IFC-ECB-Bank of Spain Conference: "External statistics in a fragmented and uncertain world"

12-13 February 2024

Challenges and opportunities presented by generative AI in official statistics[1]

Giannakouris Konstantinos and Jean-Marc Museux,
Eurostat;

and Cimpermann Miha,
Jozef Stefan Institute

[1] This contribution was prepared for the conference. The views expressed are those of the authors and do not necessarily reflect the views of the European Central Bank, the Bank of Spain, the BIS, the IFC or the other central banks and institutions represented at the event.

Challenges and opportunities of generative AI in official statistics

Museux Jean-Marc[1], Giannakouris Konstantinos[2] , Cimpermann Miha[3][4]

Abstract

The paper illustrates the challenge and opportunities of embracing Generative AI, particularly Large Language Models (LLMs), in the broad context of the imperative for innovation in the European Statistical System (ESS) to meet evolving demands for timelier, detailed statistics amidst a rapidly changing data and technology landscape (ref Eurostat-Cros). Building on the HLG-MOS collaborative white paper on LLM for Official Statistics (ref UNECE) and on Eurostat own exploration of LLM technologies for easing access to EU statistics and metadata, the paper explores their potential applications and associated challenges. The authors delve into key concepts of Generative AI, detailing LLM capabilities, promising use cases, and recent advancements.

Recognizing the need for a cautious approach, the paper discusses risks and challenges related to ethics, transparency, accuracy, bias, and privacy in LLM adoption. It introduces the concept of In-Context Learning (ICL) as a mechanism to balance the flexibility of LLMs with the need for integrating verified, up-to-date information, addressing challenges associated with fine-tuning and controlled usage.

In the Eurostat case, the paper explores the potential of LLMs to enhance user access to official statistics by providing a natural language - based interface, enabling richer interaction e.g., through chatbot, and offering advanced and semantically powered search for data based on human language. The authors share initial learnings and recommendations based the preliminary experience, emphasizing the role of prompt engineering techniques, the importance of non-standard metrics for validation, the value of open-source alternatives, and the importance of a user-centric approach.

Concluding with broader recommendations for exploring and deploying generative AI solutions in official statistics, the paper advocates for continued assessment, user-centric design, and a staged approach in the dynamic LLM ecosystem. Overall, the paper contributes to the discourse on leveraging Generative AI to transform official statistics, recognizing its potential while addressing the intricacies of implementation and other considerations.

[1] Eurostat, Unit A.5 Methodology and Innovation for Official Statistics

[2] Eurostat, Unit A.5 Methodology and Innovation for Official Statistics

[3] Jozef Stefan Institute, Department for Artificial Intelligence

[4] The views expressed in this paper are those of the authors and do not necessarily represent the official position of the European Commission.

The paper concludes with the promotion of several collaborative initiatives that should contribute to deepen our understanding of LLMs and realising proof of concepts of most promising use cases in the future.

Keywords: Generative AI; Large Language Models (LLMs); European Statistical System (ESS); Official Statistics

JEL classification: C81, O30

1. Introduction - The need for innovation in official statistics

Innovation is not an option for the European Statistical System (ESS). It is a necessity. Today's business environment — defined by competition from a large number of private data providers, an ever faster changing data ecosystem, and a rapid rise in increasingly urgent demands for new, timelier and more detailed statistics — makes innovation an even higher priority for ESS partners. In recent years, the producers of official statistics have been faced with increased demands from their users to be more responsive and agile especially in times of crises, to expand the range of statistics, improve timeliness and level of detail. The current situation calls for ESS members to expand their use of digital technologies, widen the use of data sources like new digital sources, and apply new technologies, such as artificial intelligence. The ESS members are also pursing more innovative practices and create conditions for successful innovation so that they can more quickly anticipate and respond to the challenges ahead. Novel digital technologies provide opportunities to address these challenges. Taking advantage of new technological and methodological developments — such as artificial intelligence and machine learning, privacy enhancing techniques, smart devices, methods for data integration, geospatial capabilities, and data analytics are crucial for the Official Statistics to maintain its key role of provider of Trusted Smart Statistics to citizen (Ricciato, 2020) in this new realm. These strategic considerations and other related issues have been addressed in the ESS Innovation Agenda (Eurostat, 2023) adopted but the 27 heads of Statistical Organisations of the EU in February 2023. The ESS Innovation Agenda proposes a strategic approach to innovation as well as a series of activities that aim at enabling and accelerating innovation, conducting and supporting innovative projects leading to concrete outcomes meeting user demands. The implementation of the ESS Innovation Agenda requires a constant horizon scanning to identify new trends and on-going activities within and outside official statistics and streamlined processes to turn them into the production of new and experimental statistics and the improvement of statistical processes.

Artificial intelligence (AI) is an area of strategic importance and is recognised a key driver of social and economic development. It is transforming every walk of life. The magnitude of this socioeconomic change is coming rapidly. In this context, Generative AI is undoubtably one of the recent and striking markers of the evolution of Artificial Intelligence. Materialised with the outspring and never precedent adoption by a broad public, the large language models-based technologies are on the way to revolutionise many industries by enabling the harnessing of the power of new emerging AI technologies.

In the following sections we will develop the generative AI case illustrating the challenge and opportunities it could raise for official statistics and the ESS in particular in incorporating these new feature in statistical production It is not meant to provide a technical review on the topic but next sections will provide the essential elements to understand what LLMs are and explore the risks associated with their use in the context of official statistics and the certain options to mitigate these risks.

Many of the concepts and challenges exposed here are borrowed from the recent work carried out at the UNECE level under the HLG MOS initiative, in particular the release in December 2023 of the white paper on LLM (UNECE, 2023) to which some

of the authors of the current paper contributed. It also builds on a very first explorative study carried by Eurostat in the domain with the purpose of launching of a proof of concept to demonstrate the potential of LLM and generative AI in the context of dissemination of EU statistics and metadata. Notions and definitions also come from Eurostat effort to provide a comprehensive review of LLM features in the context of Official Statistics – this document, still unpublished, is referenced as a forthcoming paper.

Despite its rapid evolution and deployment, Generative AI is still in its infancy and its adoption by statistical organisation will require a pace wise approach addressing the challenge it creates for their integration in quality framework which proved and will remain a critical asset of official statistics while recognizing the need to adapt it to the new data ecosystem context. The views expressed by the authors of this paper do not represent the official views of Eurostat. It is an attempt to present in a coherent manner few of the key issues opportunities raised by generative AI in the context of official statistics.

2. Generative AI – key concepts and standard use cases

Generative AI is a general umbrella term used to refer collectively to a range of emerging techniques and tools having in common to generate new content or data that is not explicitly derived from existing examples and data. This can include generating text, images, music, and more.

A key representant of Generative AI family is the wide language model, also known as a Large Language Model (LLM). LLMs are an advanced type of artificial intelligence technology, that has been trained on vast amounts of text to "understand" and generate human language in a sophisticated way. These models are designed to capture natural language structures, patterns, and relationships, allowing them to perform a variety of tasks related to automatic natural language processing (NLP).

