



Generative AI and official statistics:

the project of the UNECE High-Level Group for the Modernisation of Official Statistics

Disclaimer: The views and opinions expressed in this presentation are those of the author and do not necessarily reflect the official policy or position of UNECE

Authors:

Amilina Kipkeeva, InKyung Choi, Olivier Sirello,
Vytas Vaiciulis

The UNECE HLG-MOS Project on Generative AI

The HLG-MOS, established by the CES, leads the modernization of official statistics through collaboration among national and international organizations.

The project on generative AI focuses on governance, project management, technical aspects and explores prominent use cases.

Key objectives include sharing experiences, mapping AI's impact on statistics, and developing best practices for responsible AI implementation.



Collaborative Effort

The project serves as a platform for national and international organizations to share experiences, map the footprint of AI, and discuss its opportunities and limitations.



Strategic Considerations

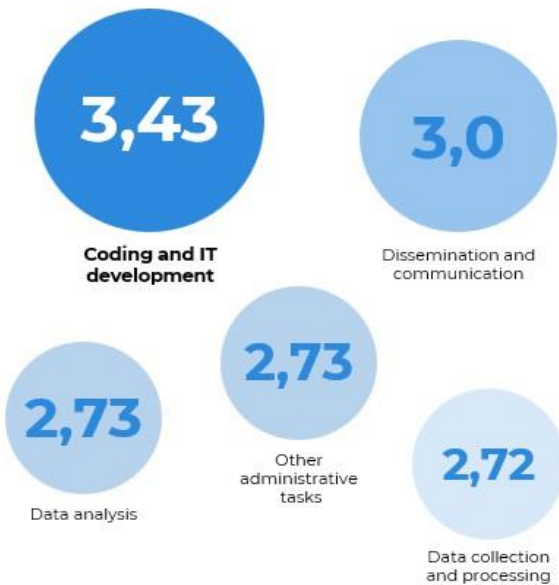
The project investigates governance and ethical considerations, as well as practical aspects like prompt engineering, infrastructure, and tools.

