

THE STATE OF ARTIFICIAL INTELLIGENCE IN PUBLIC AUDIT

EVIDENCE FROM SELECTED
COUNTRIES AND THE
EUROPEAN UNION

OECD ARTIFICIAL INTELLIGENCE PAPERS

May 2026 **No. 58**



**Funded by
the European Union**

OECD Artificial Intelligence Papers

The state of artificial intelligence in public audit

Evidence from selected countries and the European Union



This work is issued under the responsibility of the Secretary-General of the OECD and does not necessarily reflect the official views of OECD Member countries.

This document was produced with the financial assistance of the European Union. The views expressed herein can in no way be taken to reflect the official opinion of the European Union.

This document, as well as any data and map included herein, are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

© OECD 2026



Attribution 4.0 International (CC BY 4.0)

This work is made available under the Creative Commons Attribution 4.0 International licence. By using this work, you accept to be bound by the terms of this licence (<https://creativecommons.org/licenses/by/4.0/>).

Attribution – you must cite the work.

Translations – you must cite the original work, identify changes to the original and add the following text: *In the event of any discrepancy between the original work and the translation, only the text of the original work should be considered valid.*

Adaptations – you must cite the original work and add the following text: *This is an adaptation of an original work by the OECD. The opinions expressed and arguments employed in this adaptation should not be reported as representing the official views of the OECD or of its Member countries.*

Third-party material – the licence does not apply to third-party material in the work. If using such material, you are responsible for obtaining permission from the third party and for any claims of infringement.

You must not use the OECD logo, visual identity or cover image without express permission or suggest the OECD endorses your use of the work.

Any dispute arising under this licence shall be settled by arbitration in accordance with the Permanent Court of Arbitration (PCA) Arbitration Rules 2012. The seat of arbitration shall be Paris (France). The number of arbitrators shall be one.

Abstract

This paper examines how public audit institutions are exploring the use of artificial intelligence (AI) to strengthen oversight and improve audit processes. Drawing on consultations with 15 institutions across 14 countries and the European Union, it reviews emerging AI applications in areas such as anomaly detection, document processing, knowledge management and predictive risk assessment. The findings show that while AI adoption in public audit remains at an early stage, experimentation is expanding and many institutions are integrating AI within broader digital transformation efforts. However, a gap remains between pilot projects and scalable operational deployment. Key challenges include fragmented data systems, limited internal technical expertise and evolving governance frameworks. Strengthening data governance, digital infrastructure and internal development capacity will be critical for audit institutions seeking to responsibly scale AI while maintaining transparency, accountability and public trust.

Acknowledgements

This paper was prepared by Maria Eugenia Heyaca and Andrea Pallotta with input from Helene Wells, Sara Minguez, Rita Guelzim and Matteo Frison, under the leadership of Elsa Pilichowski, OECD Director for Public Governance (GOV), and Nejla Saula, Acting Head of GOV's Anti-Corruption, Integrity and Open Government Division (GOV/ACIOG). In addition to these contributors, other OECD colleagues reviewed the report and provided valuable comments, including, Jamie Berryhill, András Hlács, Delphine Moretti and Alastair Swarbrick.

This paper was developed in response to an engagement with the nine Austrian regional audit institutions (*Landesrechnungshof* – LRH) of Burgenland, Carinthia, Lower Austria, Salzburg, Styria, Tyrol, Upper Austria, Vienna, and Vorarlberg. Their interest in how artificial intelligence is being applied in audit internationally helped shape the focus of this work and reflects their broader engagement with emerging approaches to public sector oversight.

The authors would like to thank the following institutions for their valuable participation in the consultations:

- *Controladoria-Geral da União* (CGU) or Comptroller General of the Union, Brazil
- *Contraloría General de la República* (CGR) or Comptroller General of the Republic, Costa Rica
- *Cour des comptes* or Court of Audit, France
- *Bundesrechnungshof* or Federal Audit Office (FAO), Germany
- *Corte dei conti* or Court of accounts, Italy
- Board of Audit and Inspection (BAI), Korea
- European Court of Auditors (ECA), European Union
- *Algemene Rekenkamer* or Court of Audit, Netherlands
- National Audit Office of Norway, Norway
- Accountant General's Department (AGD), Singapore
- Auditor-General's Office (AGO), Singapore
- National Audit Office (NAO), United Kingdom
- And the three additional institutions that contributed their knowledge and experience but preferred remaining anonymous.

The authors are grateful for their insights and contributions, which have been essential in shaping the findings and recommendations of this paper.

The OECD Secretariat appreciates the support of Thomas Giacoletto, project co-ordinator at the European Commission. This project was funded by the European Union via the Technical Support Instrument (TSI), and implemented by the OECD, in co-operation with the European Commission via the Reform and Investment Task Force (SG REFORM).

This paper benefitted from consultation with the OECD Working Party on Public Integrity and Anti-Corruption and the OECD Public Governance Committee.

Table of contents

Abstract	3
Acknowledgements	4
Abbreviations and acronyms	7
1 Executive summary	8
AI in public audit: An emergent but rapidly evolving landscape	8
Infrastructure and capacity: Bridging the gap between pilots and production	8
Data governance: The foundation that AI cannot afford to lack	8
Shared challenges and emerging good practices	9
2 Overview and state of play	10
3 Use cases	14
Anomaly detection and risk assessments	14
AI for classification in audit and oversight	16
Predictive and preventive audit models	18
Search, retrieval and knowledge management	19
Intelligent Document Processing (IDP) and data extraction	21
Generative AI for drafting, summarising and translating	23
Visual and spatial data analysis	25
4 Infrastructure and capacity	30
Hardware infrastructure	30
Software environment and stack	31
Data infrastructure and governance	32
Development capabilities	34
5 Measuring impact	36
6 Challenges and strategic considerations	38
Ethics, explainability and bias	38
Organisational culture, capacity and supporting adoption	39
Data quality as a strategic asset	40
Governance and security	40
Managing the pace of change and reducing third-party dependencies	40
Balancing innovation, long-term stability and sustainability	41

7 Good practices and success factors	44
Transparency and explainability	44
User and stakeholder involvement: facilitating bottom-up innovation	44
Management and leadership support	45
Institutional governance and oversight	47
Effective training and knowledge sharing	47
Strategic partnerships	48
8 Future directions	50
9 Conclusions	52
Glossary	54
References	57

Abbreviations and acronyms

AI	Artificial intelligence
ML	Machine learning
RAG	Retrieval-augmented Generation
CNN	Convolutional Neural Network
ETL	Extract-transform-load
GPU	Graphics Processing Unit
IDP	Intelligent Document Processing
IoT	Internet of Things
IoU	Intersection over Union
NLP	Natural Language Processing
OCR	Optical Character Recognition
RPA	Robotic Process Automation

1 Executive summary

AI in public audit: An emergent but rapidly evolving landscape

Artificial intelligence is reshaping public sector oversight. Between March and July 2025, the OECD consulted 15 audit institutions across 14 countries and the EU to assess current AI adoption in public audit. The results point to growing institutional commitment: two-thirds have a formal AI strategy, 80% have internal AI guidelines, 87% offer staff training and the same proportion have at least one tool in active production. Yet maturity levels vary widely and many deployments remain at the pilot stage.

The evidence suggests that, while AI in audit is still maturing, institutions are deliberately investing in the preconditions for wider uptake: building the governance, skills and technical foundations that broader adoption will require.

Infrastructure and capacity: Bridging the gap between pilots and production

AI applications in public audit, spanning anomaly detection, predictive modelling, intelligent document processing and generative AI, demand robust, secure and interoperable technological environments. The consultations reveal a considerable gap between institutional ambition and available infrastructure, with hardware, software and data ecosystems varying sharply across participating institutions.

Although several SAIs have made meaningful progress in building AI-ready infrastructure, a pronounced divide persists between early experimentation and the scalable, production-grade environments needed for AI to fulfil its potential. Bridging this divide will be among the defining investment priorities of the coming years, especially as audit institutions move from piloting to integrating these tools into core processes.

Data governance: The foundation that AI cannot afford to lack

Across all consulted institutions, data quality, accessibility and governance emerged as decisive factors shaping the success of AI initiatives. Fragmented systems, limited standardisation and weak interoperability constrain scalability and undermine confidence in AI outputs.

Global standards increasingly require public institutions to maintain data environments that are secure, interoperable and auditable. As the AI field evolves, data governance and infrastructure will remain the primary enabler of digital innovation. Oversight institutions that build for long-term interoperability and governance, not just short-term use cases, will be best positioned to adopt more sophisticated solutions in the years ahead. As audit bodies increasingly rely on AI to guide findings and controls, the integrity of the data on which these tools are built will also become even more critical.

Shared challenges and emerging good practices

The participating institutions span a wide range of digital maturity, resource levels and organisational models. Despite this diversity, common obstacles recur: fragmented data systems, limited in-house technical expertise, and the absence of shared metrics for measuring AI impact. The report nonetheless documents genuine progress, cross-border collaboration, inclusive staff engagement and the emergence of internal governance frameworks tailored to responsible AI deployment.

Despite these constraints, there is strong evidence of innovation and institutional commitment across the participating organisations. Moving forward, continued investment in internal development capacity, robust data governance frameworks and international knowledge sharing will be essential to ensure that AI applications in public audit deliver meaningful impact while remaining aligned with public sector values of transparency, accountability and trust.

2 Overview and state of play

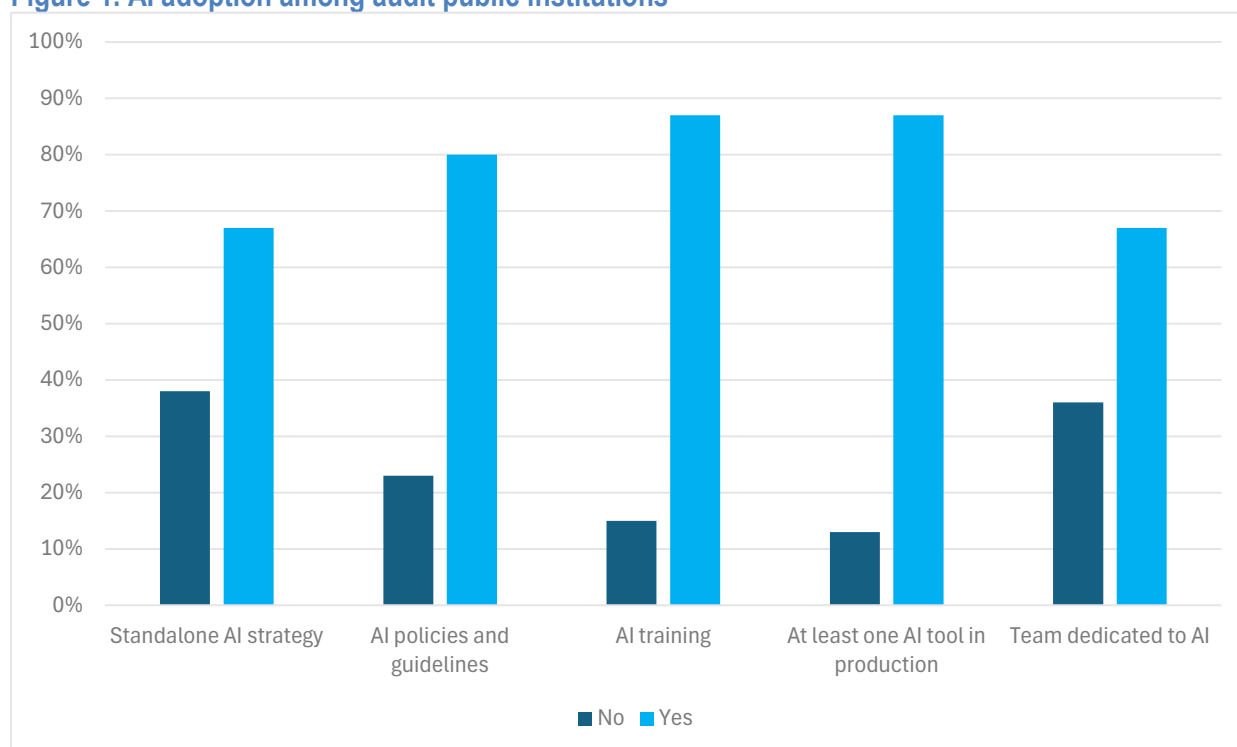
Artificial intelligence (AI) is rapidly transforming the public sector, offering new tools for oversight, efficiency and accountability. In fact, OECD (2025^[1]) work has found enhanced accountability and anomaly detection to be one of the top benefits of using AI in government. Public audit institutions, alongside other public oversight bodies and relevant private sector actors, are increasingly exploring how AI can enhance their ability to achieve their mandates. While the potential is significant, so, too, are the challenges, which range from capacity constraints to ethical concerns and the need for clear governance frameworks.

Between March and July 2025, the Organisation for Economic Co-operation and Development (OECD) conducted consultations with 15 institutions in 14 countries (Brazil, Costa Rica, France, Germany, Italy, Korea, the Netherlands, Norway, Singapore, the United Kingdom, as well as three additional countries whose institutions preferred to remain anonymous) and the EU to assess the current state of AI adoption in public audit. The group primarily comprised supreme audit institutions (SAIs) but also included a multinational accounting firm and a public supervisory authority. The findings presented here also draw upon ongoing work and partnerships of the OECD with other institutions, offering a broader perspective on emerging practices.

This report constitutes the main output of these consultations and related analytical work. It summarises the key findings and insights gathered and is organised as follows: Chapter 2 explores a range of AI use cases currently piloted or implemented across institutions, illustrating the practical applications and benefits of AI in the audit sector. Chapter 3 examines the IT infrastructure and capacity building that underpin these AI deployment efforts. Chapter 4 focuses on attempts to measure the impact of AI initiatives, while Chapter 5 delves into the key challenges and strategic considerations that institutions face in integrating AI technologies. Chapter 6 highlights good practices and critical success factors emerging from early adopters' experiences, and finally, Chapter 7 offers a perspective on the future of AI in public audit, outlining emerging trends and potential future developments, and concludes with the findings and reflections of the report.

The results demonstrate that the audit profession is undergoing a transformation, where experimentation is often coupled with strategic intent, though maturity and integration levels still vary significantly across institutions. Figure 1 summarises the results of the survey that the OECD conducted on the current state of AI adoption across the consulted organisations.

Figure 1. AI adoption among audit public institutions



Source: Consultations conducted by the OECD with 15 institutions in 14 countries and the European Union between March and July 2025.

