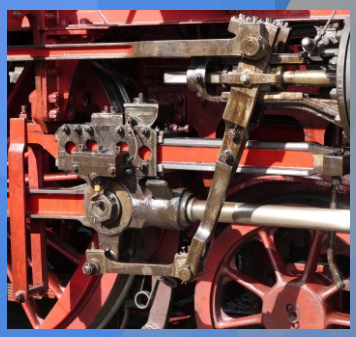
INTERNATIONAL SCIENTIFIC JOURNAL

INDUSTRY 4.0

ISSN (PRINT) 2534-8582
ISSN (WEB) 2534-997X

YEAR XI ISSUE 2/2026






① ② ③ ④

PUBLISHED BY
SCIENTIFIC TECHNICAL UNION OF MECHANICAL ENGINEERING
"INDUSTRY 4.0", BULGARIA

INTERNATIONAL SCIENTIFIC JOURNAL

INDUSTRY 4.0

YEAR XI, ISSUE 2 / 2026

ISSN (PRINT) 2534-8582, ISSN (WEB) 2534-997X

PUBLISHER

SCIENTIFIC TECHNICAL UNION OF MECHANICAL ENGINEERING "INDUSTRY 4.0"

108, Rakovski Str., 1000 Sofia, Bulgaria
tel. (+359 2) 987 72 90,
tel./fax (+359 2) 986 22 40,
office@stumejournals.com
WWW.STUMEJOURNALS.COM

EDITOR IN CHIEF

Prof. D.Sc. Georgi Popov, DHC, Technical University of Sofia, BG

All articles are published after peer review by two independent reviewers.

EDITORIAL BOARD

Members:

Acad. Igor Bychkov, RU
Cor. member Alexey Beliy, BY
Cor. member Svetozar Margenov, BG
Prof. Alexander Afanasyev, RU
Prof. Alexander Guts, RU
Prof. Alshyn Altybayev, KZ
Prof. Andrzej Golabczak, PL
Prof. Bobek Shuklev, MK
Prof. Boris Gordon, EE
Prof. Branko Sirok, SI
Prof. Claudio Melchiorri, IT
Prof. Cveta Martinovska, MK
Prof. Dale Dzemydiene, LT
Prof. Dimitar Yonchev, BG
Prof. Dimitrios Vlachos, GR
Prof. Dobre Runchev, NM
Prof. Dragan Perakovic, HR
Prof. Galina Nikolcheva, BG
Prof. Galina Zhavoronkova, UA
Prof. Gerard Lyons, IE
Dr. Giovanni Pappalettera, IT
Prof. Henrik Carlsen, DK
Prof. Idilia Bachkova, BG
Prof. Idit Avrahami, IL
Prof. Inocentiu Maniu, RO
Prof. Iurii Bazhal, UA
Prof. Jürgen Köbler, DE
Prof. Jiri Maryska, CZ
Prof. Katia Vutova, BG
Prof. Lappalainen Kauko, FI
Dr. Liviu Jalba, RO
Prof. Luigi del Re, AT
Prof. Majid Zamani, DE
Dipl.-Kfm. Michael Grethler, DE
Prof. Michael Valasek, CZ
Prof. Milija Suknovic, RS
Prof. Miodrag Dashic, RS
Prof. Mohammed Taleb Obaidat, JO
Prof. Murat Alanyali, TR
Prof. Nadežda Jankelová, SK
Prof. Nafisa Yusupova, RU
Prof. Nina Bijedic, BA
Prof. Ninoslav Marina, MK
Prof. Olga Zaborovskaia, RU
Prof. Pavel Kovach, RS
Prof. Petar Kolev, BG
Prof. Peter Korondi, HU
Prof. Peter Sincak, SK
Prof. Petra Bittrich, DE
Prof. Predrag Dasic, RS
Prof. Radu Dogaru, RO
Prof. Raul Turmanidze, GE
Prof. René Beigang, DE
Prof. Rozeta Miho, AL
Prof. Salloom A. Al-Jaboori, JO
Prof. Seniye Ümit Oktay Firat, TR
Prof. Sreten Savicevic, ME
Prof. Sveto Svetkovski, MK
Prof. Tomislav Šarić, HR
Prof. Vasile Cartofeanu, MD
Prof. Vjaceslavs Bobrovs, LV
Prof. Vladyslav Alieksieiev, UA

TECHNICAL EDITORS:

M. Sc. eng. Radoslav Daskalov, Bulgaria
M. Sc. eng. Oleg Mihailov, Bulgaria

C O N T E N T S

TECHNOLOGICAL BASIS OF "INDUSTRY 4.0"

Information Theory for Medical Data Fusion
Magdalena Punceva, Ninoslav Marina 46

Artificial intelligence approaches for modeling nonlinear dynamical systems
Bregan Isufi, Alfred Daci 51

Imagined pedestrian encounters: analyzing world-model rollouts in a reinforcement learning driving agent
Stefan Zlatinov, Ilija Mizhimakoski, Gorjan Nadzinski, Mile Stankovski 58

Application of artificial intelligence in pre-university education – case study schools in the municipality of Kamenica – Kosovo
Selami Klaiqi, Alban Rakipi 63

DOMINANT TECHNOLOGIES IN "INDUSTRY 4.0"

Preparation and Optical Properties of Gelatin-SiO2 Hybrid Coatings on Glasses for Photovoltaic Applications
Stancho Yordanov, Tatyana Ivanova, Shaban Uzun, Vladimir Petkov, Ina Gospodinova 69

BUSINESS & "INDUSTRY 4.0"

Investigation of Positioning Accuracy Improvement Methods for Sigfox IoT Devices through WiFi MAC Address Analysis
Masaya Watanabe, Takuya Nanjo, Tomoo Sigehuzi, Satoru Tada, Masayuki Yamauchi, Takeshi Nishiyone, Hironori Wakaoka, Shuji Oji, Naoki Kabemoto, Hiroki Mori ,Takeshi Tanaka, Katia Vutova 72

