Irving Fisher Committee on
Central Bank Statistics

BIS

Generative artificial intelligence in central banking

Douglas Araujo, Giuseppe Bruno, Adam Cap, Juri Marcucci, Rafael Schmidt, Olivier Sirello, Bruno Tissot[1]

Executive summary

Central banks are increasingly exploring and using *generative artificial intelligence* (generative AI)[2] to support their activities, such as economic and monetary analysis, statistical production, financial supervision and payment oversight. *Large language models* (LLMs) in particular – a class of generative AI that can analyse and produce text by mimicking human intelligence – have unlocked unprecedented opportunities to deal with textual data, of which central banks are both heavy users and producers.

Continuous and rapid innovations are further expanding the capabilities and quality of generative AI's outputs. First, methods such as *retrieval-augmented generation (RAG)* can improve the accuracy and reliability of LLM-produced content, for example by drawing on verified sources and specialised knowledge. Second, a growing area of interest is the use of *small language models* (SLMs), which are more limited and cheaper versions of LLMs tailored to specific tasks. Third, and more generally, *agentic AI* systems can enhance task automation and assist decision-making in business processes, for example by orchestrating complex workflows.

With the proliferation of new data solutions and competing work priorities, **one question for central banks is how to make the most of these promising yet often embryonic AI advances in an efficient, effective, ethical and safe way**. Experience so far has highlighted the importance of adequately managing innovation, not least to balance its benefits with the associated risks. Three distinct areas of focus have emerged:

- The first relates to the design and implementation of **proper governance frameworks** for managing the associated risks and harnessing AI effectively and responsibly. Central banks have already taken significant steps in this endeavour,

[1] Respectively, Research Economist, Central Bank of Brazil (douglas.araujo@bcb.gov.br); Director, Economics and Statistics Directorate, Bank of Italy (giuseppe.bruno@bancaditalia.it); Senior Macroeconomic Analyst, Monetary and Economic Department (MED), Bank for International Settlements (BIS) (adam.cap@bis.org); Economist, Economics and Statistics Directorate, Bank of Italy (juri.marcucci@bancaditalia.it); Head of IT, MED, BIS (rafael.schmidt@bis.org); Senior Statistical Analyst, MED, BIS (olivier.sirello@bis.org); Head of BIS Statistics, MED, BIS, and Head of the IFC Secretariat (bruno.tissot@bis.org).

The views expressed here are those of the authors and do not necessarily reflect those of the BIS, the IFC, the Bank of Italy or any of those institutions mentioned.

The authors thank Michele Bullock, Michael Machuene Manamela, Gloria Peña, Luís Teles Dias and colleagues from the Reserve Bank of Australia, Banco de Portugal and South African Reserve Bank for helpful comments and suggestions. They are also grateful to Alison Arnot for editorial review and Márcia Cavalinhos for publication support.

[2] Expressions formatted in *italics* are further defined in the Glossary in the annex.

leveraging their long-standing expertise in data issues. Yet AI governance initiatives are still evolving, often reflecting the need for new skills and organisational strategies to support organisation-wide innovation.

- The second area of focus is on technical aspects, reflecting the need to **secure adequate IT resources as well as to adopt interoperable data processes and systems** to further advance the use of generative AI both within and across organisations. Open software tools and well-established information standards such as the Statistical Data and Metadata eXchange (SDMX) initiative can play a decisive role in this regard. They can ensure that data are used properly by AI and, more broadly, preserve the quality of the reference information produced by authoritative sources, such as central banks.
- Finally, promising yet often fast-paced technological advancements serve as a good reminder of the **importance of cooperation and knowledge exchange across central banks as well as with the other stakeholders involved in the data ecosystem**. Such collaborative initiatives can yield significant benefits, including sharing best practices and raising awareness of the challenges posed by AI. They also help to optimise the use of resources in the central banking community, by facilitating opportunities for co-investment in areas such as data-sharing techniques or open source software, to make data AI ready.

1. Introduction

Central banks are increasingly using generative AI to support their various activities (BIS (2025), IFC (2025a)). AI is commonly understood as computer systems capable of tasks that normally require human intelligence, while its recent generative form (also known as generative AI) produces new information content based on patterns learned from existing data.[3]

The versatility of most *generative AI* applications makes them suitable for a broad range of central banking activities that rely on extensive data usage. These typically span not only economic and monetary analysis and forecasting, supervision and payment oversight but also financial education and consumer protection. In particular, central banks have been active adopters of LLMs – a prominent class of generative AI models able to generate human-like text (IFC (2022, 2023), Araujo et al (2023)). Generative AI can also assist central banks in their role as data producers, by collecting, processing and disseminating reference statistics.

Yet, in practice, **the use of AI is not without challenges, especially related to data quality and IT infrastructure**. Regarding the first aspect, the main concern lies in preserving the high quality, credibility and integrity of central banks' data, while ensuring their accuracy, explainability and interpretability as well as addressing broader privacy, security and ethical considerations (Dilip et al (2026)). Turning to IT aspects, the infrastructure needed typically demands secure, scalable and high-performing solutions, often entailing additional costs and complex trade-offs – for example related to the use of cloud-based services, information protection and power

[3] See the Glossary for a list of more exhaustive definitions.

capabilities. Fortunately, a number of solutions are being continually developed and refined to address those challenges, ranging from models requiring less power to privacy-preserving methods. But keeping pace with these various evolutions can be cumbersome, adding uncertainty and volatility that may detract from the long-term foresight often characterising central bank activities.

To balance such opportunities and challenges, central banks have been advancing a number of strategies. First, the development of agile, resilient and clear governance frameworks has proved instrumental for mitigating the operational and ethical risks associated with generative AI (IFC (2021, 2025a, 2025b)). Second, central banks have been further upskilling their workforce towards AI-savvy profiles, while also capitalising on the broad expertise of their staff based on interdisciplinary background, scientific rigour and attention paid to data quality. Third, the increasing yet often still experimental use of generative AI has underscored the benefits of international cooperation, notably for exchanging best practices, raising awareness of the cross-cutting issues posed by AI and developing shared solutions.

Against this backdrop, the IFC, along with the Bank of Italy, has been organising recurrent data science workshops to review developments in the big data ecosystem and the ongoing adoption of innovative data techniques. The fourth edition in 2025 emphasised the **exploration of generative AI and its potential applications in central banking**. The overview of this *IFC Bulletin* provides a bird's eye view of the latest developments related to generative AI (section 2) and discusses related applications of interest to central banking (section 3). The various projects already being advanced also suggest a number of ways to make the most of the opportunities offered by this technology (section 4), especially regarding governance aspects, IT resources, information standards and data quality, as well as international cooperation.

