts and policymakers and highlighted the importance of identifying key future partners.

Developing a common voice. Participants flagged the importance of jointly advocating for the use of AI technology, establishing agreed principles for its responsible use, and building capacity across the sector. Participants noted the lack of an agreed lexicon to discuss these issues, with collaboration seen as an important first step in developing shared definitions and holding productive conversations.

Emergent networks. Participants highlighted the existence of several networks designed to facilitate co-operation in this space. These include the [Evidence Synthesis Infrastructure Collaborative](#) (ESIC), a collective supported by the Wellcome Trust and designed to advance the use of AI and KM in evidence synthesis; OECD’s [DAC Network on Development Evaluation](#) (EvalNet), a network of organisations working to

strengthen global evaluation systems and capacities; [ALIVE](#), an alliance of organisations and individuals from across the evidence ecosystem building and testing a collaborative approach to living evidence partnerships; and the [MERL Tech initiative](#), a collaborative working with social sector organizations to achieve thoughtful tech-enabled program design and implementation and convening communities of practice to share learning and advance the responsible use of technological advancements in monitoring and evaluation. Participants were also encouraged to strengthen and expand [existing informal connections between individuals and organisations](#) working on AI in evaluation synthesis.

Principles and standards

With AI tools being developed and rolled out at pace, participants voiced concerns that the necessary focus on standards and rigor for utilising AI in evaluation and evidence synthesis may be lost. To safeguard against this, and to build and maintain trust in the use of AI, participants advocated for the creation of cross-organisational principles and guidelines and debated the validity of imposing standards for the responsible use of AI.

["There's a need for both principles and standards to engender trust in the use of AI"](#)

Fundamental principles. Participants agreed on a series of high-level principles including transparency and accountability, fairness and inclusivity, data protection and privacy, and validity and reliability. Participants also underlined the fundamental importance of agility and collaboration, while also advocating for the adoption of a human rights-focused approach to the use of AI which places dignity and rights at the centre of technological development.

Drawing on existing frameworks. Although a number of guidance and policy documents covering the use of AI in evaluation exist, many of these are institution-specific while awareness of those that are publicly available is low. Meanwhile, more established frameworks for evaluation and evidence synthesis have often not been adapted to account for the emergence of AI technologies. Participants suggested taking stock of pre-existing guidelines covering the use of AI as a starting point in developing sector-wide policies, and discussed the potential to adapt and modify wider guidance for evaluation and evidence synthesis to cover developments in this area.

Guidelines or standards? Participants distinguished between the development of broad guidelines and the imposition of more rigorous standards, discussing the relative

merits and disadvantages of each. Although standards can set clear boundaries and help the market tailor tools to the specific needs of the evaluation community, concerns were raised over their potential to dampen innovation and to quickly become outdated, while also requiring a high level of consensus to be considered legitimate. In contrast, more general guidelines were seen as both flexible and adaptable and capable of being developed with greater efficiency, while broad agreement on the principles that should underpin the use of AI in evaluation and evidence synthesis is already in existence.

Supporting different actors at different stages. While participants voiced a desire to work towards the production of comprehensive guidance covering the use of AI in evaluation and evidence synthesis, the importance of tailoring this to those operating at different points in the evidence ecosystem was flagged. Specific guidance designed to suit the needs of implementors, brokers, and evidence users, accompanied by supportive checklists and tools for their practical application, was seen as having the potential to allow those at different points in the research cycle to maximise the utility of AI. Meanwhile, the value of producing uniform guidance for suppliers bidding for, as well as commissioning, evaluative work was also highlighted. Many tenders are currently unclear on how AI might be used as part of project delivery, and a detailed set of guidelines may allow bidders to be more transparent about their planned use of AI.

The human role in AI technology

Participants highlighted the indispensable role of humans in ensuring the quality and reliability of output produced by AI and continually improving its effectiveness. The role of AI as complementing rather than replacing humans was also discussed, as was the enduring importance of human relationships and skillsets that cannot be replicated by technology.

“AI needs some guardrails, albeit at relatively high level, to make it designed by people and for people”

Human-in-the-loop. The challenges inherent in using AI for evaluation and evidence synthesis underscore the importance of retaining human oversight in its design and implementation. Although this is particularly crucial during the early stages of the incorporation of AI into the evidence ecosystem, robust systems of human supervision must be maintained as tools become more widely used across the sector. Participants emphasised the need to find a balance between achieving maximum efficiency and the

importance of critical oversight, with an agreed distinction between the role of humans and AI seen as a first step towards achieving this.

AI as supplementing human intelligence. Concerns over the impact of AI on job security for those working at different points in the evidence cycle were noted. However, participants firmly positioned AI as complementing rather than replacing human intelligence. Rather than being viewed as a threat to employment, participants emphasised the potential for AI to increase the number of evidence-related jobs by allowing people's skills to be enhanced through technological advancement; AI tools were seen as not only maximising the efficiency and impact of the work of researchers, but also as enabling those with differing degrees of skills and experience to understand and interact with complex evidence as part of their role.