Innovations in industrial enterprises: the role of AI, IOT, and automation in enhancing competitiveness
Slava Petkova 79

SOCIETY & „INDUSTRY 4.0"

AI ethics education as a tool to minimize risks in medicine
Nikoleta Leventi, Alexandrina Vodenicharova, Vidin Kirkov 83

AI and digital ethics in the age of generative systems-principles, standards and accountability across cultures
Ognen Firfov 87

Legal regulation problems in transport logistics: challenges and prospects
Irakli Kochadze, Ana Mgeladze 90

Integrating ai into insolvency procedures, challenges and opportunities in Albania
Teuta Hoxha 93

Review of the Development of Modern Green Energy Generators and Systems
Raul Turmanidze, Giorgi Popkhadze 96

Information Theory for Medical Data Fusion

Magdalena Punceva[1], Ninoslav Marina[2]
MIT Univerzitet SkopjeSkopje, Macedonia[1], University of Information Science and Technology Skopje, Macedonia[2]

Abstract: *Data fusion is the process of integrating multiple heterogeneous data sources to produce more accurate, comprehensive, and useful information than any single source alone. For the medical data fusion, this information may come from imaging, genomic data, clinical records, and physiological signals. It has emerged as a cornerstone of modern precision medicine. Information theory provides a rigorous mathematical framework for quantifying uncertainty, measuring information gain, and can thus be used to optimize fusion strategies across diverse clinical contexts. This paper presents a review of information-theoretic approaches suitable for medical data fusion, covering: fundamental concepts (entropy, mutual information, Kullback-Leibler divergence, rate-distortion), theoretical frameworks (information bottleneck, transfer entropy, partial information decomposition), and their application across major clinical domains. We synthesize recent advances in fusion-based architectures, discuss the critical challenges of uncertainty quantification, and provide practical guidelines for implementing information-theoretic fusion in clinical settings. Through a systematic analysis, we identify key challenges, including parameter sensitivity, missing modalities, and clinical interpretability, to outline promising directions for future research. This review aims to provide clinicians, researchers, and developers with a comprehensive understanding of how information theory can transform medical data for improved diagnostic accuracy, prognostic precision, and personalized patient care.*

KEYWORDS—*INFORMATION THEORY, DATA FUSION, ENTROPY, MUTUAL INFORMATION, RATE-DISTORTION THEORY, INFORMATION BOTTLENECK, TRANSFER ENTROPY, PARTIAL INFORMATION DECOMPOSITION, DEEP LEARNING.*

I. INTRODUCTION

Modern healthcare generates unprecedented volumes of heterogeneous data: high-resolution medical images (CT, MRI, PET), molecular profiles (genomics, proteomics, metabolomics), continuous physiological signals (ECG, EEG, EMG), and structured clinical data from electronic health records. Each data type provides a partial view of patient health, and thus their integration by medical data fusion holds the promise of a holistic understanding that can transform diagnosis, prognosis, and treatment selection [Duan et al., 2024, Salvi et al., 2024]. However, the fusion of medical data presents some fundamental challenges: heterogeneity (data types span continuous signals, categorical variables, images, and unstructured text), high dimensionality (genomic and imaging data often contain plenty of features), missing modalities (real-world clinical data frequently have incomplete measurements), temporal dynamics (patient status evolves over time, requiring longitudinal integration, interpretability (clinicians require transparent, explainable fusion results).

Information theory is the mathematical study of the quantification, storage, and communication of a particular type of mathematically defined information. The field was established and formalized by Shannon in his seminal work [Shannon, 1948]. It proposed a mathematical framework for the theory of communication; however, it soon found a wide range of other applications. It is now at the intersection of mathematics, statistics and computer science, and has applications in diverse fields ranging from electrical engineering to neurobiology. Shanon's paper is quite influential as it defines the fundamentals of digital communications and information theory. At its core, information theory quantifies uncertainty, measures the quantity of information, and characterizes the information content between two signals. In addition, information theory offers a fundamental approach to cryptography and cybersecurity. Such quantities are essential to understand how different medical data sources can be combined to inform clinical decisions.

Key information-theoretic concepts for medical data fusion, among others, include:

- **Entropy** $H(X)$ - Quantifies uncertainty in a random variable X, thus in the context of medical data it can be used, for example, to assess diagnostic uncertainty before and after a test.
- **Mutual Information** $I(X; Y)$ - Measures shared information between variables, the two random variables X and Y, and is ideal for identifying redundant tests or discovering novel biomarker relationships.
- **Kullback-Leibler (KL) divergence** $D_{KL}(P \parallel Q)$ - Quantifies the difference between probability distributions, more precisely, it measures how much a probability distribution Q is different from a probability distribution P. It is essential for validating clinical models.
- **Rate–distortion function** $R(D)$ is a central concept in rate-distortion theory, which deals with lossy compression. It quantifies the trade-off between the bit rate R used to represent a source and the distortion D incurred when reconstructing the source from the compressed data. It is the minimum rate R at which a source can be represented while ensuring the average distortion does not exceed D. This is important for medical data since we can define the acceptable level of quality of the medical measurements.
- **Directed information** $I(\mathbf{X}^n \rightarrow \mathbf{Y}^n)$ is an information-theoretic measure that quantifies the information flow from the random sequence $\mathbf{X}^n$ to another random sequence $\mathbf{Y}^n$. Directed information has applications to problems where causality plays an important role, such as in situations where feedback from $\mathbf{Y}^n$ to $\mathbf{X}^n$ is important. It is arguably important for medical applications, as the feedback for some type of patient parameters might improve the accuracy of the measurement of that specific parameter.

In order to use some of these concepts that are important in information theory, many scientists introduced and used several concepts that help in various aspects of data fusion, feature selection and machine learning.