Two-thirds of the institutions surveyed (67%) report having adopted a formal AI strategy. For those without a stand-alone AI strategy, leadership directives or operational objectives set within data or Information Technology (IT) directorates often serve a similar function. In many cases, AI initiatives are also embedded within broader digital transformation or innovation strategies, reflecting a holistic view of technological change.

Of the consulted institutions, 80% have established internal AI guidelines or policies. However, the nature and level of detail of these documents vary considerably. Some institutions provide only broad or summary statements outlining general principles limited to the responsible use of generative AI, while others have developed more comprehensive materials. These in-depth documents may include specific procedures, examples of use, or tailored advice addressing the needs of technical and non-technical staff.

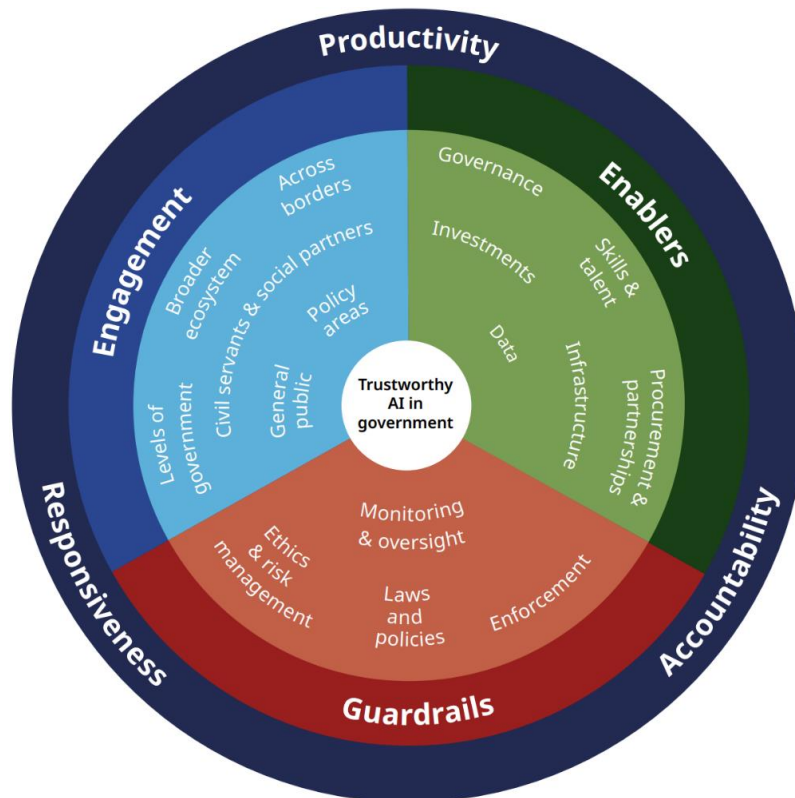
Capacity building emerged as a key priority. A sizable 87% of institutions currently offer AI-related training to general staff. However, the depth and structure of training vary too: in some cases, learning opportunities are limited to ad-hoc sessions, seminars, or conference-style formats, while others have implemented more structured programmes that address the use of different AI techniques, including large language models (LLMs), at varying levels of technical complexity.

The deployment of AI tools is also well under way. 87% of institutions reported having at least one AI tool currently in production, meaning integrated into a workflow. While for some, this may involve a single use case, such as using AI for text summarisation to support technical teams, others have already implemented a suite of AI-enabled solutions, a topic further explored in the subsequent section on use cases. That said, a large number of the tools encountered remain in pilot stages, reflecting a broader trend across functions of government (OECD, 2025^[1]). This denotes a risk-aware approach, but also several challenges including building the right infrastructure, acquiring the necessary skills, measuring impact to justify further investment, and finding ways to ensure a trustworthy use of AI tools.

Finally, 67% of institutions have established dedicated AI teams. These teams play a central role in coordinating development, deployment, and governance of AI systems, and often serve as internal centres of expertise. These teams remain predominantly small, typically comprising no more than three people.

Together, these findings reveal a growing institutional commitment to trustworthy AI, albeit with varying degrees of strategic alignment, capacity and practical deployment. As the next sections will explore in more detail, the use of AI in public audit institutions is not only advancing but diversifying, offering valuable lessons for both current adopters and those at earlier stages of their AI journey. These insights reflect and deepen the observations outlined in the OECD's *Governing with Artificial Intelligence* (2025^[11]) report, particularly regarding the cautious and fragmented nature of AI adoption by public institutions. Consistent with the OECD's previous findings, infrastructure limitations, regulatory ambiguity and skills shortages remain the major constraints. The emphasis on capacity building, especially through targeted staff training, aligns with the OECD's call for strengthening enablers such as workforce development and digital literacy. Overall, the findings of this report seek to help audit institutions understand and implement important aspects of the OECD Framework for Trustworthy AI in Government (Figure 2) through a lens that specifically takes into account their unique context and needs.

Figure 2. OECD Framework for Trustworthy Artificial Intelligence in Government



Source: OECD (2025), *Governing with Artificial Intelligence*, https://www.oecd.org/en/publications/governing-with-artificial-intelligence_795de142-en.html.

3 Use cases

Building on the insights gathered during the consultations, this section highlights specific use cases of AI developed or deployed by public audit institutions. These practical examples illustrate how AI tools are applied across various audit functions, from anomaly detection and fraud prevention to risk assessment and workflow automation. By showcasing these applications, the aim is to provide a clearer understanding of the tangible benefits associated with integrating AI into public audit practices.

The key areas in which AI is being applied include:

- Anomaly detection and risk assessments
- AI for case classification in audit and oversight
- Predictive and preventive audit models
- Search, retrieval and knowledge management
- Intelligent Document Processing (IDP) and data extraction
- Generative AI for drafting, summarising and translating
- Visual and spatial data analysis
- Other uses: internal productivity, communication, and AI governance tools

Anomaly detection and risk assessments

Audit institutions are increasingly leveraging AI to detect anomalies, assess risk levels and enhance the targeting of audit resources. New tools are being integrated across financial, compliance and performance audits, and may be applied either as part of continuous monitoring or within the scope of specific audit engagements. In many cases, AI is used to complement or replace traditional audit sampling methods, allowing for more precise and scalable identification of transactions or entities that warrant further scrutiny.

Such applications typically involve the use of machine learning (ML) algorithms - such as Random Forest, clustering techniques or autoencoders – to identify outliers in large datasets, including financial, procurement or other administrative records. Both supervised^{†1} and unsupervised learning[†] approaches are used to train the algorithms, though the latter tends to be favoured. In fact, unsupervised learning avoids the need for labelled datasets[†] and is better suited to capture previously unknown or evolving patterns that would not be identified as anomalies by supervised models trained on historical labelled data. An example of such a use is seen at Singapore's Auditor-General's Office (AGO) is presented in Box 1.

As an example of a common focus area for this type of use case, public procurement is inherently vulnerable to bribery, bid rigging, and collusion because of the large financial interests at stake and the frequent close interaction between public officials and private suppliers. Therefore, strong oversight is essential and the use of AI in procurement audits offers significant potential to enhance detection, improve efficiency, and strengthen integrity by systematically identifying risks and anomalies at scale. Public

[†] Terms that are defined in the glossary are marked at their first occurrence in the text with a † symbol. The glossary provides definitions for each marked term.

procurement is an area where there is growing use of AI in audit, also given the availability of large volumes of digital data and the widespread adoption of e-procurement platforms (though data quality such as completeness and standardisation remains a challenge). For example, Costa Rica's Comptroller General of the Republic (CGR) has deployed ML-based anomaly detection to strengthen its procurement oversight function. Algorithms such as Random Forest, robust components and local outlier factor are applied to data from the national e-procurement system to identify unusual procurement patterns. This approach enhances traditional sampling methods, providing more targeted insights into potentially irregular activities.

In financial audits, anomaly detection models focus on general ledger and journal entries to identify potentially fraudulent or misclassified transactions. Several institutions have developed AI systems that combine AI risk scores with traditional risk indicators, enabling auditors to prioritise high-risk transactions while maintaining auditor judgement in the decision-making process. The institutions have integrated these models into audit workflows, with active feedback mechanisms and performance metrics such as model precision and accuracy. Institutions are also integrating these tools with business intelligence platforms to support real-time risk assessment when possible. Further step could involve furthering feedback loops with data owners – in particular finance ministries, for continuous improvement of central government systems.

SAIs are also exploring advanced techniques such as graph-based models to identify complex anomalies that arise from unusual relationships or interaction patterns across multiple accounts or entities. By representing accounting data as networks of interconnected elements, these models can reveal irregular structures that are difficult to detect using traditional analysis methods. Beyond graph-based approaches, other advanced techniques are also being applied in practice. Norway's National Audit Office (NAO) for instance uses neural networks to support auditors' work related to fraud risk. They have built this tool with an awareness of international auditing standards, ensuring that the models not only enhance detection of irregularities like duplicate entries and unexplained reversals, but also provide interpretable and visually accessible outputs that auditors can trust and effectively incorporate into their audit processes (Huang et al., 2024^[2]).

In the context of financial audits, a key challenge is differentiating genuine misstatements or fraudulent transactions from benign anomalies within journal entries, as many outliers may be appropriate transactions (e.g. an unusually large year-end journal entry might appear suspicious but could simply reflect a legitimate one-time adjustment following an asset sale). A tool that produces an excessive incidence of false positives carries efficiency costs, as audit resources are expended on investigating legitimate transactions. Conversely, a high rate of false negatives undermines audit effectiveness by allowing potentially fraudulent activity to remain undetected. Accordingly, models should be calibrated to optimise the trade-off between investigative efficiency and assurance quality.

The availability of large, labelled transaction-level datasets is a prerequisite for training supervised ML models for risk classification. However, given the limited availability of labelled data, some institutions are using unsupervised learning techniques, such as K-means clustering. Unsupervised learning models group similar transactions and flag those that deviate from typical behaviour, allowing SAIs to detect potentially fraudulent activity even when no prior labelled examples of fraud are available.

In the UK, the National Audit Office (NAO) has developed an anomaly detection tool utilised in the context of financial audit work to identify high-risk transactions. This system employs a blend of analytical indicators and unsupervised machine learning techniques to assign a risk rating to all journals. Those journals that receive the highest risk ratings are subsequently subject to review and testing by auditors. In instances where auditors determine that certain high-risk journals do not need to be tested, they are required to provide a clear justification or conduct a supporting analysis to substantiate their decision.

Across all implementations, challenges include achieving high accuracy, ensuring data quality, defining appropriate sample sizes, balancing AI-generated findings with human judgment and monitoring model drift[†] over time. Despite these, the integration of anomaly detection and risk assessment tools represents a significant advancement in public audit methodology, enhancing both efficiency and effectiveness.

Box 1. Singapore's Auditor-General's Office (AGO): Application of unsupervised machine learning in auditing government financial statements

Traditional anomaly detection in Government Financial Statements (GFS) depends on pre-defined rules, which can overlook unusual transactions that deviate only slightly from expected norms. To enhance this process, an unsupervised ML technique – “Isolation Forest” – was employed by the AGO to detect anomalies more effectively within journal and payment records.

The method involves dividing the accounting data into multiple targeted groups, allowing each subset to be analysed separately. Within each group, the model identifies irregular entries by detecting patterns that differ significantly from typical behaviour. The results of these analyses help prioritise which subsets and transactions warrant further investigation by audit teams, who use visualisation tools to guide their review. The approach places a strong emphasis on interpretability, enabling auditors to understand the nature of the flagged entries rather than relying on complex or opaque modelling.

Subsets of the data are systematically ranked to highlight where the most significant anomalies lie. This allows audit personnel to allocate resources more effectively and efficiently by concentrating on entries considered to carry higher risk.

Visual representation plays a crucial role in the functionality of the tool, with scatterplots coloured by financial year to help auditors discern noticeable trends or concentrations of anomalies. This allows them to determine if irregularities correspond to specific periods in the year or if they stem from consolidation of journal entries.

This data-driven approach enables more precise identification of risk areas and supports a more focused and effective audit process.

Source: Singapore's Auditor-General's Office (AGO), 2025.

AI for classification in audit and oversight

Similarly to anomaly detection uses, classification models play a critical role in detecting misstatements, fraud or classification errors in accounting data, as well as categorising other audit-relevant documents and reports. By assigning data points or documents to predefined categories based on learned patterns in the data, these models support auditors in prioritising cases or transactions that warrant further scrutiny.

Singapore's AGO used to rely on a fixed rule-based approach to detect anomalies in journal entries, which may overlook misclassification cases unknown to the audit teams. To overcome this, a supervised machine learning model was implemented to identify potential misclassifications. This model leverages historical

data, including free-text descriptions, to predict account code classifications for current period journal entries. An automated evaluation step then ranks journal entries by their likelihood of misclassification, helping auditors prioritise transactions with higher risk for further review. This method improves upon static rules by recognising more complex and previously unseen patterns.

Norway's NAO has applied classification models beyond financial audits. For example, a ML model was developed to classify information and communication technology (ICT) related criminal cases from a large volume of reported incidents, identifying which cases pertain to ICT offenses. This demonstrates how classification extends to law enforcement and oversight contexts. For more see Box 2.

In a project of the Comptroller General of the Union (CGU) of Brazil, the Malha Fina tool employs a binary classification model to automatically evaluate accountability reports from the National Cinema Agency (ANCINE), determining whether the reports indicate any signs of financial irregularities or not, and providing recommendations for further human review when potential issues are detected. During the project, 3,627 accountability reports were analysed, representing a 150% increase in ANCINE's analytical capacity. The Malha Fina tool has generated savings of BRL 803 million (EUR 129 million equivalent) in operating costs, resulting in a 29-year reduction in analysis time (CGU, 2025^[3]).

Overall, classification models enable audit institutions to systematically and efficiently manage large volumes of data across multiple domains, supporting targeted audit planning, early detection of irregularities and informed decision-making.

Box 2. Norway's National Audit Office (NAO): ML for the identification of police efforts related to ICT crime

A ML model was developed by the Norway's NAO to identify which reported criminal cases were associated with ICT-related offences. The primary objective was to determine the volume and types of ICT crimes within a larger set of reported incidents. With over 300,000 cases available in text format, this initiative aimed to classify each as either ICT-related or not.