2. Generative AI: a bird's eye view of recent developments

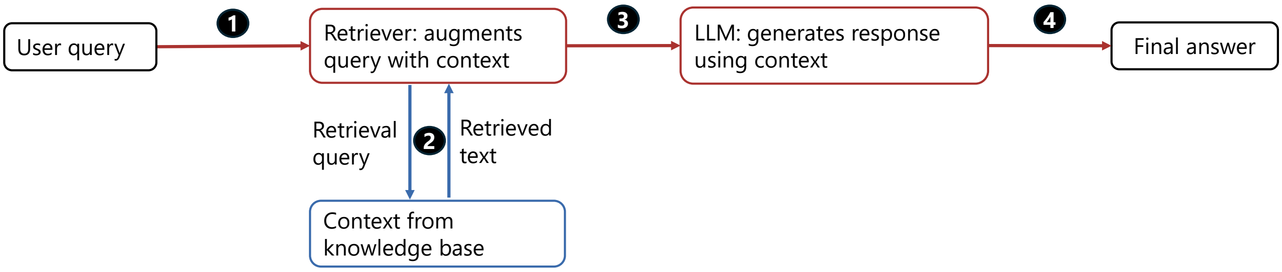
Recent and continuous technological advances in generative AI have opened new opportunities that are particularly relevant for central banks. First, RAG systems have emerged as a possible solution to meet the growing demand for tailoring LLMs' outputs to organisation-specific knowledge. Second, the development of *agentic AI* allows for flexible IT programming, supporting task automation solutions based on natural language capabilities instead of "traditional", hard-coded instructions. Third, SLMs effectively reduce the computational burden that is typical of large AI models, while safeguarding the relevance of their outputs.

Retrieval-augmented generation (RAG) systems

RAG systems can enhance the quality and reliability of LLMs' outputs by integrating selected specialised documents forming the input to these models. This is in contrast with general purpose models, which typically answer user questions after first being trained on publicly available knowledge. In practice, LLMs are able to base their answers on specific and/or internal documents selected during the initial setup phase, forming the "*knowledge base*" of the RAG (Graph 1).[4] As a result, when a user poses a question, the system will provide information derived exclusively from the customised knowledge base.

RAG simplified architecture model

Graph 1



Source: Authors' elaboration.

This approach has at least four advantages. First, it can significantly improve the quality, reliability and relevance of AI-generated answers by restricting the knowledge base, for instance to only verified and authentic sources. This makes RAGs well suited for use cases that require dealing with specialised and possibly restricted information – for example, financial supervisors may be interested in accessing data based on non-public bank assessments.[5] Second, RAGs can help control the quality of input information and thereby reduce the risk of false but seemingly true answers ("hallucinations"). This is a significant limitation of general purpose LLMs given the risk of producing faulty output because of incorrect or poor-quality input (the "garbage in garbage out" phenomenon). Third, they allow for more traceability, as input sources can be explicitly referenced and authenticated. Finally, RAG-based approaches can be simple and effective to deploy, although they require adequate IT infrastructure to operate. This advantage is especially notable in comparison to other techniques that similarly try to control the quality of models' outputs, for example by fine-tuning their input or customising their specifications – often entailing complex requirements and a large amount of computing resources.

Agentic AI

AI agents are language model (LM)-based autonomous and specialised software systems designed to automate tasks involving complex reasoning and decision-making. They can dynamically interact with data, users and external systems to perform various tasks such as workload scheduling, answering user queries and filtering emails.[6] While a key advantage of RAGs is allowing for improved accuracy

4 See Lewis et al (2020) for a more exhaustive description of RAG systems.

5 Specialised information may also encompass vast collections of archival records, which central banks often curate. RAG systems may be useful to effectively leverage such resources, as recently shown by the Reserve Bank of Australia (Bullock (2025)).

6 While AI agents typically have a moderate and task-specific scope, "agentic AI" is characterised by greater autonomy in handling complex use cases. Its development has been further supported by the so-called Model Context Protocol (MCP), an open source universal interface to help AI agents better use information resources and software tools (see Box A).

and factual reliability in models' outputs by supplying relevant context at query time, they react to specific user queries and are limited to generating informed responses. AI agents, by comparison, can orchestrate processes and tools, execute various tasks and, in certain cases, feature learning capabilities that are essential for supporting decision-making. These extend beyond – at least to date – the simpler operations performed by most generative AI models that focus on transforming input information.

AI agents can be particularly useful in the context of central banking. For example, in the area of statistical compilation, one agent in charge of quality checks can help automate the detection of potentially erroneous values based on past patterns and quality checks. Meanwhile, another AI agent can estimate missing values, for instance based on historical trends and previous observations. Finally, a third AI agent may be used to check the accuracy of these estimates, supporting human validation tasks as needed.

Despite its advantages, experience suggests that **agentic AI requires properly designed workflows as well as human oversight** to be fully effective. On the one hand, it appears to work well with simple and modular workflows, for instance to optimise the use of various LMs' capabilities (Anthropic (2024)). On the other hand, rigorous human oversight, systematic testing, adequate validation and training as well as monitoring of agentic systems often prove critical to obtain high-quality outputs. Well-governed data and AI management processes therefore appear particularly important, as they play a decisive role in ensuring that security, privacy, accountability and broader ethical concerns are met.

Small language models (SLMs)

SLMs are language models that are significantly smaller than LLMs, trading off size for specialisation. Their limited number of parameters makes them faster, more efficient and less computationally intensive than larger models. As a result, SLMs are more conveniently fine-tuned to specific tasks. Their specific focus can help provide more relevant responses, while still having the ability to reply in natural language similar to larger LLMs.[7]

SLMs can offer notable advantages in the context of central banking, for at least two reasons. First, because of their relatively limited size, they can be fine-tuned locally, for example using on-premises IT infrastructure. This is particularly important for mitigating risks related to third-party dependency or confidentiality breaches. Second, SLMs excel in various specialised or detailed tasks that are essential components of many traditional workflows in central banks, such as in the areas of statistical compilation or micro-level supervision. Moreover, SLMs can be combined with LLMs in multi-agent frameworks, where SLMs focus on domain-specific tasks, while LLMs handle broader, general purpose requests.

7 SLMs are trained through various techniques, such as "pruning" – ie where the architecture and size of the original LLM are simplified and reduced – and "knowledge distillation", where knowledge from a larger, more powerful "teacher" LLM is transferred into the smaller model. This latter technique has been used, for instance, to train the model DeepSeek-R1 (DeepSeek-AI et al (2025)), achieving comparable performance to OpenAI models but at a fraction of their size and cost.

3. Generative AI applications in central banking: main use cases

Generative AI has emerged as a versatile tool for supporting central banks across various tasks. Reported experience suggests that most use cases relate to four main areas: (i) economic and monetary analysis; (ii) forecasting exercises; (iii) statistical production; and (iv) supervision and oversight activities.

