UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE

CONFERENCE OF EUROPEAN STATISTICIANS

Expert Meeting on Statistical Data Editing

7-9 October 2024, Vienna

Editing metadata with generative AI

Olivier Sirello (Bank for International Settlements, Switzerland)

olivier.sirello@bis.org

I. Introduction[1]

1. The advent of generative AI holds great potential for the editing of metadata in official statistics. Producing and processing metadata are typically demanding tasks for compilers, often involving considerable manual effort and meticulous review, making it both complex and resource-intensive. However, the rapid development of generative AI offers promising opportunities for optimising metadata editing and improving compilers' efficiency. Large language models (LLMs), in particular, can help to generate and refine text with human-like proficiency due to the sophisticated capability of transformers to capture word relationships.

2. This work sheds light on the BIS Metadata AI Editor, a project developed by the BIS Data Bank which leverages AI assistants to edit time series metadata. Made by statisticians for statisticians, the solution has been designed to be light in terms of implementation, highly modular and customizable to respond to the changing business' needs. It also supports the Statistical Data and Metadata eXchange (SDMX) standard which streamlines the editing process and integration within the statistical pipeline.

3. The paper is structured as follows: first, it sheds light on the fundamental role played by metadata in official statistics, not least their critical task to turn data into information. Second, it presents the BIS Metadata AI Editor, its components, workflow, and interface. Finally, it discusses the limitations, challenges and risks associated with editing metadata with generative AI.

[1] This paper draws on the "White Paper on Large Language Models for Official Statistics", High-Level Group for the Modernisation of Official Statistics (December 2023) as well as Bogdanova et al (2024) presented at twelfth Biennial Conference of the Irving Fisher Committee on Central Banks Statistics (IFC).

The views expressed here are those of the author and do not necessarily reflect those of the Bank of International Settlements.

The author thanks Bilyana Bogdanova and Magdalena Erdem for helpful comments and feedback. Frank Herrmannsdoerfer, Yannis Seemann and Krzysztof Zdanowicz are also thanked for their excellent collaboration for developing the BIS Metadata AI Editor.

II. The fundamental role of metadata in official statistics

A. Metadata transform data into information

4. **In official statistics, metadata hold a pivotal role for accessing, analysing, interpreting, and communicating data**. According to the Generic Law on Official Statistics (article 4), metadata are “means data and other documentation that describe statistical data and statistical processes in a standardised way by providing information on data sources, methods, definitions, classifications and data quality”. In that, they provide contextual information which is crucial for transforming raw data into meaningful insights. Ultimately, they enable and promote the understanding of statistics, as well as its competent interpretation by providing details about the data (Garrett, 2024).

5. **More specifically, structural metadata provide the semantic context necessary for the identification, understanding, and interpretation of the data**. For instance, a simple numerical figure, such as "100", lacks inherent meaning without the accompanying metadata that specify its context, such as the measurement unit, the period, and other characteristics required for its identification. In a way, metadata give and shape the meaning to data, transforming pure numbers into information that can be understood and used effectively. This semantic context is indispensable for data users to correctly understand and identify the data.

6. **In addition, reference metadata are also essential for describing the statistical process**. They provide details about the entire lifecycle of data, including data collection methods, compilation techniques, processing procedures, and dissemination practices. This information is invaluable for evaluating various aspects of the data, such as its accuracy, completeness, reliability and, ultimately, trustworthiness. For instance, metadata detailing the sampling methodology or processing allow assessing the presence of any potential biases. More fundamentally, by providing transparency into the statistical process, metadata enable stakeholders to understand how data were manufactured and to critically evaluate their quality and relevance.

7. **Further, metadata are a necessary condition for the exchange and sharing of data**. They ensure that data are communicated in a manner that is consistent, accurate, and accessible across different systems and platforms. Without proper metadata, data-sharing becomes problematic, as there would be no standardized way to interpret or utilise the data across diverse contexts. Metadata standardise data formats, definitions, and classifications, facilitating seamless data integration and interoperability (UNECE, 2024a).