To train the model, auditors manually labelled 1,072 crime reports, creating a reliable dataset for supervised learning. Natural Language Processing (NLP) techniques were applied to the case descriptions, with emphasis on detecting terms such as "internet", "hacking" and "computer" to help the model recognise relevant patterns. Several algorithms were trialled: Naïve Bayes delivered unsatisfactory performance, while Random Forest and Gradient Boosting showed significant overfitting, where the model performs well on training data but poorly on unseen data. Ultimately, a Support Vector Machine (SVM) was selected for its superior balance of performance and generalisability.

SVM successfully categorised 21,500 out of 334,544 total cases as ICT-related. This revealed a much higher number of ICT crime incidents than originally recorded in police systems, particularly in financial crime categories. The discrepancy highlighted the value of using ML tools to uncover hidden trends within vast textual datasets.

Though not an audit goal for the Norway's NAO, this model could also serve as a practical aid for the police, enabling more informed resource allocation and strategic planning. The audit findings were well-received by law enforcement, demonstrating the value of data-driven decision support. Importantly, the project was carried out entirely in-house, with no reliance on external consultants, ensuring cost-effectiveness and full control over the process.

Source: Norway's National Audit Office (NAO), 2025.

Predictive and preventive audit models

A growing number of oversight institutions are exploring the use of AI and ML models to shift from reactive audits and oversight – focused on identifying issues after they occur – to more preventive approaches. Predictive models are designed to help anticipate risk events or detect early warning patterns, supporting auditors in focusing efforts on areas with higher potential for non-compliance or system failures. By doing so, these tools support earlier intervention and more strategic audit planning, allowing oversight bodies to focus resources on areas where problems are most likely to emerge.

These tools are typically based on classification models or predictive algorithms that analyse large volumes of structured and/or unstructured data[†]. Data sources may include for example procurement, planning documents, financial records, information on beneficial ownerships, contract databases and infrastructure project data. The models produce prioritised risk assessments that help auditors and oversight bodies target their research and controls more effectively. A detailed example of such a system used by Italy's Court of Accounts (*Corte dei conti*) is presented in Box 3. While AI enhances the speed and scale of risk analysis, so far auditors remain responsible for final decisions and validations.

Another example of application includes the use of AI in Brazil's CGU, where audit units have developed tools to assist in the preventive review of procurement processes. One such tool, developed as a pilot project, uses LLMs to assess the feasibility of IT contracts and their compliance with legislative frameworks. Early results show that AI-assisted assessments of dozens of procurement processes can be performed significantly faster and at low cost.

In another example in Italy, predictive models were also developed to support the early detection of risks in public procurement and infrastructure projects. Using data from the National Anti-Corruption Authority (ANAC) and other public procurement databases, the system produces risk indicators to prioritise oversight of maintenance and infrastructure projects. This approach supports real-time/continuous auditing by highlighting subprojects with higher risk levels, allowing audit institutions to focus resources where they are most needed.

Although still emerging, predictive and preventive audit models illustrate the potential of AI to enhance proactive risk management, reduce the likelihood of future failures and improve the efficiency of audit planning. However, it is important to remain aware that over-reliance on predictive models may lead to overlooking emerging risk factors that are not yet reflected in the training data, potentially reducing a model's accuracy over time.

Box 3. Italy's Court of Accounts: Predicting municipalities in financial distress

A Machine Learning (ML) approach enhanced by domain expertise

The financial health of municipalities is a critical concern for public administration as financial distress can lead to the inability to provide essential services, thereby affecting the well-being of communities. In Italy, local governments are legally obligated to maintain operations even in the face of financial crises, making early detection and intervention vital. To address this challenge, the Italian Court of Accounts has developed RoMoLo, an experimental ML model designed to predict financial distress in municipalities.

The development of RoMoLo involved the construction of a comprehensive dataset encompassing 7 094 municipalities over a seven-year period (2016–2022). The dataset contains 37 attributes derived from heterogeneous data sources, including financial statements, demographic information and institutional indicators. The data was reweighted before training the model: as there are only a small percentage of municipalities in financial distress, their weight in the data was adjusted to avoid biases.

The attributes included in the dataset were selected based on their relevance to financial health and were informed by the expertise of accounting judiciary professionals. They encompass various aspects of municipal operations, such as demographic characteristics, geographical location, investment levels, financial autonomy, indebtedness, expenditure rigidity and debt management practices.

A “bankruptcy risk” variable is produced by the predictive model; it is this feature that enables the classification of each municipality into a risk category based on the occurrence of severe financial imbalances. The five levels of risk are:

- Low Risk: has never experienced either pre-distress or financial distress.
- Medium-Low Risk: has experienced at least one instance of pre-distress.
- Medium Risk: has experienced at least one instance of financial distress.
- Medium-High Risk: has experienced both pre-distress and subsequently fiscal distress.
- High Risk: has experienced multiple instances of financial distress.

To support this classification, several ML models were tested, including Gaussian Naive Bayes, Logistic Regression, Random Forest, Support Vector Machine, Gradient Boosting and a custom ensemble approach. Logistic Regression was selected due to its strong performance and robustness in capturing the relationship between predictors and the defined risk categories. RoMoLo achieved a sensitivity (“recall”) of 1, meaning it correctly identified all municipalities in financial distress without any false negatives. However, the model also produced a notable number of false positives, suggesting that while the model is highly effective at detecting municipalities in distress, it may also flag some municipalities that are not actually, and necessitate further investigation and validation.

The primary objective of RoMoLo is to serve as an early warning system of municipalities at risk of financial distress for the Court of Audit. By providing timely predictions, the model enables policymakers and auditors to intervene proactively, implementing corrective measures before a crisis escalates.

RoMoLo represents a significant advancement in the application of ML to public sector financial audit. The results of this study indicate that using ML models in combination with the knowledge of accounting judiciary experts can support early detection of financial distress, leading to better outcomes for communities.

Source: Italian Court of accounts (Corte dei conti), 2025.

Search, retrieval and knowledge management

Audit institutions are increasingly applying AI to enhance document search, semantic understanding, and the retrieval of audit-relevant knowledge or legislation. Knowledge management remains a time-consuming activity, often requiring auditors to sift through vast volumes of text-based materials, including auditee data, standards, policies, and regulations. AI tools in this area act as “auditor helpers,” streamlining access to relevant information through automated document discovery, semantic search techniques[†], and question-answering capabilities over large repositories of content. Even when outputs are not perfectly accurate, these systems generally outperform manual methods, which rely on exact word or phrase matching, surfacing relevant content more quickly and consistently. By understanding the semantic meaning behind user queries, they are able to handle synonym or phrasing variations and complex search expressions, as well as the automation of tasks such as summarisation and theme extraction, thereby supporting auditors in faster analysis and decision-making. This represents a lower-risk application of AI that enhances efficiency and supports higher adoption rates within audit institutions.

Beyond improving information retrieval, AI can contribute to transforming how audit institutions organise, interpret, and apply accumulated knowledge. Advanced language and semantic models can identify patterns and connections across previous audits, regulatory updates, and emerging risks helping auditors derive insights that might otherwise remain hidden in dispersed documentation. Combined with AI-driven summarisation and classification tools, they can support maintaining up-to-date audit methodologies, guidance and training resources. This capability strengthens institutional learning, promotes greater consistency across audit teams and ultimately enhances the quality, agility, and responsiveness of public sector auditing.

Audit institutions have adopted a dual approach in implementing these systems: some are using commercial off-the-shelf tools, while others are developing custom models trained on audit-specific terminology and content. Key technologies include semantic search engines, Retrieval-Augmented Generation (RAG) pipelines, vector databases, and domain-adapted LLMs.[†]

In Germany, the Federal Audit Office (FAO) has deployed a commercial semantic search system adapted to its needs. This system aims to support auditors in navigating large volumes of documents in various formats. It combines full-text search, metadata enrichment and filtering capabilities. The system also allows for Python-based scripting to support advanced search scenarios. It has been in active use for over a year and is reported to significantly reduce the time spent on information retrieval.

Extending beyond document retrieval, AI-powered chatbots and virtual assistants are increasingly used to support auditors internally with guidance, queries, methodology, audit standards and technical help. These interactive tools facilitate communication by allowing auditors and staff to pose natural language questions and receive context-aware answers, helping streamline workflows and improve onboarding. Examples of such applications can be seen across various institutions:

- The European Court Auditors (ECA) is considering an internal chatbot designed to support audit methodology, providing auditors with easier access to methodological information, checklists, quality standards and guidance. They are also testing the integration of Microsoft 365 Copilot to enhance productivity by extracting information from tools like Outlook, Teams or SharePoint, always ensuring that data protection requirements are respected.
- In Costa Rica's Comptroller General of the Republic (CGR), the audit office uses LLM-powered notebooks as internal assistants, enabling staff to query databases and institutional procedures related to human resources, IT management and procurement. This application has improved internal knowledge management and consulting, while also accelerating the onboarding processes of new staff by allowing efficient access to accurate and up-to-date institutional information. Although these applications are not directly part of audit work, they contribute to more efficient resource management, save auditors' time, and promote the standardisation and diffusion of institutional knowledge, ultimately supporting higher-quality and more consistent audit processes.
- Additional examples are presented for Brazil in Box 4 and France in Box 6.

As audit institutions grapple with large and often fragmented knowledge repositories, AI tools for search, retrieval, knowledge management and interactive audit support offer a pragmatic entry point into AI adoption. While challenges remain – including the need for quality-controlled document corpora, careful prompt design, data security and user trust – such tools are proving effective in improving internal efficiency and supporting more informed audit work.

Box 4. Brazil's Office of the Comptroller General of the Union (CGU): AuditPesquisa

At Brazil's CGU, a sophisticated chat-based tool has been introduced to enhance the way audit reports are examined and interpreted. This tool is specifically designed to facilitate access to a vast repository of audit information by allowing users (e.g. auditors) to engage with it through natural language, creating a seamless and intuitive interaction. Rather than manually navigating dense, lengthy documents, users can ask questions as they would in conversation, and the system responds with clear, relevant answers drawn directly from the reports. This makes it easier to summarise key findings, identify risks, recognise patterns across multiple audits, and pinpoint important issues flagged in past assessments.

The tool's functioning is grounded in the Retrieval-Augmented Generation (RAG)[†] framework, which functions in a four-stage process:

1. The user submits a query.
2. The system enters a retrieval phase, in which it scans its audit report database (explained below) to locate the most pertinent segments.
3. This is followed by an analysis phase, during which the OpenAI API LLM technology interprets the retrieved material.
4. A well-contextualised response, tailored to the original query is produced.

The use of advanced AI technologies, such as ChatGPT, underpins the system's ability to generate coherent and precise outputs.

This tool is now in operation; however, its development was not without its challenges. The primary difficulty lay in the creation and standardisation of the audit report database. A total of 3,189 publicly available audit reports could be included in the initial dataset. It was necessary to break them down into their core sections, typically the Introduction, Audit Findings, Recommendations and Conclusion, as these are often the focus of user queries. The data preparation process also involved the removal of non-essential elements such as tables, graphs, figures and disconnected statements, which could otherwise confuse or mislead the system during the analysis and generation stages.

In addition to content extraction, the system also collected a broad range of metadata associated with each report, including information about the audited body, the managing or supervising unit, geographical location, service type and category of audited activity. This metadata plays a crucial role in enhancing the precision of the tool by enabling the use of contextual filters, thereby increasing the relevance and accuracy of the answers provided. Users can narrow their searches by department, government body, region, audit category, or service type, allowing for a highly targeted and efficient information retrieval experience. This capability turns what was once a labour-intensive process into a streamlined task, reducing the time needed to interpret key findings and enhancing the decision-making process. Ultimately, this innovation represents a major advancement in the application of AI to auditing, supporting transparency, operational efficiency, and strategic evaluation across the sector.

Source: Brazil's Office of the Comptroller General of the Union (CGU), 2025.

Intelligent Document Processing (IDP) and data extraction

Intelligent Document Processing (IDP) refers to the use of AI to automatically extract structured data from unstructured or semi-structured sources, such as Portable Document Formats (PDFs), scanned documents, images and handwritten forms. These systems combine technologies such as optical

character recognition (OCR), natural language processing (NLP) and LLMs, to classify documents, identify key entities and convert free text into exploitable, structured formats.

In the context of public audit, IDP tools are gaining attention for their ability to streamline high-volume, repetitive tasks such as the review of receipts, invoices, contracts and expense reports. These documents are currently mostly analysed manually, often using sampling methods. IDP offers the potential to automate these processes at scale, enhancing accuracy, reducing time spent on document handling and enabling full-population testing.

In contrast to broader knowledge management systems, IDP tools are designed for targeted extraction of specific fields or values, for example, names, dates and amounts, from varied document formats. When integrated into broader audit workflows, IDP systems can improve verification processes, support early anomaly detection, and reduce the administrative burden on audit teams.

In the Italian Court of Accounts for example, audit teams are experimenting with automated extraction of data from PDFs and image files to support population-level testing of public expenses. These efforts focus on automatically identifying and extracting text and numerical values from expense justification documents. While currently performed manually, the objective is to fully automate this process.

Box 5. Italian Court of Accounts – GiusBERTo

One of the earliest AI experiments at the Italian Court of Accounts was GiusBERTo, a Proof of Concept designed to automate the anonymisation of sensitive personal data in judicial documents. Trained on thousands previous Court decisions, GiusBERTo recognises and redacts personally identifiable information such as locations, names, and dates.

GiusBERTo underwent thorough testing using standard anonymisation metrics and the results showed its practical effectiveness when tested with real judgements that had already been anonymized. This proves the tool's ability to ensure compliance with privacy laws, while still maintaining the necessary transparency in public records.

Building on this foundation, further efforts are underway to enhance anonymisation. With advancements in technology and the growing influence of Generative AI, the model has been redesigned to utilize state-of-the-art features, incorporating human-in-the-loop methods that allow users to validate and correct outputs as needed. Additionally, the tool is integrated with cloud-based platforms like Amazon Web Services (AWS) to improve scalability and performance, therefore it is now ready for testing by end users.

As the Court's AI capabilities develop, the initiative is expanding beyond basic anonymisation toward advanced text mining techniques that support a more thorough verification of public expenses. This evolution demonstrates the Court's commitment to leveraging AI for improved governance and accountability in public spending.

Source: Italian Court of Accounts (Corte dei conti), 2025.

Elsewhere, IDP is under exploration to support the analysis of civil servants' records, using AI-models to extract structured data, such as names or academic degrees, from documents on file and converting them into machine-readable formats. This process allows for faster processing and integration of document-derived information into other analytical systems.

