Wiadomości Statystyczne. The Polish Statistician, 2024, *69*(10), 75–88 https://doi.org/10.59139/ws.2024.10.5
Dyskusje. Recenzje. Informacje / Discussions. Reviews. Information

The role of Artificial Intelligence when generating official statistical data

Rola sztucznej inteligencji w generowaniu oficjalnych danych statystycznych

1. Introduction

In today's data-rich world, the field of statistics has undergone a profound transformation, adapting to the complexities and challenges posed by modern data environments. The rise of big data offers opportunities but also poses challenges for statisticians. Traditional statistical methods have proven insufficient in terms of managing the scale and complexity of the constantly increasing volumes of datasets. Modern statistics, however, has adapted to these changes through scalable algorithms, distributed computing frameworks and parallel processing techniques to analyse massive datasets efficiently (Yung et al., 2018). Moreover, statisticians emphasize the importance of data pre-processing, dimensionality reduction and feature engineering to yield meaningful insights from large and complex data sources (Chu & Poirier, 2015; United Nations Economic Commission for Europe [UNECE], 2021; Yung et al., 2018).

In the digital age, the production of official statistics is undergoing a transformative evolution thanks to the integration of Artificial Intelligence (AI) technologies. From data collection to analysis and reporting, AI has been playing a pivotal role in revolutionising how official statistics are generated.

The integration of AI in official statistics has transformed traditional approaches to data collection, processing and analysis. AI algorithms have enabled automated data collection from diverse sources such as surveys, administrative records and sensor networks, enhancing the efficiency of statistical data collection processes. Furthermore, machine learning techniques have been instrumental in automating data processing tasks, facilitating the production of valuable insights from vast and heterogeneous datasets (Chu & Poirier, 2015; UNECE, 2021; Yung et al., 2018).

Quality control and assurance guarantee the reliability and integrity of official statistics. AI-powered anomaly detection algorithms play a critical role in identifying outliers, errors and inconsistencies within datasets, enabling statisticians to rectify discrepancies and uphold data quality standards (Boukerche et al., 2020). Additionally, Alghushairy et al. (2020) highlight the importance of AI in detecting unusual patterns or discrepancies, allowing for proactive interventions to maintain the credibility of statistical outputs.

Predictive analytics has emerged as a cornerstone of modern statistical practice, enabling statisticians to forecast future trends and make informed decisions based on historical data. Survey optimisation and design represent another area where AI technologies have made significant contributions to generating official statistics. Machine learning algorithms optimise survey designs by identifying relevant variables, minimising respondent burden and improving response rates (Braaksma & Offermans, 2021). AI-driven approaches reduce the costs and improve the efficiency of surveys, ensuring the representativeness and reliability of statistical estimates.

In addition to enhancing efficiency and accuracy, AI also plays a crucial role in addressing ethical and privacy concerns in the production of official statistics. The legal and ethical implications of using AI and machine learning techniques in statistical production emphasise the importance of transparency, accountability and fairness. The application of differential privacy techniques protects individuals' privacy while enabling the analysis and dissemination of aggregated statistics.

In conclusion, the application of AI in official statistics represents a paradigm shift in collecting, processing and analysing data. By harnessing the power of AI, statistical agencies can improve the efficiency, accuracy and timeliness of statistical production processes, ultimately empowering evidence-based decision-making and policy formulation.

Processing data into actionable information is the core function of National Statistical Offices (NSOs) around the world, with their role being critical in shaping national and international policy decisions. Governed by the Fundamental Principles of Official Statistics (FPOS), set forth by the United Nations Statistical Commission, NSOs are tasked with capturing, analysing and disseminating official national statistics ensuring their reliability and international comparability (United Nations [UN], 2014). However, as technological advancements continue to redefine data collection and analysis methods, NSOs face the challenge of integrating new technologies while upholding the principles of accuracy, reliability and impartiality.