LLM uses techniques from machine learning, deep learning, natural language processing (NLP) and computational linguistics. In a narrow view, the LLM models can be seen as stochastic parrots: that is highly efficient at stitching together words according to probability and generating convincing language, without any fundamental understanding of its meaning. Compared to more traditional NLP framework, LLM models a have significantly improved their efficiency by providing additional features such as:

- Attention Mechanism: It allows the model to focus on different parts of the input text differently, enabling it to capture contextual elements.
- Transformer Architecture: Introduces self-attention mechanism where it weighs input tokens based on their relevance.
- Training: LLMs are trained on vast datasets, sometimes encompassing parts of the internet. They learn patterns, structures, and facts from this data.
- Fine-tuning: After being pre-trained on large datasets, LLMs can be additionally trained (fine-tuned) on specific datasets. This enables models to be optimized for specific use cases and /or to perform specific tasks extremely well.

The standard use cases of Natural Language Processing in general and LLM are to enable machines to understand and work with human language in a variety of ways, such as:

- Text Analysis: Extract useful information from textual documents, such as identifying named entities (names of people, organisations, places), detecting relationships between entities, etc.
- Automatic translation: Automatically translate text from one language to another, for example, translate text from English to French.
- Summarization: Create concise, informative summaries of longer documents or articles.
- Speech Recognition: Converting human speech into written text, which allows machines to process and understand voice commands.
- Text generation: Create text automatically, whether for content writing, dialog generation, or even computer code creation.
- Sentiment analysis: Determine the emotion or opinion expressed in a text, such as detecting the positivity, negativity or neutrality of a comment.
- Chatbots and virtual assistants: Create computer programs that can interact with users in a natural way, answering their questions and providing information.
- Discourse analysis: Analysing language in a larger social or cultural context to understand intentions, implications and hidden meanings.
- Information Extraction: Identifying and extracting specific information, such as dates, locations, amounts, etc., from unstructured text.

In the last two years, the field of LLMs has witnessed a meteoric rise, revolutionising natural language processing and understanding. The rapid evolution of this subject in this time is evident in the myriad of breakthroughs and innovations.

LLM services and capabilities are widely available on the "market". Beyond, ChatGPT suite developed by Open AI, many solutions and framework exist. While the pace of evolution is very much driven by the commercial players (Google, Amazon, Microsoft), a broad range of open-source solution are developed and maintained by active and open communities.

The recent advancements and the sheer capabilities of these models see for instance. "Sparks of Artificial General Intelligence'' paper on GPT-4 [v2] have demonstrated the potential of LLMs, highlighting their ability to perform tasks that were once thought to be exclusive to human cognition. In this era, where data is abundant and the need for accurate interpretation is paramount, LLMs offer a promising avenue to transform the way official statistics are understood and utilised.

LLM have already proven value in the context of Official statistics. The HLG MOS white paper on Generative AI provides a few illustrative examples:

LLM have a great potential for assisting, improving the efficiency and the quality of regular workplace tasks. Beyond the sheer use of ChatGPT web apps and locally deployed specific instances for day-to-day task, it is expected that these features will be integrated in standard desktop suite soon.

More specifically LLM have already revealed high potential for boosting some of the supporting capabilities to produce official statistics. In particular:

- Statistics Canada is using LLMs to generate textual descriptions from a table or a series of numbers which can be tailored to different audience segments, including policymakers, journalists, and the general public. This could greatly simplify the work of analysts and communication experts by providing initial drafts that human experts could work on. LLMs can also assist in automating the creation of charts and graphs, although this area is still under exploration.
- The Bank of International Settlements is exploring the use of LLM to produce metadata to document their open data.
- The Irish Central Statistics Office has built up a prototype based on chat GPT to assist coding and translating between programming languages (SAS & R).: LLMs have shown to significantly enhance the efficiency and effectiveness of programmers and analysts by helping streamlining and optimising code development, providing code snippets and translating them.
- The Australian Bureau of Statistics is using LLM to update and maintain statistical standard and generate draft text descriptions to assist human experts in updating statistical classification systems.
- Eurostat has developed algorithms to automatic classify occupations (ISCO) extracted from On Line Job Advertisement and build thesaurus and description ESCO digital skills taxonomy in the context of ESS Web Intelligence HUB .i.e. a set of methodological and technological capabilities for harvesting (web scrapping) and processing data obtained from the web.

3. Taking benefit from LLM while addressing risks and challenges.

Navigating the expansive landscape of LLMs presents a myriad of challenges and risks that demand careful consideration. Therefore, along the developments for maximising the output of LLM, it is necessary to consider these risks. The HLG-MOS white paper on generative AI described a series of risks that need to be taken into consideration when implementing LLMs in statistics. The key elements are summarised below providing some highlights on the practical way to mitigate them. Readers would refer to the original paper (ref UNECE) for a more thorough presentation.

Ethics and Legal aspects

LLMs pertain to the large family of AI solutions. The use of AI solution raises general issues regarding public trust and social licence to operate. Use of LLMs may lead to reputational damage and infringement is in case mistakes or generated content closed to copyrighted material are propagate to the public.

Governmental organisations have above all to abide to their respective regulatory environment that are being issued (see for instance the proposal for the EU Artificial

Intelligence Act (https://data.consilium.europa.eu/doc/document/ST-5662-2024-INIT/en/pdf). There is thus an urgent need to develop specific policies at the level of organisation for the development and use of AI solutions expanding privacy and statistical confidentiality provisions.

There are multiple relevant ethical frameworks for the use of AI, all of which cover similar principles. It is worth referring the attempt of UNECE – HLG-MOS task team to release draft guidelines for Responsible AI relevant at the level of the industry (ref UNECE 2). This framework needs to be adapted and instantiated in the context of LLM. Broadly, when using LLMs, the major considerations therein are the protection of human rights and the need for human-centred oversight and authority. It requires that a suitably qualified human should always be in the loop, and have a final say on any LLM generated output.

Transparency and explainability

Algorithms used in commercially available LLMs are rarely shared as open source, leading to considering them as black boxes. In addition, as data used for training the foundational model is not controlled; the number of parameters involved is gigantic; the objective function cannot be made explicit as they are convoluted in the depth of the deep neural network layers and many steps and ad hoc intervention are involved prior to the release, make these models not directly explainable. Transparency and explainability mostly revolve in exposing and sharing the script and the pipelines that are used to prompt the LLM foundational model. Therefore, their use should be supported by clear communication on the impact on users and the rationale and benefit for using them. In particular, metadata should contain information as to how that material was produced and any limitations it has, including potential for bias, hallucinations, etc.

Accuracy and bias

One of the major issues in LLMs use is "hallucinations", which stands for model generating inaccurate or completely not true information. It presents one of the key issues in using LLM in chatbots for example. LLMs have a documented tendency to make up facts and propagate biases inherited from their generic training on a massive dataset, usually containing billions of words from diverse sources such as internet.

Although recent model developments are focused on integrating back loop with fact checking capabilities to minimize the effects of hallucination, care should be taken to evaluate them and address biases in LLM outputs. Beyond the involvement of expert to validate and correct outputs as frequently encountered, new measures of accuracy based on semantic closeness need to be generalised and implemented to assess global performance in generating coherent outputs. Few examples of accuracy measures are described in the white paper.

The second main drawback of LLM is their inability to handle adequately numerical information. In the current settings, LLM are not mean to handle numerical data in the way that specialised algorithms and statistical/mathematical models do. Automated and reproducible data analysis and mathematical tasks are out of reach of LLMs. As a work around, numerical data can be transformed in text that can be

further embedded as an input to the model. Basic mathematical operations are emulated by using patterns expressed in textual examples on which the models have been trained but without the accuracy and the reproducibility of for instance R and Python scripts. Despite recent evolution of generic models to handle logical operation, LLMs do not have an internal framework of logical computation rendering them suboptimal (for instance in terms of technical resources and data governance) to undertake complex mathematical operations nor to understand complex mathematical prompts.

