Journal of Physical Education and Sport ® (JPES), Vol. 25 (issue 3), Art 68, pp. 638 - 645, March 2025
online ISSN: 2247 - 806X; p-ISSN: 2247 – 8051; ISSN - L = 2247 - 8051 © JPES

Original Article

Sustainable data-driven framework and policy recommendations for enhancing sports promotion using generative and explainable Artificial Intelligence

TSUYOSHI YOSHIOKA[1], YUSUKE MORIKURA[2], TOSHIRO IZUMI[3], TADASHI WADA[4]
[1, 2, 3]Faculty of Humanities and Social Sciences, Teikyo Heisei University, Tokyo, JAPAN
[4]School of Science and Engineering, Kokushikan University, Tokyo, JAPAN

Published online: March 31, 2025
Accepted for publication: March 15, 2025
DOI:10.7752/jpes.2025.03068

Abstract:
Problem Statement: Sports are fundamentally crucial for improving physical health and strengthening community bonds. However, the current policy framework faces several challenges, including considerable imbalances in sports promotion across regions and age groups, insufficient data, and the absence of scientific evidence to guide policy design. These issues hinder the realization of fair and sustainable sports policies. *Approach:* This study proposes a data-driven framework that integrates generative and explainable artificial intelligence. This framework leverages generative artificial intelligence models wrapped in the synthetic data vault library to address data scarcity and bias. By analyzing statistical data related to sports promotion, the proposed framework would facilitate comprehensive policy recommendations that account for regional characteristics and intergenerational differences. *Purpose:* This study aims to establish a scientific foundation to support two primary policy goals: enhancing competitive performance and promoting health, thereby improving the fairness and effectiveness of sports policies via this framework. *Results:* This study evaluated the potential of generative artificial-based data supplementation in designing exercise programs tailored to regional characteristics and generational diversity through hypothetical applications. The use of explainable artificial intelligence enhances the transparency of the analytical results, increasing the credibility of policy recommendations. This framework offers a starting point for the theoretical examination of the applicability of data-driven approaches as a foundation for policy design. *Conclusion:* The proposed framework addresses challenges in sports policy, with potential applicability in other fields, such as healthcare and education. In addition, by addressing ethical issues, including bias and privacy protection, enhancing the reliability of data-driven policy designs is essential. Future research is expected to validate the effectiveness of the framework through empirical studies, thereby contributing to the realization of sustainable policies based on scientific evidence.
Keywords: Performance enhancement, health promotion, policy development, artificial intelligence applications, synthetic data, explainable AI

Introduction

Sports promotion policies play an important role in enhancing competitive performance and public health, serving as an important foundation for community revitalization and the design of sustainable policies (Eime et al., 2013; Heath & Parra, 2012). However, current sports policies face substantial challenges, including imbalanced sports participation across regions and age groups, which undermine the fairness of sports policies. For example, rural areas lack adequate sports facilities, whereas in urban areas, busy lifestyles hinder sports participation (Kokolakakis et al., 2020; Ruseski et al., 2011). Furthermore, the importance of the strategic placement of sports facilities in influencing competitive performance has been highlighted (Izumi et al., 2023). In addition to these structural challenges, demographic changes such as aging populations and widening urban-rural disparities have further intensified the issue, making it increasingly challenging to ensure equal access to sports opportunities (Zheng & An, 2015). Despite these pressing concerns, traditional sports policies have often relied on subjective judgments and administrative precedents rather than empirical evidence, limiting their ability to address region-specific characteristics and intergenerational differences effectively (Bar-Eli et al., 2024). This lack of data-driven approaches has led to policies that, in some cases, prioritize specific age groups while inadvertently excluding younger or older populations, thereby undermining both fairness and long-term sustainability (Faß & Schlesinger, 2019; Hyde et al., 2020). A key barrier to implementing more equitable sports policies is the shortage of scientifically valid and comprehensive data, which prevents policymakers from designing targeted and effective interventions that address diverse needs (McNoe et al., 2022). Consequently, an innovative data-driven approach based on scientific evidence is urgently needed to modernize sports policies, enhance fairness, and improve their overall effectiveness (Cerin et al., 2018; Osipov et al., 2016; Yue, 2022).