Traditionally, NSOs have relied on labour-intensive methods for generating official statistics. However, the exponential growth of data, coupled with the emergence of AI technologies, necessitates a shift towards more machine-intensive and automated strategies (Chu & Poirier, 2015; Julien et al., 2020; UNECE, 2021). Machine learning algorithms unlock the potential to streamline tasks such as data editing, imputation, categorisation and coding.

Nevertheless, while adopting AI technologies, NSOs are faced with a set of complex considerations. On the one hand, AI-driven automation of statistical tasks promises real-time insights and enhanced productivity which allows NSOs to keep pace with the ever-increasing demands for timely data. On the other hand, concerns regarding the accuracy, reliability and interpretability of AI-generated outputs must be carefully addressed. The trade-off between real-time information and the quality of statistical outputs is a critical issue that NSOs must navigate as they integrate AI into their operations (Chu & Poirier, 2015; Julien et al., 2020; UNECE, 2021).

Furthermore, the data infrastructure, institutional capacity and regulatory frameworks which shape the technological development and strategies of NSOs vary across countries. These factors can influence the adoption and implementation of AI technologies. Understanding the differences and similarities between NSOs worldwide provides significant insights essential for developing data-driven strategies in the context of the digital transformation in the public sector.

In conclusion, NSOs play a pivotal role in processing data into actionable information of a national and international significance. When adopting machine-intensive and automation strategies based on AI technologies, NSOs must balance the pursuit for real-time information with the principles of accuracy, reliability and impartiality to harness the transformative potential of AI while upholding their mandate to deliver high-quality official statistics.

2. AI integration and the Fundamental Principles of Official Statistics

The integration of AI into the realm of official statistics entails both opportunities and challenges. As NSOs navigate this technological transformation, they must uphold the FPOS while harnessing the potential of AI.

One of the FPOS is impartiality, ensuring that statistics are produced without political or ideological influence. AI contributes to impartiality by automating processes, minimising human intervention and reducing the potential for bias. For example, AI algorithms can standardise data collection methods, eliminating subjective decision-making and ensuring consistency across data sources (United Nations Statistics Division [UNSD], n.d.). By removing human biases, AI helps maintain the objectivity of statistical analyses, reinforcing the credibility of official statistics.

Another FPOS is reliability which states that data should be accurate, consistent and free from errors. AI enhances reliability by automating quality control processes and detecting anomalies within datasets. Machine learning algorithms can identify inconsistencies or outliers that might indicate data errors, enabling statisticians to promptly address any issues (UNSD, n.d.). Through continuous learning and adaptation, AI systems improve the accuracy and reliability of statistical outputs, bolstering confidence in their integrity.

Official statistics must remain relevant and timely to meet the evolving needs of policymakers, businesses and the public. AI enables statisticians to analyse vast amounts of data in a short time, providing nearly real-time insights. Predictive analytics powered by AI can forecast future trends, enabling proactive decision-making and policy formulation (UNSD, n.d.). Moreover, AI facilitates the integration of diverse data sources, enriching official statistics with contextual information and enhancing their relevance to stakeholders.

Transparency and accountability are FPOS that underpin the credibility of official statistics. While AI introduces complexity into statistical processes, it also enables greater transparency through documentation and traceability. AI algorithms, data sources and decision-making processes can be documented, ensuring transparency and reproducibility (UNSD, n.d.). Additionally, accountability mechanisms can be implemented to monitor AI systems and address any biases or errors that may arise during statistical production.

Ethical considerations are paramount in the integration of AI into official statistics, particularly concerning privacy protection and data confidentiality. AI techniques such as differential privacy can be employed to anonymise data while preserving its utility for statistical analysis. Moreover, statisticians must adhere to ethical guidelines and regulations to ensure the responsible use of AI in statistical production, safeguarding individuals' rights and maintaining public trust.

As AI becomes increasingly integrated into official statistics, it is essential to uphold the FPOS of impartiality, reliability, relevance, transparency and accountability. By leveraging AI technologies responsibly, statisticians can enhance the quality, integrity and timeliness of official statistics while addressing emerging challenges in data collection, analysis and dissemination. Through collaboration, innovation and adherence to ethical standards, AI and official statistics can work synergistically to provide valuable insights and support evidence-based decision-making in a data-driven world.

