

Irving Fisher Committee on
Central Bank Statistics



Enhancing metadata with generative AI: the case of the BIS statistics

Bilyana Bogdanova, Magdalena Erdem, Bianca Ligani and **Olivier Sirello**

XII Biennial of the Irving Fisher Committee on Central Bank Statistics

Basel, 22 August 2024

Metadata in the age of AI: an introduction

- Metadata are critical yet costly
 - They are **critical**: they give shape and meaning to data
 - They are **costly**: time consuming, hard to maintain and standardise
- Yet, the data deluge and the rise of AI mark a new epoch:
 - Data are **everywhere**
 - Data **more and more generated, used and interpreted by machines**, not just humans
 - Data are **undifferentiated** and meaningless if not transformed into **information**
- What is the **role** and **place of metadata in the age of AI**?
 - Could generative AI help producers and users to achieve better metadata?
 - What is our role as official statisticians in central banks?

Outline

- I. Generative AI as a user and producer of metadata
- II. From theory to practice: the case of BIS statistics
- III. Towards an active role of official statistics for a virtuous circle between AI and metadata

I. Generative AI as a user and producer of metadata

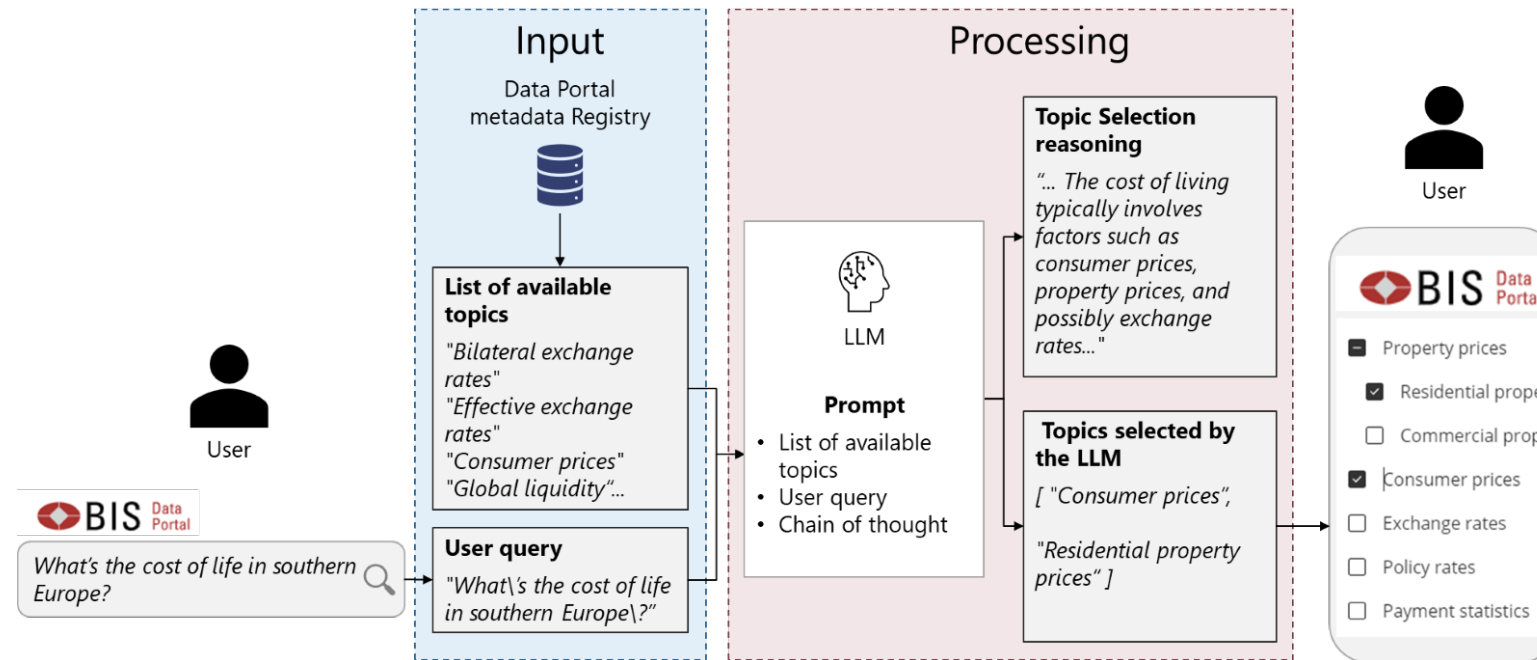
- The relationship between generative AI and (meta)data is circular and symbiotic
 - It is **circular**: generative AI *produces* content which is then *used* by the technology in a continuous loop or cycle
 - It is **symbiotic**: generative AI cannot work without metadata: they inform models to interpret the data
- **Two major implications:**
 - **Statisticians initiate the loop**: as providers of ground truth, any inaccuracies would propagate in a never-ending loop, with consequences for data misuse and misinformation
 - In the age of AI, machines should be trained to *properly* understand the data disseminated by statistical offices: **from machine-readable to machine-understandable data**

II. From theory to practice: the case of BIS statistics

- **Producing metadata** with generative AI: the BIS Metadata AI Editor
 - Check metadata of time series, for example compilation, methodology description to ensure proper enforcement of editing rules
 - Leveraging OpenAI assistants: custom instructions and tailored knowledge base
 - Leveraging SDMX end to end: from ingestion to dissemination
 - Simple, modular and agile
 - By statisticians for statisticians
 - Framework rather than solution approach

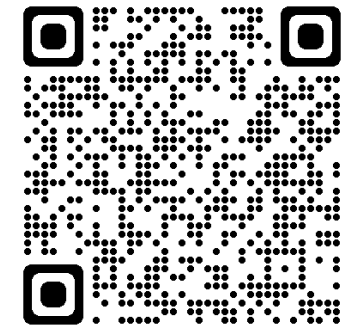
II. From theory to practice: the case of BIS statistics

- **Using metadata** with generative AI: improving the BIS Data Portal search
 - NLP-based techniques to improve the search: search with meaning, not keywords



 **BIS** Data Portal

Scan me



III. Towards an active role of official statistics: for a virtuous circle between AI and metadata

- **What is the role of official statisticians in this context?**
 - **Be aware that this circle is not free of limitations and risks**
 - Black box, lack of transparency, hallucinations and repetitions
 - Ethical considerations, including language divides and propagation of bias
 - Confidentiality
 - Governance and stewardship

III. Towards an active role of official statistics: for a virtuous circle between AI and metadata

- **Use frontier technologies *strategically*** for the development of official statistics
- A proactive approach is needed to **keep** and **feed the virtuous circle between generative AI and metadata**
 - Moving into the age of AI: **data modelling** or *how* data are
 - This calls for:
 - More, finer and better metadata: be **accurate**
 - More dissemination and communication: be **relevant**
 - More, stronger and better literacy of users, both humans and machines: to **ensure trust** in official statistics



Thank you

Bilyana Bogdanova - bilyana.bogdanova@bis.org

Magdalena Erdem - magdalena.erdem@bis.org

Bianca Ligani - bianca.ligani@bis.org

Olivier Sirello - olivier.sirello@bis.org