However, LLMs have proved to be quite efficient in translating of text to code, from semantic description into underlying programming concepts. This enables creating SQL or like SMDX queries (see IMF) to query the data repository directly from text taking benefit of the semantic content of the related metadata. With accurate implementation and suing a combination of text-to-text and text-to-SQL LLM representation, numerical values can therefore be retrieved automatically and the LLM can deliver a textual interpretation of the data [UNECE] as response to the user.

Privacy, data protection and data governance

Rights to privacy, data protection apply to both LLM training data and the information supplied as prompts. It requires to uplift the data governance in place to identify wide variety of sources that are activated and enable audits when errors and bias are detected. In addition, the wide availability of LLMs has proved to increase the risk of disclosure of protected information. As in most cases LMMs data sources provenance and accuracy are not under the control of the statistical offices, they must engage with the service provider to get minimal metadata and assurance the training sources uphold these rights. In the same line, inputting LLM models with sensitive information should be prohibited if safeguard measure (technical, organisational …) regarding their use by the LLM system are not put in place.

Generative AI and In-Context Learning

On one side, it is now recognised that the direct fine tuning of model can hardly be achieved by a single organisation. The main reasons are the difficulty to compile of vast amounts of "corporate" labelled data, the constant change and the flow of new information that will make the model already outdated when released, the lack of accessibility and openness of foundation LLMs available through API's only, the prohibitive processing costs and the restrictive licenses accompanying most powerful commercial solutions. In many cases, adjusting the model is not sustainable and effective and might not ensure the adequate quality.

On the other side, the need to limit and constraint the answers of the model to verified, fresher and cooperate information while benefiting of the flexibility of LLM is a key requirement. For instance, for chatbot assisting user to access official statistics indicators, a key requirement is to provide answer that use only open data and metadata from the dissemination systems.

The In-Context Learning (ICL) has been introduced as a practical framework to achieve a good trade-off between large scale fine tuning of LLM using corporate data and uncontrolled use of LLM exposing the user to hallucination and structural bias. In-

context learning is a mechanism to constraint LLMs towards specific results. It consists of providing the model with context when sending a request. First, ICL framework is trained on a limited set of examples to form a demonstration context. These examples are usually written in natural language templates. Then, ICL concatenates a query question and a piece of demonstration context together to form a prompt, which is then fed into the language model for prediction. A practical approach to leveraging ICL involves constructing a vector database that archives the embeddings of corporate documents. When a user submits a prompt, pertinent document vectors are swiftly retrieved from the database, supplying them as context to the model. The framework incentivises the model to respond with the answers obtained from the knowledge corpus that fed as contextual information together with the query of the model.

ICL has become the main method of applying the LLM technology for specific use cases. It is an efficient mechanism to feed specialised external knowledge to the models, to be able to provide answers in the specific domain. Research has shown that through ICL, LLMs capabilities can be expanded to execute more complex tasks, such as basic mathematical reasoning. All these functionalities rely on prompt engineering techniques, i.e., the careful and assisted design of input prompts to guide the model's output in desired directions, to make functions that work well for the specific use case.

For official statistics, leveraging LLMs in combination with ICL can be determinant in the processing the vast amounts of textual data, extraction meaningful insights, and support data-driven decisions using already available assets (Website, Publications Dissemination database, metadata systems) while at the same time providing the necessary safeguards to steer the response of LLM and control the drift of the model.

4. Improving access to Official Statistics: the Eurostat use case

Despite the significant modernisation effort National and International Statistical Offices have put transitioning their dissemination system taking benefit of state-of-the-art digital solutions, users who seek to extract insight from statistical data and related statistical information today, face several challenges. First, searching for the 'needle' of relevant data in the huge 'haystack' of Eurostat's dissemination environment (i.e., dissemination databases, statistics explained articles, statistical publications, press releases, etc.) can be time-consuming and burdensome, also due to the limitations of existing search interfaces and the need to process a great amount of context information along the process. Second, when dashboard interfaces are available, the analytics and visualisation functionalities may be limited to a small set of pre-defined ones that not always match the one required by the user. Third, even basic data analysis and visualisation tasks that do not fall into the set of pre-defined dashboard options require the users to program queries and scripts in some formal language (SQL, R, python etc.) Limitations include the complexity of 'search interfaces', limited search options, lack of prior knowledge of the structure of the dissemination environment, data overload or lack of user support for troubleshooting, or general guidance on how to use the database effectively and

interpret the results of searches. Moreover, for the analytics and visualisation, users would need to develop/use SQL/R/Python/ or other software to get insights of the data. accessing the latest, most relevant information available, integrating them with other sources required some effort (navigating through data and metadata system) and skills (statistical literacy) from users.

The development of metadata management systems accompanying dissemination databases provide to a certain degree data insights but their activation for automated analysis and visualisation possibilities are still very much unused. Advanced metadata system could link the metadata with the data they described and provide a rich semantic context to be used to search for across data silos. Effort to organise them as knowledge graph have proved of little impact on user experience.

LLMs can largely improve this situation and help improve the quality of data provision to users – through, for example,

- Providing a semantic aware interface to search statistical data using natural (human) language: questions in natural language are translated in the database query and returned in the form of graphs or table with statistical annotation and storytelling. There is no need to formulate complex queries in formal languages (e.g., SQL, R, Python, etc.) or program scripts in analytics tools.
- Enabling a richer interaction: LLMs can be used to engage in a dialogue with users to clarify their information needs and refine queries can result in more accurate and relevant responses.
- Providing different options for reply: users can choose how they receive statistical information from LLMs, for instance, summarised reports, in-depth analyses or raw data.
- Data interpretation assistance: LLMs can help users interpret complex statistical data by providing explanations, visualisations, and context and the scripts for retrieving data and carry out detailed analysis in the own environment.

It is hoped that in the long term, specialised generative application combining the many capabilities of the LLM ecosystem can contribute to the analysis the content of the data and provide deeper insights into the content combining structured and unstructured information, such as identifying relationships between concepts or detecting patterns in the data in multiple languages.

First learnings and recommendations for exploring and deploying generative AI solution.

The integration of LLMs into existing workflow systems unveils a significant transformative potential. The integration of LLMs into workflow systems not only optimises efficiency but also opens avenues for innovative applications across diverse domains. By seamlessly incorporating LLMs into established workflows or developing new ones, NSIs will automate repetitive tasks, leading to significant efficiency gains and streamlined processes. Their integration through their Application Programming Interface (API) enhances the adaptability of the application by allowing taking benefit on the development and innovations. This also facilitates a more stepwise and coherent approach, enabling to build generic capabilities serving a broad spectrum of applications.