IDP represents a promising frontier for audit automation. Key advantages include enhanced document coverage, reduced manual effort and the potential to apply analytics tools consistently across full datasets.

However, institutions must still navigate challenges such as OCR quality – improving rapidly but not yet fully reliable – data privacy and the need for domain-specific fine-tuning of models.

Generative AI for drafting, summarising and translating

Oversight institutions, and public organisations at large, are also increasingly making use of LLMs to support or automate text generation including drafting audit documents, summarising technical material, preparing scripts for public communication and translating content across languages. In all cases, AI tools act here again as assistants rather than final decision-makers by producing first drafts or suggestions that are reviewed and validated by auditors or communications staff.

The appeal of generative AI beyond the increase in productivity is its ability to produce text that is consistent in tone and terminology. At the same time, concerns remain around the accuracy of generated outputs, especially in high-stakes contexts. This is particularly relevant when translating technical audit subjects, where jurisdiction-specific terminology or language nuances may lose precision or meaning across contexts. Most institutions mitigate these risks by maintaining multi-layered review processes and ensuring clear human oversight.

At the ECA, Microsoft 365 Copilot is being tested by a group of users for a variety of use cases, including generating draft text. The deployment of this tool has been guided by an internal Advisory Committee on AI, which helped establish appropriate terms of use. The tool can be a valuable asset for speeding up early-stage drafting of audit documentation. However, to address the risk of errors in AI-generated text, human oversight and verification are essential, and the users remain ultimately responsible for the text produced. The risks are further mitigated by the institution's strong quality control and multi-level review procedures.

In Costa Rica, generative AI is being used to support the drafting of scripts for public-facing communication materials. With the aid of LLMs, the audit office has produced preliminary scripts for video briefings and citizen-oriented publications, which are then reviewed and validated by communications officers. By automating part of the scriptwriting process, the institution has strengthened the clarity, consistency and agility of its public messaging.

Both Costa Rica's CGR and the ECA are also exploring the use of generative AI for summarisation of large sets of documents. In Costa Rica, LLMs help condense lengthy technical documents and generate early drafts of working papers and audit findings, which are subsequently reviewed and validated by auditors.

A further example of how LLMs are used by the Italian Court of Accounts to anonymise official documents is presented in Box 5.

Generative AI is also playing a growing role in translation and document review. At the multinational accounting firm, AI models are used to translate audit documentation across multiple languages, a significant operational challenge in international audit work. All translated outputs are subsequently reviewed and quality-assured by internal subject matter experts to ensure accuracy and appropriateness. The firm has also developed an internal generative AI-based tool, informally known as the "Review Agent". Originally designed as a persona simulating an auditor, this tool supports staff in reviewing audit reports by flagging unclear sections, identifying contradictions, or suggesting improvements. The use of the tool is optional, and while it cannot address complex interpretative issues, such as the application of specific legal or regulatory frameworks, it is considered useful for improving consistency and readability across audit files.

A further use of generative AI to support drafting and summarisation is presented in Box 6.

Box 6. French Court of Audit- ChatJF: Development of a sovereign generative AI assistant

In response to growing institutional interest in AI and the need for secure, sovereign digital tools, an in-house prototype named ChatJF was developed by the French Court of Audit. Its objective is to explore the practical applications of LLMs in financial jurisdictions, focusing particularly on document summarisation, translation, paraphrasing, data analysis and research. The initiative also sought to ensure full compliance with legal and regulatory constraints, especially those related to General Data Protection Regulation (GDPR) and professional confidentiality.

Initial use cases were identified through collaborative workshops with staff and included document summarisation, writing and translation assistance, documentary research, data table analysis and transcription. While these use cases have been identified, not all have been fully implemented; their development will proceed according to established priorities. Subsequently, more advanced applications were added to the list including content generation, infographics creation and audit preparation assistance. As usage expanded, the platform has started to support conversational agents and RAG, enabling users to upload documents and query them directly.

ChatJF was developed using open-source technologies to ensure full control over data and system architecture. Its three core components are an open-source LLM manager; OpenWebUI, a locally hosted web interface modelled after ChatGPT and an add-in for Microsoft Word to facilitate seamless integration.

The system currently integrates multiple LLMs include for example Meta's LLaMA 3 and a model from Mistral, with model switching possible depending on the task. To ensure performance and adaptability, model inference runs on graphic processing units (GPUs) rented from a French cloud provider, while the interface and embedding computations are handled internally on virtual machines.

One notable innovation is the integration of ChatJF into Microsoft Word, enabling features such as summarisation, rewriting and translation. This was developed in-house using TypeScript, HTML and CSS, and connects via individual API keys to the main platform. The plugin supports Office 2019 and aligns with the organisation's broader productivity tools, facilitating AI adoption among users.

The system is currently in the prototype stage but is actively used, with around 300 unique users recorded over 30 days and about 100 regular users. Approximately 20% of the organisation's staff has engaged with the tool. While the platform is still being tested, users are already automating transcription, content drafting and structured document analysis. Adoption has required significant training and communication efforts to manage expectations and encourage proper use.

Security and data sovereignty were central to the project's design. All components are hosted on internal servers or with certified external providers. No data leaves the institutional perimeter, ensuring that sensitive information, subject to regulatory secrecy, is processed in full compliance with legal frameworks.

The French Court of Audit roadmap includes transitioning from prototype to production, duplicating infrastructure, integrating user support systems and expanding Microsoft Office integration. Exploratory work will continue speech recognition and automated visualisation tools. Through ChatJF, the French Court of Audit demonstrates a scalable approach to generative AI, embedding it within core workflows while preserving control over data and infrastructure.

Source: French Court of Audit (Cour des comptes), 2025.

As generative AI continues to evolve, it is offering audit and oversight institutions meaningful support with high-volume language-based tasks. These tools help accelerate internal workflows, reduce time spent on drafting and summarising and improve clarity in external communication. Nevertheless, institutions continue to take a measured approach, with emphasis on validation, transparency and ensuring that AI remains firmly in a supporting role.

Visual and spatial data analysis

While less widespread, AI applications for analysing visual and spatial data, including images, satellite footage, and drone-captured content, show promise in supporting performance and compliance audits. These tools can provide new forms of evidence for assessing physical infrastructure, monitoring land use or verifying policy implementation on the ground. However, their adoption is constrained by several factors, including legal uncertainties around the use of Internet of Things (IoT) devices such as drones, limitations in data quality and the high costs associated with capability development.

These tools are particularly relevant where traditional audit methods face physical or logistical constraints. AI can help automate the interpretation of visual data, using methods such as computer vision and deep learning models, such as Convolutional Neural Networks (CNNs)[†] to detect patterns, classify objects or assess compliance remotely. While the potential is significant, these approaches are often still in the experimental phase.

In Costa Rica, a research initiative aims to build a locally produced, low-cost IoT camera system for monitoring public construction projects. The system integrates artificial intelligence to identify construction progress patterns, supporting early detection of delays or irregularities by public auditors. This initiative represents a promising example of localised, AI-driven infrastructure monitoring (for more, see Box 7).

Box 7. Costa Rica's Comptroller General of the Republic (CGR) dIAra – Monitoring of infrastructure projects

The dIAra project represents a pioneering initiative to integrate AI with Internet of Things (IoT) technology to enhance the public audit of infrastructure projects. Developed jointly by Universidad Cenfotec and Costa Rica's CGR with support from the German Cooperation Agency, dIAra addresses the challenge of monitoring public construction works in real time, providing auditors with automated, data-driven insights that enhance oversight effectiveness and transparency.

At its core, dIAra is a monitoring system that supports public auditors by automating the collection and analysis of visual data from construction sites. Low-cost IoT cameras, assembled locally with microcontrollers and 3D-printed components, capture images of ongoing works. These images are analysed by an AI-powered computer vision model trained to identify key site elements such as workers and equipment. The AI's output is structured into reports that indicate the presence and activity levels on the site, enabling auditors to detect work stoppages or irregularities requiring further investigation.

The system complements traditional audit methods, which depend on periodic site inspections and manual reporting. Such approaches can be resource-intensive and limited in scope, especially for large-scale or geographically dispersed projects. By contrast, dIAra enables continuous and objective monitoring without the need for an auditor's physical presence.

From a technical perspective, the development of dIAra reflects a commitment to technological sovereignty and open innovation. The entire system, including hardware designs, software code, AI training datasets, and deployment documentation, is open source. This openness ensures that the platform is adaptable to local needs, repairable by local technicians, and free from dependency on third

party vendors. Such a model fosters sustainable technological growth within Costa Rica and can be readily adapted by other countries or institutions with similar oversight challenges.

A critical innovation in dIAra is its AI component, specifically a model known as Modelo_1_29, which was trained exclusively on images collected from the monitored sites. This model demonstrated strong performance, achieving an average Intersection over Union (IoU) score of 82.7 percent and a sensitivity rate of 85.4 percent. These metrics indicate reliable identification of objects such as workers within site images under the conditions of the training dataset. The AI system automatically generates structured output files that summarise detected objects along with confidence scores and spatial coordinates. This enables efficient data analysis and supports auditors in tracking daily site activity and progress patterns.

While the model's accuracy is promising, the project acknowledges important limitations. Since the AI was trained on a specific dataset from a single site, its performance may decline when applied to other construction environments with different visual characteristics or operational conditions. Variations in lighting, equipment, or worker attire could affect the model's accuracy. Nonetheless, the open-source and modular nature of the system facilitates retraining and customisation, enabling future deployments to refine the model for diverse contexts. The project is still under development and requires further definition of its scalability.

dIAra illustrates how AI and IoT technologies can be strategically applied to modernise public infrastructure audits. It offers a cost-effective and replicable solution that improves monitoring precision and frequency, enhances institutional oversight capabilities, and contributes to national technological development. By combining automated visual recognition with structured data outputs, dIAra advances a model of transparent, efficient, and scalable audit that other public institutions worldwide may consider in their efforts to ensure responsible management of public resources.

Source: Costa Rica's Comptroller General of the Republic (CGR), 2025.

Norway's NAO explored the use of satellite imagery in a performance audit focused on food security. The objective was to assess how much arable land was being used for food production. Using Sentinel-2 satellite data, which is updated every eight days, auditors applied deep learning models based on CNNs to classify crop types. However, persistent visibility obstruction of the fields by clouds affected the consistency of the data. While commercial satellite imagery was available, it was considered prohibitively expensive. Due to these constraints, the findings were not fully integrated into the final audit report, though the pilot demonstrated the potential of AI-supported geospatial analysis for future use.

As these examples show, while AI offers new ways to integrate visual and spatial data into audit work, its operational use is still limited. Legal frameworks around the deployment of IoT devices for monitoring vary significantly across jurisdictions and reliable data access remains a key barrier. Across many jurisdictions, privacy and data-protection laws impose limits on how IoT and AI monitoring systems can collect and share information, particularly when activities of individuals may be inferred. Requirements related to consent, transparency and data security further constrain large-scale monitoring and can affect the admissibility or reliability of sensor-derived evidence in public audits. Nonetheless, these applications point toward a growing interest in expanding the sources of evidence available to public audit institutions through advanced data analysis techniques. Other uses: internal productivity, communication, and AI governance tools

Beyond core audit functions, AI is increasingly being adopted to support audit institution's internal productivity, facilitate communication, and strengthen governance-related processes. While these applications do not directly contribute to audit work, they offer practical benefits for efficiency, knowledge management, and institutional agility.

Enhancing internal capacity and productivity

Several institutions are deploying AI tools to streamline internal processes and boost staff productivity. The ECA for example, is currently evaluating the feasibility of integrating AI components into existing Robotic Process Automation (RPA)[†] solutions to support routine tasks related to audit. The UK NAO is also exploring the use of generative AI tools such as Microsoft 365 Copilot, mainly for document review, helping auditors manage administrative tasks more efficiently, and other personal productivity applications. Use of these tools in core audit workflows is still at a pilot stage.

Supporting communication and public engagement

AI is also deployed to support institutional communication, particularly in producing content for the public. As noted earlier (See Generative AI for drafting, summarising and translating), Costa Rica's CGR has used generative AI to draft scripts for social media content and television briefings. These tools support agility in content production and are always reviewed by communications officers before publication.

At Korea's Board of Audit and Inspection (BAI), AI-powered sentiment analysis tools have been piloted to monitor media coverage of their audit findings, automating what was traditionally manual work. This helps understand how audit topics are received and discussed by the public, identifies distorted or inaccurate coverage and ensures alignment with original press releases by the BAI. This tool classifies media reports by tone (negative, neutral or positive) and measures three key aspects:

- Similarity: How closely media reports reflect the press releases.
- Accuracy: How correctly issues are reported based on the press releases.
- Sentiment score: The overall tone of the media reports.

About 16,000 articles from different media outlets were analysed. Similarity was measured using cosine similarity, accuracy via named entity recognition (NER) and sentiment scores classified reports into negative, neutral, or positive categories (for more see Box 8).

AI for governance, oversight, and policy analysis

AI is also emerging as a tool for governance and policy analysis. At the ECA, AI was tested in the context of an audit task to compare legislative documents. Human oversight and verification of the results remain pivotal, but this approach offers new possibilities for analysing the legislative process and understanding its institutional impact.

Box 8. Korea's Board of Audit and Inspection (BAI): An AI-based methodology for analysing internal audit results

Korea's BAI is developing "Guidelines for Establishing Internal Audit Plans" to enhance audit efficiency and inter-agency cooperation. Currently, around 200,000 audit records are stored in the Public Audit Portal, but analysis remains manual, limiting scalability and efficiency. To address this, BAI initiated a project to classify internal audit cases using AI and assess self-audit practices across local governments. The project focuses on the 2024 audit results and includes four stages: 1) Developing classification standards, 2) Data collection and preprocessing, 3) Case analysis, and 4) Building an AI-based classification model. The Audit and Inspection Research Institute and the Bureau of Proactive Administration Services and Public Sector Internal Audit carried out the project in collaboration.

1. Classification standards development

A classification framework was developed to standardise BAI's internal audit case categorisation, recognising that classification may vary according to each researcher's perspective. The framework incorporated feedback from provincial audit departments and field practitioners to ensure objectivity and completeness.

2. Data collection

Due to access limitations, a custom crawler was developed to collect 24,954 publicly available audit reports (HWP, PDF). These were processed for structured analysis.

3. Case analysis

Around 1,200 reports were manually reviewed to extract key features and create training data. Argument mining (an NLP technique that extracts argumentative structures from text) was used to obtain logical reasoning from unstructured text. Data was cleaned (removing duplicates, correcting errors, etc.) and standardised into a uniform format before use.

