

Application of Artificial Intelligence in official statistics in the context of the Fundamental Principles of Official Statistics

Zastosowanie sztucznej inteligencji w statystyce publicznej w kontekście *Fundamentalnych zasad statystyki publicznej*

1. Introduction

The introduction of Artificial Intelligence (AI) into the proceedings involving the collection, processing and presentation of statistical data opens up new horizons, but also poses new challenges. AI, as a powerful technological tool, can significantly contribute to accelerating and facilitating statistical processes by automating many tasks. For example, machine learning algorithms can be used to analyse large data sets, identify patterns and forecast trends, which can considerably increase the efficiency of statistical agencies.

Nevertheless, the advent of AI poses additional risks to the integrity, clarity and impartiality of statistical data. Automating statistical operations using AI involves potential dangers. It may result in the overreliance on algorithm-generated outcomes, disregarding the necessity for the critical examination and interpretation of data. Furthermore, AI systems are susceptible to many types of mistakes and biases, resulting in the risk for inaccurate or prejudiced findings.

The paper is a form of an experiment, as it was written with the use of one of the modern technology tools, i.e. AI. ChatGPT 3.5 created 90% of the content. The study is meant to determine the performance of AI in terms of preparing a scientific article dealing with issues relating to the FPOS. The writers' responsibility, on the other hand, was restricted to generating prompts (questions), logically evaluating the responses provided by ChatGPT 3.5, organising them into the format of a scientific article, compiling a reference list and preparing an introduction and a summary. It should also be emphasized that the article was generated in May 2024. Since then, AI-related tools have developed significantly, although the principles of working with AI tools such as prompt engineering or response verification (hallucination checking) have not changed. The Figure illustrates one of the questions that AI was asked to answer.

Figure. Sample question to ChatGPT 3.5**Prompt**

You are a scientist drafting a scientific article on the application of artificial intelligence in official statistics, guided by the Fundamental Principles of Official Statistics. I will provide topics for individual paragraphs and you will generate their content. Use a formal academic style and adopt a moderately creative approach in crafting the responses.

First topic: What are the new challenges and tasks for official statistics in the context of the Fundamental Principles of Official Statistics arising from the widespread adoption of new information technologies, the growth of the information society, the global knowledge-based economy and global and regional political and social crises?

ChatGPT answer

Here is a draft paragraph on your first topic, written in a formal academic style and aligned with the Fundamental Principles of Official Statistics:

New Challenges and Tasks for Public Statistics in the Context of Contemporary Technological and Societal Transformations

The landscape of official statistics is undergoing a profound transformation under the combined pressures of rapid technological advancement, the proliferation of the information society, the emergence of a global knowledge-based economy, and the recurrence of global and regional political and social crises. These developments present both opportunities and challenges for official statistics, particularly when viewed through the lens of the Fundamental Principles of Official Statistics (United Nations, 2014). Principle 1, which asserts the relevance, impartiality, and equal access to official statistics, is especially salient as statistical offices must now contend with an unprecedented volume and velocity of data generated through digital platforms, sensor networks and other non-traditional sources. The integration of such data necessitates new methodological frameworks, rigorous quality assurance processes and heightened attention to ethical considerations to ensure that the resulting statistics remain trustworthy and fit for purpose.

Note. The questions and answers were first created in Polish and then translated into English.

Source: authors' work based on ChatGPT 3.5.

It should be mentioned here that the free version of ChatGPT 3.5 produces responses of a limited length. An outline of the article was initially created, then refinement questions were constructed to further explore the previously obtained replies.

Within the realm of data collection, AI may be applied to autonomously gather and analyse information from a variety of sources, therefore expediting research procedures and expanding the amount of accessible data. However, using AI for this purpose might result in the development of 'black boxes', i.e. systems founded on intricate algorithms that are challenging for humans to comprehend and manage. This can consequently lead to a deficiency in openness and accountability in the procedures of data collecting.

AI may also be utilised to analyse and understand extensive data sets in order to identify and understand patterns and trends. Nevertheless, there is a potential danger that AI algorithms will produce outcomes influenced by unintentional biases or erroneous assumptions, resulting in distorted conclusions and judgements. Moreover,

the use of AI in data processing might pose challenges for human comprehension and validation, resulting in a lack of openness and impartiality in data representation.