The most sound and interesting concepts are the following:

- **Information Bottleneck** is a technique designed for finding the best tradeoff between accuracy and complexity when summarizing a random variable X with an observed relevant variable Y, given their joint probability distribution $P(X,Y)$. Applications include distributional clustering and dimension reduction, as well as deep learning. The information bottleneck can also be viewed as a rate distortion problem, with a distortion function that measures how well Y is predicted from a compressed representation T compared to its direct prediction from X.
- **Transfer entropy** is a non-parametric statistic measurement of the amount of directed transfer of information between two random processes. Transfer entropy from a process X to a process Y is the amount of uncertainty reduced in future values of Y by knowing the past values of X given past values of Y. Transfer entropy has been used for estimation of functional connectivity of neurons social influence in social networks and statistical causality between two processes. Hence, it has a great potential in evaluation of the causes of various diseases while processing medical data. Transfer entropy is a finite version of the directed information.
- **Partial Information Decomposition (PID)** decomposes multivariate information into redundancy, uniqueness, and synergy, thus providing theoretical guidance on fusion strategies. This measure has been introduced to generalize the pairwise relations between two variables described by, for

example, mutual information, to interactions between multiple variables.

This paper summarizes the existing approaches on information-theoretic concepts in the field of medical data fusion. Our contributions are focused on the following methods: (1) a systematic presentation of fundamental information theoretic concepts with clinical interpretations; (2) detailed analysis of theoretical frameworks, including partial information decomposition for medical fusion (3) comprehensive coverage of clinical applications across several medical domains; (4) a framework for evaluation of current challenges and future research directions; (5) practical guidelines for implementing information theoretic fusion in clinical settings.

Data fusion combines data from multiple and usually heterogeneous sources in order to obtain improved data quality and reduce data uncertainty. It is also essential for feature extraction and predictive analytics, and leads to reduced cost for the corresponding data processing. Data fusion has been used in various domains including sensor data fusion, geographical information systems, bio-informatics, and business intelligence. In this work, our focus is on data fusion techniques that can primarily increase data quality and help in data processing, feature selection and decision making in medical applications.

The remainder of this paper is organized such that Section II presents the mathematical foundations of information theory and their clinical interpretations. Section III introduces theoretical frameworks for multimodal medical data fusion. Section IV provides a comprehensive review of clinical applications. Section V addresses challenges and future directions, while Section VI concludes the review with a summary of the key contributions.

II. MATHEMATICAL FOUNDATIONS

In the current section, we elaborate on several theoretical concepts that show potential to be applied in the context of medical data fusion.

A. Entropy

In information theory, *entropy* is a measure of the uncertainty of a random variable. Thus, in medical data fusion, it can be used to quantify clinical uncertainty.

Definition II.1 (Shannon Entropy).

For a discrete random variable X with probability distribution p(x), the Shannon entropy is

$$\mathrm{H}(X) = -\sum_{x\in\mathcal{X}} p(x)\log p(x), \qquad (1)$$

and is measured in bits, when the logarithm is base 2.

Clinical interpretation: Entropy may quantify diagnostic uncertainty. A patient presenting with chest pain might have H(Diagnosis) = 2.1 bits when considering eight possible diagnoses ($2^3 = 8$); after an ECG, entropy might drop to H(Diagnosis | ECG) = 1.2 bits, reflecting reduced uncertainty, or in other words, more certainty, hence, a more precise diagnosis. In this particular case, with eight possible diagnoses, the situation with the highest uncertainty, or highest entropy, is when all eight diagnoses are equiprobable. In that case, the probability of each case is 1/8, and if we replace it in (1), we get an entropy of $\log_2(8) = 3$ bits. More generally, the maximal entropy in a case with N outcomes is $\log_2(N)$.

B. Mutual Information

Mutual information of two random variables is a measure of the mutual dependence between the two variables. Thus, we could use mutual information to measure shared information between any two random variables.

Definition II.2 (Mutual Information).

The mutual information between variables X and Y is

$$\mathrm{I}(X;Y) = \sum_{x,y} p(x,y)\log\frac{p(x,y)}{p(x)p(y)}. \qquad (2)$$

These are the properties relevant to clinical applications:

- $\mathrm{I}(X;Y) \geq 0$ with equality iff X, Y are independent,
- $\mathrm{I}(X;Y) = \mathrm{I}(Y;X)$ (symmetry),
- Non-linear relations missed by correlation,
- Invariant to monotonic transformations of data.

Clinical interpretation: Mutual information measures how much a diagnostic test reduces uncertainty about a disease. For example, I(Troponin; ACS) = 0.82 bits indicates that the diagnostic test "Troponin measurement" provides substantial information about the disease acute coronary syndrome (ACS). The Troponin test is a crucial tool in the diagnosis of Acute Coronary Syndrome (ACS). It is a specific marker for myocardial necrosis and is preferred for diagnosing myocardial infarction due to its high cardiac specificity. If the mutual information between another test and ACS is lower, then we can conclude that the Troponin test is preferred.

C. Kullback-Leibler Divergence

Kullback-Leibler (KL) Divergence is a type of statistical distance, or in other words, a measure of how much an approximating probability distribution Q is different from a true probability distribution P. It was introduced by Kullback and Leibler [1951]. Sometimes it is called relative entropy or I−divergence, like in Csiszar [1975]. Consider two probability distributions, a true P and an approximating Q. Often, P represents the data, the observations, or a measured probability distribution, and distribution Q represents a theory, a model, a description, or another approximation of P. However, sometimes the true distribution P represents a model and the approximating distribution Q represents (simulated) data that are intended to match the true distribution. The KL Divergence is then interpreted as the average difference of the number of bits required for encoding samples of P using a code optimized for Q rather than one optimized for P.

Definition II.3 (KL Divergence).