Corresponding Author: TSUYOSHI YOSHIOKA, E-mail: yoshiokatsuyoshi@gmail.com

Artificial intelligence (AI) has recently emerged as a promising solution to address these challenges, enabling data-driven decision-making in the field of sports, while its applicability to sports policy design has been suggested (Zhou et al., 2024). For example, Liu et al. (2024) argue that leveraging generative AI effectively addresses data imbalance and gaps, thereby improving the quality of data necessary for policy design. Recently, the rapid development of generative AI has emerged as an innovative solution to overcome data scarcity and bias (Paproki et al., 2024; Peppes et al., 2023). Generative AI develops new data based on existing datasets, representing a powerful tool for data supplementation and expansion (Lan et al., 2023). Notably, generative AI has the potential to produce synthetic data that comprehensively reflect the diverse range of athletic abilities of sports participants, from novices to elite athletes, beyond mere regional or generational differences. This enables the formulation of diverse policy recommendations tailored to various target groups. Furthermore, generative AI data can be analyzed using AI, providing insights to inform policy recommendations. However, when the basis for AI-driven recommendations is unclear, decision-makers and stakeholders may find it challenging to understand the results (Janssen et al., 2020). In this context, explainable AI (XAI) serves as an effective technology for ensuring the validity and interpretability of the outcomes produced by AI models (Holzinger, 2019). By leveraging XAI, clarifying the rationale behind AI-generated analyses becomes possible, thereby reinforcing the legitimacy of policy recommendations and enhancing policy-design decision-making. Agbozo et al. (2024) noted the substantial potential of generative AI and XAI as a basis for decision-making in sports, including performance prediction and strategic planning. Nevertheless, concerns remain regarding the quality of synthetic data generated by AI and interpretability of analysis results, emphasizing the importance of ensuring transparency and scientific validity when employing AI techniques in policy-making processes.

This study proposed a data-driven framework that leverages generative AI to address challenges, such as data scarcity and bias, while employing XAI to enable highly transparent analyses. This framework aims to scientifically support policy recommendations tailored to regional characteristics, intergenerational differences, and varying levels of athletic ability. Specifically, this study employs generative AI methods wrapped in the synthetic data vault (SDV, n.d.) library to statistically supplement and generate data related to the distribution and utilization of sports facilities, regional demographic indicators (such as age, gender, and income), and residents' health parameters (such as body mass index (BMI) and exercise frequency). In addition, XAI techniques, notably SHapley Additive exPlanations (SHAP, n.d.) and Local Interpretable Model-agnostic Explanations (LIME, n.d.), are applied to ensure transparency of the generated analytical results. For example, SHAP analysis quantifies the contributions of factors such as facility proximity or exercise frequency on athletic performance and health outcomes, whereas LIME provides localized explanations of policy impacts within specific demographic contexts. Through these techniques, policymakers can gain clear insights into AI-driven recommendations, facilitating well-informed and transparent policy decisions. The proposed data-driven framework is expected to address challenges in sports policy and contribute to the formulation of sustainable and inclusive initiatives. Moreover, this study aimed to establish a novel framework applicable to other fields, such as healthcare, education, and local governance.

Material & methods

This section elaborates on the methodological approach of the data-driven framework proposed in the present study, focusing on its potential to address key challenges in sports policy design. The proposed framework enables evidence-based policy recommendations by collecting, supplementing, and analyzing various data relevant to sports policy design. By addressing data scarcity using generative AI and ensuring the transparency and interpretability of the analysis results through XAI, this framework offers a distinct advantage over conventional policy design methods. This study explored the framework's potential as a novel foundation for data-driven policy design.

Construction of data generation and analysis framework

The data-driven framework proposed in this study aims to strengthen the scientific foundation of sports policy design by integrating data synthesis using generative AI with a transparent analysis framework using XAI. Detailed steps involved in constructing this analytical framework using the data-driven framework are listed below, as illustrated in Figure 1.