4. AI model development

An ensemble model, combining Decision Tree and Naive Bayes algorithms for increased prediction accuracy, was trained to classify audit cases. The model outperformed individual models by harnessing complementary strengths: specifically, Decision Trees are more adept at managing complex variable relationships, while Naive Bayes models offer probabilistic predictions. Additionally, the Llama language model was used to automatically label and summarise the non-training dataset. The project aims to compare the ML ensemble model and the LLM to determine which method performs best, with validation currently underway.

A prototype was developed to demonstrate the practical application of the model and to verify feasibility. The prototype was designed to reflect the interests of actual BAI users by enabling comparisons of internal audit status across regions and by providing more detailed analyses of individual audit cases.

Source: Korea's Board of Audit and Inspection (BAI), 2025.

These use cases show a rich diversity of potential applications of AI across audit institutions, be it using ML, neural networks, NLP or generative AI. While most initiatives remain at the pilot or exploratory stage, the breadth of ideas and experimentation demonstrates that organisations recognise the transformative potential of these approaches.

Across all use cases, human auditors remain firmly at the centre of the process from conceptualisation to testing and adoption. This emphasis reflects both practical considerations (ensuring acceptance and trust

in early deployments) and a deliberate response to current challenges related to algorithmic transparency and explainability, which will be discussed in a later section. AI is thus being positioned as a tool to augment, rather than replace, professional judgement.

The consultations further indicated that AI's impact can be amplified when combined with complementary technologies such as user-friendly interfaces, data visualisation, robotic process automation and IoT applications. Such integration enables more efficient data handling, richer analytical insights and usability, and enhanced audit workflows.

At this stage, the most rapid progress is observed in the use of AI for audit support functions such as document analysis or risk sampling, rather than in generating direct audit findings. Compared with other oversight or inspection bodies, SAIs tend to employ AI less frequently for ex-ante risk assessment or large-scale data analytics. This difference reflects, in part, their ex-post mandates and the uneven availability of digital data across audited entities.

Overall, the findings suggest that while AI in audit remains emergent, the foundations for broader adoption are being actively explored. The next section examines the infrastructure, capabilities and institutional arrangements being developed to support these growing AI ambitions.

4 Infrastructure and capacity

This chapter presents a cross-institutional analysis of the digital and AI capabilities of the consulted public sector audit institutions. Drawing exclusively on self-reported data, the assessment focuses on four key dimensions: hardware infrastructure, software environments, data management and governance, and development capabilities. The goal is to understand how audit institutions are preparing and positioning themselves to leverage AI in support of their core mandates by identifying common approaches and areas of divergence.

Hardware infrastructure

The hardware infrastructure across the consulted institutions varies considerably in terms of maturity, scalability and alignment with AI deployment goals. A recurring theme is the effort to balance technological experimentation with stringent data protection requirements. While some organisations have invested heavily in local digital infrastructure (e.g., on-premises data storage and computational power), others have opted for cloud-based services from commercial vendors or hybrid models to accommodate evolving AI needs.

Several institutions have opted for building and maintaining infrastructure in-house. Dedicated computing environments, including advanced workstations and GPU servers, have been procured to support AI experimentation and development. These setups often include isolated environments with high-speed internal networks, which allow for secure experimentation with ML models while maintaining a controlled perimeter. In a few instances, physical rooms have been set aside for this purpose. These rooms are setup so that internet access is allowed, but internal network connectivity is restricted, reflecting a deliberate risk-aware effort to promote innovation without compromising system security.

Indeed, across institutions, the availability of hardware for experimentation has been seen as a driver of internal innovation. Teams are often encouraged to trial new AI ideas in secure environments, with some providing access to tools such as LM Studio or PrivateGPT for running local large language models. These tools offer a “sandbox” experience, giving data teams and auditors an opportunity to test capabilities without the risk of data leakage or dependency on commercial APIs.

Some organisations are starting to invest in specialised hardware to support advanced AI workloads. This includes building GPU clusters[†] for training or fine-tuning language models internally. In some cases, both cloud-based and on-premise solutions are being trialled at the same time in order to find the most efficient long-term setup.

High-performance GPUs like the NVIDIA H200 or L40 are used by some of the consulted organisations, highlighting a shift toward more advanced AI applications beyond basic automations or reporting/dashboarding. While some institutions purchase their own hardware, other institutions prefer to rent GPUs (i.e. access GPU capacity through the cloud) from commercial cloud providers. This approach offers flexibility and scalability, especially for those still shaping their AI strategies or needing temporary compute power. Renting also allows organisations to experiment with different model architectures without being locked into specific hardware investments. However, this approach can introduce cost variability and operational dependencies that must be carefully managed.

A majority of organisations have expressed concerns around data sovereignty and rules on legal jurisdictions for data hosting and processing, which strongly influence their approaches. Some are explicitly avoiding global commercial cloud services due to geopolitical, data protection or legal considerations, preferring local data centres or regionally governed platforms. Even among those leveraging big international cloud providers, the preference is often for infrastructure held within their national or continental borders.

Complementing these trends, SAIs at the early stages of AI adoption typically rely on shared government infrastructure or lightweight virtual environments for their initial use case testing and prototyping. One notable example involves a secure virtual desktop system designed specifically for auditors, which allows remote access to encapsulated data analysis tools. This reflects a broader pattern where hardware decisions are driven not only by performance needs, but also by compliance and risk management considerations.

Only a small number of institutions reported clear plans to pursue actions that would require them to expand or upgrade their hardware infrastructure in the near future, such as implementing Spark clusters (a type of cluster computing platform) or building out data lakes to support more intensive computation. This suggests that while AI experimentation is accelerating, significant investments in hardware infrastructure are still limited to a few forward-leaning institutions.

Generally, those further ahead in their AI journey exhibit more mature infrastructure setups. These organisations are not only equipped with the physical hardware to run sophisticated models but have also developed practices that allow them to do so safely, efficiently and in compliance with existing regulatory frameworks. Conversely, organisations that rely heavily on external hosting or general-purpose infrastructure may, in some instances, find themselves constrained as AI use cases become more complex and data intensive. That being said, organisations that have a strategic approach to using commercial tools have been able to successfully build AI suites without the necessity to add any on-premise infrastructure, relying on cloud-based access to tools, computing power and data storage through commercial partners.

It is a best practice to ensure that infrastructure choices are aligned with the specific needs of current and desired AI workloads, which typically require both high compute availability and robust data security measures. Institutions that aim to scale their AI capabilities may benefit from developing hybrid architectures that combine the flexibility of the cloud with the control and compliance of on-premises systems. This would allow for both experimentation and production-grade deployment in a way that balances cost, security and operational agility.

Overall, while a number of SAIs have laid a solid foundation for AI-enabling infrastructure, there remains a significant gap between early-stage experimentation and scalable, production-ready environments. Future investments will likely need to address this gap, especially as audit institutions move from piloting AI tools to integrating them into core processes.

Software environment and stack

A diverse range of software tools and platforms are being utilised across institutions to support digital transformation and AI integration. The software ecosystems differ in their depth and specialisation, often reflecting the maturity of each organisation's broader digital strategy. In general, there is a clear trend toward leveraging commercial productivity suites and open-source frameworks, alongside selective exploration of more specialised AI software environments.

Nearly all institutions rely on established commercial platforms for productivity and collaboration, with Microsoft 365 and Google Workspace being the most common. These tools often come bundled with integrated AI assistants such as Microsoft 365 Copilot, which are increasingly being tested or deployed at scale. Institutions that have incorporated these features report improvements in workflow efficiency, though

some have imposed limitations to restrict their use to public or non-sensitive data. These guardrails are usual to address concerns around data leakage, particularly where content might be processed externally or used for third-party model training.

Open-source frameworks such as Python, Jupyter Notebooks and associated data science libraries are widely adopted, especially among teams engaged in hands-on analytics and automation. Several audit institutions maintain GitHub repositories for version control, internal collaboration and project tracking. These practices indicate a level of software development maturity that supports iterative experimentation and code reusability. In some cases, codebases are being built and maintained internally as strategic assets, laying the foundation for shared AI components across different audit functions.

More advanced organisations have begun incorporating AI-specific tools and platforms such as LM Studio, PrivateGPT, or RAG architectures in order to run LLMs in controlled environments – either for testing purposes or for limited production use – while maintaining greater control over data privacy and performance. Some organisations have also explored fine-tuning or training smaller language models tailored to audit-related terminology and regional languages. These initiatives reflect a growing recognition of the value of domain-specific AI tools over general-purpose models.

Cloud-native services are also part of the software landscape. Some institutions seeking more complex data processing and model deployment are experimenting with AWS Bedrock or integrating with Palantir Foundry. However, the adoption of commercial AI platforms remains cautious, often limited to pilot programs or isolated deployments. Institutions continue to be sensitive to issues of data residency and vendor lock-in, and many prefer software solutions that guarantee data ownership and operational transparency.

Security and compliance remain key drivers in software selection. Institutions have adopted policies to ensure that no private data is transmitted externally or used to retrain third-party models (such as Microsoft 365 Copilot). Some have implemented local policies or Chief Information Security Officer (CISO) oversight to approve which tools may be used, particularly when they interface with sensitive datasets. Using LLMs through interactive tools like online notebooks (e.g. Jupyter, Databricks) or secure, self-contained environments (e.g. Docker) that run in a web browser shows how organisations are balancing ease of use with the need to maintain control over data and systems.

Data infrastructure and governance

A critical foundation for any successful AI or digital transformation initiative is the quality, accessibility and governance of data. Among the institutions consulted, maturity levels vary widely: some have invested heavily in robust data lakes and structured pipelines, while others remain in the early stages of aligning disparate systems and establishing data governance standards, as emphasised in the OECD's (2019^[4]) work on data governance in the public sector.

Several organisations have built centralised data lakes that consolidate information from diverse internal and external sources, including financial records, corporate registries, government procurement data and other high-value datasets. In one institution, the data lake has grown to exceed 44 terabytes – i.e. large enough to hold millions of documents - supporting hundreds of users and underpinning nearly all audit activity. These more advanced environments are elevated by metadata management systems, automated risk indicators and robust data mining capabilities. Such institutions have often spent years, sometimes over a decade, to lay the technical and organisational groundwork for this level of data integration.

There are also numerous examples of institutions building systems to bridge structured and unstructured data. Where documents arrive in formats such as PDFs or scanned images, Python scripts and custom extract-transform-load (ETL) routines[†] are used to extract the content and structure it for analysis. In one case, audit teams use custom-built tools to transform raw files into structured tables, allowing auditors to

work with diverse datasets without the need for sophisticated analytical tools. This reflects a practical, use-case-driven approach to data infrastructure, where efficiency and audit-readiness take precedence over reliance on a generalized data architecture.

Data governance policies also vary. Several institutions have formal data strategies, often developed internally or in partnership with external advisors. These strategies typically emphasise data quality, standardisation and secure access. A few institutions have already begun developing governance layers around their AI tools, ensuring that human oversight is maintained for algorithmic outputs.

A recurring challenge across multiple organisations is the fragmentation of data sources. Legacy systems, siloed datasets and inconsistent formats are common issues. One institution, for example, faces overlapping databases and misaligned structures, particularly in contracts and public records. This highlights difficulties with horizontal integration, noting that even datasets from the same ministry can differ in scope and architecture. Adding to this complexity, SAs often rely on data they do not own (e.g., data from entities they are auditing), meaning they cannot control its quality. Moreover, audit topics are subject to frequent variation, resulting in significant changes to the datasets required for analysis. As a result, data integration and cleaning frequently occur after the fact, requiring substantial manual effort. These challenges hinder automation and limit the scalability of any type of analytics. In response, some institutions/governments are investing in standardisation and centralisation initiatives. One example involves a common audit management system under development, designed to align workflows across all agencies and improve the consistency of incoming data. Another institution is testing a system that integrates dashboards and analysis scripts into a single environment, allowing for smoother transitions from raw data ingestion to final output. These initiatives show a growing understanding that digital transformation depends as much – if not more – on consistent data infrastructure as it does on novel AI applications.

Efforts are also being made to ensure that data quality is maintained throughout the lifecycle. This includes not only the technical aspects of ingestion and transformation but also organisational processes that assign responsibility for data integrity. In a few institutions, the role of Chief Data Officer or an equivalent has been introduced to drive centralised data policy and oversee data cataloguing, reflecting a broader institutional shift toward treating data as a strategic asset.

While some institutions are still in the process of forming data governance teams or policies, others are experimenting with more advanced data architectures. These include vector databases[†] for use with large language models, as well as integration with services such as OpenSearch and semantic indexing protocols. Such developments are still relatively rare, but they suggest a transition toward more AI-native data infrastructures.

In general, SAs that have placed an early emphasis on data governance appear to be progressing more quickly in AI deployment. Their structured, accessible and well-documented datasets provide a solid foundation for ML and automation. Conversely, institutions with fragmented or undocumented data sources often find that even basic analytics require disproportionate effort, limiting the potential for scalable AI tools.

Global standards increasingly call for public institutions to maintain data environments that are secure, interoperable and auditable. As the AI field evolves, strong data governance and infrastructure will remain the primary enabler of digital and AI-based innovation. Institutions that build for long-term interoperability and governance, not just short-term use cases, will be best positioned to adopt more sophisticated solutions in the years ahead. As audit bodies increasingly rely on AI to guide findings and controls, the integrity of the data on which these tools are built will become even more critical.

Development capabilities

The ability of institutions to effectively develop and deploy AI and digital solutions is closely tied to their internal capabilities, organisational structures and access to specialised expertise. Across the cohort consulted, there is significant variation in how development work is organised, ranging from entirely in-house teams to hybrid models that rely on external consultants or research and partnerships. This variation reflects differing levels of AI maturity, budget constraints and strategic priorities.

Several institutions have opted for in-house development as their primary model, often driven by concerns over data security, control of intellectual property and long-term sustainability. These institutions typically maintain dedicated data science or innovation teams tasked with developing, testing and deploying AI tools directly aligned with audit needs. In such cases, development teams work closely with auditors to help identify and understand their problems and goals and determine whether an AI use case, or another approach, would be a practical and optimal solution. These teams are often composed of a mix of data scientists, developers and domain experts, sometimes supported by internal innovation labs or centres of competence that serve as institutional hubs for experimentation.