The Eurostat experience collected so far through the landscaping and exploration of LMMs in the context access and dissemination of EU statistics indicates furthermore:

- The importance of prompt engineering techniques. The careful and programmatic design of input prompts to guide the model's output in the desired directions is at the core of any solution that works in real-life scenarios. This new discipline requires new skills and need to be developed at scale in the organisation prior to any development LLMs based solutions.
- It is necessary to design non-standard process including pre-validation of outputs by user and integrating in design domain expert for validation and finetuning of prompt system to achieve the desired output and quality. For instance, system can ensure that the data match end users request, by prompting the to validate the query parameters on the output prior to submitting the final query results. System should be able to segregate the numerical aspects of the query and the semantic and contextual aspects and to trat them separately but coherently. To gain in efficiency and confidence new quality metrics based on semantic similarity and factual correctness need to be developed enabling systematic benchmark of the different solutions and allowing gaining confidence of the level of uncertainty and mistakes generated by the system.
- Many valuable alternatives to commercial LLM exist and have been tested. Several open source and free alternatives are available and proved to be functionally equivalent to large models offering decent performance at a lower cost and enabling transparency and reproducibility. Commercial solutions still tend to have better performance and can be used as benchmark without prohibitive costs. Open-source alternative available can be deployed locally so to facilitate compliance to security policies. The open source models are available on open access platforms (such as replicate.com), provided as a service for the public user, while the models can also be deployed locally (on-premisse) with appropriate infrastructure to run them (GPU).
- In Context Learning is an effective way to constraint and integrate external information considering the need for constant adaptation to new information and the relatively high cost of fine-tuning foundational model. This should lead to a significant reduction of errors in practical applications.
- Adopting a user centric approach when designing solution does not only contribute probing the social expectance but also assessing actual benefit of the solution and helping in fine tuning the trade-off between uncertainty and usefulness of the system.
- Continue to assess and value the possibility to gradually develop industry model that are trained and finetuned with the core semantic assets and values of official statistics.
- Integrating and deploying LLM capabilities requires a careful and stagewise approach. The LLM ecosystem is fast-changing where new LLMs are being listed basically on weekly basic. Therefore, when experimenting and realising LLM applications, a layered and modular architecture is fundamental, to enable model-agnostic architecture. This important aspect enables to potentially change the LLM model used I application or use multiple models for different functionalities.

5. Generative AI – the way forward

Generative AI and ESS innovation: The One stop shop on AI/ML (AIML4OS)

The use of Artificial Intelligence/Machine Learning to produce official statistics is one of the strategic domains that need to be developed further and where coordinated ESS actions could be beneficial. It plays an important role in developing innovative solutions with respect to statistical products and processes, allowing for more timely production of official statistics and better response to user needs.

The ESS Innovation Agenda endorsed by ESSC in February 2023, has clearly recognised AI/ML as a key domain for cross cutting developments for processing new data sources and supporting complex processes. It involves the development and deployment of AI/ML to extract information for unstructured data, to assist in labour intensive activities, to improve data integration and users' experience. It requires pooling expertise to provide support to business initiatives exploring, for example, the use of data from sensors, satellite images, and web scraping to enhance statistics portfolio Several NSIs have already embarked on various pilot projects testing and implementing AI/ML methods and tools in various domains leading to new or improved statistical products. Therefore, collaboration and co-creation is likely to boost development in this domain for creating the conditions for bringing these approaches into statistical production across the ESS.

The goal of the one-stop-shop for AI/ML is to guide and support ESS organisations, towards experimentation and eventually the use of AI/ML solutions for official statistics production (products and processes). On the one hand, to realise added value by aiming at providing effective and tailored guidance and assistance to the ESS organisation. On the other hand, to continue developing knowledge and provide mock up and standard for implementing use cases to produce official statistics -based on AI/ML solutions.

The One Stop Shop on AI/ML will be single entry point for ESS staff involved in innovation to identify concrete opportunities and to be guided and supported in deploying AI/ML solutions within adequate methodological and implementation frameworks in view of developing new statistical products and processes taking advantage of AI/ML state of the art research and developments. The scope of AI/ML applications is broad and includes existing applications in production pipelines as well as experimental and testing applications. Guidance should be adequate for basic, intermediate and advanced AI/ML implementation levels.

In addition, the One Stop Shop on AI/ML will continuously develop, maintain, and evolve a coherent set of relevant capabilities (including methodologies, guidelines, sandboxes, labelled data, processes, methodological, implementation and quality frameworks for implementing AI/ML-based solutions in official statistics across the ESS. It will provide a platform/hub providing a single gateway for ESS staff to access the relevant capabilities. It will provide support and guidance for the integration and maintenance of relevant AI/ML-based solutions in ESS organisations through training and active and efficient support. Fostering reuse of new solution, the consortium should build on open-source solutions and their active communities to provide a range of tools that can be leverage. The AIML4OS will implement new form of agile

cooperation allowing sharing ideas, experiences, success stories and lessons learned to stimulate innovation. Special attention will be given to facilitate the transition from development and experimentation of AI/ML-based solutions to industrialisation and production.

Beyond the already well-established use cases such as automatic classification editing and imputation and satellite image processing, LLM in the context of NLP will certainly receive much attention and the project should contribute to mature solutions than can be scale up at the level of the industry.

The case for international cooperation

As already stated, LLMs are still in their infancy and remain an uncharted field for Official Statistics. The skills and knowledge are scattered. No one fits all solution is expected and the pipelines to be designed will be realised by the combination of specialised modules to adapt to real life situation.

Customisation and fine-tuning of LLMs are pivotal activities for ensuring quality and performance of LLMs in statistical applications. Tailoring the LLM for statistics involves customising its parameters and compiling rich information corpus to be embedded in models and queries to ensure its relevance and accuracy within specific context of statistics.

In this context, the importance of data curation, metadata descriptions and semantic enrichment (providing more verbose descriptions of tables and columns), has a very significant impact on performance of models and ability to understand the data represented in the tables. Using rich descriptions of the tables enables for models to build a cohesive representation of Official Statistics dissemination system. Work is needed to prepare and optimise the way contextual data descriptions and definitions, together with vocabularies, to enable more advanced AI functionalities.

Part of these activities can more efficiently be performed at the level of Official Statistics industry leveraging common standards and assets such SDMX. It should also involve the active engagement of the community of users as social license and output validation are essential. By activating a community around the LLM, official statistics organisations and developers can create an environment conducive to knowledge sharing and collaboration. This collaborative approach not only enhances the model's capabilities but also ensures that it remains attuned to the evolving needs and challenges within the statistical community.

An important aspect is collaboration with statistical domain subject matter experts. It is instrumental in enhancing the effectiveness of LLMs for specific use cases. By actively engaging with experts in specific fields, LLMs can be improved with respect to accuracy and relevance. Seeking input from these experts also serves as a crucial validation mechanism, ensuring that the LLMs' output aligns (e.g., concerning quality) with the requirements and of specialised fields.

In the area of scalability and performance optimisation, despite the current trend is to development of much smaller model, very often provided with open-source license delivering core LLM capabilities, the requirements for a technical infrastructure remain significant. Bearing the cost of managed solutions provided by standard cloud providers may not be sustainable in the long run. It is crucial to achieve economies of

scale, effectively sharing resources and infrastructure. This collaborative approach would also help to moderate the environmental impact of LLMs, addressing concerns related to computational resources, energy consumption, e-waste, hardware production, carbon footprint of data centres, and the need for more sustainable practices.

In this context, international cooperation based on voluntary participation as initiated at UNECE level is essential waiting for more sustainable and structured approaches. In particular, the HLG MOS is launching in 2024 a project on generative AI. It capitalises on the momentum generated by the work conducted for white paper on generative AI. Participation is not limited to statistical organisations; it includes government agencies and research/academia. It should involve the sharing of use cases to share experiences and insights, not only for text based LLM output but also other types of multi-modal generative AI use (e.g., image generation). It will encourage co-development of solution(s) on common targeted use cases: that are of common interest for statistical organisations (e.g., chatbot), The project should contribute to accumulating and sharing experience in the community possibly addressing concrete recommendations on the use of generative AI for official statistics.