8. **No doubt that the importance of metadata is amplified by the increasing volume and diversity of data available today**. With the rise of machine learning and AI, producers can tap into the abundance of alternative sources, prompting the need for robust data integration and linkage frameworks and pipelines (UNECE, 2024b). Here metadata play a key role, not least to inform users about the usability of data and their correct interpretation. Metadata also play a critical function in enhancing data discoverability and accessibility. They provide a necessary layer for the development of data standards and protocols that ensure data can be easily found, accessed and, perhaps more importantly, reused. In this regard, the FAIR Guiding Principles for scientific data management and stewardship stress the importance of creating metadata components which can be actionable by machines (Wilkinson et al, 2016).

9. **More fundamentally, comprehensive, and transparent metadata are critical for the reliability, credibility, and trustworthiness of official statistics**, aligning with the Fundamental Principles of Official Statistics and the Principles Governing International Statistical Activities.

Specifically, the third principle of the Fundamental Principles of Official Statistics stresses the need for statistical agencies to “facilitate a correct interpretation of the data, […] to present information according to scientific standards on the sources, methods and procedures of the statistics”.[2] Along the same vein, the Committee for the Coordination of Statistical Initiatives recalls in its Principles Governing International Statistical Activities that “concepts, definitions, classifications, sources, methods and procedures employed in the production of international statistics are chosen to meet professional scientific standards and are made transparent for the users”. It also emphasizes the good practices of “documenting and publishing concepts, definitions, classifications and metadata used by the organization [… and] how data are collected, processed and disseminated by the organization”.

B. Moving into the age of AI: the critical need for sound metadata management and processing

10. **Achieving sound metadata management, particularly to better integrate them in the statistical production chain, is not a new topic**. In fact, this question has been widely debated in the 1990s, notably with the fast-evolving information technology landscape and, not least, with the advance of the Internet. Key issues at that time were the need of achieving integrated meta-information modelling approaches, particularly through metadata standardization (Froeschl, 1999). This led to the creation of sematic links to information units held in data repositories as well as designing standards to ensure the data integrity (Gillman and Appel, 1994).

11. **Yet in a context increasingly marked by an avalanche of data and the fast advance of AI, metadata will play an even greater critical role**. In fact, the exponential growth in data sources and types over the past decade has highlighted the importance of metadata in giving meaning to the avalanche of undifferentiated data. Perhaps even more importantly, as we move into the era of AI and computational superpower, metadata will play a vital function in informing models about the data (BIS, 2024). They will fundamentally ensure that the innovative technology can correctly interpret and, more importantly, understand the data they are trained on (Dupriez et al, 2024). Due to the circular relationship between AI and metadata, inaccuracies in the metadata can create a vicious cycle where models are trained on erroneous data, leading to the production of unreliable content, which can be then used to train future models with similar flaws (Bogdanova et al, 2024).

12. **Against this backdrop, effective metadata management necessitates robust quality frameworks achieved through systematic and automated editing**. Ensuring high-quality metadata involves efficiently curating and updating metadata, which includes modifying, updating, or verifying information related to the data sets. This process requires deciding when and how to adjust values, reflecting changes in data sources or business processes, correcting inaccuracies, and adding supplementary information. Unlike data processing, which is clearly defined in the statistical business process,[3] metadata editing across spans the entire value chain -

[2] See Resolution 68/261 adopted by the General Assembly on 29 January 2014 (A/RES/68/261).

[3] The Generic Statistical Business Process Model (GSBPM) includes the subphase of “editing and impute” (5.4) under the “Process” phase. This subphase is defined as follows: “Where data are considered incorrect, missing, unreliable or outdated, new values may be inserted or outdated data may be removed in this sub-process. The terms editing and imputation cover a variety of methods to do this, often using a rule-based approach. Specific steps typically include: determining whether to add or change data; selecting the method to be used; adding/changing data values; writing the new data values back to the data set, and flagging them as changed; producing metadata on the editing and imputation process”.

from design to dissemination - since data properties may typically change along the chain. This underscores the need for greater automation of the metadata editing process.