3. AI in official statistics

The role of AI in official statistics is multifaceted and continuously evolving. Below are some key aspects of AI's contribution to official statistics.

3.1. Data collection and processing

AI technologies such as natural language processing (NLP), computer vision and machine learning algorithms are used to automate processes of data collection from various sources, including administrative records, surveys, sensor networks and social media. These technologies can prove helpful in extracting relevant information from unstructured data sources and in streamlining data processing tasks (Barrachina et al., 2009; Koch, 2016; Silver et al., 2018). Official statistics uses AI methods for classification as well as for recognition, estimation and/or the imputation of relevant characteristic values of statistical units (UNECE, 2021).

One of the key advantages of AI in data collection is its ability to automate the extraction of information from unstructured data sources, such as text documents, images and videos. NLP techniques enable AI algorithms to parse and analyse textual data, extract key insights and categorise information based on predefined criteria. For example, AI-powered chatbots can engage with respondents in natural language conversations to collect survey responses or gather information from social media posts and online forums. Similarly, computer vision algorithms can analyse images and videos to identify relevant data points, such as demographic information or spatial features, enabling statistical agencies to incorporate multimedia data into their analyses. Moreover, machine learning algorithms can be trained on historical data to classify observations into predefined categories, such as industry sectors, occupation codes or demographic groups. This automated classification process not only accelerates data processing but also reduces the potential for human error and bias. Additionally, AI algorithms can perform recognition tasks, such as optical character recognition (OCR) for digitising paper-based documents or speech recognition for transcribing audio recordings, further streamlining data collection and processing workflows. AI methods also enable statistical agencies to leverage alternative data sources and innovative data collection techniques, such as web scraping, sentiment analysis and social media mining. By harnessing the vast amounts of digital data generated by online platforms and digital devices, AI-powered data collection methods complement traditional survey-based approaches, providing additional insights into societal trends, consumer behaviour and economic indicators. For example, sentiment analysis algorithms can analyse social media posts to gauge public sentiment towards specific topics or products, providing real-time indicators of public opinion and behaviour.

3.2. Quality control and error detection

As datasets continue to grow in size and complexity, the task of identifying and correcting data errors becomes increasingly challenging for NSOs and other statistical agencies. Traditional manual approaches to quality control are often time-consuming and labour-intensive, involving a thorough inspection of every data point. In this context, AI techniques offer valuable tools for automating quality control processes and detecting anomalies, such as data errors, outliers and inconsistencies in large and diverse datasets. AI algorithms can identify patterns and deviations through such techniques as clustering, classification and regression. As a result, statisticians are able to prioritise investigation and corrective action, ultimately enhancing the quality and reliability of statistical outputs (Alghushairy et al., 2020; Boukerche et al., 2020).

Furthermore, AI techniques enable predictive modelling approaches that anticipate and pre-emptively address data quality issues before they propagate through the analytical process. By training models on historical data and learning from past errors, AI algorithms can predict the likelihood of future data anomalies and proactively implement preventive measures. For example, predictive models can detect abnormal data patterns, missing values or inconsistencies and automatically trigger alerts or corrective actions, such as data imputation or outlier removal.

3.3. Forecasting and predictive analytics

Forecasting and predictive analytics provide insights into future trends and potential outcomes, allowing for effective policy decision-making and strategic planning. In the realm of official statistics, machine learning models used for forecasting have emerged as powerful tools applied for predicting accurate and reliable values of socio-economic indicators. By analysing historical data patterns, relationships and trends, machine learning models, as opposed to traditional statistical methods, can generate forecasts for key indicators such as GDP growth, employment rates, inflation, population trends and more (Aburto & Weber, 2007; Wickramasuriya et al., 2019). Machine learning offers a diverse array of modelling techniques that can be tailored to different forecasting tasks and data types. From traditional time series analysis methods like ARIMA (Auto Regressive Integrated Moving Average) to more advanced algorithms such as neural networks, ensemble methods and gradient boosting, machine learning provides statisticians with a versatile toolkit for building forecasting models that can manage complex data structures and non-linear relationships (Aburto & Weber, 2007).