For two probability distributions P and Q over the same space

$$\mathrm{D}_{\mathrm{KL}}(P||Q) = \sum_{x\in\mathcal{X}} P(x)\log\frac{P(x)}{Q(x)}. \qquad (3)$$

These are the properties of the KL Divergence:

- $\mathrm{D}_{\mathrm{KL}}(P \parallel Q) \geq 0$, with equality iff $P = Q$,
- Asymmetric: $\mathrm{D}_{\mathrm{KL}}(P \parallel Q) \neq \mathrm{D}_{\mathrm{KL}}(Q \parallel P)$,
- Not a true distance metric but a divergence.

Clinical applications of KL Divergence:

- Population shift detection: Comparing current patient distribution to the historical baseline,
- Model validation: Detecting when clinical prediction models become unreliable,
- Multi-center trial monitoring: Identifying sites deviating from protocol,
- Diagnostic test evaluation: Quantifying how much a test shifts diagnostic probabilities.

The following fundamental relationships hold between mutual information and entropy, as well as mutual information and KL divergence:

$$\mathrm{I}(X;Y) = \mathrm{H}(X) - \mathrm{H}(X|Y), \qquad (4)$$

$$\mathrm{I}(X;Y) = \mathrm{H}(Y) - \mathrm{H}(Y|X), \qquad (5)$$

$$\mathrm{I}(X;Y) = \mathrm{H}(X) + \mathrm{H}(Y) - \mathrm{H}(X,Y), \qquad (6)$$

$$\mathrm{I}(X;Y) = \mathrm{D}_{\mathrm{KL}}\left(P(X,Y)||P(X)P(Y)\right). \qquad (7)$$

In the above equations, $\mathrm{H}(X,Y)$ is the joint entropy of the random variables X and Y, $\mathrm{H}(A \mid B)$ is the conditional entropy of A when B is given, and $P(X,Y)$ is their joint probability distribution.

Equation (4) is particularly important clinically, since it shows that mutual information is exactly the reduction in uncertainty about X gained by knowing Y, which is the information gain from a diagnostic test. KL divergence remains fundamental in ensuring that fused multimodal representations preserve meaningful information while maintaining statistical consistency across different medical imaging modalities and clinical data sources.

D. Rate-Distortion Function

Rate-Distortion (R-D) theory, pioneered by Claude Shannon, addresses a key question in lossy data compression. Given a source signal *X*, it explores what is the minimum number of bits per symbol *R* required to represent *X* such that the average distortion between *X* and its reconstructed version $\hat{X}$ does not exceed a given threshold *D*. This fundamental limit is called *rate-distortion function*, and is denoted *R*(*D*).

Definition II.4 (Rate-Distortion function).

The Rate Distortion function R(*D*) *is defined as the minimum of the mutual information* I(*X; $\hat{X}$*) *over all conditional probability distributions p*($\hat{x}$|*x*) *that satisfy the distortion constraint* E[*d*(*X*, $\hat{X}$)] ≤ *D:*

$$R(D) = \min_{p(\hat{x}|x)):\mathbb{E}[d(X,\hat{X})]\leq D} \mathrm{I}(X;\hat{X}). \qquad (8)$$

Here, *d*(*X*, $\hat{X}$) is some distortion metric (e.g., Mean Squared Error (MSE) for continuous signals or Hamming distance for discrete data). R(D) is a convex, non-increasing function, signifying that to achieve lower distortion, a higher bit rate is necessary. The curve effectively defines the boundary of what is theoretically achievable. Applying R-D theory to fused medical data is more complex than applying it to a single modality. The goal shifts from compressing a single signal to compressing a set of correlated signals (modalities) in a way that preserves the clinically relevant information contained in their joint structure. The theoretical insights from R-D analysis can inform the development of practical compression algorithms for medical data fusion. Neural networks, particularly autoencoders with a rate-distortion loss function, can be trained to learn an optimal, non-linear transformation of the multi-modal data. The loss function is designed to minimize a weighted sum of the rate and the task-specific fusion distortion, effectively learning to approximate the R(D) function for the given data and clinical task. In summary, Rate-Distortion theory provides an indispensable, rigorous foundation for designing efficient and clinically effective medical data fusion systems.

E. Directed Information

While conventional information-theoretic measures such as mutual information capture symmetric statistical dependencies between variables, they are fundamentally limited in their ability to distinguish causal relationships. In the context of medical data fusion, where the goal is often to understand the directional influence between modalities or from modalities to clinical outcomes, a more precise framework is required. Directed information, originally developed by Massey [1990], provides a framework to quantify the causal influence from one stochastic process to another, respecting the temporal or logical flow of information.

Definition II.5 (Directed information).

Directed information measures the amount of information that flows from a source process $\mathbf{X} = (X_1, X_2, ..., X_n)$ *to a target process* $\mathbf{Y} = (Y_1, Y_2, ..., Y_n)$ *in a causal manner. It is formally defined as*

$$\mathrm{I}(\mathbf{X} \to \mathbf{Y}) = \sum_{i=1}^{n} \mathrm{I}(\mathbf{X}^i; Y_i | \mathbf{Y}^{i-1}), \qquad (9)$$

where $\mathbf{X^i} = (X_1, X_2, ..., X_i)$ *represents the past and present of the source up to time i, and* $\mathbf{Y^{i-1}} = (Y_1, Y_2, ..., Y_{i-1})$ *represents the past of the target before time i. The key distinction from mutual information is the conditioning on the past of the target, which ensures that only predictive information that is not already contained in the target's own history is attributed as directed information.*

This formulation captures the intuitive notion of causality: **X** causes **Y** if knowing the past of **X** helps predict the next value of **Y** better than knowing only the past of **Y** alone. The relationship between directed information and mutual information is given by

$$\mathrm{I}(\mathbf{X} \to \mathbf{Y}) + \mathrm{I}(\mathbf{X} \leftarrow \mathbf{Y}) = \mathrm{I}(\mathbf{X}; \mathbf{Y}), \qquad (10)$$

where I(**X** ← **Y**) represents the information flow in the opposite direction. This decomposition reveals that mutual information can be viewed as the sum of information flowing in both directions. The application of directed information to medical data fusion addresses several critical challenges that arise when integrating heterogeneous clinical data sources. Modern healthcare generates extensive longitudinal data, including continuous physiological monitoring, repeated imaging studies, and sequential laboratory measurements. Directed information provides a principled framework for understanding how these temporal processes influence one another. For example, in critical care settings, one might ask how changes in electrocardiogram (ECG) patterns direct information to subsequent blood pressure measurements, or how medication administration times influence physiological responses.