13. **Yet despite continuous advances in the automation of metadata processing, their editing remains resource-intensive**. In fact, this task often requires manual review by statisticians typically involving several checks, ranging from the application of basic rules, such as capitalisation, formatting, and layout, to more advanced tasks such as grammar, spelling and syntax checks, verification of the fluency and consistency across the text and review of the overall logical flow. Depending on the complexity of the text to be checked, a statistician may spend several minutes to validate the value of a given attribute limited to around 1000 characters for each time series (UNECE, 2023).

III. Leveraging generative AI for editing metadata: the BIS Metadata AI Editor

A. Using AI assistants as metadata editors

14. **The Bank for International Settlements (BIS) regularly disseminates statistics through the BIS Data Portal on major financial and macroeconomic indicators**, including central bank statistics, credit to the non-financial sector, exchange rates, international banking statistics and property prices. Data are usually released with extensive metadata attached to the observation, time series or data set. The dissemination of these metadata, including information about methodology, collection, coverage, and sources, is instrumental to promote and enhance the understanding of the BIS statistics. Leveraging the ISO standard SDMX, metadata are embedded in a consistent, orchestrated, and homogenous way across several features in the BIS Data Portal, such as dashboards, tables, and the glossary. This approach prevents content duplication and helps users to navigate through complex information quickly and efficiently (Lambe and Park, 2024).

15. **To enhance and efficiently editing metadata of time series, the BIS DataBank has been developing AI assistants** that respond to specific sets of instructions to edit the metadata fields. More specifically, assistants use OpenAI's models through the Application Programming Interface (API) with specific instructions to execute a certain number of tasks. In practice, an assistant can access several tools simultaneously, such as a file search function, or other user-defined functions. Assistants can also use persistent threads. These enhance the development of the AI application by saving the message history and trimming it when the conversation becomes too long for the model's context length.

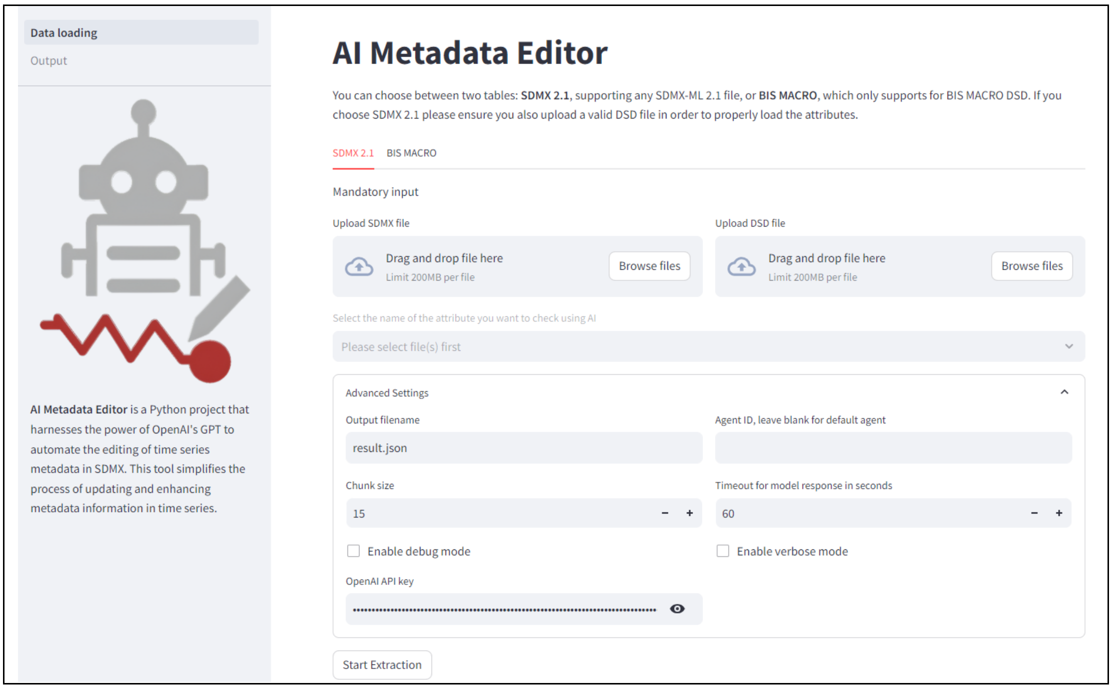
16. **Perhaps more importantly, assistants are highly customisable, both in terms of knowledge base and instructions**. For instance, on a practical level, statisticians can easily share files with the assistant to augment its knowledge base. Similarly to Retrieval-Augmented Generation (RAG), this functionality allows incorporating external information, hence helping the AI assistant to reference specific user-provided data. This feature is critical for maximizing the accuracy of outputs. For example, by uploading files containing detailed guidelines, policies, or domain-specific knowledge, statisticians ensure that assistants can adhere to these instructions in their responses. This allows assistants to deliver answers that are not only accurate but also aligned with specific requirements and standards. Moreover, this functionality enables assistants to handle specialized queries more effectively. By leveraging the additional data from uploaded files, assistants can in fact offer nuanced insights and detailed information that would otherwise be beyond their scope.