Among the more advanced organisations, in-house development has led to the creation of custom AI tools specifically tailored to audit tasks. Examples include systems for risk prioritisation, price anomaly detection, document summarisation, contract review and even RAG chatbots to support interpretation of auditing standards. These solutions are often built using open-source frameworks or internal platforms, allowing for a high degree of customisation. Institutions pursuing this route have emphasised the importance of integrating development cycles into the broader organisational workflow, rather than treating digital tools as standalone technical projects.

In contrast, several institutions operate under hybrid models that combine internal expertise with external support, generally through procurement of development and consultancy support or through partnerships with business or other research organisations. This is generally aligned with approaches in government organisations more broadly (OECD, 2025^[1]). These arrangements often arise where internal teams are small or still developing and external partnerships provide access to specialised skills or accelerated project delivery. In such cases, internal staff typically focus on project oversight, integration and long-term maintenance, while external consultants handle initial development or scaling. Collaborations with academic institutions, AI labs or large consulting firms are common, especially where institutions aim to build capacity over time or address complex technical requirements.

In institutions where dedicated AI teams do not exist, development is often driven by motivated individuals or small groups of staff with technical skills, frequently operating under a bottom-up innovation model. In such environments, staff members, often more junior employees or recent external hires, identify opportunities for automation or efficiency gains and propose digital tools to address them. Management support for such initiatives varies, but in a few cases, it has led to the emergence of informal innovation clusters that eventually evolve into formalised roles or departments. In these contexts, management support is essential both in setting an organisational tone that values innovation and in actively enabling the identification, validation, and implementation of high-quality ideas, allowing bottom-up innovation to move beyond ad hoc experimentation (OECD, 2025^[1]).

Nevertheless, the absence of structured development teams can pose significant challenges. Institutions lacking in-house capability may struggle to move beyond pilot projects, especially if they lack the internal resources to maintain or scale solutions developed externally. In such contexts, even promising proof-of-concepts risk becoming obsolete due to changing software environments, unmaintained codebases or staff turnover. Moreover, reliance on external partners without clear knowledge transfer plans or ensuring ownership of intellectual property rights and data ownership may inhibit the institution's ability to evolve or adapt its tools in response to shifting audit priorities.

Some institutions have attempted to address this by embedding development capacity into existing departments, such as IT, data analytics or innovation offices. In one case, two distinct teams, one focused on software development and data engineering, the other on analytics, are being considered for a merger to improve coordination and reduce duplication. Other SAs are still in earlier stages, working with external research institutes to develop foundational capabilities. In these cases, the long-term objective is typically to build internal expertise, even if early progress is heavily reliant on external partnerships.

Across the institutions examined, the most successful AI initiatives are those where technical expertise, domain knowledge and leadership support are tightly integrated. Development capability is not simply a matter of staffing or technical infrastructure, but of institutional alignment, governance and a shared commitment to innovation.

The analysis reveals a diverse landscape of digital maturity across the participating institutions. While some organisations have made significant investments in centralised data infrastructure, secure development environments and in-house AI capabilities, others are still laying foundational elements or relying on external partnerships to build capacity. Common challenges include fragmented data systems, limited internal specialist resources and the need for clearer governance around AI deployment. At the same time, there is strong evidence of innovation, collaboration and institutional commitment to integrating AI into public audit functions. Moving forward, continued investment in internal development teams, robust data governance and cross-sector learning will be essential to ensure that AI solutions are both impactful and aligned with public sector values.

5 Measuring impact

Measuring the impact of AI adoption in public audit is emerging as a key challenge, aligning with broader trends with regard to AI in government (OECD, 2025^[1]). While there is growing interest in understanding the return on investment from AI tools, the relationship between their use and tangible outcomes is often indirect, delayed or difficult to isolate. In many cases, AI tools, and generative AI models in particular, enable new types of tasks that were previously impractical, such as systematically exploring large volumes of documents or generating content outside of auditor working hours. These forms of productivity improvement are valuable, but not easily quantifiable. Similarly, improvements in audit quality or depth may be real but are hard to capture with current metrics.

This difficulty is also heightened by the fact that AI deployment and associated upskilling are still in early stages across most SAls. As a result, institutions have yet to develop comprehensive frameworks to assess the impact of these tools. In many cases, experimentation and piloting are ongoing and lessons learned are being integrated gradually. In the absence of robust metrics, SAls may consider incorporating structured qualitative feedback (e.g. user satisfaction surveys, auditor stories) to capture perceived value and usability. Comparative benchmarking with peer institutions or previous audit cycles could also help create informal but actionable indicators of progress.

Nonetheless, some organisations are beginning to define more practical indicators of advancement and value. One institution, for example, is tracking metrics such as:

- the percentage of the institution's departments actively using AI tools,
- the frequency of AI tool usage,
- time-to-value (the time it takes to see results from implementation),
- the number of staff trained, and/or
- the replacement of legacy processes by AI-supported workflows.

These remain a first layer of assessment of the widespread usage of AI within core functions; however, they do not yet constitute necessarily advanced analysis of the ROI or the impact of the use of AI.

Early gains have also been noted in terms of timeliness with knowledge management and writing assistance. In the consulted institutions where LLMs are trained on standardised audit documentation, the use of generative AI has enabled the rapid writing of first draft reports, resulting in significant time savings compared to traditional writing approaches. The ability to process materials outside of regular working hours, including automating processes overnight for example, has also been cited as a meaningful productivity boost.

More ambitious targets have also emerged. One organisation reported a clear objective to reduce its audit workforce by 50 positions, leveraging AI and automation to achieve this transformation. While such goals are uncommon, they illustrate how some institutions are exploring AI not just for incremental improvement, but for structural changes in audit operations.

Beyond the public sector, experience from private audit firms also offers insight. One firm highlighted that, through a combination of AI tools and process standardisation, audit turnaround times for auditees have dropped from 60 days to as little as 10 to 15 days. However, in this case, technology alone is not responsible: it's the combination of tools, the standardisation of data and reengineering of workflows, and

the retraining of staff that, combined, allow these gains. This reinforces the notion that real impact stems from the alignment of tools, processes and people, rather than the deployment of technology in isolation.

To help governments better overcome challenges in this area, the OECD is developing guide to measuring impact and demonstrating returns on AI investments (OECD, forthcoming^[5]).

6 Challenges and strategic considerations

This chapter outlines the key challenges and strategic directions for AI adoption in public auditing, as emerged from the consultations. Institutions consistently raised concerns about ethics, explainability, and model bias. Many also described organisational inertia as a discouraging factor for innovation. Data governance and security also surfaced as a prominent challenge, amplified by the need to keep pace with rapid technological developments and managing reliance on external providers. Together, these perspectives point toward strategic directions centered on flexibility and adaptability in both technical implementation and organizational frameworks.

Ethics, explainability and bias

In the consultations, organisations consistently raised concerns around ethics, integrity and transparency, especially in relation to AI-driven tools. The need to uphold public trust requires that AI used in audits is not only effective but also defensible, transparent and explainable. Several organisations cited risks of hallucinations, algorithmic bias and a lack of explainability in AI outputs; while different in nature, these risks could all pose threats to the credibility and validity of audit findings.

For instance, in consultations with the Netherlands Court of Audit, it was highlighted that current LLMs could better reflect different languages and cultures beyond those aligned with major producers. Language barriers compound these challenges, limiting the applicability of such models in non-English contexts. To address this, the Netherlands Court of Audit is awaiting the development of a Dutch language model, which aims to provide greater linguistic and cultural relevance, thereby enhancing fairness and trustworthiness in AI-assisted audits. Such an example is GPT-NL, which is investing around EUR 13.5 million provided by the Netherlands Ministry of Economic Affairs and Climate (EZK) to train a model (OECD, 2025^[1]).

To mitigate related challenges, many organisations are today focused on applying AI in areas that support auditors and improve productivity, but where, like with humans, a certain margin of error is acceptable. For example, AI can help review large volumes of documentation more efficiently than humans alone, without audit findings relying directly on the AI's output. The final responsibility remaining with the auditors is also a common mitigation measure to ensure all outputs are verified and handled with the appropriate level of professional judgement and scepticism. As one interviewee noted “hallucinations are a feature, not a bug”; given that it is currently impossible to eliminate them entirely, auditors must be trained to identify, critically assess and validate AI-generated outputs.

Several institutions highlighted the absence of audit standards for handling evidence generated or analysed by AI. This mirrors broader trends where governments organisations lack actionable guidance on how to use AI in a trustworthy manner in specific functions and domains (OECD, 2025^[1]). In practice, this means that AI-generated insights are today subject to the same critical scrutiny as those derived by traditional methods, but without clear guidance on how to validate or document their reliability. This lack of standards creates uncertainty, both for auditors, for quality assurance functions and for the public.

In response to questions on output transparency, Norway NAO has also initiated a research partnership with a university to advance work on explainability of AI. Algorithm explainability is also addressed today by working with models that provide some transparency by generating reports that explain their outputs and their sources (for more see Chapter 7. Good practices and success factors).

Open data practices which involve making data, and software code, freely available to the public, helps ensure transparency and accountability. By publishing anonymised datasets and codes, organisations allow the independent verification of results, thereby promoting trust in the process. Adopting open data standards, providing clear documentation of data sources and methodologies, and establishing independent review mechanisms can further promote transparency and trust. The OECD AI Principles (2024^[6]), endorsed by 48 countries and the European Union, offers value-based principles and recommendations for policymakers for the trustworthy use of AI, including the topics discussed in this section. The OECD Framework for Trustworthy AI in Government (Figure 2) provides a lens for applying these principles in the public sector.

Organisational culture, capacity and supporting adoption

AI adoption remains in the early stages in many public audit bodies consulted. Several organisations reported resistance to change and slow adaptation to AI and data-driven approaches in general, particularly among audit staff not trained sufficiently in the effective use of AI tools.

Effective user training is essential for the successful adoption of AI tools in auditing. Poor training may lead to inappropriate use and fears and risk aversion that impede adoption. Prompt design (i.e. how to craft instructions in plain language to elicit a response from an AI system) came up as a topic in several conversations, with the challenge of training users to design effective prompts for chatbots and LLMs. Indeed, poor prompt design not only reduces a tool's effectiveness, but can also lead to user frustration and slower/lower adoption rates. To address this, SAs and other audit organisations are developing training programmes for their staff (sometimes mandatory) and offering continuous learning. User training appears currently focused on the use of generative AI, while upskilling in other areas of AI remains mostly focal to IT and data experts.

The role of leadership and governance was highlighted in helping support change, which OECD (2025^[1]) work has raised a critical for successful AI adoption in any government organisation. The UK NAO in particular described an active approach to AI governance through:

- establishing an AI Governance Board to identify, assess and manage risks arising from new AI technologies and use cases
- recruiting a new AI and Data Analytics Director
- introducing assessment and review processes related to the use of new technologies such as AI within the system of quality management
- providing training and guidance materials for practitioners on the responsible use of AI.

Norway NAO anticipates cultural transformation challenges, identifying leadership as essential to drive organisational change and supporting the upskilling of auditors. They also highlighted an approach to change management through change agents to facilitate an internal culture shift to be more data-driven.

Data quality as a strategic asset

While the increasing availability of data allows the use of technology, data quality remains a fundamental challenge in the adoption of AI. Audit-specific tasks require structured, clean and often standardised data, yet many SAIs struggle with fragmented, inconsistent or unstructured datasets, particularly in text-heavy audit domains. Inconsistent data formats across auditees, years, departments and institutions can lead to errors in AI-driven audits as data limitations affect model accuracy and precision.

A private auditing firm noted that data harmonisation is a priority, with pilot models for data cleaning and establishing workflows. For SAIs, the challenge is even more complex as they typically do not own the data they audit. Instead, they must rely on the data management practices of other government entities or apply post-hoc tools to prepare the data for analysis. To overcome these challenges, organisations are investing in data preprocessing and cleansing tools, establishing data governance frameworks and, where possible, promoting data standardisation to enhance data quality and usability. To avoid bottlenecks, alignment between data coming into the organisation system and audit management tools or systems is critical, especially for inconsistent data handling such as timestamps.

Governance and security

Security is a critical concern in the adoption of AI for auditing. As powerful commercial AI products become more widely available and affordable, many SAIs have identified the use of cloud-based solutions as a particular challenge, especially in relation to data security and compliance with legislative and regulatory frameworks.

Security and compliance are also central to data governance efforts with several organisations operating under strict requirements that limit the use of personal or sensitive data in AI tools. In some cases, adherence to regulation and policies leads to local deployment of AI models to avoid the legal risks associated with transmitting data across borders or into third-party systems.

For example, Germany's FAO implemented a strict data ownership principle. Data owners can decide by themselves, up to a certain level of confidentiality, on the usage of data in external cloud-based tools like Microsoft Copilot. Others have deployed secure internal environments, such as encapsulated virtual desktops, to maintain control over data access and data residency (the physical or geographical location of the data). For some consulted institutions, privacy and geopolitical concerns have prompted a broader reconsideration of cloud solutions. For example, in Norway NAO, a shift towards on-premises infrastructure is being driven in order to reduce dependence on foreign-based cloud service providers.

However, use of commercial platforms is still common. Some are leveraging strong contractual agreements with major providers such as Microsoft to ensure data protection and integrity. Across all institutions, implementing robust data security measures and adhering to national or regional data protection regulations and laws, if applicable, are viewed as essential to maintain data security and compliance in the development of AI.

Managing the pace of change and reducing third-party dependencies

The rapid evolution of AI tools poses a strategic dilemma for SAIs. Technologies adopted today may become obsolete or unsupported soon after, raising concerns about legacy systems, long-term maintenance and sustainability and continuous staff training requirements. One institution described this dynamic as an "arms race" between public audit bodies and technology providers, where maintaining relevant capabilities requires constant investment and vigilance.

The pace of change carries not only financial risks but also strategic and operational consequences. Frequent tool turnover or upgrades can lead to fragmentation, duplication of efforts and what some referred to as “tool sprawl”, a situation where too many disconnected tools create confusion and inefficiencies rather than clarity.

Several SAls expressed concern about relying too heavily on commercial platforms. Commercial platforms allow audit institutions to keep pace with technological evolution, facilitate access to competence and technical solutions and ease maintenance. However, short product life cycles, shifting vendor priorities and data security challenges can leave SAls vulnerable to disruptions or being locked in with a specific provider.

To address this, some SAls such as in Korea are seeking to move to a model of in-house development to reduce reliance on third-party providers and regain control over key systems. However, this requires significant computational resources, staff with technical expertise and maintenance capacity, which many SAls currently lack. Fine-tuning LLMs, for example, to fit audit-specific tasks is heavily constrained today by the available computational power and expertise.

