Advancing Sustainable AI: Research Perspectives from the CINI-AIIS Lab at Federico II

Antonino Ferraro[3], Antonio Galli[1], Lidia Marassi[1], Stefano Marrone[1], Vincenzo Moscato[1], Antonio Elia Pascarella[1,2] and Carlo Sansone[1]

[3]*Department of Information Science and Technology, Pegaso University, Piazza Trieste e Trento 48, Naples, 80132, Italy*
[1]*University of Naples Federico II, Via Claudio 21, Naples, 80125, Italy*
[2]*Latitudo 40, via Emanuele gianturco 31, 80146 Naples, Italy*

Abstract

Artificial Intelligence (AI) is emerging as both a powerful enabler and a critical challenge in the context of sustainability. As AI technologies become increasingly pervasive, they introduce complex environmental trade-offs—particularly in energy-intensive applications like generative models and deepfake systems. This paper explores key dimensions of sustainable AI through three application domains. First, it examines the ecological footprint of generative AI, highlighting the need for energy-efficient model architectures and the integration of sustainability into ethical and regulatory frameworks. Second, it addresses the use of AI in biofuel research, presenting challenges in modeling biomass pyrolysis and the limitations of data imputation techniques on predictive accuracy and interpretability. Finally, it introduces a graph-enhanced deep learning approach for smart waste management in urban areas, combining spatial embeddings with temporal modeling to forecast bin fill levels more accurately. Together, these case studies emphasize the need for environmentally aware AI systems that are both technically robust and aligned with sustainable development goals.

Keywords

Synthetic data, Ethics, Human-Centred AI, Waste Management, Generative AI

1. Introduction

Artificial Intelligence (AI) is increasingly shaping the evolution of multiple sectors, offering not only enhanced automation and advanced decision-making capabilities, but also contributing in meaningful ways to sustainability and environmental resilience. From energy optimization to waste reduction and the development of clean technologies, AI is emerging as a critical enabler in the pursuit of global climate and sustainability goals. Its capacity to analyze vast and heterogeneous datasets, uncover latent patterns, and support predictive and adaptive strategies makes it particularly valuable in managing the complexity and variability typical of environmental systems.

One of the most impactful areas of application is the optimization of energy systems. Through Machine Learning (ML), AI supports the modeling and forecasting of resource consumption, enables early detection of anomalies in power generation infrastructures, and assists in balancing supply and demand in renewable energy grids. These capabilities are especially important as renewable energy sources, such as photovoltaics and wind, introduce greater variability into the energy landscape. Moreover, in the field of bioenergy, AI facilitates the integration and interpretation of complex experimental data to predict yields from biomass conversion processes. This is particularly relevant in pyrolysis and fermentation, where input characteristics and process conditions vary significantly. Even in the presence of incomplete or noisy data, ML algorithms can provide reliable estimations, contributing to

Ital-IA 2025: 5th National Conference on Artificial Intelligence, organized by CINI, June 23-24, 2025, Trieste, Italy
*Corresponding author.
†These authors contributed equally.
antonino.ferraro@unipegaso.it (A. Ferraro); antonio.galli@unina.it (A. Galli); lidia.marassi@unina.it (L. Marassi); stefano.marrone@unina.it (S. Marrone); vincenzo.moscato@unina.it (V. Moscato); antonioelia.pascarella@unina.it (A. E. Pascarella); carlo.sansone@unina.it (C. Sansone)
0009-0006-8134-5466 (L. Marassi); 0000-0001-6852-0377 (S. Marrone); 0000-0002-1079-7741 (A. E. Pascarella); 0000-0002-8176-6950 (C. Sansone)

the design of more efficient and adaptive biofuel production pipelines.

Deep Learning (DL), a subset of ML, extends these capabilities by handling high-dimensional, multimodal inputs such as images, time series, and spatial data. In the context of smart cities, DL is transforming traditional infrastructure services. A prominent example is smart waste management, where DL models trained on spatio-temporal data can anticipate bin fill levels and optimize collection routes. By incorporating both spatial relationships and temporal trends, these models help municipalities reduce fuel consumption, minimize overflow events, and improve service responsiveness—ultimately contributing to cleaner, more livable urban environments.

However, the increasing adoption of AI also raises important questions about its own environmental footprint. The development and deployment of large-scale AI models, particularly generative architectures like transformers or deepfakes, demand substantial computational resources. Training such models often involves the use of energy-intensive data centers, which may rely on non-renewable energy sources and require additional cooling infrastructure. As a result, the environmental cost of AI itself—especially in high-performance computing contexts—has become a pressing concern. This has led to growing interest in sustainable AI practices, including energy-efficient model design, low-impact training procedures, and the shift toward greener cloud infrastructures powered by renewable energy.

In this paper, we present a selection of applied research activities conducted at the University of Naples Federico II, within the framework of the CINI-AIIS Lab. The projects discussed span multiple domains—ranging from energy forecasting and biofuel production to smart urban services and sustainable AI system design. Through these initiatives, we aim to highlight how AI can be harnessed not only as a tool for technological advancement, but also as a strategic asset in building environmentally conscious, scalable, and socially responsible solutions for a sustainable future.