17. **In this context, leveraging AI assistants can be helpful to balance the need for high-quality metadata with their resource-intensive curation**. First, assistants are relatively easy to configure and customise. For example, the assistant can be preconfigured directly from the OpenAI platform. Here, statisticians can specify the model, the system instruction as well as add more custom functions. Second, they can upload files to augment the knowledge base of assistants. Finally, assistants can be easily integrated within an application through API calls.

B. The BIS Metadata AI Editor in practice

18. **Built upon assistants, the BIS Metadata AI Editor is an application developed by statisticians for statisticians to process time series metadata**. The application has three major components. First, it leverages the Statistical Data and Metadata eXchange standard (SDMX): at the very start, users can upload an SDMX file containing the metadata, for example specified in a given attribute at the time series level. Once the processing has been completed, the application generates an SDMX file with the results of the metadata processing. Integration with SDMX is key to streamline the workflow as well as to ensure high modularity with existing statistical pipelines, which typically rely on SDMX to ingest and validate data.

19. **Additionally, statisticians can provide the application with their own assistant** through a unique ID and generate/edit the metadata attributes based on a set of custom instructions. This is particularly relevant when a statistician seeks to adjust the system instructions of the assistant, for example, to ensure metadata consistency in a specific manner tailored to a given data set or sets of series only.



AI Metadata Editor is a Python project that harnesses the power of OpenAI's GPT to automate the editing of time series metadata in SDMX. This tool simplifies the process of updating and enhancing metadata information in time series.

AI Metadata Editor

You can choose between two tables: **SDMX 2.1**, supporting any SDMX-ML 2.1 file, or **BIS MACRO**, which only supports for BIS MACRO DSD. If you choose SDMX 2.1 please ensure you also upload a valid DSD file in order to properly load the attributes.

Output filename: result.json

Agent ID, leave blank for default agent:

Chunk size: 15

Timeout for model response in seconds: 60

□ Enable debug mode

□ Enable verbose mode

OpenAI API key: ••

Figure 1: the BIS Metadata AI Editor's web interface (UI)

20. **To facilitate the workflow, the application comes with command-line and web interfaces**. More practically, the web user interface is based on Streamlit and has been designed

to be very straightforward and easy to maintain (Figure 1).[4] Here users are presented with a homepage, containing the input parameters such as uploaders, assistant ID, metadata field to edit and other advanced parameters to tune the process. The page also allows statisticians to enter their own API key, which is stored in the application cache for convenience. Additionally, it offers more advanced features, including options for debugging.

21. **Specifically, the metadata editing process is orchestrated in several key steps** (Figure 2). First, the user specifies the requirements through the system instructions of the assistant. Typical instructions may ask for applying English capitalisation rules. There could also be specific rules, such as to ensure the precise spelling of names associated with central banks and other institutions, following the official names of the BIS shareholders. This is a crucial step to prevent any potential error that may compromise the integrity of the metadata. Other instructions may also include shortening the text in case it exceeds the counter limits. In fact, limitations on the number of characters per metadata field are usually common and can be easily handled by the assistant.