AI may be used in official statistics to automatically generate reports, charts, analyses and publications that provide information and data. This can greatly improve the efficacy and efficiency of statistical communication procedures. On the other hand, a potential risk arises that automatically created reports might contain mistakes and inaccuracies or be excessively brief, potentially resulting in the dissemination of false information or the misinterpretation of data by the recipients. The utilisation of AI in data presentation might also lead to the deprivation of context and comprehension, hence restricting the usefulness of the data to the end users.

This article analyses the above-mentioned challenges and explores the potential conflicts between the use of AI and the rules set out in the Fundamental Principles of Official Statistics (FPOS). The main aim of the article is to present an analysis of the potential advantages and disadvantages that arise from the application of AI in the collection, processing and presentation of statistical data. Additionally, the paper provides strategies to mitigate these risks and assure the adherence to the FPOS when dealing with statistical data.

2. FPOS – basic information

The United Nations Economic Commission for Europe (UNECE) developed the FPOS in 1992 for the UNECE region. Subsequently, in 1994, the United Nations Statistical Commission universally implemented them. The FPOS were accepted by the UN Economic and Social Council in 2013 and then ratified by the General Assembly in January 2014. This acknowledgement at the highest level emphasises the significance of dependable and unbiased statistical data for the process of making decisions. These principles include (United Nations Statistics Division, 1994):

- Relevance impartiality and equal access: official statistics are the key element of the provision of information to democratic societies, offering data on different areas of life to governments, the economy and the public. In order to meet the criteria of usefulness and respect the right to information, official statistics must be compiled and made available by national statistical institutes in an impartial manner;
- Professional standards and ethics: national statistical institutions must make decisions based on widely-understood professionalism, taking into account scientific principles and professional ethics when collecting, processing, storing and presenting statistical data;
- Accountability and transparency: to facilitate the correct interpretation of data, national statistical institutions should present information on the sources, methods and procedures in accordance with scientific standards;

- Prevention of misuse: national statistical institutions have the right to respond to misinterpretations and abuses of statistics;
- Sources of official statistics: statistical data can come from a variety of sources, such as statistical surveys or administrative registries. National statistical institutions should select sources considering the quality, timeliness, costs and the workload respondents need to take on;
- Confidentiality: individual data collected by national statistical institutions are confidential and used only for statistical purposes;
- Legislation: the legal provisions governing the operation of statistical systems are open to the public;
- National coordination: collaboration between national statistical institutions and other authorities is essential for the coherence and efficiency of the statistical system;
- Use of international standards: the use of international definitions, classifications and methods supports the consistency and efficiency of statistical systems;
- International cooperation: collaboration at an international level contributes to the improvement of official statistics systems around the world.

Various research centres worldwide, tasked with handling official statistics, provide their own guidelines and standards (e.g. National Research Council, 2013; Office of Management and Budget, 2014; Statistical Office of the European Communities, 2011). These guidelines should serve as a reference for organisations engaged in the collection, processing and presentation of statistical data. Along with the FPOS, they aim to guarantee the superior quality of statistical data and their trustworthiness, coherence and comparability. The most important of these guidelines and standards include:

- Objectivity: statistical data should be collected, processed and presented in an objective manner, regardless of personal preferences, views or interests;
- Credibility: statistical data should be based on trustworthy and verified sources of information and data collection methods to ensure their reliability;
- Representativeness: samples should be representative of the population they are meant to reflect. For this reason, the sample should be properly sized to accurately reflect the characteristics of the general population;
- Consistency: statistical data should be consistent across time and space, meaning that data collection methods, definitions and classifications should be the same at different times and places;
- Confidentiality and privacy: personal information should be protected in accordance with the applicable privacy laws. Institutions that collect statistical data are required to maintain the confidentiality of personal information;

- Transparency: the processes of collecting, processing and presenting statistical data should be transparent and accessible to the public. This allows a better understanding of the methodology and results;
- Institutional independence: institutions dealing with official statistics should operate independently of political influence or private interests in order to maintain their credibility and objectivity;
- User consultation: statistical institutions should consult data users about their needs and provide them with relevant statistical information.