The ongoing importance of human relationships. As AI tools continue to become more sophisticated, evidence suggests that their ability to respond to complex demands and mimic human interaction is increasing. However, participants emphasised the irreplaceable function of human relationships in producing evidence and maximising its uptake. The ability of experienced evaluators to interpret and communicate evidence pragmatically was flagged as a key example of the ongoing value of human relationships in the evidence cycle, as was the capacity of evidence brokers to react to rapidly changing political contexts and maintain relationships with decisionmakers in complex and nuanced environments.

AI and KM driving evidence uptake

A key aim of developing high-quality AI tools and effective KM approaches is to increase evidence uptake. Technological advances can contribute to this by generating more accessible and timely insights, and by freeing up resources that can be channelled into dissemination activities. While the potential of AI in this area is significant, participants emphasised the continued importance of advocating for evidence use and building close relationships with decision makers.

"We're doing all of this because we want evidence to be used"

Freeing up resources. The appropriate use of tailored AI tools combined with effective KM has the potential to allow for significant efficiency gains throughout the research cycle. Participants highlighted that the time and money saved by employing AI technology effectively can be diverted to encouraging evidence use, with researcher capacity freed up to focus on disseminating findings more widely and building relationships with decision makers.

Supporting synthesis impact. Technological advances offer opportunities for researchers to synthesise evidence more efficiently. However, these technologies can also be harnessed to allow policymakers to understand and interact with the product of these efforts. Participants flagged the potential for AI tools to close the loop between evidence production and evidence uptake by increasing policymakers' access to digestible information on what works.

Producing timely insight. Policy decisions often need to be made quickly, responding to rapidly changing situations in complex environments. There is usually limited time to access and consider all the available evidence. The ability to produce timely insights offers the opportunity for researchers and evidence brokers to help decision makers take an evidence-informed position, and robust KM offers the opportunity to base responsive insights on a more comprehensive and inclusive evidence base. AI tools were also seen as allowing researchers to anticipate the questions that evidence users might ask in advance, enabling the production of relevant insights ahead of time.

Shaping the demand side. The advantages of co-generating evidence alongside governments and policymakers were highlighted, with participants demonstrating that programme impact can be increased by several orders of magnitude when decision makers are involved in evidence production from the start. Participants positioned AI as enabling the development of processes and frameworks to help those on the demand side establish and communicate their requirements to evidence producers, and allowing researchers to provide robust insights that are tailored to those needs. AI and KM were also seen as having the potential to lead to behaviour change among both researchers and evidence users, enabling researchers to adopt a relationship-centric approach and focus on building trust and credibility and allowing policymakers to act as champions for evidence use.

Understanding evidence users. Despite technological advances in evidence synthesis, the challenge of ensuring that evidence is used effectively by policymakers remains significant. Participants agreed that practical steps should be taken to understand the needs of the end-user of synthesis products and AI tools, and how incentives can be formulated to ensure policymakers engage constructively with evidence.

Next steps

Participants agreed on a series of actions to drive progress. These focused on the development of a diverse and inclusive evidence base, the production of principles and minimum standards, the role of good knowledge management practices, and the importance of fostering connection, collaboration, and communication.

Diverse and inclusive evidence base. Curating a more coordinated global system of evidence synthesis that incorporates different methodological approaches, local and indigenous knowledge, and increased evidence from the global south.

Principles and minimum standards. Mapping and raising awareness of existing guidance, and developing common principles and standards for AI-enabled evidence synthesis that are based on consensus and acknowledge the pace of change.

Knowledge management practices. Developing a knowledge management framework for development and foreign policy organisations supported by a maturity model tool and a common taxonomy, and engaging with librarian and archivist communities.

Connection, collaboration, and communication. Widening access to existing data repositories and creating a ‘repository of repositories’ to enhance opportunities for evidence synthesis, working together to align funding opportunities, facilitating peer learning, and developing a common voice.

Maintaining momentum. Arranging regular follow-up opportunities to track progress and solidify new relationships, increasing awareness of the role of AI in evaluation synthesis among evidence users and policymakers, expanding networks, and disseminating key outcomes at a series of breakthrough events throughout 2025.

Ian Montagu

Wilton Park | March 2025

Wilton Park reports are brief summaries of the main points and conclusions of a conference. The reports reflect rapporteurs’ personal interpretations of the proceedings. As such they do not constitute any institutional policy of Wilton Park nor do they necessarily represent the views of the rapporteur. Wilton Park reports and any recommendations contained therein are for participants and are not a statement of policy for Wilton Park, the Foreign, Commonwealth and Development Office (FCDO) or His Majesty’s Government.

Should you wish to read other Wilton Park reports, or participate in upcoming Wilton Park events, please consult our website www.wiltonpark.org.uk.

To receive our monthly bulletin and latest updates, please subscribe to www.wiltonpark.org.uk/newsletter

Wilton Park is a discreet think-space designed for experts and policy-makers to engage in genuine dialogue with a network of diverse voices, in order to address the most pressing challenges of our time.

enquiries@wiltonpark.org.uk

Switchboard: +44 (0)1903 815020

Wilton Park, Wiston House, Steyning,
West Sussex, BN44 3DZ, United Kingdom

wiltonpark.org.uk