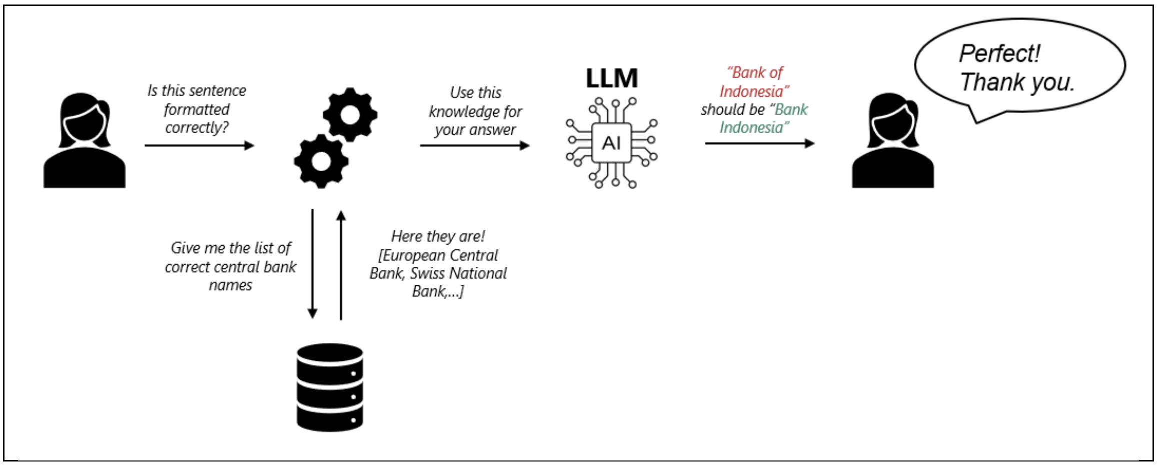

Is this sentence formatted correctly?
Use this knowledge for your answer
LLM
AI
"Bank of Indonesia" should be "Bank Indonesia"
Perfect! Thank you.
Give me the list of correct central bank names
Here they are! [European Central Bank, Swiss National Bank,...]

Figure 2: the BIS Metadata AI Editor's workflow

22. **As a second step, the process goes through a comprehensive formatting clean-up**. This includes the removal of double, trailing, or any other extraneous characters, thereby ensuring a standardised and polished text. While this step may also be achieved using regular expressions, the added-value of leveraging an AI assistant is the ability to preserve metadata consistency thanks to a throughout examination of the text. This involves not only cross-referencing within but also across attributes of time series to ensure uniformity. For instance, we apply consistently "data are sourced" instead of "the sources of these data are". Other examples may involve applying consistently date formats or replacing abbreviations.

23. **Finally, once the AI assistant completes the editing, the application redirects users to the 'output' page**. This page serves two main purposes. First, it displays the tracked changes, highlighting the modifications made by the AI assistant compared to the original metadata. This is crucial for ensuring traceability from input to output and for validating the changes. Secondly, and perhaps more importantly from an operational perspective, the web interface allows users to modify the generated output or fully revert the changes. The displayed content includes both the original and formatted results. Once all revisions are validated by the statistician, the application generates an SDMX file ready for ingestion.

[4] Streamlit is an open-source framework to develop data applications in Python (see streamlit.io).

