**UNECE High-level Group for the
Modernisation of Official Statistics**

modern:stats
by HLG - MOS

Business Case for "Generative AI for Official Statistics"

> This business case was prepared by Applying Data Science and Modern Methods Group, Blue Skies Thinking Network and UNECE, and is submitted to the HLG-MOS for their approval.

Type of Activity			
☒	New project	☐	Extension of existing project

Purpose

In recent years, statistical organisations have utilised machine learning (ML) to enhance the efficiency of resource-intensive tasks, such as classification and coding, editing and imagery analysis. ML has also enabled the production of new statistical products, such as sentiment indicators derived from social media posts. Some statistical organisations started exploring the application of ML in other parts of the GSBPM, but it is too early to understand the full potential of ML.

While these applications are primarily focused on improving decision-making efficiency (e.g., classification of textual job description into occupation codes, identification of outliers), the introduction of ChatGPT in late 2022 marked a significant shift in the field and increased the number of potential use cases of AI for official statistics.

Large Language Models (LLMs), which ChatGPT is based on, have unique capabilities which set them apart from traditional ML applications, such as being able to generate texts and engage in human-like conversations. Traditional ML applications, on the other hand, are primarily focused on assisting humans in prediction-tasks rather than creating content. Moreover, the introduction of user-friendly services based on LLMs (e.g., ChatGPT by OpenAI, Bard by Google) has lowered the barrier to generative AI beyond the research community to the wider public. These developments have substantial implications for statistical organisations, affecting their workflows and product dissemination methods.

In response to these new developments, the HLG-MOS Modernisation Groups - Blue Skies Thinking Network (BSTN) and Applying Data Science and Modern Methods (ADSaMM) Group - jointly initiated a white paper on the use of LLMs for official statistics in mid-2023. This report, currently in the finalisation stage, explore key issues such as the implications for official statistics, use cases within statistical organisations, associated risks, and provides some recommendations and strategic considerations. In parallel to the drafting of the white paper, there have been a series of "show-and-tell" sessions to demonstrate the concrete examples of ChatGPT/LLM application (e.g., SAS to R translation, using LLM for dissemination, using ChatGPT for updating classification system)

Building on the current works from the LLM white paper, the "Generative AI for Official Statistics" Project aims to investigate further on the strategic considerations arising when statistical organisations want to use generative AI effectively and responsibly (e.g., governance, open-source models), as well as co-develop concrete solutions.

Description of the project and the Work Packages/sub-activities

Given the fast-changing nature of this field, it is important to maintain the agile approach. Below are some activities that are considered particularly important at the current time. These activities

are generally based on the work conducted by the HLG-MOS white paper draft team. However, the choice of activities is left somewhat open at this stage, with the selection of activities to be finalised if the project is launched.

Activity 0. Scoping and re-orientation if needed

Activity 1. Sharing use cases: the project will organise regular online meetings to share practical use cases and insights, not only for text based LLM output but also other types of multi-modal generative AI uses (e.g., image generation). Participation is not limited to statistical organisations; it includes government agencies and research/academia provided its relevance to official statistics.

Activity 2. Co-development of solution(s) on common use cases: the project will identify use cases that are of common interest for statistical organisations (e.g., chatbot), and develop open-source solutions.

Activity 3. Compiling practices and concrete recommendations: based on the activities 1-2 as well as the HLG-MOS white paper, the project will work on the concrete recommendations on the use of generative AI for official statistics. It is essential to focus on a few themes that are particularly relevant to statistical organisations, instead of aiming for developing a comprehensive guidebook.

Deliverables and timeline
Activity 0: updated project proposal document (January 2024) **Activity 1:** compilation of presentations and slides a shared GitHub repository of generative AI services developed by statistical organisations, if applicable, to be continuously updated throughout the project period (Jan. – Dec. 2024) **Activity 2:** open-source solutions based on generative AI (Jan. – Dec. 2024) **Activity 3:** recommendations on the use of generative AI for official statistics (Jan. – Dec. 2024)
Offices/Countries committed
Currently, no offices or countries have formally committed, but we expect sufficient interest based on recent activities in this area.
Alternatives considered
An alternative would be no action. There might be other international initiatives setting up on the topic, but it would be a missed opportunity not to respond to the quickly growing demand by leveraging existing works and our established network of experts.
How does it relate to the HLG-MOS vision and other activities under the HLG-MOS?
Generative AI is an emerging technology that could change the way statistical organisations work in a transformative way, thus the activity is highly in line with the vision of HLG-MOS. The project also has a close relation with other activities under the HLG-MOS such as the HLG-MOS Machine Learning Project (2019-2020), the joint collaboration with ONS for the Machine Learning Group (2021-2022) and Machine Learning Workshop (2023).
Proposed start and end dates

Start: January 2024	End: December 2024