III. THEORETICAL FRAMEWORKS FOR MULTIMODAL DATA FUSION

In this section, we describe three frameworks that are motivated by the mathematical foundations of information theory and are proven to be useful in multimodal data fusion in general, hence in the medical applications that rely on multimodality.

A. Directed Information

The Information Bottleneck (IB) principle provides a theoretical foundation for learning efficient representations. The information bottleneck method is a technique designed for finding the best tradeoff between accuracy and complexity (compression) when summarizing a random variable *X*, given a joint probability distribution *P*(*X*,*Y*) between *X* and an observed relevant variable *Y*. It was introduced by Tishby et al. [2000]. Applications include distributional clustering and dimension reduction, and more recently, it has been suggested as a theoretical foundation for deep learning. It generalized the classical notion of minimal sufficient statistics from parametric statistics to arbitrary distributions, not necessarily of exponential form. It does so by relaxing the sufficiency condition to capture some fraction of the mutual information with the relevant variable *Y*. The information bottleneck can also be viewed as a rate distortion problem, with a distortion function that measures how well *Y* is predicted from a compressed representation *T* compared to its direct prediction from *X*. This interpretation provides a general iterative algorithm for solving the information bottleneck trade-off and calculating the information curve from the distribution *P*(*X*, *Y*). Let the compressed representation be given by a random variable Z_θ, which is a learned representation of the input *X*. The algorithm minimizes the following functional with respect to the conditional distribution $\theta = P(Z_\theta|X)$. The information bottleneck is defined as

$$\min_{\theta} \left(\mathrm{I}(X; Z_\theta) - \lambda \, \mathrm{I}(Z_\theta; Y) \right), \qquad (11)$$

where λ is a Lagrange multiplier that controls the trade-off between compression and prediction.

The following clinical applications were identified for the use of information bottleneck: learning robust features invariant to artefacts and patient variations, filtering out irrelevant information while preserving diagnostically relevant content, and improving generalization across different hospitals and populations.

B. Transfer Entropy

Transfer entropy (TE) measures directional information flow between physiological signals. It was proposed by Schreiber [2000] and is a powerful concept that adds a dynamic, directional layer to the information-weighted sensor fusion we discussed earlier. While fusion focuses on combining data at the same time for a reliable snapshot, transfer entropy helps us understand the flow of informa-

tion over time between different signals. It is a statistical measure that quantifies the directed (asymmetric) flow of information from one time series to another. In simpler terms, it tells you how much the past values of a source variable (e.g., heart rate) help predict the future values of a target variable (e.g., blood pressure), over and above what the target's own past could predict. If including the source's past reduces uncertainty about the target's future, we say information has flowed from source to target. This makes transfer entropy particularly useful for understanding complex, interconnected systems. Let's for $i < j$, denote any process (A_i, A_{i+1}, ..., A_j) by the vector $\mathbf{A}_i^j$. Also, denote the process ($B_{-\infty}$, ..., B_j) by the vector $\mathbf{B}^j$. The most common general definition of transfer entropy from a process $\mathbf{X}^t$ to a process $\mathbf{Y}^t$ is

$$T_{\mathbf{X}\to\mathbf{Y}} = \mathrm{I}\left(Y_t; \mathbf{X}_{t-1}^{t-L} | \mathbf{Y}_{t-1}^{t-K}\right), \qquad (12)$$

where I(C;D|F) is the conditional mutual information between C and D given F, Y_t is the future state of the target variable at time t, $\mathbf{X}_{t-1}^{t-L}$ represents the past states of the source variable (from $t-1$ to $t-L$), and $\mathbf{Y}_{t-1}^{t-K}$ represents the past states of the target variable (from $t-1$ back to $t-K$), which we are conditioning on. In essence, this definition reads that the transfer entropy from **X** to **Y** is the information shared between the future of **Y** and the past of **X**, given that we already know the past of **Y**. To make this practical for calculation, transfer entropy is usually expressed using Shannon Entropy H. If we simplify by assuming the processes are first-order Markov chains (i.e., we only need to consider the immediate past $t-1$, to predict the next state), the equation becomes

$$T_{\mathbf{X}\to\mathbf{Y}} = \mathrm{H}\left(Y_t | Y_{t-1}\right) - \mathrm{H}\left(Y_t | Y_{t-1}, X_{t-1}\right), \qquad (13)$$

where H(Y_t| Y_{t-1}) is the entropy (or uncertainty) of current Y_t given its own immediate past, H(Y_t| Y_{t-1}, X_{t-1}) is the entropy of the future of current Y_t given both its own past and the immediate past of **X**.

By subtracting the second term from the first, we see exactly how much the uncertainty of **Y** is reduced by knowing the past of **X**. This reduction is, in fact, the transfer entropy. Transfer entropy can also be defined as the Kullback-Leibler divergence between two probability distributions

$$T_{X\to Y} = \mathrm{D}_{\mathrm{KL}}\left(P(Y_{t+1}, |Y_t, X_t) || P(Y_{t+1}|Y_t)\right). \qquad (14)$$

The fundamental goal of TE is to detect causality or directional influence. It is an asymmetric measure, meaning the information flow from **X** to **Y** is different from the flow from **Y** to **X**. This allows us to map the directional influences within a network, which is crucial for identifying which variable is driving changes in another. Transfer entropy is actively being used to solve practical problems such as the prediction of patient deterioration in intensive care units (ICUs).