Economic and monetary analysis

One of the most promising areas where generative AI can support central banks is in **economic and monetary analysis, which is a fundamental area supporting the conduct of their monetary and financial stability policies**. Concrete applications include text analysis, assessing economic news and sentiment and performing classification and summarisation tasks.

First, **regarding text analysis**, **LLMs are unlocking new opportunities for monitoring economic and financial developments** by leveraging a wider range of economic indicators (Aprigliano et al (2023), Marcucci (2024)). For example, they facilitate a more structured analysis of press sources to provide timely insights on financial market developments, as shown by the Bank of Korea. These models can also support analysis of the real economy, with arguably greater detail and agility than traditional methods (Dahlhaus et al (2025)). The Danmarks Nationalbank, for instance, uses LLMs to analyse job postings and track movements in employer demand for specific skills over time, offering a more nuanced view of the labour market. Similarly, the Bank of Japan and the Hokkaido University have deployed LLMs to monitor climate-related news, providing useful insights in terms of international developments, regulatory aspects and business practices. News and other similar text-based sources can also inform risk assessments that can be out of the scope of traditional monitoring exercises. A notable example relates to episodes of social unrest, as shown by the Federal Reserve Board, which leverages voluminous data from the Global Data of Events, Language and Tone (GDELT) project.[8]

Second, **generative AI can help users gain a better grasp of "soft" indicators, such as economic sentiment and expectations**. This can be especially relevant for central banks when analysing economic agents' behaviour, for instance, to assess the influence of their policy communication and monitor feedback dynamics. For example, the National Bank of Romania developed a language sentiment index for measuring the impact of financial news, using OpenAI's ChatGPT complemented with a model specialised in financial and monetary policy statements (Araci (2019), Evdokimova et al (2023)). Similarly, a joint study by the Bank of France and the University of Lille investigates financial news related to surprises and expectations close to monetary policy meetings, relating them to the volatility of foreign exchange rates. Expectations by economic agents are also – and very often – captured by surveys, which often produce large volumes of unstructured text that need to be adequately integrated into policy frameworks. Case in point, the Reserve Bank of

[8] The GDELT aims to ensure comprehensive coverage of global news media to monitor events in near real time through sophisticated algorithms; see Leetaru and Schrodt (2013) and gdeltproject.org.

Australia shows that generative AI can help process these data with greater efficiency and ease.

Third, **another important aspect when dealing with vast amounts of textual documents is to be able to extract meaningful information for policy purposes.** This calls for adequate classification and summarisation processes, something that generative AI excels at (Zhang et al (2025)). Regarding classification, the Deutsche Bundesbank has developed an internal tool based on LLMs to help users categorise documents easily. The approach, using a bootstrapping resampling technique, allows for greater agility and lower computational and analytical costs. Turning to the capacity of generative AI to summarise documents, work by the Bank of Italy illustrates the usefulness of LLMs. Yet this research also finds the value of AI-generated summaries relative to the ones written by humans may vary, depending on users' literacy. For instance, individuals with higher educational backgrounds and greater familiarity with the Economic Bulletin published by the central bank tend to prefer human-written summaries.

Forecasting exercises

A key task of central banks that can typically be supported by generative AI is **forecasting short- and medium-term economic developments, including through nowcasting techniques** to make predictions about the current dynamics of key economic indicators.[9] For example, research by the BIS shows that the *transformer* architecture that underpins LLMs can be effectively adapted to both nowcast and forecast variables such as GDP, inflation and unemployment. Relatedly, the BIS Innovation Hub's Project Neo uses AI to nowcast the Swiss economy, leveraging a wide palette of data sources, including on retail sales, railway transport and health insurance, that can be accessed more rapidly than official statistics. Alternative data can also help to predict more specific developments outside the economic domain, such as the incidence of influenza, as demonstrated by the Bank of Italy using web-based information.

In addition to allowing for more timely forecasts, LLMs can also improve their accuracy, by integrating "traditional" macroeconomic time series with unstructured information like text (IFC (2025b)). For instance, Bangko Sentral ng Pilipinas shows that indicators built from domestic news can improve the performance of inflation forecasts. This reflects in particular the value of the information content of central bank communication as conveyed by the press, as argued by Araujo et al (2024). Joint work by the Bank of Israel, University of Tokyo, Korea's KDI School and Sogang University finds that the analysis of monetary policy statements – collected in a dedicated database encompassing 51 jurisdictions from 1990 to 2024 – can help to better predict the financial cycle.

Yet, in practice, **the effectiveness of generative AI in forecasting will often depend on the specific use case and data available**. Traditional statistical approaches may sometimes be more suitable, especially when considering the costs and/or risks associated with the use of innovative techniques. Reflecting this point, the approach followed by Bank Indonesia – applying standard clustering methods

9 Nowcasting can also be helpful in predicting recent dynamics of economic indicators, especially when these are published with significant lags.

using models such as random forests to group borrowers – has proved well suited to capture external debt dynamics, especially in comparison with new methodologies.

Statistical production

Central banks are leveraging generative AI to support the production of statistics, a task that has become increasingly important over recent decades (Dilip et al (2026)).[10] A first key use is to simplify classification tasks, for instance by identifying products in price data,[11] mapping codes between different nomenclatures[12] and updating statistical classifications (UNECE (2023a)). A second strand of applications relates to data curation and quality assurance, including the optimisation of anomaly detection workflows. Metadata editing in particular appears to benefit considerably from LLMs' capability to generate and check text correctness, highlighting generative AI's potential to enhance existing editing processes (Sirello et al (2025)). A last and growing application of generative AI relates to the structured extraction and processing of unstructured data, such as text and images. For instance, generative AI allows for the compilation of experimental statistics on environmental, social and governance (ESG) indicators, derived from corporate reports and satellite imagery (IFC (2025c)).

Another important part of central banks' statistical work relates to dissemination, which can also be effectively supported by generative AI. One noteworthy application developed by the Central Bank of Malaysia is deploying RAG systems to improve users' access to internal and external statistical databases, including those from international organisations and the Malaysian statistical office. This tool also identifies and plots the relevant data, simplifying search and exploration. Another promising area is the development of LLM-based chatbots that facilitate data queries, such as the IMF's StatSChat and StatGPT projects (Farias (2025), Tebrake et al (2026)).

Statistical data retrieval can be further automated through the use of AI agents, as shown by the Bank of Thailand and the European Central Bank. Specifically, a chain of specialised LM–based software systems addresses the user's query all the way to the final response. For instance, a sample workflow related to the question "What was the inflation rate in Germany and Thailand last year?" first transforms the query into "Inflation rate in Germany and Thailand in 2024", then decomposes it into sub-queries (eg one for each country). A third agent would then construct specific requests[13] to retrieve the data from the original databases. The workflow could further include an agent to judge the quality of the response and another one to generate the answer, which is also subject to validation (Graph 2).