One of the key advantages of machine learning-based forecasting is its ability to provide real-time or near-real-time predictions. By continuously updating models with the latest data, machine learning algorithms can generate forecasts that reflect the most current trends and developments, enabling policymakers to make timely and informed decisions. Real-time forecasting is particularly valuable in fast-paced environments where rapid responses to changing conditions are essential (Wickramasuriya et al., 2019). Moreover, machine learning-based forecasting enables planning and risk management analysis by simulating various future scenarios and assessing their potential impact on socio-economic indicators. These include forecasting economic growth, predicting unemployment rates, or projecting demographic trends (Wickramasuriya et al., 2019). Machine learning models thus help policymakers evaluate different policy options, anticipate potential risks and develop contingency plans to mitigate adverse outcomes (Aburto & Weber, 2007).

3.4. Survey optimisation and design

The integration of AI algorithms into survey design represents a transformative leap in the field of official statistics, offering unparalleled opportunities for the optimisation of data collection processes. With the ever-growing complexity of data and the need for accurate timely insights, AI-driven approaches are revolutionising the way surveys are designed, conducted and analysed. This integration not only enhances the efficiency and effectiveness of survey methodologies but also improves the quality and reliability of statistical outputs.

One of the key advantages of AI-driven survey optimisation is its ability to streamline the selection of variables and determination of sample sizes. Traditional survey design processes often rely on manual selection methods, which can be time-consuming and prone to bias. AI algorithms, on the other hand, can analyse large datasets to identify the relevant variables and predictors that influence survey outcomes. By leveraging techniques such as feature selection and predictive modelling, AI algorithms can identify the most informative variables to include in the survey instruments, ensuring that the surveys capture the essential information while minimising respondent burden. Additionally, AI algorithms can determine optimal sample sizes and sampling techniques based on the desired levels of precision and statistical power, allowing statisticians to design surveys that achieve reliable and representative results (Cateni et al., 2013).

Another significant benefit of AI-driven survey design is the ability to create personalised survey instruments adjusted to individual respondent characteristics. Traditional surveys often follow a one-size-fits-all approach, which may not fully capture the diverse perspectives and experiences of respondents. AI-driven survey tools, however, analyse demographic data, past survey responses and other relevant information to tailor the survey questions, formats and delivery methods to individual respondents. This personalised approach increases engagement and response rates.

Furthermore, AI algorithms enable an adaptive survey design, allowing survey instruments to evolve in real-time based on the incoming responses and feedback. Typically, traditional surveys have a static structure, which may not account for changes in respondent behaviour or any emerging trends. AI-driven survey tools, on the other hand, can dynamically adjust the survey content, skip patterns and question sequencing to optimise the respondent experience and data quality. Adaptive survey design minimises respondent fatigue, reduces survey dropout rates and maximises the efficiency of data collection efforts.

In addition to enhancing survey design and data collection processes, AI-driven approaches also strengthen quality control measures and response validation mechanisms. NLP techniques enable AI algorithms to analyse open-ended responses, identify sentiment and detect semantic inconsistencies. By automating response validation processes, AI-driven survey tools streamline data cleaning efforts and improve the overall quality of the survey results.

Moreover, AI-driven predictive analytics play a crucial role in non-response adjustment. Machine learning algorithms analyse demographic, behavioural and contextual data to predict response propensity and identify factors associated with non-response bias. Predictive analytics enable statisticians to adjust the survey weights and account for non-response bias, so that the survey results accurately reflect the characteristics of the target population.