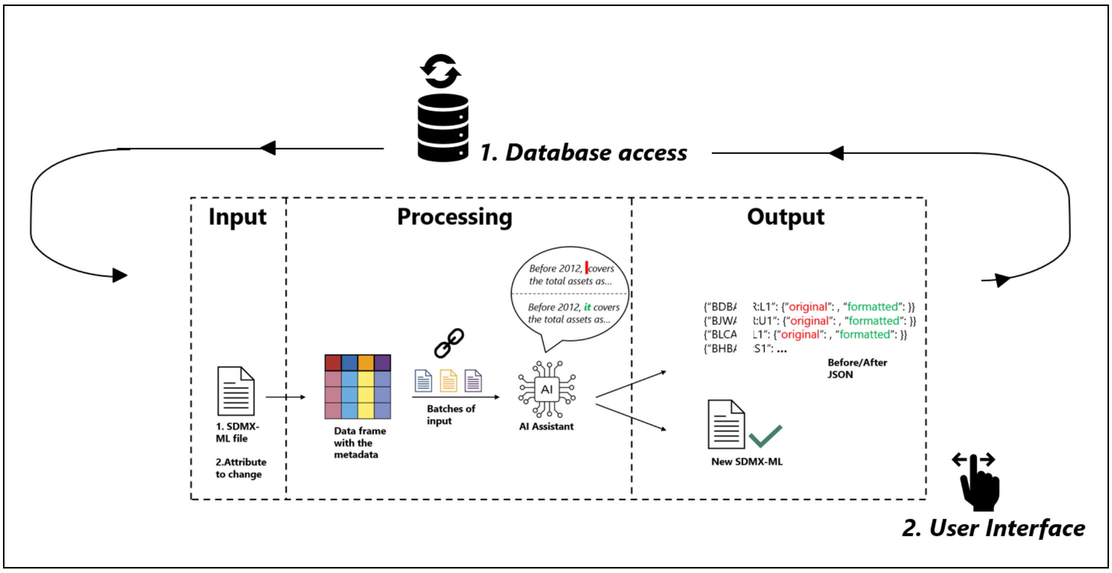

1. Database access
Input
Processing
Output
Before 2012, covers the total assets as...
Before 2012, it covers the total assets as...
{"BDB/ R:L1": {"original": , "formatted": }}
{"BJW/ R:U1": {"original": , "formatted": }}
{"BLCA L1": {"original": , "formatted": }}
{"BHB/ S1": ...
Before/After JSON
1. SDMX-ML file
2.Attribute to change
Data frame with the metadata
Batches of input
AI
AI Assistant
New SDMX-ML
2. User Interface

Figure 3: the BIS Metadata AI Editor's workflow

24. **More broadly, the BIS Metadata AI Editor is agile, modular and opportunity-driven**. Given the fast pace of AI developments, these principles have been fundamental to drive the project ahead. By avoiding to "reinvent the wheel" and assembling existing solutions under a common framework, the BIS Metadata AI Editor has been extremely fast to develop, with less than three months to deliver the proof of concept. Another significant benefit of this approach is the combination of low initial costs and high customisation. By keeping implementation and deployment costs low, the project could benefit from quick decision-making while also minimising the risk of failure. Similarly, integrating the application with existing OpenAI accounts within the organisation can ensure seamless access and usage while making sure that the application stays up to date. Finally, perhaps more importantly, the solution has been opportunity-rather issue-driven. Although the initial goal has been to increase productivity to edit large volumes of metadata, the project has also enabled a deeper exploration of other, similar generative AI solutions, hence creating new opportunities beyond the initial scope, including in terms of capacity development.

IV. Overcoming risks, limitations, and challenges through robust governance frameworks

A. Addressing risks: the need for comprehensive mitigation strategies

25. **Although editing metadata with generative AI presents several benefits, it also features some limitations, risks, and challenges**. First, on a practical level, the most critical barrier is the confidentiality restrictions that may apply to the use of the API. In practice, an important limitation of the application is that only public metadata can be currently exposed to the OpenAI API. Another practical limitation includes the maximum number of tokens to be passed per request or minute. This could especially be problematic for editing large sets of series, typically above a thousand per request. Yet the application has been designed to optimize the requests to be passed to the API so to ensure a smooth and fast execution of the tasks, through asynchronous calls, while also avoiding hitting the limits. Finally, the use of the OpenAI assistants

also involves some costs, as requests to the API are billed. However, these costs remain low with a reasonable number of requests and depending on the model selected.