C. *Partial Information Decomposition (PID)*

The Partial Information Decomposition (PID) framework was introduced by Williams and Beer [2010] as an extension of information theory that aims to generalize the pairwise relations described by information theory to the interaction of multiple variables. It has been shown that PID cannot follow the core properties and axioms of classical information theory. While mutual information is able to capture non-linear statistical relationships between two random variables, the PID framework may quantify multivariate statistical dependencies in a more general way. For two source variables X_1, X_2 and a target variable Y, the mutual information I(X_1, X_2;Y) can be decomposed into four components:

$$\mathrm{I}(X_1, X_2; Y) = \mathrm{U}(X_1; Y \backslash X_2) + \mathrm{U}(X_2; Y \backslash X_1) + \mathrm{S}(X_1, X_2; Y) + \mathrm{R}(X_1, X_2; Y), \qquad (15)$$

where U(X_1;$Y \setminus X_2$) is the unique information from X_1 to Y which is not in X_2, U(X_2;$Y \setminus X_1$) is the unique information from X_2 to Y which is not in X_1, S(X_1, X_2;Y) is the synergistic information that is in the interaction of X_1 and X_2 about Y, and R(X_1, X_2;Y) is the redundant information present in both sources. PID has been applied to diverse fields. Liang et al. [2023] applied PID to information fusion for medical data.

Clinical Implications of PID Components may be summarized in the following examples, where these measures could be used:

- **High redundancy** (R): Multiple tests provide overlapping information. In that case we consider eliminating redundant tests to reduce cost and patient burden.
- **High uniqueness** (U(X_1;$Y \setminus X_2$), U(X_2;$Y \setminus X_1$)): Each modality contributes distinct information, hence both are needed for comprehensive assessment.
- **High synergy** (S): Information emerges only from combination, it requires sophisticated fusion approaches (early or intermediate fusion).

Zhang et al. [2025] conducted the first comprehensive evaluation of PID metrics across diverse biomedical data using seven different modality pairs. The data modalities studied in this paper include clinical radiology, pathology, and genomic data. Future Research Directions may involve hybrid frameworks combining PID with deep learning for end-to-end interpretable fusion and causal PID extending decomposition to capture causal relationships between modalities.

IV. *Clinical Applications of Information-Theoretic Fusion*

There is an enormous potential for the use of information theory for the purpose of medical data fusion. Next, we sumarize the medical applications in which information-theoretic fusion, may improve the outcome of the patient examination.

Neurology and Neuroimaging are promising use cases, as Alzheimer's disease diagnosis and multi-center clinical trial monitoring for brain functioning. Multimodal fusion combining MRI, PET, Cerebrospinal fluid (CSF) biomarkers, and cognitive tests has transformed Alzheimer's diagnosis. Main findings by [Zhang et al., 2020] conclude that amyloid PET and CSF provide redundant information, tau PET provides unique prognostic information, and cognitive tests provide information almost independent of biomarkers. Speaking about multi-center clinical trial monitoring, traditional methods for analyzing brain connectivity often rely on correlations across a group of subjects, which averages out individual variability. For a single patient's PET scan, it compares the statistical distribution of signals in every pair of brain regions using the symmetric Kullback-Leibler divergence defined as DKL(P‖Q) + DKL(Q‖P). By collecting data from different sites and measuring some of their features, the authors conclude that some particular sites have patients with more advanced disease. It may allow early detection and may enable corrective action.

Cardiovascular Medicine is another important area where the information-theoretic approach may improve the diagnostics and treatment. It is used in heart failure prognosis and multi-sensor intensive care unit (ICU) monitoring. Wu et al. [2025] reviewed multimodal deep learning for heart failure prognosis using: admission notes (unstructured text), clinical tabular data (labs, vitals), electrocardiogram (ECG) waveforms, and echocardiography images. Integration improved prognostic precision, with moderate redundancy. By fusing information with an understanding of context and signal quality, it can filter noise, reduce cognitive load, and alert clinicians to the most critical events with higher confidence. Intelligent systems assess the quality and context of each signal to create a fused, more accurate estimate. The improved monitoring comes from the enhanced signal reliability, context-aware alarms, multi-dimensional anomaly detection, efficient mobility assessment, and comprehensive respiratory monitoring.

Other areas where information theory may be used are oncology and sepsis prediction. There are several emerging applications that include digital twins for personalized medicine and wearables for continuous monitoring. Artificial Intelligence (AI) powered aging digital twins are proposed to combine genetic data, medical history, continuous physiological monitoring (e.g., from wearable sensors or

other standard consumer electronics), lifestyle factors, and similar. Information theory guides which data sources are most informative for individual health trajectories. The integration of consumer wearables with clinical monitoring systems enables seamless transition between in-hospital and ambulatory monitoring, early detection of deterioration after discharge, longitudinal tracking of physiological baselines, and personalized alert thresholds based on individual patterns.