References

Eurostat. (2023) 'ESS innovation agenda' https://cros.ec.europa.eu/ess-innovation-agenda

UNECE. (2023) 'HLG-MOS White paper on LLM for Official Statistics'. https://unece.org/sites/default/files/2023-12/HLGMOS%20LLM%20Paper_Preprint_1.pdf

UNECE. (2024) 'International Framework on Responsible AI/ML for Official Statistics'. forthcoming publication

Ricciato, F, Wirthmann, A. (2020) "Trusted Smart Statistics: How new data will change official statistics." *Data & Policy*, 2: e7. doi:10.1017/dap.2020.7

Eurostat. (2024) 'An introduction to Large Language Models and their relevance for statistical offices', Statistical working papers, forthcoming publication.

The Potential of Generative AI in Official Statistics

External statistics after the pandemic: addressing novel analytical challenges
Madrid, 12-13 February 2024

Museux Jean-Marc
Chief Enterprise Architect
- Eurostat Unit A.5
jean-marc.museux@ec.europa.eu

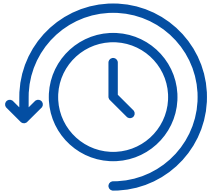
GIANNAKOURIS Konstantinos (Eurostat) & CIMPERMANN Miha (JSI)

Outline

- Acknowledgements
- Innovation of official statistics
- Generative AI – Large Language Model
- Risks and challenges
- Eurostat use cases (In Context Learning)
- Future perspectives:
  - International Collaboration
  - ESS One-Stop-Shop on AI/ML

European Commission

The Need for Innovation



Timelier, more detailed statistics needed to meet user demands

Rapid societal and technological changes require more frequent and more granular statistics

European Statistical System must be more responsive and agile and resilient to crises

ESS needs to expand use of digital technologies and new data sources

Innovation is vital for ESS to meet evolving user needs in a rapidly evolving data ecosystem

ESS Innovation Agenda

Goals

- **Strengthen the ESS ability to respond rapidly to new and urgent user needs**
- **Augment products and service portfolio for meeting policy needs**
- **Realize efficiency gains to free up resources**
- **Strengthen resilience to shocks and adapt to societal change**

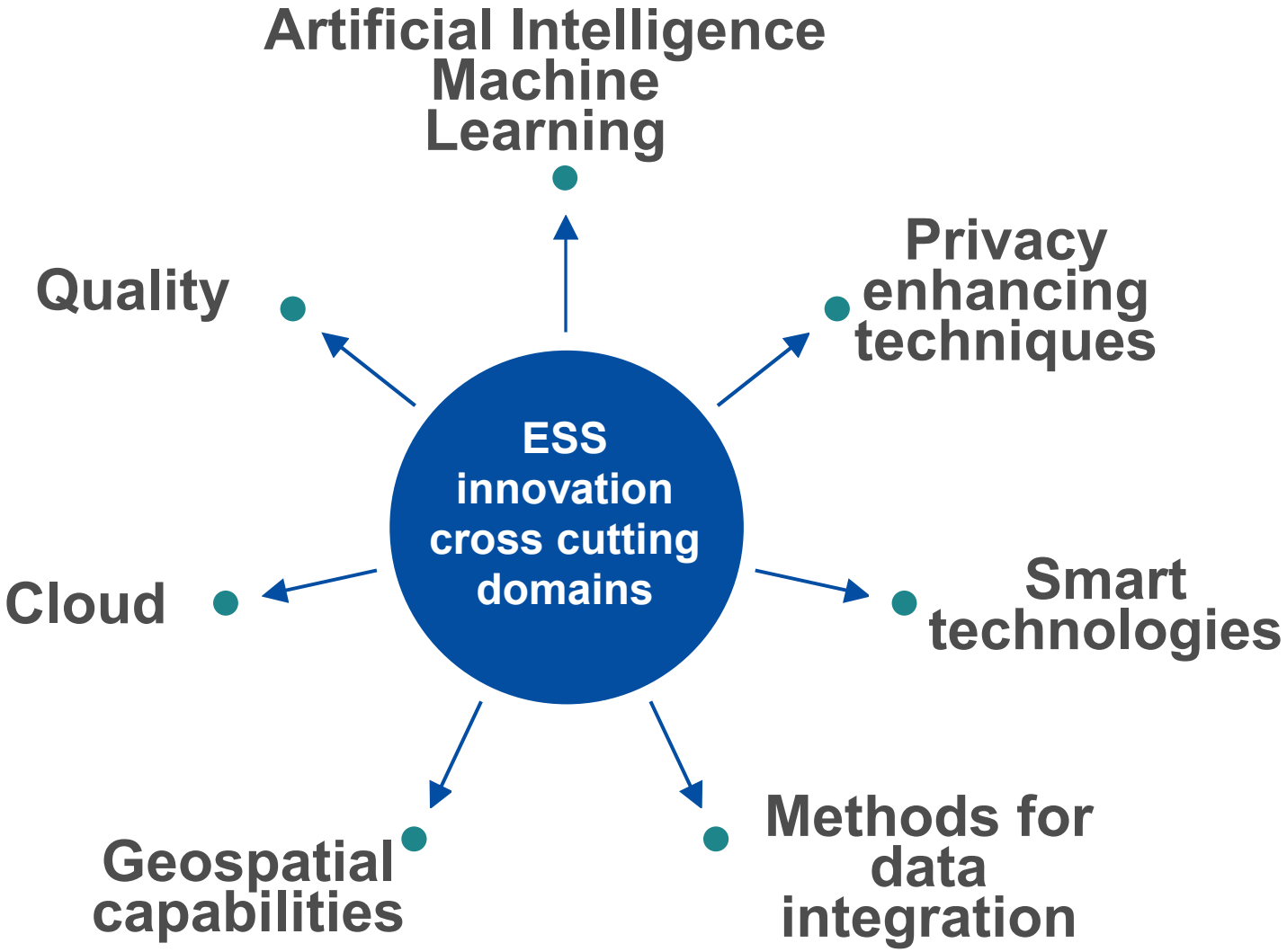
Means

- Leverage key innovation **drivers and enablers**, offered by the recent technological developments
- Promote internal and external **cooperation**
- '**Cross-domain**' thinking and working
- Focus on implementing in statistical production