26. **Turning to operational risks, the core one relates to the hallucinogenic AI**, including the possibility that the solution may not work as intended or cannot be accessed, leading to disruptions in the workflow. In fact, hallucinations arise when a large language model generates incorrect content due to misinterpreting patterns, often caused by insufficient training data or biases. This issue is particularly severe with metadata, as incorrect outputs can be reintroduced into the system as inputs, such as those used by statisticians. This creates a continuous cycle that contaminates future AI models trained on flawed metadata with the potential to undermining knowledge (Garrett, 2024). Additionally, the risk of repetition is problematic because AI-generated content tends to replicate the relationships found in its original training data. Ultimately, the dissemination of inaccurate metadata would lead to severe reputational damage.

27. **In addressing these risks, a number of mitigation measures are in place to ensure reliability and accuracy of edited metadata**. Firstly, human oversight is key. Pratically, this means that statisticians can review both the original text and the changes made by the BIS Metadata AI Editor. Crucially, statisticians also retain the control to reject any AI-generated suggestions partially or fully, thus performing manual edits when necessary. Further, human intervention extends to a thorough review of the final metadata to identify and correct potential errors, for example caused by hallucinogenic AI. Secondly, metadata versioning plays a critical role. Each piece of metadata is in fact versioned to monitor changes over time and maintain a detailed audit trail. Finally, it is worth noting that metadata editing is a one-time process. Unlike data that may require continuous updates, key time series metadata fields such as titles, compilation descriptions, or methodological guidance do not change very often. Consequently, metadata editing is an occasional task, not a continuous one, allowing for further and deeper validation layers to enhance accuracy and reliability.

B. The need for data governance frameworks for overcoming limitations and challenges

28. **More broadly and at higher level, a key concern with the use of generative AI in official statistics is the lack of transparency**. Referred to as the "black box" issue, the challenge arises from the fact that the algorithms and models used in generative AI are often complex, making it difficult for users to understand the processes behind results (UNECE, 2024c). In this context, the traceability of the inputs to the outputs is key to achieve. Recent and ongoing initiatives by the High-Level Group on the Modernisation of Official Statistics have emphasized these challenges in official statistics, underlining the necessity of transparency to ensure that AI systems are reliable and trustworthy in statistical applications (UNECE, 2023).

29. **Additionally, generative AI's dual role as both a user and producer of metadata introduces further complexities**. If not properly governed, there could be a heightened risk of data misuse; mis and disinformation. Metadata can be exploited unethically or inappropriately, leading to issues such as the manipulation of AI training processes. Specific threats include model poisoning, where malicious data is introduced to corrupt the AI's functioning, and prompt attacks, which manipulate the inputs to elicit harmful outputs. These vulnerabilities highlight the importance of stringent governance and oversight in the use of generative AI.

30. **Another set of limitations involves ethical concerns that arise from the deployment and use of AI systems**. One major issue is language dependency. AI systems are often trained on data from specific linguistic or cultural contexts, which can significantly impact their performance when applied to different contexts. For example, a substantial portion of training data sets are in

English (OECD, 2024). This creates considerable challenges for applications in regions where English is not the primary language, as the AI may not understand local nuances, idioms, or cultural references.

31. **Legal concerns, not least copyright and confidentiality breaches, are also crucial aspects to be aware of**. Copyright concerns arise regarding whether content produced by LLMs is identical to proprietary inputs, such as articles published elsewhere. This raises questions about intellectual property rights and the potential for plagiarism, including unintentional, which could have legal implications and affect reputation. Second, in terms of confidentiality, metadata often do not receive the same level of protection and governance as the primary data. This discrepancy can lead to breaches of confidentiality and privacy. Metadata, while not always considered sensitive, can contain information that, if improperly handled, might compromise the privacy of individuals or organizations. Furthermore, the use of metadata to train AI systems introduces questions about unauthorized data access and the best security protocols to mitigate these risks. Since metadata can be as revealing as the data themselves, unauthorized access could lead to significant privacy violations.

32. **The cumulative effect of these risks may result in significant reputational damage and prompts for "algorithm auditing and assurance"**. Organisations leveraging generative AI, especially in official statistics, must contend with the risk of reputational damage and, ultimately, public mistrust (Araujo et al, 2024). For example, increased dependence of statistical offices, including central banks, on LLMs may increase operational risks, potentially also compromising their ability to fulfil their mandate (Doerr et al, 2022). Ethical lapses or failures in AI performance due to cultural and linguistic insensitivity can also damage an organization's credibility and reliability. To guarantee the safety, legality, and ethics of AI algorithms, new activities are emerging to monitor and assess algorithm performance. These efforts focus on auditing algorithms to ensure they comply with the organisational, national, and international standards (Koshiyama et al, 2024)).