Balancing innovation, long-term stability and sustainability

Amid these challenges, institutions such as Singapore’s Accountant General’s Department (AGD) are searching for models of development that strike a balance between experimentation with operational stability while keeping in mind the costs of digital transformation.

Ensuring the long-term viability and scalability of AI solutions remains a critical challenge, particularly the ability to maintain and evolve systems internally throughout the AI system lifecycle. This challenge can be mitigated through modular technologies and designs with interchangeable components, which enable faster upgrades, smoother updates, and more efficient expansion, thereby enhancing flexibility and minimising downtime. Embedding structured feedback loops within tools’ development, use and refinement cycles further strengthens adaptability, ensuring that tools continuously evolve in response to user needs and contextual changes. Complementing this, sustained internal capacity-building is essential to equip institutions with the technical and organisational capabilities needed to manage and adapt AI solutions over time. Without adaptive design and continuous feedback, even technically sound tools risk irrelevance in real audit workflows.

To support these long-term stability and sustainability efforts, it is crucial when possible to incorporate new tools in an integrated and complimentary manner with infrastructure in software already in use or to take well-considered change management activities to replace existing approaches. Doing so helps ensure seamless integration into, or transformation of, existing workflows while maximising the value of prior investments. We can find an example of this in the UK NAO, where a proof of concept has successfully layered LLMs onto an existing tool to search the back catalogue of published audits (see Box 9). The success of these integrations is closely dependent on the level of adaptability of the systems and procedural flexibility.

Box 9. The UK National Audit Office (NAO): Integrating LLMs for semantic search and response generation

At the UK NAO, the integration of LLMs into an existing in-house application, the “Back Catalogue Analyser”, is currently at a pilot stage. This enhancement aims to introduce semantic search and generative AI-based natural language responses into a tool that currently relies on traditional keyword-based search methods.

The existing application serves as a back-catalogue search tool to locate and analyse past publications. While effective in its previous form, the tool is limited to searching for exact words or phrases within report texts. The planned upgrade represents a significant improvement, allowing for more intuitive and intelligent information retrieval.

By incorporating LLMs, the tool will be capable of understanding the semantic meaning behind user queries, thus facilitating more accurate and context-aware search results. This enhancement will also enable the generation of natural language responses, allowing users to engage with the system in a conversational manner. As a result, users will be able to retrieve relevant information with greater efficiency, without the need to reformulate queries or worry about precise wording.

The integration of LLMs is expected to deliver several key benefits, including a reduction in missed relevant content due to synonym or phrasing variations, improved handling of complex search expressions and the automation of tasks such as summarisation and theme extraction from report texts. This will not only streamline the research process, but also significantly reduce the manual effort currently required for tasks such as summarisation and theme extraction.

Source: UK National Audit Office, 2025.

7 Good practices and success factors

As audit institutions worldwide explore the integration of AI into auditing practices, a variety of approaches have emerged in the consultations that may offer valuable insights for organisations considering similar paths. The institutions consulted shared some key success factors and good practices. These insights, drawn from diverse institutional experiences, illustrate potential organisational, technical and cultural strategies for leveraging AI while emphasising audit integrity and stakeholder trust.

Transparency and explainability

As previously discussed, transparency and explainability of AI results has emerged as a foundational element in the adoption of AI tools, and in fact represent one of the five value-based principles in the OECD AI Principles (2024^[6]). SAls emphasised the importance of transparent AI practices and explainable AI outputs to maintain trust, but also to facilitate adoption by auditors. Maintaining an audit trail and having clarity on the reasoning behind AI outputs is essential, and institutions repeatedly noted that current models tend to lack full explainability. One institution stressed that when having to choose between different models, they always prioritise explainability over performance.

Some institutions address explainability challenges through how they organise their work. First and foremost, as seen, a large majority of use cases presented do not by themselves produce audit outputs but rather focus on efficiency enhancement.

At Brazil's CGU, their chatbot model has been developed to display the source report used to generate its response, thereby improving traceability. Similarly, the ECA is exploring the use of retrieval-augmented generation (RAG) architectures to ensure that AI outputs are strictly grounded in the institution's own documentation and data, strengthening the connection between inputs and AI-generated content.

Importantly, user training has emerged as the most effective safeguard to date, ensuring the sound use of AI tools. Training programmes focus on the appropriate use of AI, emphasising continued professional scepticism regarding AI-generated outputs, and reinforcing that AI should not be solely relied upon to derive audit findings. Many institutions are committed to ensuring that staff maintain vigilance by consistently verifying AI outputs against original source materials. As a best practice in some institutions, auditors must also systematically disclose AI has been used to generate outputs further supporting transparency.

User and stakeholder involvement: facilitating bottom-up innovation

Across consultations, multiple approaches to the development and adoption of AI have been observed. Some institutions adopt a bottom-up strategy driven by motivated users; others are guided by leadership and senior management's strategic direction, while many combine both approaches. In all cases, a strong culture of continuous user and stakeholder involvement appears crucial to success.

The level of involvement of users, which are almost always auditors, in the tool development process varies in reality. In some institutions, auditors play a key role in identifying operational pain points and defining

technical requirements. This participatory model ensures that digital tools remain aligned with real-world needs and are more readily adopted by frontline staff. In other cases, auditors are themselves beginning to write scripts or adapt existing code, particularly in organisations where digital literacy is more widespread, or where recently graduated auditors are entering the workforce with strong digital or coding skills. This democratisation of AI development activities reflects the increasing accessibility of data science tools and a growing expectation that professionals across disciplines engage directly with digital workflows.

Bottom-up innovation has been a key driver of AI adoption in institutions such as Costa Rica's CGR and the ECA. In Costa Rica, where the institution is rather young, an organic approach to innovation emerged. Staff members from diverse roles who demonstrated motivation and promising ideas were granted paid access to a commercial AI tool to explore and apply. Choosing an off-the-shelf, paid platform helped boost adoption by addressing concerns around data privacy and security that are often associated with free tools (for more see Box 10).

Similarly, at the ECA, the initial rollout of AI focused on empowering motivated individuals across a range of roles. Early AI users, with their enthusiasm and diversity of perspectives, were instrumental in helping to cultivate a bottom-up process of idea generation and experimentation.

In the context of participatory innovation, agile methodologies have been shown to enhance user engagement in digital projects. For example, the Netherlands' Court of Audit has taken a very agile approach to tool development whereby they create prototypes, share them with the auditors/users to test, receive feedback and improve the tools in an iterative cycle that enhances digital solutions development.

Singapore's AGD takes a gradual approach to building capabilities in digitalisation and use of AI technology for internal audit. The internal audit professionals are encouraged to embark on pilot initiatives or use cases to explore potential benefits. If there is value to scale success cases, this is done by sharing of good practices and tools across the profession.

Appointing product owners with prior audit experience to lead the development of AI pilots and tools was identified as another success factor. These hybrid profiles (audit + digital) help bridge the gap between technical AI development and the practical needs of audit functions. In the absence of those hybrid profiles, bottom-up innovation can still be effectively fostered through a pragmatic approach: connecting individuals who have specific needs with those who can develop or provide solutions. Open feedback mechanisms, featuring rapid response cycles and a 'no-blame' culture, combined with realistic expectations and ongoing support, foster continuous improvement and strengthen an effective bottom-up innovation process.

Management and leadership support

Strong leadership and active management support are critical accelerators of successful AI adoption, especially in situations where innovation is hindered by risk aversion or bureaucratic inertia. Leadership helps set the tone for experimentation, allocating resources and legitimising the use of new technologies, and digital exploration. Top-down support further ensures that AI projects are aligned with the institutional goals.

In the case of Costa Rica, for example, senior management actively endorsed AI projects, creating a permissive environment that encourages creativity, experimentation and calculated risk-taking (see Box 10).

The level of institutional ambition also shapes AI development trajectories. While some organisations take a cautious, incremental approach – focusing on scaling successful pilots or building infrastructure for experimentation – others have set bolder targets. One institution, for example, has set a target to reduce audit headcount through AI implementation, reflecting a bold commitment to reimagining traditional

workflows (see Measuring impact). Others have set more incremental goals, such as scaling successful pilot projects or building stable platforms for ongoing experimentation.

Box 10. Costa Rica's Comptroller General of the Republic (CGR): Facilitating adoption

Bottom-up innovation

At Costa Rica's CGR, an organic innovation culture emerged, with staff across different roles encouraged to explore AI applications. All personnel have access to AI capabilities integrated within the institution's productivity platform, which has facilitated widespread experimentation. Free tools are also explored, but only after review by the IT unit and the Information Security Officer, and always in line with institutional guidelines on the use of artificial intelligence. These safeguards help mitigate risks related to confidentiality, data protection, and responsible use.

Crucially, this bottom-up culture is embedded within a broader governance framework. The audit office has established a dedicated Innovation and Learning in Oversight unit, which leads experimentation initiatives and manages organisational learning. It has issued formal Guidelines for the Management of Experimentation Initiatives, which define principles such as co-creation, prototyping, and tolerance for failure, while requiring structured processes for risk assessment, validation, and systematisation of results. In parallel, the institution has also adopted official Guidelines for the Ethical and Responsible Use of AI, which establish principles of responsibility, transparency, fairness, respect, and privacy, as well as specific duties for staff, the IT unit, and the Information Security Officer in approving and monitoring tools.

Together, these frameworks ensure that innovation in Costa Rica's CGR is both open and responsible: encouraging staff-driven experimentation while safeguarding ethical standards, institutional integrity, and the sustainability of AI adoption.

Leadership support

In Costa Rica, senior management has played a central role in embedding innovation into the institution's strategy and culture. The Institutional Strategic Plan 2021-2026 explicitly identifies *learning and innovation* and *technology as a transformative tool* as strategic drivers, with clear objectives and measurable targets related to artificial intelligence, data-driven oversight, and organisational cohesion.

This top-level commitment is reinforced by formal governance instruments. The audit office has issued Guidelines for the Management of Experimentation Initiatives, providing a structured framework for prototyping, co-creation, and risk-managed innovation, and Guidelines for the Ethical and Responsible Use of AI, which establish principles of transparency, security, and institutional responsibility. Together, these instruments ensure that experimentation with AI is both encouraged and safeguarded.

Leadership in CGR has thus legitimised digital exploration, allocated resources, and created a safe environment for staff to innovate, in line with the institution's strategic objectives. This alignment between leadership vision, strategic planning, and operational guidelines has fostered a culture that encourages creativity, calculated risk-taking, and sustainable adoption of AI in public audit.

Source: Costa Rica's Comptroller General of the Republic (CGR), 2025.

Institutional governance and oversight

As AI initiatives grow in scale and complexity, institutions are increasingly formalising internal governance mechanisms to coordinate efforts and ensure strategic alignment. Councils or steering groups have been established in several organisations to review project proposals, oversee implementation, develop guidelines and policies, and monitor ethical and legal compliance. These bodies, often composed of representatives from leadership, IT, audit and legal functions, serve both as champions and safeguards for accountability. One example includes the creation of both an AI council and Executive council at an SAI that jointly oversee roadmap execution and resources allocation. These governance structures not only enhance transparency and coordination but also contribute to a more systematic and sustainable approach to digital transformation.

At the ECA, an Advisory Committee on AI - including representatives for Legal, Management, Ethics, Information Security and DPO - has been created to facilitate institutional compliance with the regulatory framework governing the use of AI, coordinate the assessment and management of AI risks, and act as a source of internal expertise on AI-related matters.

In addition to governance bodies, clear and accessible internal guidelines are essential to bridge the gap between high-level regulations, professional standards and day-to-day operational practice. Well-crafted and enablement-oriented guidelines help translate strategic objectives and compliance requirements into actionable steps, thereby supporting effective and consistent AI adoption across the institution. For example, the German FAO developed a guideline for the adoption of technology that incorporates AI. This includes data usage as well. The guideline focuses on brevity and readability. To support the practical implementation of this guideline, the FAO additionally offers AI workshops. In these workshops, employees learn the practical and theoretical fundamentals of AI and also become familiar with the compliance framework of the FAO. To date, over 300 employees of the FAO participated in these workshops.

Effective training and knowledge sharing

Effective training programmes are a critical component in the successful adoption and diffusion of AI within SAIs. The competencies to be developed across audit teams span a broad range, from technical knowledge and regulatory understanding, to curiosity, ethics and professional scepticism. A common challenge lies in bridging the gap between general training and practical application. As with other digital knowledge, without immediate use, newly acquired knowledge can quickly fade, undermining the value of the training itself.

To address this, Singapore's AGO implemented a model that requires participants to apply what they have learned after completing training. Staff are expected to initiate a project or work activity involving AI, helping to reinforce learning, encourage adoption and ensure that skills are retained through real-world application.

Beyond technical training, institutions are also embedding upskilling initiatives within broader change management strategies aimed at building a digital culture. Recognising that cultural resistance can hinder AI adoption, some SAIs are actively investing in leadership engagement, structured communication and other initiatives to shift mindsets across the organisation. Several institutions have launched internal training academies or partnered with academic institutions to deliver targeted learning in AI and data science. These efforts are designed not only to transfer knowledge but also to foster curiosity, digital confidence and long-term organisational adaptability.

At the ECA, a multi-touchpoint approach to training and knowledge sharing has been adopted to facilitate the staff's familiarisation with AI. By offering a range of engagement formats, from formal instruction to informal peer exchanges, this strategy aims to build familiarity, confidence, and practical understanding across the organisation (see Box 11).

Box 11. A holistic approach to AI training: The experience of the European Court of Auditors (ECA)

The ECA adopts a comprehensive strategy to build AI competency among staff.

Given the wide availability of generalist AI training, the ECA has shifted its focus toward coaching and targeted support, ensuring that staff develop both theoretical understanding and practical skills for AI application.

To support the broader cultural shift required for AI adoption, the ECA has combined formal training with informal, peer-based learning and knowledge sharing.

- **Official Training:** Generalist AI courses are offered through formal ECA programs and external online sources.
- **GPT Cafés:** Informal gatherings, with 20-30 internal participants for each session, focused on discussing AI topics raised by participants.
- **Knowledge Node:** An online, curated knowledge repository housing articles, guidelines, and relevant AI information for ongoing self-directed learning. Furthermore, a prompt library that will compile useful AI prompts will be included to enhance practical AI use.
- **Functional forums:** Dedicated communication channels such as a functional mailbox and dedicated Teams channels where staff can submit AI-related questions and receive timely expert responses.
- **Conferences:** Participation in internal and external conferences facilitates exposure to broader AI developments and best practices across institutions.