2. Sustainability Challenges in Generative AI

The proliferation of generative artificial intelligence (AI) technologies, particularly deepfake systems, has highlighted an underappreciated but critical dimension of their deployment: environmental sustainability. While significant attention has been devoted to ethical, legal, and societal risks, the environmental implications of AI remain inadequately addressed in regulatory frameworks such as the European Union's Artificial Intelligence Act.

Generative AI models—including large language models and deepfake generation networks—rely on vast computational resources during both training and inference phases. This heavy reliance on computational power leads to substantial energy consumption, largely driven by high-performance data centers. These centers not only require significant electricity to perform operations but also depend on advanced cooling systems to maintain operational stability, thereby amplifying total energy demand.

A notable concern is the source of the energy used to power these infrastructures. Many data centers are still fueled by non-renewable sources such as fossil fuels, which substantially contribute to carbon emissions. As a result, the carbon footprint associated with training and deploying generative AI systems has escalated dramatically in recent years [1]. For instance, models trained using hundreds of GPUs over several weeks may consume energy equivalent to that used by multiple households in a year.

Deepfake technology offers a particularly illustrative example. The generation of hyper-realistic synthetic media requires considerable computational processing for both the initial model training and the iterative content creation processes. As the visual and auditory realism of these outputs increases, so does the associated energy cost. This represents a direct trade-off between media quality and environmental degradation.

These trends pose a significant ethical challenge. Traditional discussions of ethical AI often center on algorithmic bias, transparency, and fairness [2, 3]. However, an ethically grounded AI development framework must also incorporate ecological considerations. Technological progress should not come at the expense of planetary health. The energy-intensive nature of modern AI demands a paradigm shift in how ethical AI is conceptualized and implemented.

Future directions should include the design of energy-efficient model architectures, the adoption of low-impact training procedures, and the transition of data centers to renewable energy sources. More importantly, sustainability must become a foundational element of AI governance. Until environmental impact is incorporated into mainstream regulatory discourse and technical practice, the broader goal of ethical and responsible AI will remain incomplete.

3. Artificial Intelligence for Biofuels

In the domain of sustainable bioenergy, biomass pyrolysis has emerged as a promising thermochemical route for converting agricultural and forestry residues into valuable biofuels. However, modeling pyrolysis remains challenging due to the complexity and variability of biomass feedstocks and operating conditions. In [4], AI techniques were applied to this domain by compiling and harmonizing 1137 experimental records from over 160 studies into a unified dataset—named *Pyris*—focused on fixed-bed, non-catalytic pyrolysis. The dataset includes proximate and ultimate analyses of biomass, macro-components like cellulose and lignin, process parameters such as temperature and heating rate, and the corresponding bio-liquid yield.

A major difficulty encountered in building predictive models was the high rate of missing data, especially for complex biomass properties such as lignin or cellulose content. To address this, the study explored multiple imputation techniques, including Generative Adversarial Imputation Nets (GAIN) [5]. GAIN operates by training two neural networks adversarially: a generator, which attempts to fill in the missing values, and a discriminator, which tries to distinguish real observed data from the generator's imputed values. This adversarial mechanism encourages the generator to produce realistic imputations that are statistically coherent with the observed data distribution.

Despite the sophistication of GAIN, results revealed that imputing missing data introduces noise that propagates into both predictive accuracy and model interpretability. The main regression task involved predicting the bio-liquid yield (i.e., the condensed product of pyrolysis expressed as mass percentage of the original biomass) using biomass and process features. When trained on a complete subset of about 500 records, the XGBoost model achieved a mean absolute error (MAE) of 2.28. However, this error increased to 3.45 when using the full dataset with GAIN-imputed values, accompanied by a drop in R^2 from 0.80 to 0.66, as reported in Table 1. These results highlight the uncertainty introduced by synthetic imputations, even with advanced generative approaches.

More importantly, the imputed data affected the stability and trustworthiness of interpretation tools, such as SHAP (SHapley Additive exPlanations) and Partial Dependence Plots (PDP), which are often used to extract mechanistic insights from black-box models. The study observed that interpretation results—such as the directional effect of variables like ash or cellulose—varied depending on the presence of imputed data and on the interpretability method used. This inconsistency is particularly problematic in a domain like pyrolysis, where interpretability must align with known physico-chemical principles to gain scientific credibility. These findings underscore an open challenge in the field: interpretability in AI for physical processes is a core requirement that must be grounded in domain knowledge.

To address these issues, the authors advocate for a shift towards *process-informed* or *physics-aware* AI models, which explicitly embed domain constraints—such as mass balance or kinetic laws—into learning architectures or objective functions. Such hybrid models can improve both the robustness of predictions and the plausibility of interpretations, helping bridge the gap between data-driven learning and scientific understanding.