CES Survey on Generative AI in Statistical Organizations

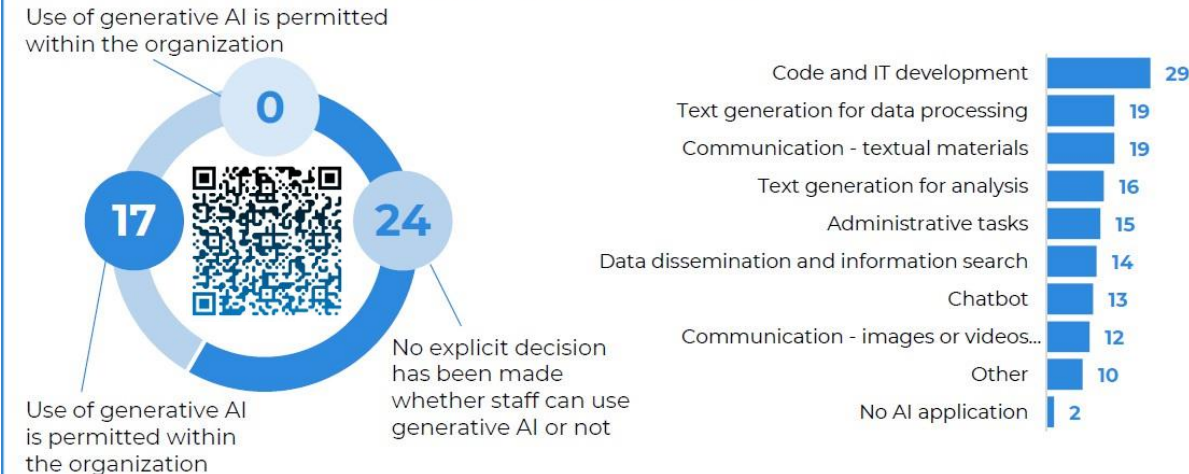
Overview of Respondents



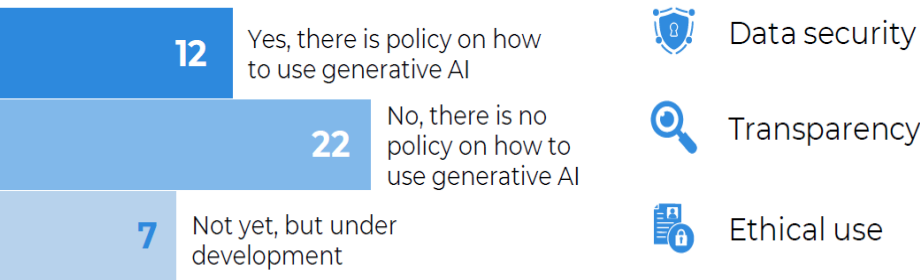
Impact of Generative AI*



Usage and Permission*



Organizational Policies and Guidelines



Challenges and Concerns*



* Weighted average score

Use Cases in Data Processing and Analysis



Metadata Editing

The Bank for International Settlements (BIS) developed AI-powered assistants for metadata editing, significantly reducing the time and effort required and enhancing consistency and accuracy.



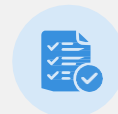
Synthetic Data Generation

LLMs are used to generate synthetic datasets for safe data exchange and internal operations, improving data confidentiality and balancing training datasets.



Statistical Classifications

The Australian Bureau of Statistics used LLMs to update occupation codes, achieving 70% quality compared to human-created lists and saving over 1600 working hours.



Report Generation

Statistics Canada experiments with LLMs to generate draft reports from Census data, creating detailed articles that are reviewed for accuracy. The process includes generating summary statistics and bilingual report translation.



Use Cases in Code and IT Development

Code Translation and Generation

The Central Statistics Office (CSO) Ireland developed the "SAS to R Transcompiler" using GPT-4, automating the translation of SAS code to R. This tool saves significant time and reduces manual correction needs.



Coding Assistants

Generative AI assists in various coding tasks, such as code generation, debugging, and documentation, enhancing code quality and robustness.



Open-Source Transition

Many statistical offices are transitioning to open-source programming languages like R and Python, facilitated by AI-powered code translators and generators.



Use Cases in Communication



Social Media Content

The UNECE Statistics Division uses LLMs to create coherent social media posts, ensuring consistency and brand recognition.



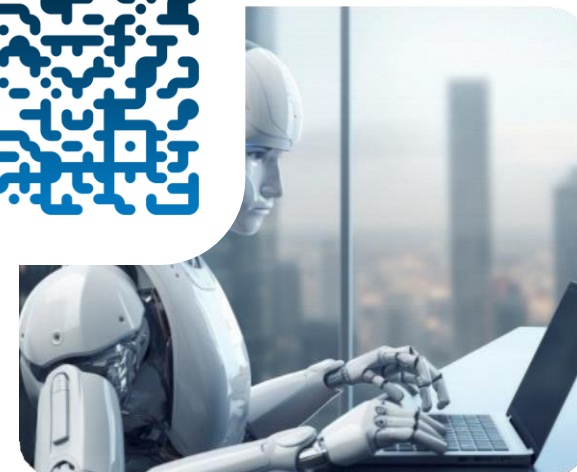
Non-Textual Content

Statistics Canada explores generating narrated content and visuals to enhance data accessibility and engagement with the public as well as experimenting with using AI for video production.



AI-Powered Chatbots

AI chatbots improve user interaction with statistical data by providing intuitive and user-friendly interfaces for answering data-related questions.



Frameworks for Generative AI in Official Statistics

Unified Framework

Establishing a unified framework for sharing experiences and best practices is crucial for the responsible adoption of generative AI in official statistics.



Governance and Ethics

Developing guidelines tailored to the needs of statistical organizations, focusing on accuracy, privacy, transparency, and trust.



Technical Considerations

Emphasizing prompt engineering strategies, quality assurance, and the selection of appropriate AI models. Criteria include scalability, performance, and cost-effectiveness.



International Collaboration

Highlighting the need for international cooperation to share knowledge, best practices, and develop common standards and guidelines for AI use in official statistics.



Thank you for attention

Amilina Kipkeeva

✉ amilina.kipkeeva@un.org