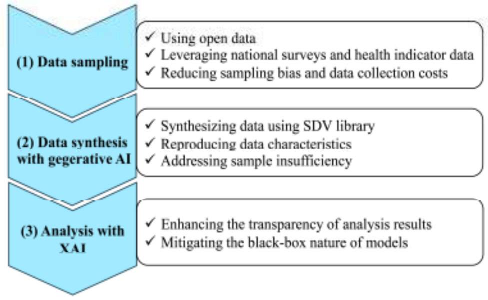

Fig. 1 Data-driven policy design framework integrating generative AI and XAI with a SDV library for data synthesis

(1) Data sampling

The proposed framework primarily uses open data and focuses on widely accessible statistical data, including national census data, sports facility distribution, regional health indicators (such as BMI, blood pressure, and frequency of physical activity), and regional demographic data. The use of open data is essential for minimizing sampling bias due to restriction to specific regions or conditions, while also enabling the design of broad policies applicable nationwide (Kassen, 2013). In addition, leveraging publicly available data helps reduce data collection costs and facilitates the development of transparent policy designs (Kassen, 2013).

(2) Data synthesis using generative AI

The SDV library was employed for data synthesis to address the issue of insufficient sample sizes for statistically significant results, and to correct biases in the collected sampling data (SDV, n.d.). Specifically, generative AI models wrapped in the SDV library were used to expand the dataset by generating synthetic data that maintains the statistical properties of the sampled data. This approach overcomes the difficulty of obtaining data that aligns with the intended objectives by generating statistically significant datasets from limited samples, thereby enabling more evidence-based analyses. SDV is a framework for managing and implementing generative AI models, including generative adversarial networks (GAN; Goodfellow et al., 2020), variational autoencoders (VAE; Kingma & Welling, 2013), and CopulaGAN (Xu et al., 2019), to facilitate efficient and flexible data synthesis. These models function as generative AI, enabling the creation of synthetic data that reflects complex statistical properties (Table 1). The application of generative AI enables the creation of data that represents the diverse socioeconomic backgrounds across regions and intergenerational variations in sports participation, thereby strengthening the scientific foundation for policy design. Furthermore, by generating synthetic data that accounts for a wide range of athletic abilities—from beginners to elite athletes—the framework supports the formulation of policy recommendations tailored to target groups.

Table 1. Characteristics of generative AI methods and applicable data types

Generative AI method	Characteristics	Examples of applicable data
GAN	Reproduces complex data structures, effective for high-dimensional and continuous data, but requires modifications for categorical data.	Spatial distributions of sports facilities, facility utilization rates, geospatial data for accessibility analysis
VAE	Learns latent representations for the smooth generation of continuous data; ideal for high-dimensional data generation.	Time-series data on athletic performance trends, behavioral patterns in sports participation, long-term exercise adherence rates
CopulaGAN	Combines the strengths of Copula theory and GAN to model mixed datasets containing both categorical and continuous variables.	Mixed datasets combining demographic data (age, gender, and income) with health and sports participation data

Table 2 outlines the three main steps of the data-synthesis process using the SDV library. Step 1: Data modeling: the statistical structure of the sampled data was analyzed to establish the foundation for designing the generated synthetic data. Step 2: Data generation: new data were synthesized using generative AI methods wrapped in the SDV library, addressing issues such as missing data and biases. Step 3: Data quality evaluation: the alignment of the generated data with the real data was assessed to confirm reliability. The use of SDV facilitates managing complex generative models in a unified manner to efficiently produce high-quality synthetic data that reflect the statistical properties of real data. This process enhances the reliability of the generated data, ensuring they meet the quality standards necessary for policy design and decision-making.

Table 2. Overview of the data synthesis process using the SDV library

Step	Description	Specific methods/indicators
Step 1: Data modeling	Analyze the statistical structure of the sampled data in detail to understand data distributions and correlations. Build the foundation for modeling the statistical properties of the entire dataset.	・ Analysis of distribution characteristics (e.g., mean, variance, skewness, and kurtosis) ・ Evaluation of correlations and dependencies between variables
Step 2: Data generation	Generate new data based on the statistical structure obtained during modeling. Select appropriate methods depending on the type and characteristics of the data to be generated, addressing issues such as missing data and biases.	・ GAN ・ VAE ・ CopulaGAN
Step 3: Data quality evaluation	Evaluate the reliability of the generated data. Compare the distributions of real and generated data to assess their alignment, ensuring data quality.	・ Jensen-Shannon Divergence: Measures differences between probability distributions ・ Kolmogorov-Smirnov Test: Tests distributional consistency ・ SDV library Evaluation Module: Calculates scores