Adhering to these core principles ensures that statistical data are valuable, dependable and beneficial for decision-making in political and socio-economic contexts. Minor variations in the principles above have led to a variety of research articles in the literature comparing these principles (Pierson, 2015). The author's work involves comparing the principles formulated by organisations in the USA, EU or UN. Additionally, the author emphasises the need of informing decision-makers and the public about the standards of official data. Furthermore, the pivotal significance of statistics in relation to governments, the economy and democracy is highlighted. The author specifically refers to statisticians as protectors of democracy.

3. FPOS and AI – challenges and opportunities

With the widespread use of modern information technologies and the growth of the information society and the global knowledge-based economy, official statistics is encountering various new challenges and responsibilities. These challenges arise due to global and regional political, social, economic and environmental crises. The most important ones are displayed in the Table.

Table. Key challenges for official statistics

Challenge	Description
Big data management	Along with the development of information technology, official statistics has to deal with vast amounts of data that can be used to analyse social, economic and environmental trends. The challenge is to collect, process, analyse and interpret these data effectively
Protection of personal data	In the digital age, the protection of personal data is becoming a key challenge for official statistics. Data privacy awareness and needs are increasing, which requires proper security measures and compliance with data protection regulations
Data availability and transparency	Modern society expects greater access to data and transparency in the process of their collection and analysis. Official statistics must provide the public with easy access to data and ensure the transparency of the methodologies and results

Table. Key challenges for official statistics (cont.)

Challenge	Description
Dynamic and flexible official statistics	In the face of rapid social, economic and environmental changes, official statistics must be dynamic and flexible in order to respond quickly to new challenges and societal needs
Multidimensional data analysis	Globalisation and the complexity of contemporary issues require the analysis of multidimensional data, covering various areas of social, economic and environmental life. Official statistics must develop new methods of data analysis that take into account this multiplicity of dimensions
Interdisciplinarity	In order to effectively deal with contemporary challenges, official statistics must be interdisciplinary, collaborating with other fields of science, such as economics, sociology, political science or environmental sciences
Responding to crises	Official statistics must be able to respond quickly to political, social, economic and environmental crises by providing relevant data and analyses that can support decision-making and action
Education and public awareness	Greater public awareness of the growing importance of statistical data is needed, as is the public's ability to interpret the data. Official statistics must work to educate the public about statistics and promote its understanding and usefulness
Application of AI	The introduction of AI generates threats to the quality, transparency and fairness of statistical data or publications. There is a risk that the AI-based automation of statistical processes may lead to excessive trust in the results generated by algorithms, bypassing the need for critical evaluation and interpretation of data. In addition, AI algorithms can be prone to different types of errors and biases, which can result in false or biased conclusions

Source: authors' work.

The challenges outlined in the Table necessitate the adoption of novel methodologies and tools, and the collaboration between various sectors of the economy, society and statistical agencies. This is crucial for official statistics to adequately fulfil its role in facilitating decision-making processes, promoting social development and effectively implementing the FPOS.

The application of AI in official statistics has the capacity to transform the methods by which we gather, handle and scrutinise statistical data on a domestic and global scale. The adoption of AI offers such opportunities as:

- automating data collection: AI can be used to automatically collect data from a variety of sources, such as the Internet, administrative registrations, or IoT (Internet of Things) sensors. AI systems can analyse vast amounts of information in a short period, accelerating the process of collecting statistical data and reducing the costs associated with gathering information manually;

- improving data quality: with AI technologies, it is possible to detect and eliminate errors in statistical data in real time. Algorithms can identify unusual or abnormal patterns in the data, allowing for a quick correction and improvement of data quality;
- data analytics: AI enables more advanced statistical analysis, such as predictive analytics, classification and data clustering. With the help of advanced machine learning and deep learning algorithms, hidden patterns in data may be uncovered, leading to a better understanding of the analysed phenomena;
- personalisation of reporting: AI systems can generate personalised reports and statistical analyses, tailored to the specific needs of the users. This allows a quicker and more effective use of statistical data in decision-making processes;
- trend detection and forecasting: AI can be used to identify trends based on historical data and to forecast future events and trends. This allows decision-makers to make more accurate and truthful decisions based on data-driven forecasts;
- process optimisation: the integration of AI in official statistics can lead to the optimisation of data collection, processing and analysis, which translates into saving time and resources and increasing the efficiency of statistical systems;
- data integration: AI can be used to integrate data from different sources and to create consistent statistical databases. As a result, it is easier to compare data from different areas and conduct comprehensive multidimensional analyses;
- automatically generated reports, data visualisations and publications: AI systems can produce reports and visualisations of statistical data automatically, making it easier to convey the results of analyses and enabling users to understand information faster.