3.5. Privacy preservation and data confidentiality

The preservation of privacy and data confidentiality are one of the main concerns in official statistics. The need to protect individual-level data is balanced with the imperative to derive meaningful insights from aggregate data. This aim is achieved through differential privacy and other privacy-preserving techniques. By leveraging AI-based methods for privacy protection, statisticians can effectively address these concerns while still harnessing the power of data.

Differential privacy has emerged as the leading approach for privacy preservation in official statistics. This framework provides a rigorous mathematical definition of privacy, ensuring that the inclusion or exclusion of any single individual's data does not significantly impact the outcome of statistical analyses. By adding carefully calibrated noise to query responses or by perturbing data during collection, the differential privacy technique allows statisticians to share and analyse sensitive information without compromising individual privacy. AI-based methods further enhance privacy preservation efforts by offering sophisticated techniques for anonymisation, encryption and data obfuscation. For example, advanced encryption algorithms can secure data both at rest and in transit, preventing unauthorised access and ensuring data confidentiality throughout its lifecycle. Similarly, anonymisation techniques such as k-anonymity and l-diversity help protect against re-identification attacks by aggregating or generalising individual attributes while not distorting the analysis at the aggregate level.

Furthermore, AI-powered data masking techniques enable statisticians to generate synthetic data that closely mimic the statistical properties of the original dataset while preserving privacy. By training generative models on the original data distribution, AI algorithms can generate synthetic data samples that maintain the key statistical characteristics without revealing sensitive information about the individual respondents. This allows for a robust analysis and modelling while mitigating privacy-related risks associated with sharing or disclosing sensitive data.

In addition to these technical approaches, privacy preservation in official statistics also relies on robust governance frameworks and regulatory safeguards. NSOs and other statistical agencies adhere to strict data protection regulations and ethical guidelines to ensure compliance with privacy laws and standards. This includes implementing rigorous data access controls, conducting privacy impact assessments and establishing protocols for data sharing and dissemination that prioritise privacy and confidentiality (El Mestari et al., 2024).

3.6. Automation of reporting and dissemination

Automation of reporting and dissemination processes is a transformative practice in official statistics which has revolutionised the collection, processing, analysis and dissemination of data. Advanced technologies, including AI, ensure that statistical agencies can streamline workflows, enhance the efficiency and accuracy of statistical information, and guarantee its timely delivery and easy access.

One of the primary benefits of automation in reporting and dissemination is that the production and dissemination of statistical outputs is fast. Traditionally, the process of collecting, processing and analysing data for reporting purposes tends to be time-consuming and labour-intensive. However, with automation technologies, including AI-powered data processing and analysis tools, statistical agencies can significantly reduce the time and effort required to generate reports and disseminate information. Automated algorithms can ingest raw data, perform complex analyses and prepare reports and visualisations in a fraction of the time it would take using manual methods. Moreover, automation allows for greater standardisation and consistency in reporting practices. By implementing predefined templates, automated reporting systems ensure that statistical outputs adhere to the established formatting guidelines and quality standards, reducing the risk of errors and inconsistencies in the reporting process, adding to the credibility and reliability of official statistics.

Furthermore, automation facilitates the customisation and personalisation of statistical information to meet the diverse needs of users. To enhance user engagement and satisfaction, AI-driven algorithms analyse user preferences, behaviour and feedback and tailor reporting and dissemination strategies accordingly. For example, statistical agencies can use machine learning algorithms to segment users based on their interests, expertise and information needs, delivering customised reports, dashboards and interactive data visualisations that are relevant and actionable. Additionally, automation enables statistical agencies to expand the accessibility and reach of official statistics through digital platforms and online portals. By using AI-driven technologies, such as natural language processing and chatbots, agencies can automate the generation of data summaries, FAQs and explanatory notes, making statistical information more understandable and accessible to non-expert users. Moreover, automated dissemination systems can deliver statistical updates and alerts in real-time via email, social media and mobile applications.

3.7. Ethical and legal considerations

As AI is becoming more pervasive in official statistics, it is essential to address ethical and legal considerations surrounding data privacy, algorithmic bias, transparency and accountability. Statisticians must ensure that AI-driven approaches comply with relevant regulations and guidelines while upholding principles of fairness, transparency and user trust.