(3) Analysis framework using XAI

Data supplemented through generative AI undergo appropriate preprocessing to transform them into an analyzable format. This preprocessing includes tasks such as the final imputation of missing values, handling outliers, and unifying data scales. Based on the preprocessed data, machine learning models were constructed, and XAI was employed to enable a highly transparent analysis. The application of XAI clarifies the rationale behind the AI-generated analysis results, thus supporting decision-making processes in policy design. In addition, XAI provides interpretable insights into the factors and influences underlying the AI model outputs, enabling policymakers and stakeholders to accurately understand the analyzed results and make scientifically informed decisions. This approach not only validates the appropriateness of AI-driven recommendations but also significantly enhances the transparency and reliability of policies.

The detailed process for integrating machine learning models and XAI into sports policy design is summarized in Table 3. Step 1: Model construction: appropriate machine learning models were built based on the goals of sports policy design, integrating both generated and real data for comprehensive analysis. This process improves reliability through model optimization and performance evaluation. Step 2: Application of XAI: explainability techniques such as SHAP and LIME were applied to the constructed models to ensure transparency of the analysis results. SHAP quantifies the contribution of each input factor to a model output (Lundberg & Lee, 2017; Lundberg et al., 2020), whereas LIME generates locally interpretable explanations for specific model predictions (Ribeiro et al., 2016). Step 3: Interpretation of analysis results: the interpreted analysis results obtained through XAI techniques are used as the basis for specific policy recommendations, including proposals for optimizing sports facility allocation using SHAP and enhancing health promotion initiatives with LIME.

Table 3. Process for the utilization of machine learning models and XAI in sports policy design

Step	Details	Examples of application
Step 1: Model construction	Build appropriate machine learning models based on the goals of sports policy design. Perform integrated analysis of generated and real data, as well as performance evaluations. Improve reliability through model optimization.	・ Regression Models: Measure the impact of sports facility allocation on health indicators (including BMI and blood pressure) or athletic performance. ・ Classification Models: Categorize the effects of specific policies on different target groups (age groups, competition levels). ・ Clustering Models: Cluster regional sports promotion statuses and resident characteristics to identify target groups for policy interventions.
Step 2: Application of XAI	Apply explainability techniques such as SHAP and LIME to the constructed models to reduce black-box characteristics and ensure transparency of analysis results.	・ SHAP: Quantify and visualize the impact of sports facility accessibility on athletic performance. ・ LIME: Provide localized explanations of the effects of exercise programs in specific regions.
Step 3: Interpretation of the analysis results	Interpret analysis results using XAI techniques to develop concrete policy recommendations.	・ Improvement of sports facility allocation: Based on SHAP results, propose facility installation in rural areas or improvements in transportation. ・ Optimization of health promotion initiatives: Adjust the frequency and intensity of exercise programs based on insights from LIME to maximize health benefits.

Framework supporting the pillars of sports promotion and policy recommendations

This study provides a foundation to scientifically support policy recommendations for the two primary pillars of sports promotion: enhancing competitive performance and promoting health. By combining data supplementation through generative AI with a transparent analysis using XAI, the framework enhances the scientific basis and reliability of policy design.

Identifying barriers to sports participation and designing measures to overcome them enhances competitive performance. In rural areas, limited facilities and poor accessibility often hinder participation. Using data supplemented through generative AI on facility distribution and usage actualizes efficient facility development plans. Conversely, in urban areas, busy lifestyles and high facility utilization rates are common barriers to participation. Flexible training programs and reservation systems are required to address these challenges.

Designing health promotion initiatives based on regional residents' exercise frequencies and health indicators (such as BMI and blood pressure) is crucial. Consequently, using generative AI to address data scarcity and bias in health-related data, exercise programs tailored to regional characteristics can be developed to provide a foundation for initiatives aimed at improving health. For instance, light exercise programs for older adults can leverage generative AI for data supplementation and XAI for transparent analysis to clarify the relationships between exercise frequency and health improvements, thereby optimizing the initiative design.

Results

This section theoretically examines the potential effects of the proposed framework, and demonstrates its utility through hypothetical applications.