[10] For additional details on how generative AI can support statistical agencies, including central banks, across the various phases of the generic statistical business process model (GSBPM) and on the associated challenges faced in the related production workflows, see Piela (2024) and UNECE (2025a).

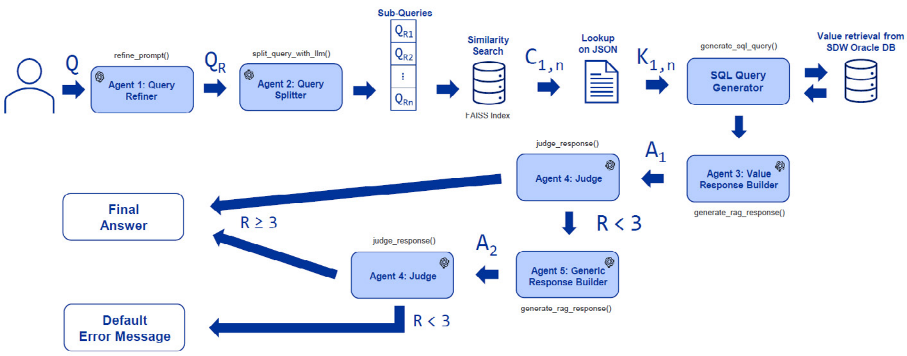
[11] For example, see Nunes and Palumbo (2025).

[12] For example mapping ISIC 4, ISIC 5 and NACE; see Pérez et al (2025). Another use case is mapping NACE 2008 to 2025; see Faria (2025).

[13] SQL, or Structured Query Language, is a widely used computer language for working with databases.

Using AI agents to facilitate access to the ECB Data Portal

Graph 2



Source: L Petracca, S De Benedictis, T Grotton and Z Hofmeister, "Supporting users in seeking data on the ECB Data Portal: a use case for retrieval augmented generation", *IFC Bulletin*, no 67, March 2026.

Supervisory and oversight activities

Central banks are often tasked, depending on national circumstances, with supervising, monitoring and overseeing payment and financial systems, including banking and insurance activities. **These tasks often require access to vast amounts of specialised internal knowledge and/or restricted documents, where generative AI can provide valuable support**. In particular, RAGs can be very useful in dealing with sensitive information that is by its nature excluded from publicly available LLMs trained on non-confidential data.

A key area relates to the **microprudential supervision of banks**, as LLMs can be particularly helpful for analysing the activities of and risks posed by financial institutions. For example, the Federal Reserve Board follows a RAG approach to efficiently extract information from the vast amount of bank documents available (eg annual reports, financial statements). Generative AI can also be useful to keep track of the often complex and frequent changes characterising financial regulation (Brookes et al (2022)). For example, the Bank of Italy and the Sant'Anna School of Advanced Studies have built a RAG using public information related to European banking regulation on liquidity risk. The system enables easier identification of relevant legal references and provides tailored contextual guidance – including through pertinent examples – to eventually assist users in gaining more accurate and meaningful insights into complex regulatory topics. Relatedly, this approach can also be effective for the **oversight of payment systems.** As a compelling use case, the Central Bank of Chile, has developed a fine-tuned LLM – specifically trained in Spanish – to track, interpret and summarise the legal documents associated with payments.

Overall, a key lesson of these various projects is that **human involvement is often required to verify the accuracy of AI-generated responses**, reflecting the importance of specialised knowledge and accountability in supervisory processes.[14]

[14] This verification is also commonly referred to as "human-in-the-loop" (HITL).

4. Unlocking generative AI's full potential in central banking: the way forward

Central banks' experience with generative AI thus far highlights that making the most of the opportunities offered by this new technology calls for: (i) developing adequate governance and securing sufficient resources; (ii) strengthening information standards and (meta)data curation; and (iii) sharing best practices especially through international collaboration.

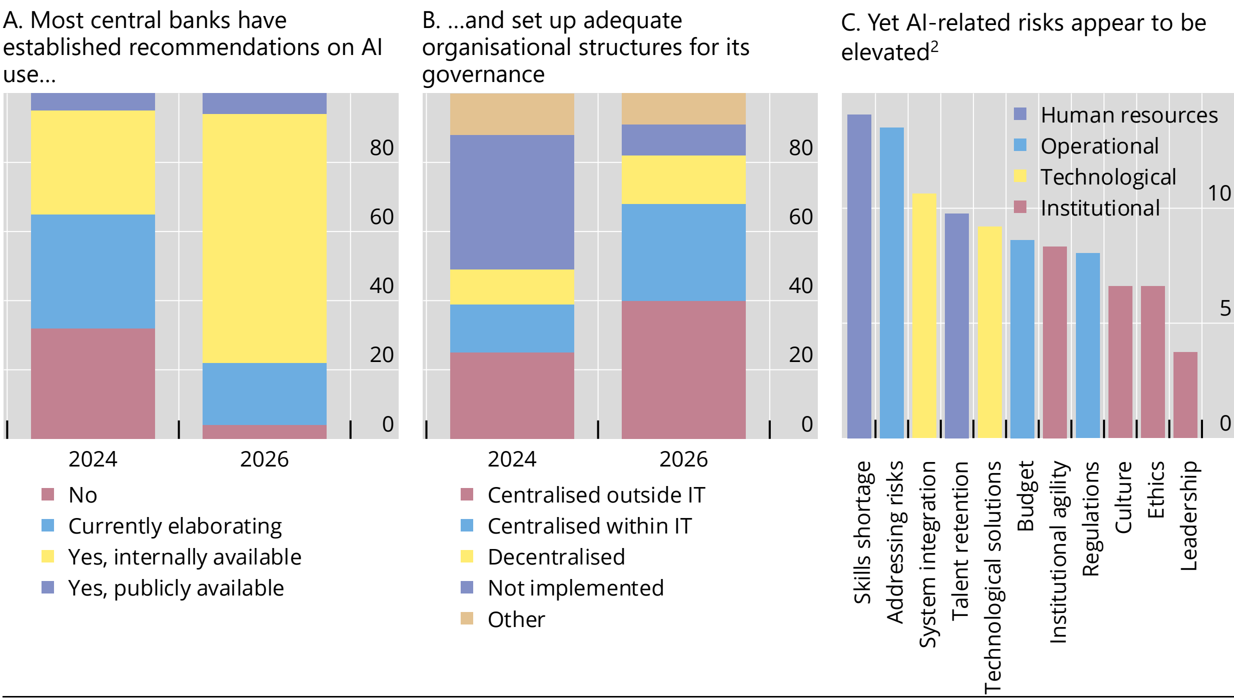
Governance and resources

The integration of generative AI into central banks' processes requires an adequate governance framework, notably to balance innovation with AI-associated risks, as well as to ensure the availability of adequate IT, human and financial resources.