Although AI holds a significant promise for enhancing the efficiency, accuracy and accessibility of official statistics, it also poses challenges that need to be carefully addressed to realise its full potential in this domain. The use of AI in processing large datasets, especially those containing personal information, raises significant ethical concerns. Issues such as privacy, consent and the potential for surveillance need to be addressed. Without stringent ethical guidelines, the use of AI in official statistics could lead to privacy violations or the misuse of personal data. This could result in a loss of public trust and legal challenges.

For this purpose, policymakers and legal experts need to collaborate with statisticians and data scientists to develop regulations and ethical guidelines governing the use of AI in official statistics. This includes addressing issues like data privacy, consent, algorithmic transparency and accountability.

3.8. Geospatial analysis and spatial modelling

Geospatial analysis and spatial modelling, powered by AI techniques such as geographic information systems (GIS), spatial analysis and remote sensing are indispensable tools in the realm of official statistics. These methods play a vital role in analysing spatial patterns, conducting regional assessments and modelling spatial relationships, thereby providing crucial insights into geographical disparities, urbanisation trends, environmental impacts and other spatially explicit phenomena. By integrating AI-driven geospatial techniques into statistical analysis, official statistical agencies can support spatial planning and policy-making with comprehensive, data-driven assessments.

One of the primary applications of AI-driven geospatial analysis in official statistics is the identification and analysis of spatial patterns and trends. By leveraging GIS and spatial analysis techniques, statistical agencies can analyse spatially distributed data, such as population demographics, land use and socio-economic indicators to identify clusters, hotspots and trends across geographic regions. This enables policymakers to gain a deeper understanding of spatial disparities and inequalities, guiding targeted interventions and resources allocation strategies to address socio-economic challenges and promote inclusive development.

Moreover, AI-driven geospatial analysis facilitates the integration of diverse datasets from multiple sources, including satellite imagery, aerial photography and ground-based sensors. By combining spatial data with statistical information such as census data and survey results, statistical agencies can conduct comprehensive spatial analyses to assess the impact of urbanisation, environmental changes and natural disasters on local communities and ecosystems. This interdisciplinary approach enables policymakers to develop evidence-based policies and interventions that address complex spatial challenges such as climate change adaptation, disaster risk management and sustainable urban development.

Furthermore, AI-powered spatial modelling techniques enable statistical agencies to simulate and forecast spatial relationships and phenomena, providing valuable insights into future scenarios and potential outcomes. By using machine learning algorithms and predictive modelling approaches, statistical agencies can analyse historical spatial data to forecast future trends such as population growth, land use changes and infrastructure development. This enables policymakers to anticipate spatial dynamics and plan for future challenges such as urban sprawl, transportation congestion and environmental degradation, ensuring sustainable and resilient spatial development.

In addition to informing policy-making at the national and regional levels, AI-driven geospatial analysis also supports local decision-making and community development initiatives. By providing spatially disaggregated data and interactive mapping tools, statistical agencies empower local authorities, non-governmental organisations and community stakeholders to identify local priorities, plan targeted interventions and monitor progress towards sustainable development goals (SDGs). This bottom-up approach to spatial planning and decision-making ensures that policies and interventions are context-specific, participatory and responsive to the needs of local communities (Hu et al., 2019; Janowicz et al., 2020; Openshaw, 1992).

4. Conclusions

The integration of AI into official statistics represents a transformative shift in how data is collected, processed, analysed and disseminated. This paradigm shift is driven by the recognition of AI's ability to tackle the complexities and challenges posed by modern data environments, including the exponential growth of data volumes and the need for real-time insights.

Through AI-driven approaches, statistical agencies can improve the efficiency, accuracy and timeliness of statistical production processes. AI technologies enable automated data collection from diverse sources, automated data processing tasks and enhanced quality control mechanisms. Additionally, AI facilitates predictive analytics, survey optimisation, and design, thereby empowering evidence-based decision-making and policy formulation.