Overview of the framework and role of generative AI

The data-driven framework proposed in this study suggests the possibility of overcoming challenges, such as data scarcity and bias, that have traditionally hindered sports policy design. Generative AI is used to create synthetic data that comprehensively reflect regional characteristics and generational differences. Based on these data, a transparent analysis was conducted using XAI, leading to the outcomes outlined in Table 4.

Table 4. Expected outcomes of the proposed framework

Category	Description	Impact
Effectiveness of data supplementation	The proposed framework supplements biases and deficiencies in existing data, providing a foundation for diverse policy designs that encompass a wide range of athletic abilities (from beginners to elite athletes).	Constructing comprehensive datasets necessary for initiatives to enhance competitive performance and promote health enables evidence-based decision-making.
Improvement in transparency	The analysis results produced by AI clarify the validity of the data generated by generative AI and rationale for policy recommendations.	Contributes to improved transparency in the decision-making process by making the validity of analysis processes and results explainable, thereby earning the trust of policymakers and stakeholders.
Applicability to policy design	Enables customizable policy recommendations tailored to regional characteristics and target demographics. The analysis is conducted using XAI based on data generated by generative AI that comprehensively reflects interregional and intergenerational differences.	Supports the realization of sustainable sports promotion initiatives through policy recommendations adapted to regional characteristics and target groups.

Hypothetical Applications

Based on the proposed data-driven framework, a theoretical examination conducted through hypothetical applications suggested that the framework could deliver the effects outlined in Table 5. These examples demonstrate how the proposed framework, through the integration of generative AI and XAI, can provide robust evidence-based policy recommendations tailored to diverse needs.

Table 5. Utilization of the proposed framework in hypothetical applications

Application area	Data supplementation with generative AI	Results from analysis using XAI	Expected policy recommendations
Enhancing competitive performance	Supplements imbalances in sports facility distribution and competitive performance data across regions. By generating statistically significant synthetic data that reflects facility quality and usage from limited observed data, the foundation for policy design is significantly strengthened.	Visualizes the impact of facility distribution and quality on competitive performance, providing scientific evidence. For instance, quantifies the importance of facility proximity and quality.	Addressing regional imbalances in sports facility distribution by improving facility quality and accessibility.
Promoting health	Supplements gaps and biases in data related to exercise frequency and health indicators (including BMI and blood pressure). Builds comprehensive datasets tailored to regional characteristics and resident demographics.	Quantitatively evaluates the effects of light exercise on health indicators. Clarifies factors such as program content and frequency that contribute to health improvements.	Optimization of exercise programs for older adults and introduction of online exercise programs for urban residents.

Discussion

In this study, we proposed a data-driven policy design framework that integrates generative AI and XAI and conducted a theoretical examination of its potential and challenges through hypothetical applications. This section re-evaluates the academic significance and practical applicability of the proposed framework while discussing the challenges identified based on insights gained from the hypothetical applications. Furthermore, it outlines the ethical and technical difficulties associated with the application of generative AI and guides future research directions and solutions to these challenges.

Academic and practical significance of the proposed framework

The data-driven policy-design framework proposed in this study, which integrates generative AI and XAI, has significant academic and practical value.

From an academic perspective, the proposed framework offers a novel approach to addressing data scarcity and bias. Generative AI can supplement missing data and generate synthetic data that retain statistical properties while reflecting target populations and regional characteristics. XAI strengthens the scientific basis of policy design by enhancing the transparency of the generated data and analysis results. The approach presented in this study goes beyond mere technical advancements and enables policy recommendations tailored to the diverse needs of target groups by enhancing competitive performance and promoting health. Notably, the integration of data generation and analysis that account for varying levels of athletic ability (including beginners, amateurs, intermediates, and elite athletes) introduces a new perspective to previous studies.

From a practical perspective, the integration of Generative AI and XAI enables the design of policies adapted to the characteristics of specific regions or groups, thereby providing concrete solutions to real-world challenges that are difficult to address using conventional methods. Notably, the data-driven framework offers a foundation for designing efficient measures to address issues such as shortages of sports facilities in rural areas or barriers to sports participation in urban areas due to busy lifestyles. Furthermore, it facilitates scientifically grounded policy recommendations tailored to different generational needs, such as light exercise programs for older adults and measures enhancing competitive performance for younger populations. Therefore, this framework provides a new pathway to address regional imbalances and insufficient support for specific groups.