Fortunately, and reflecting the strategic importance of this new technology, **most central banks have been anticipating additional resources for AI applications as well as establishing frameworks** to adequately use and govern their use within the organisation (Graph 3; Box A; IFC (2025a)). Yet significant barriers to adoption remain,

Central banks have made notable advancements in AI governance since 2024[1]

In per cent of respondents — Graph 3



[1] Results are based on the IFC survey on AI in central banks conducted in 2024 and early 2026. Figures for 2026 are preliminary at the time of publication.

Sources: IFC; authors' calculations.

Box A

AI governance in central banks: recent developments and possible ways forward

AI governance is a broad and complex system of decision rights and accountabilities encompassing the principles, responsibilities, structures and frameworks intended to enable effective and responsible use of AI technology.

Central banks have been implementing AI governance through four key components.① The first is their established expertise in data governance and management, to ensure the quality, security and integrity of the data used by AI systems (Križman and Tissot (2022)). Second, they have been developing guidelines and policies governing the use of AI, such as terms of use as well as explainability requirements. Third, central banks have been adapting or setting up dedicated structures to enable organisation-wide development of AI. Typical arrangements include centralised units (eg steering committees), federated or hub-and-spoke models (eg AI hubs) and decentralised approaches at the level of business areas. Last but not least, risk management frameworks are another important component, including through renewed information classification rules, restrictions on AI tools and audit trails.②

Looking ahead, insights from a recent survey conducted among IFC central banks suggest three areas of focus to further develop governance, with the objective of making the most of AI opportunities while minimising the associated risks.

- *Quality of data and reference information.* Given AI's reliance on data, one question is whether currently disseminated statistical information is sufficiently machine-readable to be accurately processed, interpreted and returned by AI systems. In practice, delivering guidance on what "AI-ready data" means may call for reviewing existing (statistical) data quality frameworks, at least along three lines. First, while generative AI relies on structured and unstructured data, such as text,③ existing frameworks might not properly address the quality requirements of these novel data types – including those related to ethical considerations.④ Second, AI-readiness strategies for statistical data may also require developing frameworks to responsibly use and deploy AI, for example addressing issues related to ethical purpose, accountability, environmental sustainability and robustness.⑤ Third, gaining a better understanding of AI systems, including their limitations and potential inaccuracies stemming from training data, may require promoting statistical thinking and evaluation.⑥
- *Interoperability of data processes and systems.* Statistical standards, such as SDMX, can act as AI enablers,⑦ for instance to enhance data discoverability and interpretability – putting a premium on having curated centralised metadata that are easily accessible by machines. Additionally, these standards can improve the quality of AI systems, especially when paired with a Model Context Protocol (MCP), to ensure that the outputs generated by AI are accurate and traceable (see Box B). Yet further work might be required to enhance interoperability among data systems and ensure AI systems can reliably source high-quality data through established standards.⑧
- *Securing the adequate skills and resources for AI deployment.* Ensuring resilient AI applications in central banks demands building adequate skills at the organisational level, to raise awareness among staff about AI's opportunities and risks, eg related to information protection. It also calls for vigilance to avoid excessive and uncritical reliance on AI solutions that might eventually erode human expertise – a potential phenomenon described as "skill atrophy".⑨ Turning to IT resources, the deployment of AI typically requires advanced hardware and puts a premium on having adequate infrastructures (including the use of cloud services) to allow for more secure and scalable deployment. But such investments need to consider the associated potential risks, eg in terms of costs, third-party dependency and (data) sovereignty issues.
- *Evaluating generative AI outputs against existing methods.* Although benchmarks for evaluating natural language systems are becoming increasingly available,⑩ frameworks for assessing systematically the quality of AI-generated outputs – particularly for statistical purposes – remain limited to date. Furthermore, the opaque nature of these models (often referred to as "black boxes") may undermine the ability to verify data accuracy and integrity. This lack of transparency can result in unchecked inaccuracies, with generative AI systems producing erroneous outputs that are challenging to validate through established statistical methods.⑪

① IFC (2025a). ② CGRM (2025). ③ UNECE (2025a). ④ Floridi and Taddeo (2016) and UNECE (2026). ⑤ Haddou et al (2025) and UNECE (2025b). ⑥ RSS (2026). ⑦ IFC (2025d) and Anvar (2025). ⑧ Araujo et al (2025). ⑨ OECD (2026). ⑩ Examples include the General Language Understanding Evaluation (GLUE) benchmark and the Stanford Question Answering Dataset (SQuAD). ⑪ UNECE (2025a).

most notably in securing skilled personnel and suitable IT infrastructure. Moreover, ethical and technological risks – especially related to security, privacy and dependency on external providers – are reported to stay elevated (Graph 3.C). This puts a premium on developing suitable risk management frameworks, both at the level of the organisation – eg to avoid overreliance on a few external technological companies, including by adopting open source solutions – and at the level of the financial system – eg to prevent phenomena such as content homogenisation, potentially leading to correlated decisions and amplifying risks of procyclicality, contagion and systemic financial stress (FSB (2024), Zhang (2026)).

Turning to more operational aspects, implementing generative AI in the organisation calls for **adequate computing resources**. Specifically, high-performance hardware such as graphics processing units (GPUs) are in demand for running generative AI models, as seen in the Bank of Canada's use of AI to solve complex economic models. Yet, these IT resources – particularly when deployed on premises – may entail significant investments, relating to both their purchase and maintenance. One option, increasingly being explored by central banks, is cloud-based services which may feature lower adoption costs but also present a number of important trade-offs, especially in terms of data protection and sovereignty (IFC (2025a)).

Lastly, making the most of generative AI opportunities also demands both recruiting specialised staff that are proficient in data science techniques and **advancing general AI literacy across the organisation**. One solution has been to foster interdisciplinary approaches, including by attracting adequate staff profiles combining IT, statistical and data expertise (Araujo et al (2023), OECD (2026)). Another, perhaps more practical way is to favour more staff diversity within units, allowing for the combination of dedicated AI skills with other, more subject-oriented competencies. One interesting experiment from this perspective is the Innovation Laboratory (iLab) set up by the Bank of Latvia, with **cross-functional AI working groups** mixing various profiles across the organisation.

The need for adequate documentation and data integrity

As data are the bedrock of generative AI, their rigorous documentation (ie data about the data, or "metadata") and their integrity (ie the assurance that the information is kept accurate, authentic, consistent, reliable and traceable along its entire life cycle, including through robust information standards) are **key to ensure the adequate interpretation and reliability of the outputs generated by the new techniques**.