Table 1
Performance of XGBoost on different dataset configurations

Dataset	MSE	RMSE	MAE	R^2
Complete (no missing)	17.80	4.21	2.28	0.80
GAIN-imputed	32.17	5.66	3.45	0.66

4. Waste Management in Smart Cities

An emerging direction in the application of artificial intelligence to urban infrastructure is the development of predictive systems for smart waste management, particularly in the context of smart cities [6, 7, 8, 9]. Managing waste in dense urban areas presents a multifaceted challenge, compounded by the unpredictability of waste generation patterns and the spatial distribution of collection points. Traditional machine learning models based solely on time-series analysis of individual waste bins often fail to capture the spatial dependencies that significantly influence waste accumulation dynamics. To address this limitation, a novel methodology has been proposed that integrates graph-based modeling and deep learning techniques to forecast waste bin fill levels with higher accuracy and contextual awareness.

The approach begins by modeling the network of smart waste bins as a graph, where each bin is represented as a node and spatial relationships—computed using geodesic distance via the Haversine formula are encoded as weighted, undirected edges. This representation captures physical proximity and potential influence among bins. The core innovation lies in leveraging this graph structure to generate node embeddings that encapsulate spatial dependencies. To achieve this, a random walk strategy is employed to simulate navigation through the bin network, with higher probabilities assigned to shorter spatial transitions, thus prioritizing local interactions. These walks are then used to train a skip-gram model adapted from natural language processing to learn continuous vector representations of each bin. The result is a dense embedding space in which spatially close or structurally similar bins are represented by nearby points.

These embeddings are incorporated as additional features into recurrent neural networks—specifically, Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) architectures—that also consume temporal sequences of fill-level data. By fusing spatial and temporal information, the models are capable of learning complex spatio-temporal patterns that govern waste generation and accumulation. Experimental results validate the effectiveness of the approach: baseline models relying solely on time-series data achieved a Mean Absolute Error (MAE) of 0.24, whereas the inclusion of graph embeddings improved performance to 0.21 (GRU) and 0.18 (LSTM). With further hyperparameter optimization, the LSTM model using graph embeddings achieved an MAE of 0.16, outperforming its non-embedded counterpart by a margin of 0.08. These gains were consistent even when extending the training to a full three-year dataset, demonstrating the robustness and scalability of the method.

Beyond accuracy improvements, the integration of graph-based features introduces a modular and interpretable way to encode city-specific spatial configurations. This facilitates generalization to other urban contexts, as the graph structure can be updated dynamically when bins are added, removed, or relocated. Furthermore, the method supports broader smart city use cases beyond waste management, offering a template for predictive modeling in domains such as traffic forecasting, energy consumption, and mobility systems, where spatially distributed sensors and infrastructures interact in complex ways.

This line of work highlights the growing importance of graph machine learning techniques in modeling urban environments, particularly where data is both spatially and temporally structured. By treating infrastructure elements as nodes in a graph and learning from their interactions, it becomes possible to build more accurate, adaptive, and context-aware AI systems that align with the goals of sustainability, operational efficiency, and urban livability.

Acknowledgments

This work was supported in part by the Piano Nazionale Ripresa Resilienza (PNRR) Ministero dell'Università e della Ricerca (MUR) Project under Grant PE0000013-FAIR

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

[1] P. Dhar, The carbon impact of artificial intelligence, Nature Machine Intelligence 2 (2020) 423–425.

[2] K. Siau, W. Wang, Artificial intelligence (ai) ethics: Ethics of ai and ethical ai, Journal of Database Management (JDM) 31 (2020) 74–87.

[3] B. Mittelstadt, Principles alone cannot guarantee ethical ai, Nature Machine Intelligence 1 (2019) 501–507.

[4] A. E. Pascarella, A. Coppola, S. Marrone, R. Chirone, C. Sansone, P. Salatino, Critical assessment of machine learning prediction of biomass pyrolysis, Fuel 394 (2025) 135000.

[5] J. Yoon, J. Jordon, M. Schaar, Gain: Missing data imputation using generative adversarial nets, in: International conference on machine learning, PMLR, 2018, pp. 5689–5698.

[6] M. Nilssen, To the smart city and beyond? developing a typology of smart urban innovation, Technological forecasting and social change 142 (2019) 98–104.

[7] M. Aazam, M. St-Hilaire, C.-H. Lung, I. Lambadaris, Cloud-based smart waste management for smart cities, in: 2016 IEEE 21st International Workshop on Computer Aided Modelling and Design of Communication Links and Networks (CAMAD), 2016, pp. 188–193. doi:10.1109/CAMAD.2016.7790356.

[8] J. M. Gutierrez, M. Jensen, M. Henius, T. Riaz, Smart waste collection system based on location intelligence, Procedia Computer Science 61 (2015) 120–127. URL: https://www.sciencedirect.com/science/article/pii/S1877050915030008. doi:https://doi.org/10.1016/j.procs.2015.09.170.

[9] R. Elhassan, M. A. Ahmed, R. AbdAlhalem, Smart waste management system for crowded area : Makkah and holy sites as a model, in: 2019 4th MEC International Conference on Big Data and Smart City (ICBDSC), 2019, pp. 1–5. doi:10.1109/ICBDSC.2019.8645576.