European Commission

Methodological and Technological innovations

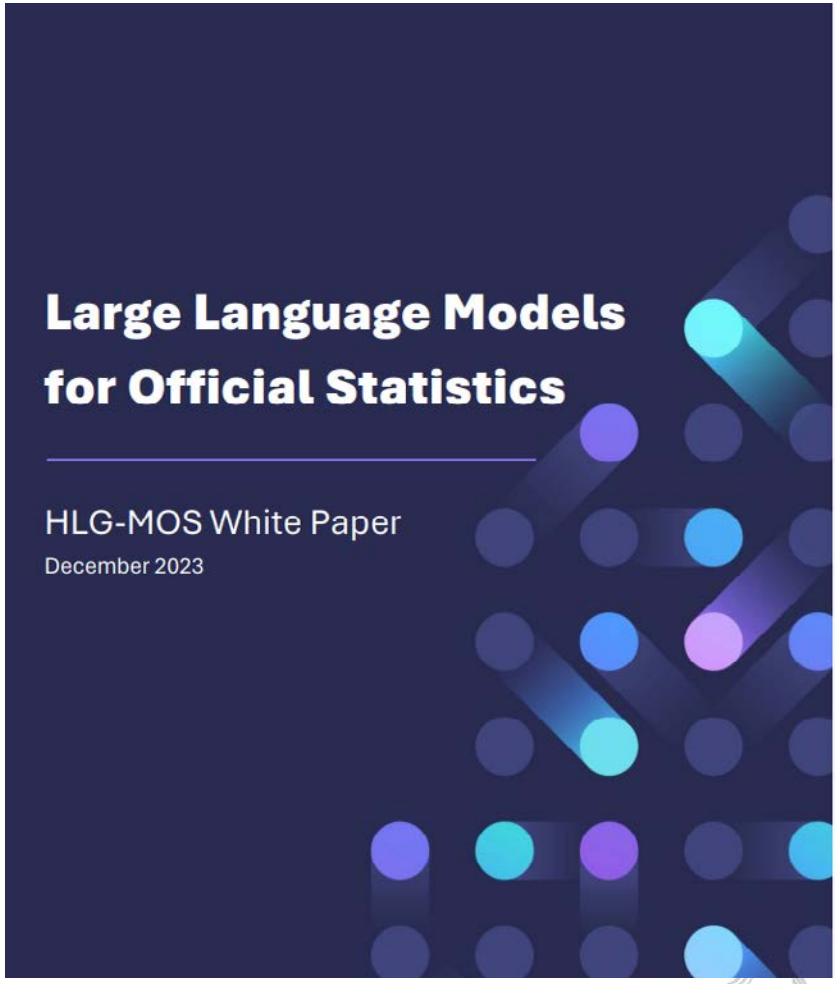
TOP priorities



European Commission

AI & Official Statistics
Improving efficiency and statistical services

- Data collection
- Data processing and cleaning
- Quantitative data analysis (imagery, web)
- Qualitative data analysis and extraction of findings
- Data and knowledge discovery
- Improving data user experience (Chatbots)
- Operational efficiency of the supporting services: translation, memos, emails, summarization



White paper on generative AI for official statistics (unece.org)