V. CHALLENGES AND FUTURE DIRECTIONS

Current challenges may be summarized into two groups: theoretical and clinical challenges. Theoretical challenges are connected to parameter sensitivity, scalability, interpretability, and missing modalities. Some of the metrics may heavily depend on the preprocessing, while information-theoretic calculations may become computationally expensive. In addition, real-world clinical data are often incomplete. Potential solutions involve sensitivity analysis, approximation methods, hierarchical fusion, and generative models. The clinical challenges are related to the regulatory hurdles across various jurisdictions. Most studies remain retrospective rather than prospective. Many clinicians require interpretable outputs for trust and adoption, and the new proposed methodologies must complement rather than disrupt existing practices. Lastly, one has to ensure that the novel models perform equally across diverse populations. There are multiple emerging technologies and opportunities. Generative AI offers powerful tools for medical data fusion, including the synthesis of the missing modalities when appropriate, augmenting limited datasets for rare conditions, creating counterfactual examples for model explanation, and generating realistic synthetic patients for simulation and training. Large Language Models promise to unlock information in unstructured clinical narratives by extracting structured information from physician notes, integrating textual data with structured Electronic Health Records, providing natural language explanations of fusion results, and enabling conversational AI for clinical decision support. Self Supervised Learning approaches may reduce dependence on labeled data by learning representations from unlabeled multimodal data, pre-training on large-scale public datasets, and fine-tuning on specific clinical tasks with limited labels. Federated learning enables multi-institutional collaboration without sharing patient data, exploring the models trained across hospitals while data remains local. It may also enable rare disease research through data aggregation. Future research directions may improve the estimation by using robust methods with uncertainty quantification. It is also interesting to explore the causal information theory by distinguishing correlation from causation in fusion. Another potential improvement may be applied with the use of dynamic information theory by tracking information flow over time. From a research point of view, a promising approach may be the use of Riemannian geometry of statistical manifolds.

VI. CONCLUSION

Information theory provides a rigorous mathematical framework that is transforming medical data fusion across the healthcare continuum. This comprehensive review has synthesized the foundational concepts, theoretical frameworks, clinical applications, and future directions of this rapidly evolving field. We have provided a systematic presentation of the main information-theoretical quantities with clear clinical interpretations and practical examples. The PID framework offers theoretical guidance on fundamental fusion questions: whether to fuse, how to fuse, and what to expect from fusion. Information-theoretic fusion has demonstrated impact across major medical domains: oncology, neurology, cardiology, and critical care. The future of IT medical data fusion lies at the intersection of these converging trends: theoretical advances, methodological innovations, clinical translation, and technological convergence. Information theory has fundamentally altered how we think about medical data fusion, moving from an ad-hoc combination of data sources to principled quantification of information content, redundancy, and synergy. The field has matured from theoretical foundations to practical applications that are already improving patient care through earlier detection, to more accurate diagnosis, and personalized treatment. Yet significant challenges remain. The parameter sensitivity of PID metrics, the gap between theoretical recommendations and empirical performance, and the substantial barriers to clinical translation remind us that this field is still in development. The most exciting advances may lie ahead, as emerging technologies like generative AI and large language models create new opportunities for IT analysis of increasingly complex, multimodal medical data. For clinicians, researchers, and developers working at this intersection, the message is clear: information theory provides powerful tools for understanding and optimizing medical data fusion, but these tools must be applied thoughtfully, validated rigorously, and interpreted in the context of real-world clinical constraints. When used appropriately, they can unlock the full potential of multimodal patient data to improve diagnostic accuracy, prognostic precision, and ultimately, improve the overall health of patients.

REFERENCES

[1] G I. Csiszar. I-Divergence Geometry of Probability Distributions and Minimization Problems. The Annals of Probability, 3(1):146–158, 1975. doi: 10.1214/aop/1176996454.

[2] J. Duan, J. Xiong, Y. Li, and W. Ding. Deep learning based multimodal biomedical data fusion: An overview and comparative review. Information Fusion, 2024. doi: 10.1016/j.inffus.2024.102536.

[3] S. Kullback and R. A. Leibler. On Information and Sufficiency. The Annals of Mathematical Statistics, 22(1):79–86, 1951. doi: 10.1214/aoms/ 1177729694.

[4] P. P. Liang, Y. Cheng, X. Fan, C. K. Ling, S. Nie, R. Chen, Z. Deng, N. Allen, R. Auerbach, F. Mah mood, R. Salakhutdinov, and L.-P. Morency. Quantifying & modeling multimodal interactions: An information decomposition framework, 2023.

[5] J. Massey. Causality, feedback and directed information. In Proceedings 1990 International Symposium on Information Theory and its Applications (ISITA). Waikiki, Hawaii, Nov. 1990.

[6] M. Salvi, H. W. Loh, S. Seoni, P. D. Barua, S. García, F. Molinari, and U. R. Acharya. Multi-modality approaches for medical support systems: A systematic review of the last decade. Information Fusion, 103, 2024. doi: 10.1016/j.inffus.2023.102134.

[7] T. Schreiber. Measuring information transfer. Phys. Rev. Lett., 85:461–464, Jul 2000. doi: 10.1103/ PhysRevLett.85.461.

[8] C. E. Shannon. A mathematical theory of communication. The Bell System Technical Journal, 27(3): 379–423, 1948.

[9] N. Tishby, F. C. Pereira, and W. Bialek. The in formation bottleneck method, 2000. URL https: //arxiv.org/abs/physics/0004057.

[10] P. L. Williams and R. D. Beer. Nonnegative decomposition of multivariate information. 2010. URL https://arxiv.org/abs/1004.2515.

[11] J. Wu, K. He, R. Mao, X. Shang, and E. Cambria. Harnessing the potential of multimodal EHR data: A comprehensive survey of clinical predictive modelling for intelligent healthcare. Inf. Fusion, 123, 2025.

[12] T. Zhang, R. Ding, K.-D. Luong, and W. Hsu. Evaluating an information theoretic approach for selecting multimodal data fusion methods. Journal of Biomedical Informatics, 167, 2025.

[13] Y.-D. Zhang, Z. Dong, S.-H. Wang, X. Yu, X. Yao, Q. Zhou, H. Hu, M. Li, C. Jiménez-Mesa, J. Ramirez, et al. Advances in multimodal data fusion in neuroimaging: Overview, challenges, and novel orientation. Information Fusion, 64:149–187, 2020. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955.