33. **Addressing these issues urgently calls for tailoring existing data governance frameworks to the specificities of official statistics**. Despite extensive efforts in recent years to develop governance frameworks for generative AI, a framework specifically tailored to organisations involved in official statistics is still missing (Choi et al, 2024). The project on generative AI for official statistics, under the *aegis* of the High-Level Group for the Modernisation of Official Statistics, precisely seeks to fill this gap by focusing on two essential aspects. First, the priority is to identify and manage AI-related risks in official statistics. This includes creating an interoperable risk management framework, potentially based on the severity of harm posed by AI systems, and clearly defining responsibilities, roles, and ownership of (meta)data assets, alongside implementing effective mitigation strategies. Additionally, the framework should address organisational structures for overseeing AI initiatives, considering both top-down approaches like steering committees and project boards, as well as bottom-up strategies that involve decentralised cross-organisational units. Overall, such governance frameworks will enable official statistics to fully reap the benefits offered of the technology while also carefully managing its risks and limitations.

VI. References

Araujo D, S Doerr, L Gambacorta and B Tissot (2024): "Artificial intelligence in central banking", *BIS Bulletin*, no 84, January,

Bank for International Settlements (BIS) (2024): "Artificial intelligence and the economy: implications for central banks", *Annual Economic Report 2024*, June.

Bogdanova, B, M Erdem, B Ligani and O Sirello (2024): "Enhancing metadata with generative AI: the case of BIS statistics", presentation at the 12th Biennial Conference of the Irving Fisher Committee on Central Bank Statistics (IFC), August.

Choi I, A Kipkeeva, O Sirello and V Vaiciulis (2024): "Generative AI and official statistics: the project of the UNECE High-Level Group for the Modernisation of Official Statistics", presentation at the 12th Biennial Conference of the IFC, August.

Doerr, S, L Gambacorta, T Leach, B Legros and D Whyte (2022): "Cyber risk in central banking", BIS Working Papers, no 1039, September.

Dupriez, O, H Fu, C Hammer and A Solatorio (2024): "The transformative role of AI for development data", *World Bank Blogs*, April.

Froeschl, K A (1999): "Metadata Management in Official Statistics – An IT-based Methodology Approach", *Austrian Journal of Statistics*, vol 28, no 2, April.

Garrett, A (2024): "The devil, the detail, and the data", *Journal of the Royal Statistical Society*, Series A: Statistics in Society, no qnae063, June.

Gillman, D W and M V Appel (1994): "Metadata Database Development at the Census Bureau", Working Paper, *UNECE Conference of European Statisticians*, Work Session on Statistical Metadata (METIS).

Koshiyama A et al (2024): "Towards algorithm auditing: managing legal, ethical and technological risks of AI, ML and associated algorithms", *Royal Society Open Science*, vol 11, 5, May.

Lambe, E and T Park (2024): "The BIS Data Portal project – delivering the next generation platform for BIS statistics", *IFC Bulletin*, no 60, April.

Organisation for Economic Co-operation and Development (OECD) (2024): "Embracing the technology frontier", *OECD Digital Economy Outlook*, vol 1, May.

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

——— (2024a): *Data governance framework for statistical interoperability*, High-Level Group for the Modernisation of Official Statistics, March.

——— (2024b): *In-depth review of linking data across domains and sources*, Conference of European Statisticians, Seventy-second plenary session, ECE/CES/2024/5, June.

——— (2024c): *Organisational aspects of implementing ML based data editing in statistical production*, High-Level Group for the Modernisation of Official Statistics, February.

Wilkinson, M et al (2016): "The FAIR Guiding Principles for scientific data management and stewardship", *Scientific Data*, vol 3, no 160018.