AI disruption: Generative Pretrained Transformer

GPT-1: 2018

117 million parameters

GPT-2: 2019

1.5 Billion parameters

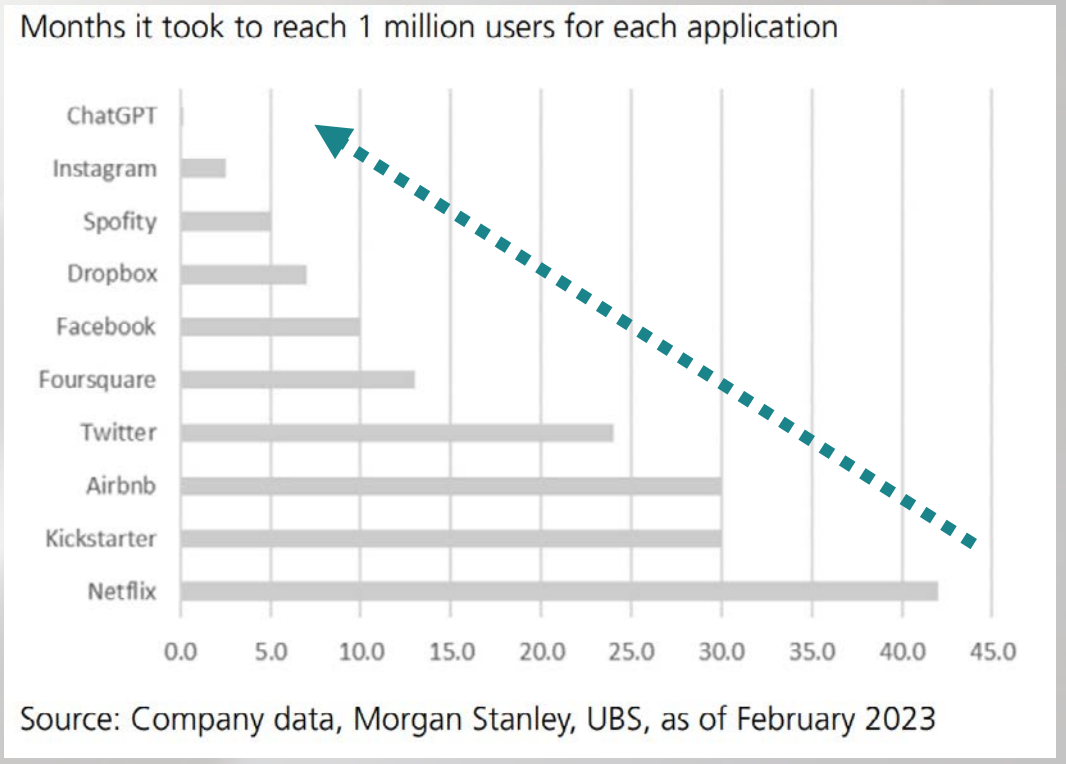
GPT-3: June 2020

175 billion parameters

ChatGPT

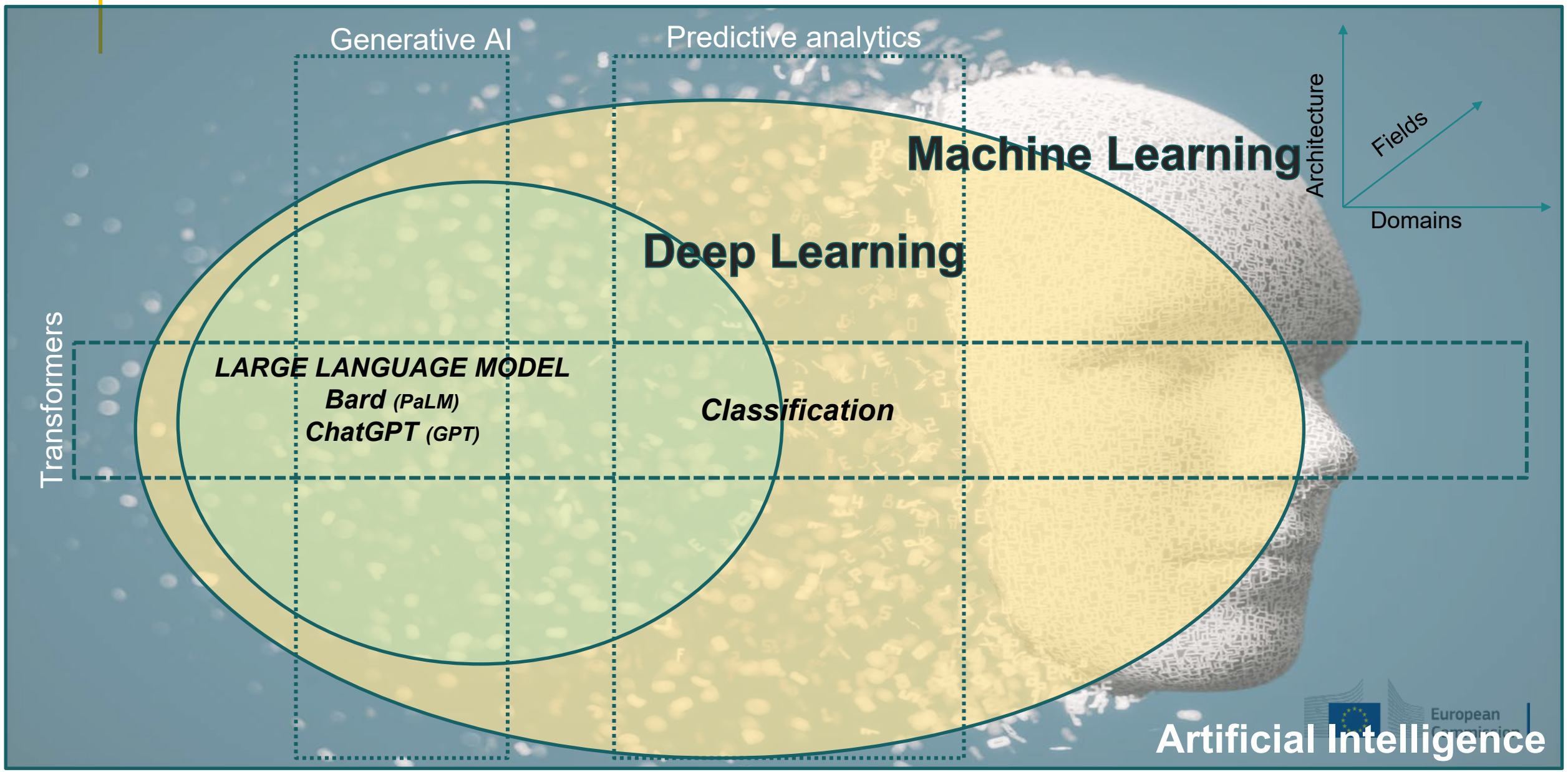
GPT4: Early 2023

100 trillion parameters



European Commission

AI landscape



Large Language Models

What are they ? Advanced AI systems trained on vast amounts of text data to generate human-like text.

How they works

Models predict the next word in the output sentence

Content is generated by a deep neural network. Parameters are adjusted during the training

Strengths

Efficient at text-related tasks, can understand nuance and generate human-quality text.

Foundational model can be fined tuned for specific tasks

Standard tasks

Analyze text, answer questions, summarize, translate and generate new text.

Limitations

Prone to hallucinations, inaccuracies, and perpetuating biases in training data.

Costs of finetuning

European Commission

Risks and Challenges

- Ethics

LLMs raise issues regarding ethics and social acceptability.

Ethical frameworks need adapting to LLMs:

human oversight and user feedback.

- Accuracy and Bias

'Hallucinations' and propagating biases from training data are key issues.

New semantic closeness measures are needed to assess LLMs.

- Privacy, Data Protection & IPR

Rights to privacy, data protection, IPR apply to LLM training data and derived user output.

Data governance needs strengthening for audit trails.

- Transparency

LLMs are 'black boxes': limited access to algorithms, input data and parameters.

Use should be supported by documentation and communication on limitations and bias potential.

Eurostat use cases (exploration)

UC Level 1: Statistical advisor

The application enables users to search for the relevant Eurostat data, based on simple text description and recommends the relevant data sources to be used.
A chat session assistant.

UC Level 2: Statistical assistant

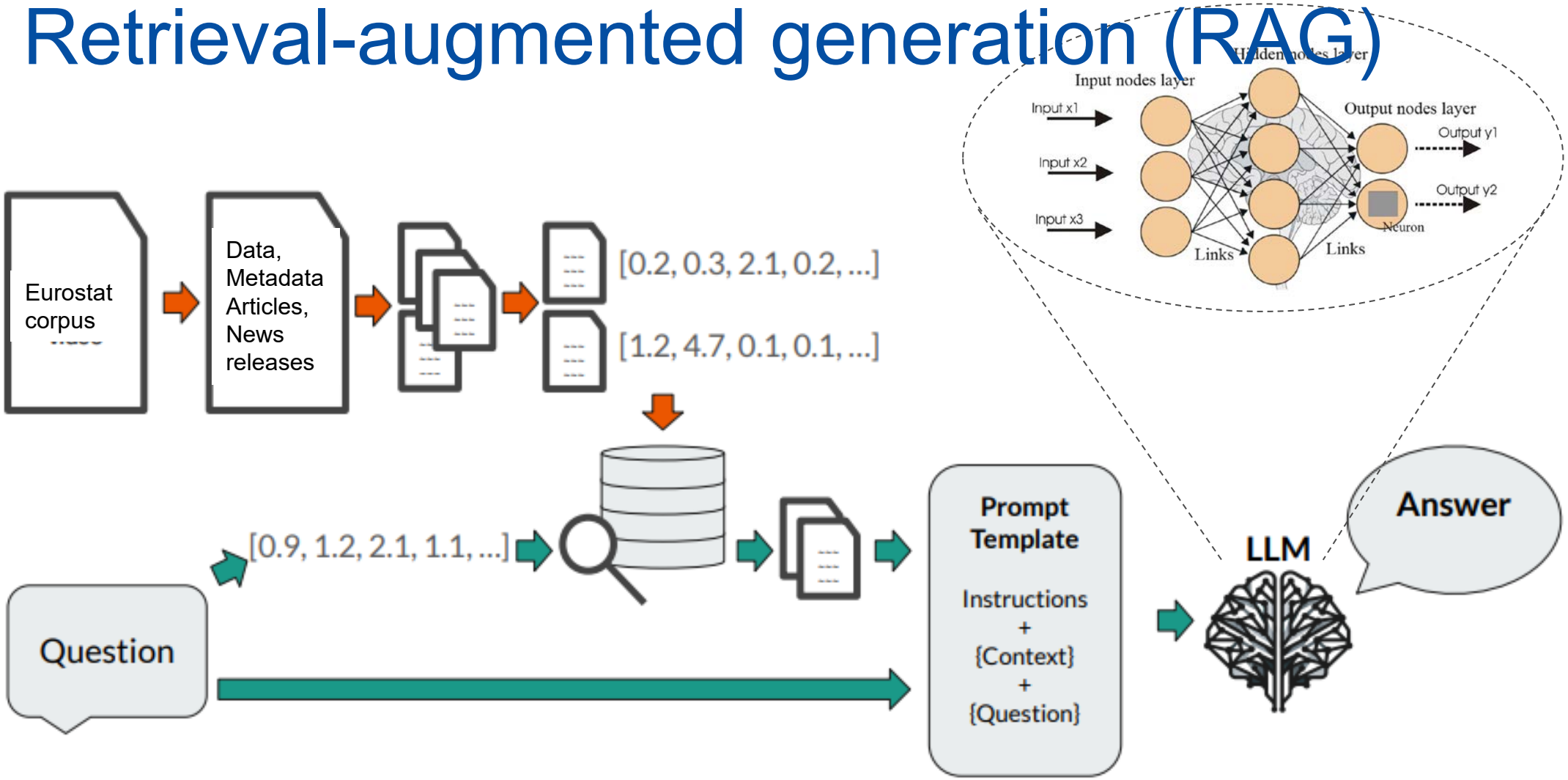
The application enables searching and retrieving relevant Eurostat data, based on problem description.
The data would be available in different formats (graphs, tables, extraction scripts) accompanied with relevant description.
Chat session + numerical data retrieval

UC Level 3: Analytical assistant

Performing analysis on top of Eurostat's data corpora, based on problem description.
Chat sessions + complex analytical data processing
.