A way to improve the accuracy and sustainability of generative AI applications is to **focus on the quality of their input data. This will help to address the so-called "garbage in, garbage out" phenomenon**, where AI outputs are polluted by low-quality inputs (Bogani et al (2022)). Hence, an important action point is to ensure the provision of rich and standardised documentation about the information provided, such as details about the sources, compilation methodology and other quality assessments (Sirello et al (2025)).[15] From this perspective, the experience of the

[15] Other noteworthy quality aspects are replicability of statistical processes and the reproducibility of outputs, for example through so-called reproducible analytical pipelines (RAPs). These are even more

European Central Bank has underlined the relevance of building a common metadata architecture in the organisation as a way to integrate and standardise documentation across various data sets. Another option is the *Model Context Protocol* (MCP), which is emerging as an open solution for connecting AI models to external systems, including specialised tools or databases. A relevant use case, increasingly explored by statistical authorities, is to develop such MCP-based solutions to retrieve data from well-documented dissemination portals, with important benefits in terms of data traceability, accuracy and interpretability (see Box B).

Box B

Improving generative AI's accuracy with MCP: benefits and challenges

The Model Context Protocol (MCP) is a technical solution enabling LLMs to interact with external tools, such as software and data sets.① This approach brings various advantages. First, it can improve LLMs' accuracy by allowing models to interact with customised resources, such as specific databases or in-house software tools. This can be a key benefit for central banks, whose specialised workflows often rely on proprietary data sets and tools. It also helps increase the accuracy and reliability of AI models by reducing the risks of misuse and inaccurate inputs. Second, MCP works as a universal and standardised protocol (like a "USB-C port") for AI resources, eliminating the need for ad hoc solutions, which can be difficult to set up and maintain. Third, MCP can access resources both over the internet and locally. This is particularly important for central banks seeking to secure the confidentiality of their internal information and limit access rights.

Yet MCP solutions also feature some important challenges. First, their adoption can be complex, making them more suited for sophisticated business cases. Additionally, there is a risk of introducing specific cyber vulnerabilities, such as unintentionally revealing restricted data. Third, this approach is still relatively new, suggesting that major changes and compatibility issues could arise. Fourth, while MCP has experienced fast-paced adoption,② its long-term viability as a "universal" standard remains to be seen and will ultimately depend on whether it will be supported by a large share of AI providers.

Despite these challenges, experience suggests that MCP can benefit central banks in performing various tasks. For one, MCP can tap into central banks' internal information without compromising its overall security or leading to (risky) third-party dependencies. Perhaps more fundamentally, MCP can be effective in ensuring that AI applications source authoritative and trustworthy data, for instance through the use of the SDMX standard.③ This can contribute to having more accurate AI-generated outputs and also ensure that the data produced by central banks are accurately sourced by AI, reducing the risk of misuse and misinformation.

① MCP was introduced by the company Anthropic in November 2024; see Emilia (2024). ② For example, a number of AI players such as OpenAI, Microsoft and Google have been adopting MCP; see openai.com. Available MCP servers can be found in so-called directories; one popular directory is pulsemcp.com/servers. ③ Case in point, MCP solutions have become available for a number of national statistical offices, such as the National Statistics Office under the Ministry of Statistics and Programme Implementation (MoSPI) in India, to enable AI assistance to query their data through application programming interfaces (APIs); see pib.gov.in.

Another, complementary objective is to advance, promote and implement **interoperable information standards**. One practical way is to leverage well-established initiatives **to make data "*AI-ready*"**, that is, to ensure their accurate consumption by LLMs (Dossé (2025)).[16] A number of specialised standards, semantic models and common vocabularies, including the Data Documentation Initiative (DDI),

compelling given the probabilistic nature of AI/ML modelling techniques and the related need to quantify uncertainty, particularly to address issues such as model overconfidence (Dumpert (2025)).

16 See the recommendations on making data "findable, accessible, interoperable, reusable and AI-ready" (or "FAIR^2"; Hill (2025)).

Dublin Core Metadata Initiative (DCMI) and the Data Catalog Vocabulary (DCAT),[17] already support both machine readability and actionability.[18] Another key standard actively used in the central bank community is the Statistical Data and Metadata eXchange (SDMX), which provides a rich information model that standardises both data and metadata through harmonised and unambiguous statistical concepts. This can be particularly effective to facilitate LLMs accessing accurate documentation, as shown by the BIS. One important feature from this perspective is that the SDMX standard can deal with various types of data, ranging from "traditional" macro time series to micro-level records and geospatial information (IFC (2025d)). In addition, beyond its usefulness as a tool for data and metadata management, SDMX can effectively support both data validation (for example the experience of the Bank of Italy) and data search (for instance through the sdmx.io initiative led by the BIS).[19]

A final focus point is to **implement active and continuous curation processes for creating, organising and improving metadata**. This is essential to enhance the authenticity, reproducibility and traceability of information, in turn ensuring adequate and transparent use by generative AI. First, the use of unique identifiers – including persistent ones such as the Digital Object Identifier (DOI) – can play an important role in identifying, locating and naming data both over time and across data sets. A related focus area is the development of attribution requirements and their related conventions, such as standardised (open) licences, disclaimers and terms of use (UNSC (2025, 2026)). Furthermore, innovative techniques such as blockchain technology can be leveraged to ensure data integrity and authenticity, particularly when handling information from official sources (Lambe et al (2025)). Finally, generative engine optimisation (GEO) tools have become of increasing interest for ensuring that official data sources are clearly and adequately referenced in AI-generated outputs provided by third-party providers.

Sharing and collaboration

Central banks' experience shows that **running collaborative projects can be a very cost-effective way to deploy or fine-tune generative AI techniques**. Yet this also requires adequate cooperation mechanisms, to allow for the needed collection and exchange of data while also securing their protection (IFC (2025d)).

One strand of options relates to leveraging existing *privacy-enhancing and privacy-preserving technologies* (OECD (2025a), Kim et al (2025)). These can be particularly useful in ensuring that AI models – which often need to be trained on vast information sets – make use of all the available data sources, while appropriately complying with privacy, copyright and data protection laws (OECD (2025b)). For

17 DDI is an international standard for describing surveys, questionnaires, statistical data files and social sciences study-level information. It encompasses standards that support the entire life cycle of data (DDI-Lifecycle) and cross-domain integration (DDI-CDI). The DCMI supports innovation in metadata design, best practices and standards for resource description. Finally, DCAT – developed by Eurostat – enables semantic interoperability, for example by supporting standardised data set descriptions.

18 For example, the World Bank has publicly released a metadata editor supporting the documentation of a wide range of structured data types, mapping them to established metadata standards and schemas (World Bank (2025)).