Insights gained from hypothetical applications

The theoretical examination of the potential impact of the proposed data-driven framework, which integrates generative AI and XAI on policy design, yielded several key insights through hypothetical applications. First, generative AI proved effective in addressing challenges such as data scarcity and bias, providing scientific evidence to support policy objectives like enhancing competitive performance and promoting health. Specifically, this approach can overcome the challenge of preparing statistically significant datasets that align with specific objectives, enabling the construction of reliable datasets based on limited sample data. Moreover, this study examined the impact of accessibility and facility quality on athletic performance, based on the distribution and usage characteristics of sports facilities across regions, to enhance competitive performance. Within this framework, synthetic data generated using generative AI demonstrated its potential as a foundation for statistically significant analysis. In addition, combining XAI techniques improves the transparency of analysis results and provides a scientific basis for policy recommendations.

Regarding health promotion, the hypothetical analysis explored the impact of exercise program frequency and intensity on health indicators, demonstrating the importance of regionally tailored initiatives. Generative AI has the potential to address data scarcity and bias, enabling the exploration of policy directions that account for differences in health indicators across regions and generations using such synthetic data. Furthermore, utilizing XAI to visualize the analysis results clearly and intuitively allows policymakers to fully understand the rationale behind them and incorporate these insights into decision-making.

The integration of generative AI and XAI into the proposed framework can enhance the transparency and reliability of decision-making processes. However, these findings are based on a theoretical framework; future studies should focus on validating these findings using real-world data to confirm the effectiveness of the framework in actual policy design through empirical studies. Establishing mechanisms to evaluate the quality and biases of the data provided by generative AI is also essential.

Challenges and future research directions

This study faced some challenges on the path to practical implementation (Table 6). Future studies should address these challenges to enhance the effectiveness of generative AI and XAI applications by establishing data collection methods that minimize bias in training data and developing standards for evaluating the quality of the generated data. In addition, algorithms to detect and mitigate biases in generated data, as well as technologies that enhance privacy protection, should be introduced to address the ethical issues associated with data usage.

Empirical studies are crucial to validating the effectiveness of this framework and improving its practical applicability. This includes demonstrating how the outcomes of hypothetical applications can be applied to actual policy design and providing concrete evidence of its practical implementation. Furthermore, exploring its applicability to other fields, such as healthcare and education, is important to leverage the potential of generative AI and XAI, offering solutions to societal challenges.

The practical utility and reliability of the proposed framework can be improved by tackling these challenges, thereby enhancing its value as a foundation for sustainable policy design.

Table 6. Key challenges and proposed solutions for applying generative AI and XAI

Type of challenge	Specific issues	Proposed solutions
Technical challenges	The quality of generated data is affected by biases and deficiencies in training data.	Explore data collection methods to reduce bias and improve the accuracy of generative models.
Quality evaluation	Methods for assessing how closely generated data matches real data remain underdeveloped.	Establish evaluation standards using statistical metrics such as Jensen–Shannon Divergence and Kolmogorov-Smirnov Test.
Ethical challenges	The risk of generating data that includes biases or discrimination.	Introduce fairness evaluation algorithms and develop techniques for correcting biases.
Privacy	Insufficient privacy protection in data usage.	Implement differential privacy technologies and strengthen anonymization methods.
Practical application	Unclear how analysis results from XAI can be utilized in policy design.	Strengthen collaboration between policymakers and technologists and systematize the utilization process.

Conclusions

This study proposed a data-driven policy design framework that integrates generative AI and XAI, focusing on its applicability to sports policy. This framework addresses challenges such as data scarcity and bias, while further enhancing the transparency and reliability of policy recommendations. Generative AI is a valuable tool for constructing scientifically grounded datasets from limited observational data. In contrast, XAI plays a crucial role in increasing the transparency of decision-making processes by presenting the analysis results in an interpretable manner. Herein, the effectiveness of the framework was demonstrated theoretically in the field of sports policy. The framework demonstrated in this study provides a scientifically grounded approach to address challenges in sports policy, particularly in enhancing competitive performance and promoting health. This method effectively integrates generative AI and XAI to generate data that is reflective of regional and generational diversity, and enhances the transparency of decision-making processes. By addressing issues such as data scarcity and bias, this framework supports the creation of more equitable policies. Future studies should focus on validating these theoretical findings using real-world data and exploring the application of this framework in other fields such as healthcare and education, where similar data-driven, transparent approaches could address critical societal challenges.