However, the adoption of AI in official statistics also presents complex considerations, including ethical, legal and privacy concerns. As AI is becoming more pervasive, it is crucial for statistical agencies to uphold principles of impartiality, reliability, relevance, transparency and accountability. By leveraging AI responsibly and ethically, statistical agencies can maintain public trust and confidence in official statistics while harnessing the full potential of AI to address emerging challenges and opportunities in the digital age.

In addition to the FPOS and ethical considerations discussed, the integration of AI in official statistics offers several notable advantages. Firstly, AI-driven approaches enable statisticians to process and analyse vast amounts of data more efficiently and accurately than through the application of traditional methods. This capability is particularly beneficial in the era of big data where the volume, velocity and variety of data sources continue to expand exponentially. By leveraging machine learning algorithms and other AI techniques, statisticians may uncover valuable insights, identify patterns and make predictions that were previously inaccessible or impractical.

Moreover, through AI technologies statistical agencies are able to adapt quickly to the evolving data sources and analytical methodologies. With the proliferation of new data types such as social media data, sensor data and satellite imagery, traditional statistical methods may no longer suffice to capture the complexity and nuances of modern datasets. AI-driven approaches offer flexible and adaptable solutions that can accommodate diverse data sources and analytical requirements.

Furthermore, AI fosters innovation and creativity in official statistics by empowering statisticians to explore new methods, techniques and applications. The interdisciplinary nature of AI, which draws from fields such as computer science, mathematics and cognitive psychology, encourages collaboration and cross-pollination of ideas. This interdisciplinary approach spurs innovation in statistical modelling, data visualisation and decision support systems, leading to the development of novel tools and methodologies.

Additionally, AI-driven approaches improve accessibility and usability of statistical information by making data more understandable, interactive and user-friendly. Through data visualisation techniques, NLP interfaces and interactive dashboards, statistical agencies can present complex information in intuitive and engaging formats that cater to diverse audiences. This democratisation of data empowers policymakers, researchers, businesses and the general public to access, understand and utilise official statistics effectively.

Overall, the integration of AI in official statistics represents a transformative shift that holds an immense potential to advance the field and address the evolving needs of stakeholders in an increasingly data-driven world. By embracing AI responsibly, statisticians can leverage its capabilities to produce high-quality, timely and actionable insights that drive evidence-based decision-making, promote transparency and accountability, and ultimately contribute to the development of the society as a whole.

References

Aburto, L., & Weber, R. (2007). Improved supply chain management based on hybrid demand forecasts. *Applied Soft Computing*, *7*(1), 136–144. https://doi.org/10.1016/j.asoc.2005.06.001.

Alghushairy, O., Alsini, R., Ma, X., & Soule, T. (2020). A Genetic-Based Incremental Local Outlier Factor Algorithm for Efficient Data Stream Processing. In *ICCDA 2020. Proceedings of the 4th International Conference on Computational Data and Analytics* (pp. 38–45). Association for Computing Machinery. https://doi.org/10.1145/3388142.3388160.

Barrachina, S., Bender, O., Casacuberta, F., Civera, J., Cubel, E., Khadivi, S., Lagarda, A., Ney, H., Tomás, J., Vidal, E., & Vilar, J. M. (2009). Statistical Approaches to Computer-Assisted Translation. *Computational Linguistics*, *35*(1), 3–28. https://doi.org/10.1162/coli.2008.07-055-r2-06-29.

Boukerche, A., Zheng, L., & Alfandi, O. (2020). Outlier detection: Methods, models, and classification. *ACM Computing Surveys*, *53*(3), 1–37. https://doi.org/10.1145/3381028.

Braaksma, B., & Offermans, M. (2021). Statistics Netherlands and AI. In *Proceedings of Statistics Canada Symposium: Adopting Data Science in Official Statistics to Meet Society's Emerging Needs*. https://www150.statcan.gc.ca/n1/en/pub/11-522-x/2021001/article/00011-eng.pdf?st=ncDUTkRG.