The use of AI in the area of official statistics offers new possibilities and enables the utilisation of statistical data in a more efficient and all-encompassing manner, thus facilitating the implementation of the FPOS. Nevertheless, it is crucial to monitor data effectively and ensure the compliance with data protection and privacy legislation to guarantee the safety and trust of statistical data consumers. The supervision of contemporary IT solutions, including AI is equally important, since its improper integration poses a threat to the successful execution of the FPOS. Specifically, the utilisation of AI can result in the creation of 'black boxes' which are systems built on complex algorithms that are challenging for humans to understand and manage. This can result in a lack of transparency and accountability in terms of the data gathering procedures and the release and interpretation of data.

4. Recommendations and discussion

Integrating AI into statistical operations requires a carefully planned strategy that aligns with the FPOS. Statistical authorities must guarantee that the utilisation of AI is transparent, fair and done responsibly. Monitoring AI processes and evaluating their adherence to the FPOS is also crucial. The difficulties pertaining to the use of AI in official statistics must be tackled within the framework of updating and interpreting the FPOS.

To mitigate the potential hazards linked to the application of AI in official statistics and to guarantee the adherence to the FPOS, statistical organisations can adopt several precautionary measures, including:

- an ethical evaluation of algorithms: before introducing AI algorithms, a thorough assessment of their impact on the quality, fairness and transparency of statistical data needs to be conducted. This evaluation should take into account potential biases and ethical issues related to the design and implementation of algorithms. It is also crucial to follow the principles of data ethics, such as the protection of privacy and fairness in the collection and processing of data;
- regular audits: conducting regular audits of AI algorithms can help identify and eliminate possible errors or biases that may arise during data processing as a way to maintain the accuracy and reliability of statistical results. As part of the audits, it is also worth analysing decision-making processes based on AI algorithms to ensure that they comply with specific ethical standards and the FPOS;
- open data and algorithms: providing access to both data and AI algorithms fosters greater transparency and accountability in statistical processes. Making datasets and algorithms available to the public enables their extensive analysis and control, ensuring the integrity of statistical data. In addition, making them accessible to a wider range of individuals and institutions is likely to stimulate innovation;
- verification of the reliability of data publications and other publications focusing on data analysis, including scientific ones: an important element in the process of using AI in official statistics is also the verification of the reliability of data publications and other data-based publications, including scientific ones. It is necessary to ensure that the published results are reliable, supported by sound methodologies and free from errors or inaccuracies. For this purpose, peer review mechanisms, critical analysis and data verification systems should be used.

The implementation of the above-mentioned measures allows for a comprehensive approach to mitigating the challenges associated with the use of AI in official statistics and strengthens its compliance with the FPOS. Moving towards more reliable processes for evaluating, auditing and sharing data can help improve the quality, reliability and integrity of statistical data in the age of AI.

The article provides a clear indication of the advantages and challenges that arise from the use of AI while retaining the adherence to the FPOS. The input of the article is in 90% AI-generated, while the structure of the article and the refinement questions further exploring the previously obtained responses were prepared by the authors. This methodology shows that the issues identified by ChatGPT 3.5 are both contemporary and significant, and the AI-generated responses themselves are transparent and logical. The explanations of the mentioned assertions are well-defined, providing the reader with information about the subject matter of the text. What is more, the solutions proposed by AI also appear to be well-founded. This proves that the use of AI may have several advantages. However, it is important to note that the authors had to validate the data sources referenced by AI as part of the careful examination of the fragments of the analysis produced by ChatGPT 3.5.

The verification of the sources indicated that the overwhelming majority of the mentioned publications are completely non-existent. An instance of this may be observed in the following articles referenced by ChatGPT 3.5:

- Trewin, D. (2007). The principles of official statistics. *Australian Economic Review*, 40(4), 446–460;
- Cox, L. H. (2013). International statistical standards and the emergence of national statistical systems: A genealogy of the United Nations Fundamental Principles of Official Statistics. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 176(2), 431–453.