The basic structure of this framework extends beyond sports policy; hence, future studies should investigate its applicability in fields such as healthcare, education, and local governance. These fields require consideration of their unique issues and requirements to optimize the framework for their specific contexts. Moreover, empirical studies in the field of sports policy are crucial to validate the effectiveness of the framework and establish a reliable data-driven approach. Furthermore, exploring its applicability in other fields could further enhance the framework, providing a basis for supporting sustainable policy designs. The findings of this study are expected to contribute to evidence-based decision-making and provide valuable support to creating scientifically grounded policies.

Acknowledgments

During the manuscript preparation, OpenAI ChatGPT-4 was used to proofread and refine the language of the manuscript. Finally, we have entrusted the English language editing service to Editage (www.editage.jp) and express our gratitude for their assistance.

Conflicts of interest

The authors report that there are no competing interests.

References:

Agbozo, E., Pandya, K., Jovanovic, P., & Suvorova, E. A. (2024). A comprehensive overview of artificial intelligence applications in basketball. *Journal of Physical Education and Sport, 24*(1), 44–52. https://doi.org/10.7752/jpes.2024.01006

Bar-Eli, M., Lidor, R., Lath, F., & Schorer, J. (2024). The feudal glove of talent-selection decisions in sport – Strengthening the link between subjective and objective assessments. *Asian Journal of Sport and Exercise Psychology, 4*(1), 1–6. https://doi.org/10.1016/j.ajsep.2023.09.003

Cerin, C., Vuillemot, R., Stolper, C. D., Stasko, J., Wood, J., & Carpendale, M. (2018). State of the art of sports data visualization. *Computer Graphics Forum, 37*(6), 663–686. https://doi.org/10.1111/cgf.13447

Eime, R., Young, J. A., Harvey, J., Charity, M., & Payne, W. (2013). A systematic review of the psychological and social benefits of participation in sport for children and adolescents: Informing development of a conceptual model of health through sport. *International Journal of Behavioral Nutrition and Physical Activity, 10*(1), 98. https://doi.org/10.1186/1479-5868-10-98

Faß, E., & Schlesinger, T. (2019). The role of individual resources, health behaviour and age perception as determinants of sports participation in older age. *Ageing and Society, 41*(3), 746–772. https://doi.org/10.1017/S0144686X19001260

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial networks. *Communications of the ACM, 63*(11)*,* 139–144 http://dx.doi.org/10.1145/3422622

Heath, G. W., & Parra, D. C. (2012). Evidence-based intervention in physical activity: Lessons from around the world. *The Lancet, 380*(9838), 272–281. https://doi.org/10.1016/S0140-6736(12)60816-2

Holzinger, A., Langs, G., Denk, H., Zatloukal, K., & Müller, H. (2019). Causability and explainability of artificial intelligence in medicine. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *9*(4), e1312. https://doi.org/10.1002/widm.1312

Hyde, E. T., Omura, J., Fulton, J., Lee, S. M., Piercy, K. L., & Carlson, S. (2020). Disparities in youth sports participation in the U.S., 2017–2018. *American Journal of Preventive Medicine, 59*(2), 239–251. https://doi.org/10.1016/j.amepre.2020.05.011

Izumi, T., Hyodo, H., Yoshioka, T., & Wada, T. (2023). The influence of strategic swimming pool facility placement on swimming performance. *Journal of Physical Education and Sport, 23*(7), 1742–1750. https://doi.org/10.7752/jpes.2023.07214

Janssen, M., Hartog, M. D., Matheus, R., Ding, A. Y., & Kuk, G. (2020). Will algorithms blind people? The effect of explainable AI and decision-makers' experience on AI-supported decision-making in government. *Social Science Computer Review, 40*(3), 478–493. https://doi.org/10.1177/0894439320980118

Kassen, M. (2013). A promising phenomenon of open data: A case study of the Chicago open data project. *Government Information Quarterly*, 30(4), 508–513. https://doi.org/10.1016/j.giq.2013.05.012