Source: European Court of Auditors, 2025.

Strategic partnerships

Strategic partnerships can also enhance institutional capabilities without the need to acquire a multitude of technical capabilities and without the complications of developing and maintaining proprietary technology. Some organisations rely on a combination of internal expertise and multiple collaborations involving industry, academia and other national or international institutions to leverage specialized knowledge and foster innovation. Recognising the rapid obsolescence of LLM versions for instance, they favour working with external providers to deliver bespoke solutions while retaining ownership of the data and the code, ensuring simultaneously innovation and control.

While globally, it is considered best practice to combine internal development with strategic external partnerships, particularly in the early stages of capability building, long-term success and sustainability depends on investing in dedicated teams, formalising development processes and aligning AI initiatives with organisational priorities. Without these structural supports, institutions risk becoming overly dependent on external actors or unable to adapt tools as needs evolve.

8 Future directions

Looking forward, the most advanced institutions are beginning to explore the next frontier of AI capabilities: “agentic” AI systems that act autonomously – with minimal human input – towards defined goals, continuous intelligence for real-time insights, reasoning models to improve transparency, and multimodal solutions capable of processing diverse data types.

While not yet using agentic AI in practice, audit institutions are actively researching how it could help reveal new insights and uncover risk areas not easily detected through traditional means. Several of them see potential within the next 3 to 5 years, with longer-term visions extending up to 20 years. One institution is exploring the use of AI agents to simulate fraud scenarios and detect document tampering, such as identifying fake certificates. Korea BAI aims for AI systems to eventually draft entire audit reports. However, these experiments are approached with caution due to quality control and trust concerns. Norway NAO envisions future tools enabling natural language or voice-based interaction with financial data and the automation of managerial review and quality assurance processes to improve efficiency.

As technology progresses, the Netherlands’ Court of Audit envisions the use of AI to support a continuous stream of data and information, enabling auditors to stay consistently informed in near real-time. Rather than relying solely on periodic reports or manual data collection, auditors could receive daily updates relevant to their sectorial focus, for example, a healthcare auditor would automatically be alerted to new developments such as parliamentary discussions, media reports, regulatory changes or emerging risks. This proactive model would allow auditors to maintain a current, comprehensive view of their domain, significantly enhancing their ability to identify issues early and respond more effectively.

In the nearer term, some organisations are focusing on reasoning models that provide transparency by showing how conclusions were reached, potentially supporting audit trail requirements. These models could record the rationale behind outputs and make AI-generated insights more accountable, potentially bringing them closer to AI outputs that meet auditing standards. Emerging use cases also touch on the integration of multimodal models (that can process and integrate information from multiple types of data), including tools that can extract information from poor-quality images, such as scanned or damaged images of birth certificates. These capabilities may significantly broaden the scope of AI’s application in document-heavy audit environments.

Other areas of focus for future applications include more advanced predictive analytics, for example developing models to detect procurement price deviations from market rates. Other institutions are looking inwards first, recognising the need to better exploit the large volumes of data already available to them before expanding into more complex AI solutions.

Looking ahead, the enhancement of analytical capabilities and the emergence of more advanced AI approaches, such as agentic AI, may have the potential to automate entire processes and associated decision-making. While adoption will be cautious, this shift may require a fundamental rethink of the current auditing methodologies and standards, as well as updates to relevant regulations and legal frameworks, to ensure that audit conclusion remains valid and reliable. The rapid evolution of these technologies further reinforces the need for critical skills and an adaptive digital culture across institutions.

Alongside these developments, the broader public-sector context is evolving, with many central government bodies, including ministries of finance and treasuries, increasingly developing in-house AI

capabilities. This creates potential opportunities for closer interaction between audit institutions and other public administrations in the area of AI. Such interaction could help limit duplication of effort, enable audit-relevant requirements, such as traceability, explainability and governance, to be considered earlier in the design of AI systems, and contribute to the accumulation of shared expertise. While this is not yet a widespread practice, it reflects a possible future direction within an increasingly interconnected public-sector AI ecosystem.

9 Conclusions

The integration of AI-supported methodologies can produce significant gains in terms of productivity (efficiency and effectiveness), responsiveness, and accountability in public auditing procedures. Advanced models enable the automatic detection of anomalies in financial data and support preventive actions by signalling risky situations. AI can also improve auditors' knowledge management by facilitating the extraction of key information from large volumes of textual documents and the compilation of structured datasets from unstructured sources. In addition, LLMs can support the drafting, summarising and translation of technical material, ensuring greater consistency, while computer vision and deep-learning models enable the assessment of physical infrastructure through the analysis of visual data.

The cross-institutional analysis highlighted that the transformative potential of AI in auditing depends as much on the underlying hardware and software infrastructure as it does on data management processes and development capabilities. Hardware and software choices reveal a trade-off between the upfront investment in autonomous solutions and the reliance on external providers. This reflects considerations of data security, compliance and vendor lock-in. While fragmented data sources and heterogeneous architectures remain significant bottlenecks, greater data quality and standardisation may be achieved through investments in preprocessing tools and clearer governance frameworks. Development capabilities vary widely, and institutions that have built in-house teams to develop tailored AI solutions appear better positioned to achieve a successful and sustainable AI transition in the long term.

Strategic considerations extend beyond infrastructure. The lack of audit standards for the use of AI in auditing creates uncertainty, while ethical and legal concerns about explainability, transparency and biases underscore the need for human oversight. In this scenario, organisational culture and hands-on training play a decisive role in accustoming users to critically evaluate AI outputs and promoting their active engagement in the development process. Successful examples also show that bottom-up innovation, when supported by strong governance and management commitment can significantly boost adoption and ensure AI solutions remain both effective and accountable.

So far, AI offers a powerful complement to traditional audit methods, but not a substitute of professional judgement. Its effective deployment calls for holistic strategies that combine infrastructure capacity, governance frameworks and staff training. By enhancing data quality, building adaptive capacities, and promoting an innovation-friendly culture, audit institutions can not only improve their operational effectiveness but also reinforce public trust in the integrity and accountability of government.

The growing use of AI tools has significant implications for internal control systems within audit institutions. As AI becomes integrated into risk assessment, data analysis and evidence generation, these tools effectively become part of the audit process and must therefore be subject to the same control principles as other critical audit systems. A strong internal control environment, including clear process ownership, responsibilities, robust data governance, access controls, validation procedures, documentation, audit trails, etc., is a prerequisite for the reliable and responsible use of AI. Where internal controls are weak or fragmented, AI is likely to amplify errors and inconsistencies rather than improve audit quality. Embedding AI within a sound system of internal control is therefore essential to ensure that technological innovation strengthens, rather than undermines, audit credibility, accountability and public trust.

Beyond their internal use of AI, audit institutions are increasingly confronted with the broader question of how AI is designed, deployed and governed across the public sector. As governments integrate AI into

core policy, financial management and service-delivery, SAs are likely to play a central role in examining how accountability, transparency and controls are maintained in AI-augmented decision-making processes. This places SAs in a position not only to audit AI-enabled systems, but also to help structure the public-sector conversation on responsibility, explainability and oversight, drawing on their mandate and experience in assessing governance arrangements. In this evolving context, auditing AI may become as much about scrutinising institutional frameworks and safeguards as about assessing individual technologies.

Glossary

The following definitions are offered to facilitate the reader's understanding and do not represent official OECD definitions, unless otherwise noted.

Application Programming Interface (API): refers to a set of routines, protocols, and tools for building software applications. An API specifies how software components should interact with each other to exchange data, features and functionality. A good API makes it easier to develop a program by providing all the building blocks. (Source: European Commission)

Cloud-native services: services made available to users on demand via the Internet from a cloud computing provider's servers as opposed to being provided from a company's own on-premises servers. (Source: European Commission)

Convolutional Neural Networks (CNNs): a type of deep learning algorithm designed to process and analyse visual data such as images and videos. CNNs use layers of interconnected nodes that apply convolution operations to detect and learn patterns like edges, textures, and shapes, making them highly effective for tasks like image recognition, classification, and computer vision. (Source: IBM)

Crawler: an automated programme that searches the Internet for new Web documents and places their addresses and content-related information in a database, which can be accessed with a search engine. (Source: European Commission)

Custom ensemble approach: techniques in machine learning that combine multiple models to improve prediction accuracy by reducing errors and increasing robustness. (Source: IBM)

Domain-adapted LLM: refers to a large language model (LLM) that has been trained or fine-tuned for a field or subject matter, so to achieve better performances on domain-specific tasks, compared to a general-purpose LLM. (Source: IBM)

Extract-transform-load (ETL): a data processing workflow that extracts data from various sources, transforms it into a suitable format, and loads it into a destination system, such as a data warehouse. (Source: IBM)

Graph-based model: a data structure or algorithm that represents information as nodes (entities) and edges (relationships), useful for analysing complex networks like social connections or knowledge graphs.

Graphics processing unit (GPU): refers to an electronic circuit specifically designed for digital image processing and to accelerate computer graphics. The ability of GPUs to rapidly execute vast numbers of calculations has made them indispensable in artificial intelligence, where they are used to manage data-heavy and computationally intensive tasks. (Source: IBM)

Intelligent Document Processing (IDP): refers to the use of AI to automatically extract structured data from unstructured or semi-structured sources, such as pdfs, scanned documents, images and handwritten forms. [definition from the report]

Internet of Things (IoT): a network of interconnected physical devices embedded with sensors and software, enabling them to collect, exchange, and act on data. (Source: Oracle)

Intersection over Union (IoU): also known as the Jaccard index, measures how much two shapes that locate the same object in a picture overlap, by comparing the areas of the predicted box and the true box. A higher IoU corresponds to a more overlap and thus to a better match. IoU is the most popular evaluation metric for tasks such as segmentation, object detection and tracking. (Source: Stanford University)

Labelled data: data points that have been annotated with one or more labels indicating their meaning, category, or context. These labels serve as reference outcomes during model training, allowing models to learn by comparing their predictions against the assigned labels.

Large language models (LLMs): refer to a class of artificial intelligence models trained on vast amounts of text data to understand and generate human-like language. (Source: Stanford University)

Machine Learning (ML): is a set of computational techniques that allows machines to learn in an automated manner through patterns and inferences rather than through explicit instructions from a human. Machine learning approaches often teach machines to reach an outcome by showing them many examples of correct outcomes. However, they can also define a set of rules and let the machine learn by trial and error. (Source: OECD)

Model drift: refers to the degradation of machine learning model performance due to changes in data or in the relationships between input and output variables. Model drift—also known as model decay—can negatively impact model performance, resulting in faulty decision-making and bad predictions. (Source: IBM)

Natural language processing (NLP): the field of artificial intelligence focused on enabling computers to understand, interpret, and generate human language in a meaningful way. (Source: IBM)

Overfitting: In machine learning, overfitting occurs when an algorithm fits too closely or even exactly to its training data, resulting in a model that can't make accurate predictions or conclusions from any data other than the training data. (Source: IBM)

Optical character recognition (OCR): a technology that converts different types of printed or handwritten text from images or scanned documents into machine-readable and editable digital text. (Source: IBM)

Production-grade: refers to software or systems that are reliable, scalable, and robust enough for real-world deployment and use in operational environments. (Source: IBM)

Retrieval-Augmented Generation (RAG): an AI technique that enables a generative model to retrieve information from targeted data sources to produce responses informed by a specified set of documents. (Source: IBM)

Robotic Process Automation (RPA): is a technology that uses software “robots” to automate repetitive, rule-based digital tasks. It mimics human interactions with systems and applications to complete processes faster and with fewer errors, improving efficiency and allowing employees to focus on higher-value work. (Source: IBM)

Self-contained environments: isolated computing setups that include all necessary software and dependencies, ensuring consistent and reproducible performance across different systems. (Source: IBM)

Semantic research: is a data searching technique that aims at understanding the contextual meaning and intent behind a user's search query, rather than only matching keywords. (Source: Google)

Spark clusters: a group of interconnected computers working together to process large-scale data using Apache Spark, a fast and general-purpose cluster computing system. (Source: Apache Spark)

Supervised machine learning: a type of machine learning where models are trained on labelled data, learning to predict outcomes based on input-output examples. (Source: IBM)

Unsupervised machine learning: a machine learning approach that identifies patterns or groupings in data without labelled outcomes, often used for anomaly detection. (Source: IBM)

Vector databases: a database optimised to store and search data represented as vectors (numerical arrays), commonly used for similarity search in AI applications like recommendation systems. Vector databases play a key role in enabling advanced AI applications by efficiently storing and searching high-dimensional data representations, known as vectors or embeddings with AI models like large language models converting text, images, or other data into these vectors to capture their meaning. (Source: Pinecone)

References

- (n.a.) (2025), *The Agentic AI landscape: Conceptual foundations, key features and illustrative examples*, [https://one.oecd.org/official-document/DSTI/DPC/GPAI\(2025\)18/en](https://one.oecd.org/official-document/DSTI/DPC/GPAI(2025)18/en) (accessed on 13 January 2026). [8]
- CGU (2025), <https://www.gov.br/ancine/pt-br/assuntos/atribuicoes-ancine/fomento/prestacao-de-contas/malha-fina>. [3]
- Huang, Q. et al. (2024), “A graph says more than a thousand journal entries: Harnessing graph autoencoder networks in auditing”, https://www.efk.admin.ch/wp-content/uploads/publikationen/fachtexte/2024_12_a_graph_says_more_than_a_thousand_journal_entries.pdf (accessed on 30 October 2025). [2]
- OECD (2025), *Governing with Artificial Intelligence: The State of Play and Way Forward in Core Government Functions*, OECD Publishing, Paris, <https://doi.org/10.1787/795de142-en>. [1]
- OECD (2024), “Governing with Artificial Intelligence: Are governments ready?”, *OECD Artificial Intelligence Papers*, No. 20, OECD Publishing, Paris, <https://doi.org/10.1787/26324bc2-en>. [7]
- OECD (2024), *Recommendation of the Council on Artificial Intelligence*, <https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449>. [6]
- OECD (2019), *The Path to Becoming a Data-Driven Public Sector*, OECD Digital Government Studies, OECD Publishing, Paris, <https://doi.org/10.1787/059814a7-en>. [4]
- OECD (forthcoming), *Governing with Artificial Intelligence: Measuring impact and demonstrating returns on AI investments*. [5]