The thorough search for the above-mentioned publications confirmed that they had, in fact, never been written. It should also be emphasized again that the article was written in May 2024. Since then, AI tools have developed significantly. Many LLM models have ‘deep search’ functions, which, however, do not allow for the complete elimination of the problem above and still require human verification.

In addition, the AI-generated replies to the human-formulated questions lacked pertinent research material related to the subject, such as the study conducted by Rozkrut et al. (2021), which explores the FPOS in relation to eight emerging sources of big data. According to the FPOS, official statistics can be generated using any data source. In order to ensure the compliance with the first FPOS (Rozkrut et al., 2021), it is imperative to utilise these sources inside national statistical systems. It is therefore evident that the mentioned publication aligns perfectly with the topic of this article, yet AI failed to include and consider it. In conclusion, the fact that AI mentioned publications which proved unreliable, unveiled one of the risks of using unverified AI-generated information and materials within FPOS-oriented official statistics.

5. Conclusions

The application of AI in official statistics which is rooted in the FPOS has its advantages, but may as well pose certain challenges. The ability of AI to automate processes, detect trends and enhance efficiency may greatly improve statistical operations. Nevertheless, there is a potential danger that algorithms might produce outcomes influenced by unconscious biases or erroneous assumptions, resulting in skewed conclusions and judgements. To mitigate these dangers and optimise the advantages of using AI in official statistics, the implementation of suitable safeguards is important as is the rigorous adherence to the FPOS. It is crucial to carry out an ethical assessment of the applied algorithms, to conduct frequent audits, make data and algorithms accessible, and ensure the credibility of data and scientific papers in this particular situation (prepared using AI).

It is vital to consider the difficulties associated with the application of AI in official statistics while updating and interpreting the FPOS. Further exploration should be conducted on the integration of AI in statistical processes, with a careful consideration of the ethical, technical and methodological factors. Furthermore, it is crucial to consistently monitor and evaluate the influence of AI on the accuracy of statistical data and to implement necessary measures if any potential issues are identified. By employing AI, statistical procedures may be enhanced and the accuracy and dependability of official data upheld.

This article serves as an experimental exploration of the controversial application of AI in drafting scientific papers, with a focus on its use in official statistics. It demonstrates that AI, when guided by appropriately formulated human instructions, can produce coherent and plausible content. However, the study identifies significant limitations in AI's capabilities that may compromise the accuracy and reliability of research outcomes, particularly in terms of adhering to the FPOS. These shortcomings necessitate rigorous human oversight and verification to ensure the validity of AI-generated outputs.

References

- National Research Council. (2013). *Principles and Practices for a Federal Statistical Agency* (15th edition). The National Academies Press. <https://doi.org/10.17226/18318>.
- Office of Management and Budget. (2014). Statistical Policy Directive No. 1: Fundamental Responsibilities of Federal Statistical Agencies and Recognized Statistical Units. *Federal Register*, 79(231), 71610–71616. <https://www.federalregister.gov/documents/2014/12/02/2014-28326/statistical-policy-directive-no-1-fundamental-responsibilities-of-federal-statistical-agencies-and>.
- Pierson, S. (2015). Official statistics principles compared. *Statistical Journal of the IAOS*, 31(1), 21–23. <https://doi.org/10.3233/SJI-150880>.

Rozkrut, D., Świerkot-Strużewska, O., & Van Halderen, G. (2021). Mapping the United Nations Fundamental Principles of Official Statistics against new and big data sources. *Statistical Journal of the IAOS*, 37(1), 161–169. <https://doi.org/10.3233/SJI-210789>.

Statistical Office of the European Communities. (2011). *European Statistics Code of Practice*. <https://doi.org/10.2785/18474>.

United Nations Statistics Division. (n.d.). *Fundamental Principles of National Official Statistics*. Retrieved April 29, 2024, from <https://unstats.un.org/unsd/dnss/gp/fundprinciples.aspx>.

Norbert Duczkowski

Szkoła Główna Handlowa w Warszawie, Szkoła Doktorska, Polska / SGH Warsaw School of Economics, Doctoral School, Poland
ORCID: <https://orcid.org/0000-0003-4543-2678>

Lubomir Słowik

Uniwersytet Jagielloński w Krakowie, Szkoła Doktorska, Polska / Jagiellonian University in Kraków, Doctoral School, Poland
ORCID: <https://orcid.org/0000-0002-9817-9576>