19 The sdmx.io website contains an ecosystem of resources and tools – including those geared towards data scientists.

instance, pseudonymisation methods can be combined with the provision of fine-grained access rights depending on the user profile and purpose, as shown by Banco de Portugal. Another possibility is using data encryption, a technique which can be resource-intensive but offers greater security. A specific method is homomorphic encryption combined with multiparty secure private computing, which allows computations to be performed on restricted data in a decentralised way and without directly exchanging them (Ricciato (2024)). For instance, a project developed at the Bank of Italy allows for implementing secure multiparty linear regressions on data owned by different institutions, each with different sets of variables for the same individuals. Finally, LLMs can assist in improving statistical disclosure control by analysing sensitive textual content and automatically determining whether the content should be flagged as confidential (Rigaud (2025)).

Another important avenue increasingly adopted by central banks is the generation of synthetic data (IFC (2025e)). This technique replaces confidential information with "seemingly true" data that preserve the statistical properties of the original set. In this way, it allows users to perform analyses without accessing sensitive raw data (UNECE (2023b)). For instance, the Bank of Canada has used a synthetic version of its payment system data to train generative AI agents and analyse intraday liquidity patterns. However, not all synthetic data can adequately reproduce the characteristics of the original data set, as shown by the experience of the Central Bank of Türkiye and Istanbul Technical University. Consistent with this observation, the Central Bank of Malaysia has developed an evaluation framework to assess the quality of synthetic data. This framework measures statistical distance, structural similarity, performance in machine learning tasks and the risk of disclosing sensitive information to produce a composite weighted score for evaluating synthetic data.

Finally, collaboration between central banks and other stakeholders involved in the data ecosystem can play a decisive role in advancing generative AI applications. For example, the BIS Innovation Hub Ellipse Data and Knowledge Platform Community (EDKP) provides a platform for central banks and financial regulators to share source code, pool expertise and jointly develop AI-based supervisory technology solutions supporting financial supervision. Collaboration can also involve actors beyond the central banking/supervisory community, such as statistical offices, international organisations, academia and the private sector, for example by sharing knowledge, software and co-investment opportunities (IFC (2025a)).

Glossary

Artificial intelligence (AI): computer systems capable of tasks that normally require human intelligence (FSB (2017)), for instance by inferring how to generate outputs from the inputs they receive (OECD (2024)). It includes applications such as machine learning (ML) and generative AI.

AI agents: systems designed to act autonomously, making decisions and taking actions based on their programming and objectives. These systems are capable of performing tasks without constant human intervention.

AI-ready data: data that are suitable for consumption by AI systems and models, such as LLMs. Key dimensions for readiness include quality, interoperability, accessibility and compliance (both ethical and legal).

Generative AI: a subset of AI that produces new content, such as text, images, audio or video, based on patterns learned from existing data (or "training data"), usually in response to a user's prompt.

Knowledge base: a repository of information that provides domain-specific or general data to AI models. Knowledge bases can include various data types, including textual documents, which can be used to provide organisation-specific context through techniques such as RAG.

Large language models (LLMs): AI models trained on vast amounts of text data with the goal of generating human-like language. These models are designed to perform a wide range of natural language processing tasks, including summarising, translating, answering questions and interacting with users.

Model Context Protocol (MCP): an open framework and standard that enables the integration of AI systems with selected external software and data sources.

Privacy-enhancing and -preserving technologies: technologies to minimise risks of privacy infringement, such as anonymisation, encryption, differential privacy, homomorphic encryption and secure multiparty computation.

Retrieval-augmented generation system (RAG): a technique aimed at improving the quality of generative AI responses by integrating relevant context from pre-defined knowledge sources to each user prompt, offering reliable and well-defined sources for the generation of responses.

Small language models (SLMs): AI models similar to LLMs, but with significantly fewer parameters that are trained on less data than LLMs. While their smaller size reduces their capabilities compared with LLMs, SLMs are less computationally intensive and can often serve more specialised roles in specific domains.

References

Anthropic (2024): "Building effective agents", *Engineering at Anthropic,* 19 December.

Anvar, E (2025): "Why SDMX matters? A community journey towards SDMX as an AI-enabler and a data mesh enabler", *IFC Bulletin*, no 64, May.

Aprigliano, V, S Emiliozzi, C Guaitoli, A Luciani, J Marcucci and L Monteforte (2023): "The power of text-based indicators in forecasting Italian economic activity", *International Journal of Forecasting*, vol 39, no 2.

Araci, D (2019): "FinBERT: financial sentiment analysis with pre-trained language models", *arXiv*, no 1908.10063.

Araujo, D, N Bokan, F A Comazzi and M Lenza (2024): "Word2Prices: embedding central bank communication for inflation prediction", *CEPR Discussion Papers*, no 19784, December.

Araujo, D, G Bruno, J Marcucci, R Schmidt and B Tissot (2023): "Data science in central banking: applications and tools", *IFC Bulletin*, no 59, October.

Araujo, D, A Cap, I Mattei, R Schmidt, O Sirello and B Tissot (2025): "Data science in central banking: unlocking the potential of data", *IFC Bulletin*, no 64, May.

Bank for International Settlements (BIS) (2025): *The use of artificial intelligence for policy purposes*, report submitted to the G20 Finance Ministers and Central Bank Governors, October.

Bogani, R, A Theodorou, L Arnaboldi and R Wortham (2022): "Garbage in, toxic data out: a proposal for ethical artificial intelligence sustainability impact statements", *AI and Ethics*, vol 3, October.

Brookes, J, M Everitt and Q-A Vo (2022): "The Basel regulatory framework: evolution of its textual complexity", *Bank Underground*, 24 August.

Bullock, M (2025): "Technology and the future of central banking at the RBA", speech at the 60th Shann Memorial Lecture, Perth, 3 September.

Consultative Group on Risk Management (CGRM) of the BIS Representative Office for the Americas (2025): *Governance of AI adoption in central banks*, January.

Dahlhaus, T, R Ellwanger, G Galassi and P-Y Yanni (2025): "From online job postings to economic insights: a machine learning approach to structuring naturally occurring data", *American Economic Association Papers and Proceedings*, vol 115, May.

DeepSeek-AI et al (2025): "DeepSeek-R1: incentivizing reasoning capability in LLMs via reinforcement learning", *arXiv*, no 2501.12948.

Dilip, A, Z Mdingi, O Sirello, B Tissot (2026): "The evolving role of central bank statistics", *IFC Bulletin*, no 66, February.

Dossé, J (2025): "Is SDMX data AI-ready? Is AI SDMX-ready?", presentation at the SDMX Global Conference, Rome, September.

Dumpert, F (ed) (2025): *Foundations and advances of machine learning in official statistics*, Springer.

Emilia, D (2024): "Anthropic releases Model Context Protocol to standardize AI-data integration", *VentureBeat*, 26 November.

Evdokimova, T, P Nagy Mohácsi, O Ponomarenko and E Ribakova (2023): "Central banks and policy communication: how emerging markets have outperformed the Fed and ECB", *Peterson Institute for International Economics Working Papers*, no 23-10, October.