European Commission

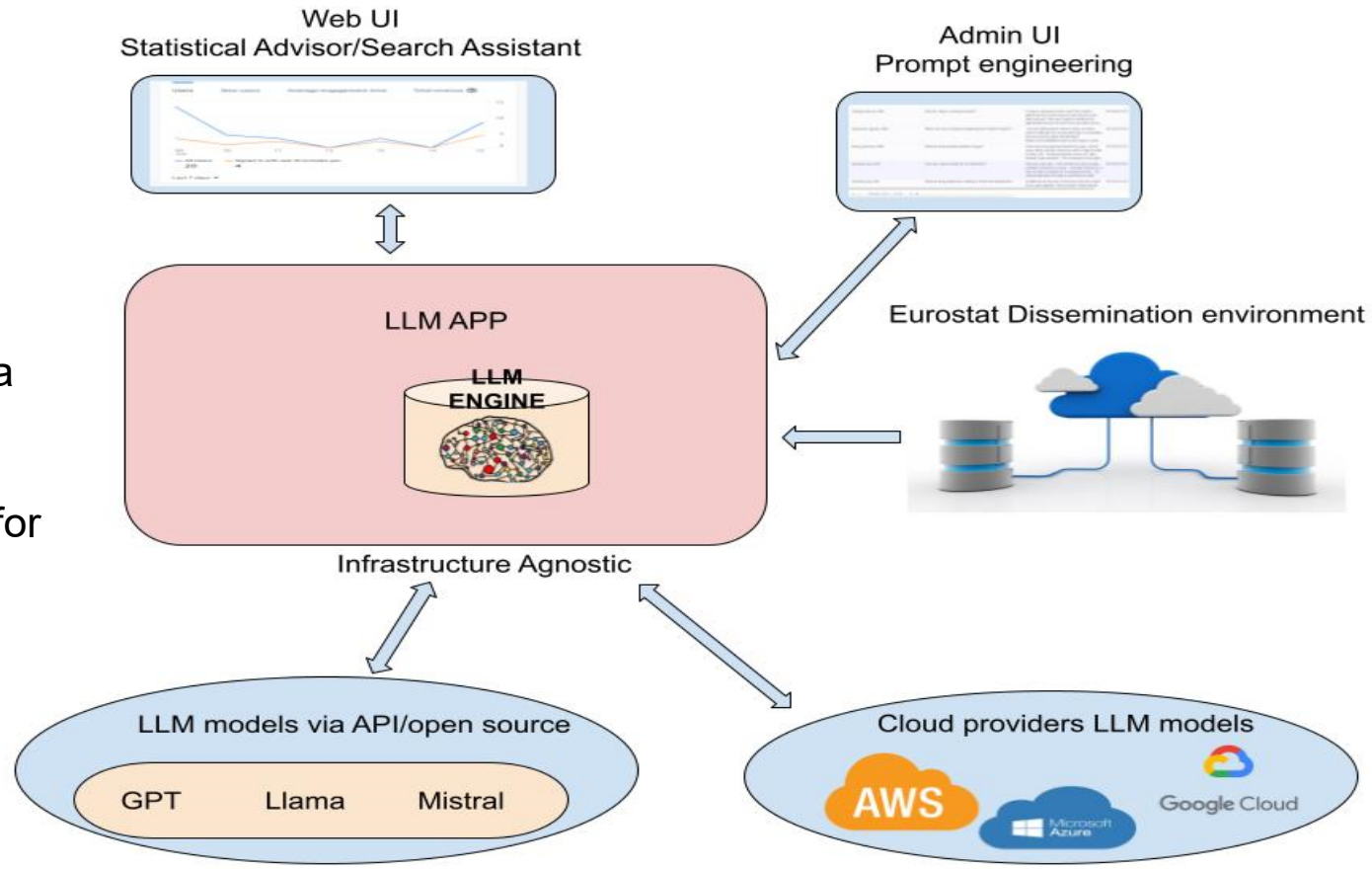
In-Context Learning (ICL)
Retrieval-augmented generation (RAG)



European Commission

Eurostat experimentation evolvable workbench

- LLM models: able to use open source models on premise and dedicated cloud-based proprietary instances such as MS Azure or AI studio
- Eurostat's assets (metadata description, catalogues, articles) embedded in the vector data base
- Web UI for the users of the Statistical assistant, for information search and retrieval accessible from Eurostat's dissemination environment
- Admin Web UI for Administrator of the system to do prompt engineering or other configurations



Eurostat aims to cautiously evaluate LLMs capabilities benchmarking the different models available

Case for international cooperation

Model Customization

Collaboration needed for model finetuning which requires data and domain expertise.

Data Curation

International cooperation to curate Official Statistics data corpus and datasets for training generative AI models.

Pooling skills

Sharing knowledge across countries will aid in customizing models for specific statistical use cases.

Infrastructure Sharing

Pooling computational resources through infrastructure sharing can help reduce the environmental impact of large AI models.

Accumulating Experience

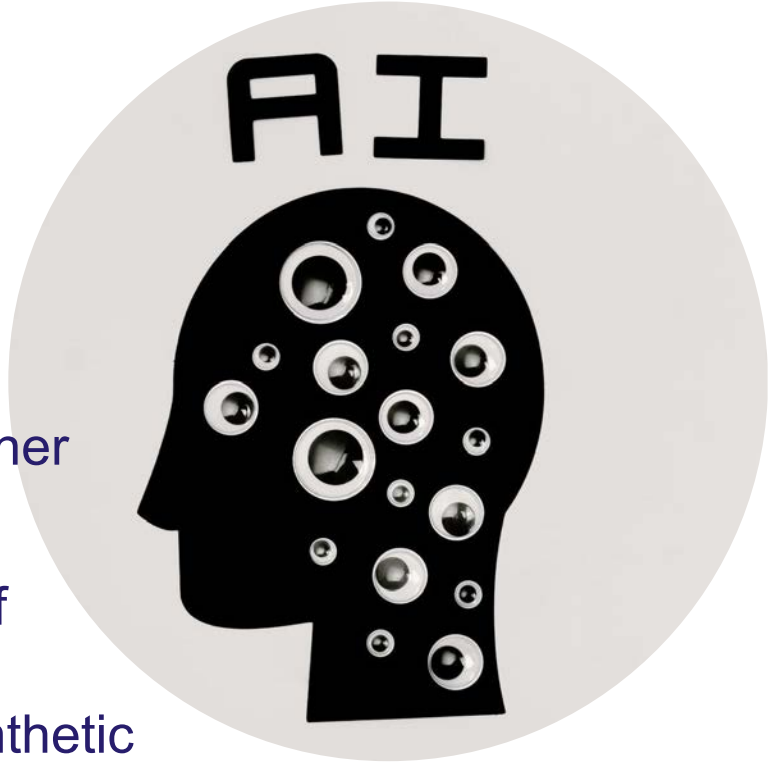
Cooperative initiatives allow accumulating experience in applying generative AI to official statistics.

Developing Use Cases

International collaboration enables exploring and developing new use cases for generative AI in statistics.

European Commission

AI/ML One-Stop-Shop



A single-entry point for NSIs staff

- Sandbox infrastructure to test and pilot AI/ML approaches together with training and coaching
- Reference implementation of AI approaches in the production of official statistics (e.g. information extraction, data editing and imputation, automatic coding, use of LLM, network analysis, synthetic data / predictive analytics)
- Communities to share ideas, experiences, success stories and lessons learned to stimulate innovation based on the use of AI/ML.
- Best practices to facilitate the transition from development and experimentation of AI/ML-based solutions to industrialisation and production

Thank you

© European Union 2023

Unless otherwise noted the reuse of this presentation is authorised under the CC BY 4.0 license. For any use or reproduction of elements that are not owned by the EU, permission may need to be sought directly from the respective right holders.

European Commission