Kingma, D. P., & Welling, M. (2013). Auto-encoding variational Bayes. *arXiv*. https://doi.org/10.48550/arXiv.1312.6114

Kokolakakis, T., Castellanos-García, P., & Lera-López, F. (2017). Differences in formal and informal sports participation at regional level in England. *International Journal of Sport Policy and Politics, 9*(4), 491–504. https://doi.org/10.1080/19406940.2017.1287757

Lan, G., Xiao, S., Yang, J., Wen, J., & Xi, M. (2023). Generative AI-based data completeness augmentation algorithm for data-driven smart healthcare. *IEEE Journal of Biomedical and Health Informatics*. https://doi.org/10.1109/JBHI.2023.3327485

Liu, S., Chen, J., Feng, Y., Xie, Z., Pan, T., & Xie, J. (2024). Generative artificial intelligence and data augmentation for prognostic and health management: Taxonomy, progress, and prospects. *Expert Systems With Applications, 255*, 124511. https://doi.org/10.1016/j.eswa.2024.124511

Local Interpretable Model-agnostic Explanations (LIME) (n.d.). Welcome to LIME's documentation. https://lime.readthedocs.io

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Proceedings of the 31st International Conference on Neural Information Processing Systems,* 4768–4777. https://dl.acm.org/doi/10.5555/3295222.3295230

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., & Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, 2, 56–67. https://doi.org/10.1038/s42256-019-0138-9

McNoe, B., Marsh, L., Venter, N., Morgaine, K., Reeder, A., & McLean, R. (2022). National sporting organisation policies: A health promotion opportunity? *Health Promotion Journal of Australia, 34*(2), 480–487. https://doi.org/10.1002/hpja.601

Osipov, A., Starova, O., Malakhova, A., Vonog, V., Zhavner, T., Salyamova, P., Struchkov, V., & Kudryavtsev, M. (2016). The modernization process of physical education of students in the framework of the implementation of the state strategy for the development of physical culture, sport, and tourism in the Russian Federation. *Journal of Physical Education and Sport*, *16*(4), 1236–1241. https://doi.org/10.7752/jpes.2016.04196

Paproki, A., Salvado, O., & Fookes, C. (2024). Synthetic data for deep learning in computer vision & medical imaging: A means to reduce data bias. *ACM Computing Surveys, 56*(11)*,* 1–37. https://doi.org/10.1145/3663759

Peppes, N., Tsakanikas, P., Daskalakis, E., Alexakis, T., Adamopoulou, E. F., & Demestichas, K. (2023). FoGGAN: Generating realistic Parkinson's disease freezing of gait data using GANs. *Sensors*, *23*(19), 8158. https://doi.org/10.3390/s23198158

Ribeiro, M., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations*, 97–101. https://doi.org/10.18653/v1/N16-3020

Ruseski, J., Humphreys, B., Hallmann, K., & Breuer, C. (2011). Family structure, time constraints, and sport participation. *European Review of Aging and Physical Activity, 8*(1), 57–66. https://doi.org/10.1007/s11556-011-0084-y

SHapley Additive exPlanations (SHAP) (n.d.). Main website. https://shap.readthedocs.io

Synthetic Data Vault (SDV) (n.d.). Main website. https://docs.sdv.dev/sdv

Xu, L., Skoularidou, M., Cuesta-Infante, A., & Veeramachaneni, K. (2019). Modeling tabular data using conditional GAN. *Advances in Neural Information Processing Systems*, *32*, 7333–7343. https://doi.org/10.48550/arXiv.1907.00503

Yue, Q. (2022). Dynamic database design of sports quality based on genetic data algorithm and artificial intelligence. *Computational Intelligence and Neuroscience, 2022,* 7473109. https://doi.org/10.1155/2022/7473109

Zheng, J., & An, R. (2015). Satisfaction with local exercise facility: A rural-urban comparison in China. *Rural and Remote Health, 15*(1), 2990. https://doi.org/10.22605/RRH2990

Zhou, T., Wu, X., Wang, Y., Wang, Y., & Zhang, S. (2024). Application of artificial intelligence in physical education: A systematic review. *Education and Information Technologies*, *29*(7), 8203–8220. https://doi.org/10.1007/s10639-023-12128-2