Faria, T (2025): "NACE Rév. 2.1 update: retraining an ML model in production using LLM", UNECE generative AI and official statistics workshop, Geneva, 12–14 May.

Farias, A (2025): "Talk2Manuals: your genAI-powered guide to statistical standards", UNECE generative AI and official statistics workshop, Geneva, 12–14 May.

Financial Stability Board (FSB) (2017): *Artificial intelligence and machine learning in financial services – market developments and financial stability implications*, November.

——— (2024): *The financial stability implications of artificial intelligence*, November.

Floridi, L and M Taddeo (2016): "What is data ethics?", *Philosophical transactions of the Royal Society*, vol 374, no 2083, December.

Haddou, M, T Seimandi and A Del Monaco (2025): "Fairness in machine learning: companion paper to the framework on responsible AI for official statistics", *High-Level Group for the Modernisation for Official Statistics report*, October.

Hill, S (2025): "FAIR^2 for official statistics: preparing data for responsible and interoperable AI use", UNECE generative AI and official statistics workshop, Geneva, 12–14 May.

Irving Fisher Committee on Central Bank Statistics (IFC) (2021): "Issues in data governance", *IFC Bulletin*, no 54, July.

——— (2022): "Machine learning in central banking", *IFC Bulletin*, no 57, November.

——— (2023): "Data science in central banking: applications and tools", *IFC Bulletin*, no 59, October.

——— (2025a): "Governance and implementation of artificial intelligence in central banks", *IFC Report*, no 18, April.

——— (2025b): "Data integration: opportunities and challenges in central banks", *IFC Guidance Note*, no 6, July.

——— (2025c): "New insights from financial statements", *IFC Bulletin*, no 65, August.

——— (2025d): "SDMX adoption and use of open source tools", *IFC Report*, no 17, February.

——— (2025e): "Data science in central banking: enhancing the access to and sharing of data", *IFC Bulletin,* no 64, May.

Kim, S, R Jansen, M Jug, D Buckley, R de Fondeville, A C Fadel, A Saxena, J Stahl and W Hsiao (2025): "A Delphi study on the role of privacy enhancing technologies (PETs) in data sharing ecosystems", *Statistical Journal of the IAOS*, vol 41, no 3, July.

Križman, I and B Tissot (2022): "Data governance frameworks for official statistics and the integration of alternative sources", *Statistical Journal of the IAOS*, vol 38, no 3, August.

Lambe, E, G Tice, R Schmidt, C Schmieder and M Rusev (2025): "Blockchain and SDMX: immutable data sharing for trust", presentation at the SDMX Global Conference, September.

Leetaru, K and P A Schrodt (2013): "Gdelt: global data on events, location and tone, 1979-2012", *ISA annual convention*, vol 2, no 4.

Lewis, P, E Perez, A Piktus, F Petroni, V Karpukhin, N Goyal, H Küttler, M Lewis, W Yih, T Rocktäschel, S Riedel and D Kiela (2020): "Retrieval-augmented generation for knowledge-intensive NLP tasks", *Advances in Neural Information Processing Systems (NeurIPS 2020)*, vol 33.

Marcucci, J (2024): "Macroeconomic forecasting with text-based data" in M Clements and A B Galvão (eds): *Handbook of research methods and applications in macroeconomic forecasting*, chapter 16, Edward Elgar Publishing.

Nunes, I and L Palumbo (2025): "Classification at scale", UNECE generative AI and official statistics workshop, Geneva,12–14 May.

Organisation for Economic Co-operation and Development (OECD) (2024): "Explanatory memorandum on the updated OECD definition of an AI system", *OECD Artificial Intelligence Papers*, no 8, March.

——— (2025a): "Sharing trustworthy AI models with privacy-enhancing technologies", *OECD Artificial Intelligence Papers*, no 38, June.

——— (2025b): "Mapping relevant data collection mechanisms for AI training", *OECD Artificial Intelligence Papers*, no 48, October.

——— (2026): *OECD digital education outlook 2026: exploring effective uses of generative AI in education*, January.

Pérez, A, C Sáez, S Gallego and A Jurado (2025): "Synthetic LLM-generated text to train small models for automatic coding", UNECE generative AI and official statistics workshop, Geneva, 12–14 May.

Piela, R (2024): "Incorporating AI into statistical standards: enhancing GSBPM with (generative) AI", UNECE ModernStats World Workshop, Geneva, 21–22 October.

Ricciato, F (2024): "Steps toward a shared infrastructure for multi-party secure private computing in official statistics", *Journal of Official Statistics*, vol 40, no 1, March.

Rigaud, T (2025): "Enhancing statistical disclosure control using large language models", presentation at the UNECE Expert Meeting on Statistical Data Confidentiality, Barcelona, 15–17 October.

Royal Statistical Society (RSS) (2026): *AI is statistics: why statistical thinking is vital for the effective, ethical and safe use of AI*, March.

Sirello, O, B Bogdanova and E Erdem (2025): "Metadata in the age of AI: the role of official statistics in securing a virtuous cycle", *Statistical Journal of the IAOS*, vol 41, no 3, July.

Tebrake, J, E B Boukherouaa, J Danforth and M N Harikrishnan (2026): "StatGPT: AI for official statistics", *IMF Departmental Papers*, vol 2026, no 004, March.

United Nations Economic Commission for Europe (UNECE) (2023a): *Large language models for official statistics*, High-Level Group for the Modernisation for Official Statistics white paper, December.

——— (2023b): *Synthetic data for official statistics – a starter guide*, January.

——— (2025a): *Generative artificial intelligence for official statistics*, High-Level Group for the Modernisation for Official Statistics report, September.

——— (2025b): *Framework on responsible AI for official statistics*, High-Level Group for the Modernisation for Official Statistics report, October.

——— (2026): *Ethics in official statistics: foundations and practices to foster public trust*, January.

United Nations Statistical Commission (UNSC) (2025): *Fostering AI-readiness and responsible redistribution of official statistics*, room document by the World Bank, Fifty-sixth session, New York, 4–7 March.

——— (2026): *Joint report of the Committee of the Chief Statisticians of the United Nations System and the Committee for the Coordination of Statistical Activities*, items for information: coordination of statistical programmes, Fifty-seventh session, New York, 3–6 March.

World Bank (2025): *Metadata editor: user and practice guide*, Office of the Chief Statistician, May.

Zhang, T (2026): "The financial stability implications of artificial intelligence and digital finance", speech at International Financial Week – Asian Financial Forum, Hong Kong, 26 January.

Zhang, H, P S Yu and J Zhang (2025): "A systematic survey of text summarisation: from statistical methods to large language models", *ACM computing surveys*, vol 57, no 11, June.