Cateni, S., Vannucci, M., Vannocci, M., & Colla, V. (2013). Variable Selection and Feature Extraction Through Artificial Intelligence Techniques. In L. Freitas, & A. P. B. Rodrigues De Freitas (Eds.), *Multivariate Analysis in Management, Engineering and the Sciences* (pp. 103–118). https://doi.org/10.5772/53862.

Chu, K., & Poirier, C. (2015). *Machine Learning Documentation Initiative*. https://unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.50/2015/Topic3_Canada_paper.pdf.

El Mestari, S. Z., Lenzini, G., & Demirci, H. (2024). Preserving data privacy in machine learning systems. *Computers & Security*, *137*, 1–22. https://doi.org/10.1016/j.cose.2023.103605.

Hu, Y., Li, W., Wright, D., Orhun, A., Wilson, D., Maher, O., & Raad, M. (2019). Artificial Intelligence Approaches. In J. P. Wilson (Ed.), *The Geographic Information Science and Technology Body of Knowledge* (pp. 1–12). University Consortium for Geographic Information Science. https://doi.org/10.22224/gistbok/2019.3.4.

Janowicz, K., Gao, S., McKenzie, G., Hu, Y., & Bhaduri, B. (2020). GeoAI: spatially explicit artificial intelligence techniques for geographic knowledge discovery and beyond. *International Journal of Geographical Information Science*, *34*(4), 625–636. https://doi.org/10.1080/13658816.2019.1684500.

Julien, C., Choi, I., Deeben, E., Yung, W., Wesley, Y., & Measure, A. (2020). *HLG-MOS Machine Learning Project*. https://statswiki.unece.org/display/ML/HLG-MOS+Machine+Learning+Project.

Koch, C. (2016). How the Computer Beat the Go Player. *Scientific American Mind*, *27*(4), 20–23. https://doi.org/10.1038/scientificamericanmind0716-20.

Openshaw, S. (1992). Some Suggestions Concerning the Development of Artificial Intelligence Tools for Spatial Modelling and Analysis in GIS. *The Annals of Regional Science*, *26*(1), 35–51. https://doi.org/10.1007/BF01581479.

Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D., Graepel, T., Lillicrap, T., Simonyan, K., & Hassabis, D. (2018). A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. *Science*, *362*(6419), 1140–1144. https://doi.org/10.1126/science.aar6404.

United Nations. (2014). Resolution adopted by the General Assembly on 29 January 2014. Fundamental Principles of Official Statistics (A/RES/68/261). https://unstats.un.org/unsd/dnss/gp/fp-new-e.pdf.

United Nations Economic Commission for Europe. (2021). *Machine Learning for Official Statistics*. United Nations. https://unece.org/sites/default/files/2022-09/ECECESSTAT20216.pdf.

United Nations Statistics Division. (n.d.). *Principles Governing International Statistical Activities*. https://unstats.un.org/unsd/ccsa/principles_stat_activities/.

Wickramasuriya, S. L., Athanasopoulos, G., & Hyndman, R. J. (2019). Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization. *Journal of American Statistical Association*, *114*(526), 804–819. https://doi.org/10.1080/01621459.2018.1448825.

Yung, W., Karkimaa, J., Scannapieco, M., Barcarolli, G., Zardetto, D., Ruiz Sanchez, J. A., Braaksma, B., Buelens, B., & Burger, J. (2018). *The use of machine learning in official statistics*. United Nations Economic Commission for Europe. https://unece.org/sites/default/files/2024-07/HLGMOS%20The%20use%20of%20machine%20learning%20in%20official%20statistics.pdf.

Arzu Taghiyeva
State Statistical Committee of the Republic of Azerbaijan; PhD student on Econometrics at Scientific-Research Institute of Economic Studies of Azerbaijan State University of Economics
ORCID: https://orcid.org/0000-0002-8031-4768. E-mail: arzuvalehqizi@gmail.com