Artificial intelligence approaches for modeling nonlinear dynamical systems

Bregan Isufi[1], Alfred Daci[2]
PhD Candidate in Engineering Mathematics, Polytechnic University of Tirana[1]
Associate Professor at the Faculty of Mathematical Engineering and Physical Engineering, Polytechnic University of Tirana[2]
bregan.isufi@gmail.com

Abstract: *Nonlinear dynamical systems arise in numerous scientific and engineering domains, including physics, economics, biology, and control theory. Their complex behavior, sensitivity to initial conditions, and possible chaotic dynamics make accurate modeling and prediction challenging using traditional analytical approaches alone. In recent years, artificial intelligence (AI) techniques have demonstrated strong potential for modeling nonlinear and complex systems through data-driven methods. This paper explores artificial intelligence approaches for modeling nonlinear dynamical systems, focusing on the integration of machine learning techniques with classical mathematical modeling. We consider representative nonlinear systems and analyze how neural networks, regression models, and hybrid AI–mathematical frameworks can be used to approximate system behavior, predict future states, and capture hidden structures in time-series data. Special attention is given to systems exhibiting chaotic behavior, where small perturbations in initial conditions can lead to significant divergence in trajectories. The study presents numerical simulations and comparative analyses between traditional mathematical models and AI-based approaches. The results highlight the advantages of machine learning methods in capturing nonlinear patterns and improving predictive accuracy, especially when analytical solutions are difficult or unavailable. Additionally, we discuss the interpretability of AI models in the context of dynamical systems and outline potential applications in engineering, intelligent control, and data-driven system identification. The proposed framework contributes to the growing intersection between dynamical systems theory and artificial intelligence by demonstrating how AI tools can support the analysis and modeling of complex nonlinear phenomena. This work aims to provide a foundation for future research on hybrid mathematical–AI methods for understanding and predicting complex systems.*

Keywords: NONLINEAR DYNAMICAL SYSTEMS, ARTIFICIAL INTELLIGENCE, MACHINE LEARNING, CHAOTIC SYSTEMS, SYSTEM MODELING, HYBRID MODELING.

1. Introduction

Nonlinear dynamical systems play a fundamental role in modeling natural and engineered phenomena [1]. From fluid turbulence and population dynamics to financial markets and control systems, nonlinear equations describe processes where outputs are not proportional to inputs and where complex behaviors such as bifurcations and chaos may arise.

Classical analysis of nonlinear systems relies on differential equations, stability theory, Lyapunov methods, phase plane analysis, and bifurcation theory [2, 11]. However, analytical solutions are rarely available for real-world systems. Even numerical integration methods, while powerful, may struggle in the presence of strong nonlinearity, noise, or partially observed states.

In parallel, artificial intelligence (AI) and machine learning (ML) have developed into powerful tools for approximating highly nonlinear mappings directly from data [3]. Neural networks, regression models, and hybrid AI–physics models now offer alternative frameworks for modeling dynamical systems.

This paper investigates how AI techniques can model nonlinear dynamical systems, especially those exhibiting chaotic behavior. We compare classical mathematical models with AI-based approaches and propose a hybrid framework integrating both perspectives.

2. Mathematical Background on Nonlinear Dynamical Systems

In this section, we introduce the fundamental mathematical definitions of dynamical systems, which constitute the theoretical foundation of our study.

Definition 2.1: A dynamical system consists of a phase (state) space P and a family of transformations

$$\phi_t : P \to P$$

where the time parameter t belongs either to $\mathbb{R}$ (continuous time) or $\mathbb{Z}$ (discrete time).

The family $\{\phi_t\}_{t\in T}$ statisfies the following properties:

1. Identity property

$$\phi_0(x) = x, \quad \forall x \in P$$

2. Group (additivity) property

$$\phi_s(\phi_t(x)) = \phi_{s+t}(x), \quad \forall x \in P,\ \{s,t\} \in T$$

Thus, a dynamical system can be understood as a mathematical rule describing the evolution of the state $x \in P$ over time.

The function $\phi_t(x)$ is called the flow of the system when time is continuous, and the iteration operator when time is discrete.

Definition 2.2: A dynamical system is said to be time-continuous if the time parameter satisfies $t \in \mathbb{R}$. In this case, the system is typically described by a system of ordinary differential equations (ODEs):

$$\dot{x} = F(x)$$

where

1. $x(t) \in P \subset R^N$,
2. $F : P \to R^N$ is a vector field,
3. $\dot{x} = \dfrac{dx}{dt}$.

The formal solution is written as

$$x(t) = \phi_t(x_0)$$

where x_0 is called the initial condition.

Time-continuous systems describe processes evolving smoothly in time, such as:

- mechanical systems,
- electrical circuits,
- population dynamics,
- fluid motion.

Under suitable smoothness conditions on F, existence and uniqueness theorems guarantee that trajectories do not intersect. This implies determinism: each initial condition determines a unique trajectory.

An important structural result is the Poincaré–Bendixson theorem, which states that chaotic behavior cannot occur in two-dimensional continuous-time systems. Therefore, for chaotic dynamics to arise in continuous systems, the phase space dimension must satisfy $N \geq 3$.

Definition 2.3: A dynamical system is called time-discrete if the time parameter satisfies $n \in$. Such systems are described by an iteration rule:

$$x_{n+1} = M(x)$$

where $M : P \to P$ is a mapping.

The evolution after n steps is given by:

$$x_n = M^n(x_0)$$

Where M^n denotes the n-fold composition of M.

Time-discrete systems frequently arise from numerical discretization of differential equations, population models, and iterative algorithms [12, 13].

Unlike continuous systems, even one-dimensional non-invertible maps can exhibit chaotic dynamics.

A dynamical system is called linear if the evolution rule satisfies the superposition principle:

$$F(ax_1 + bx_2) = aF(x_1) + bF(x_2)$$

If this property does not hold, the system is called nonlinear**.**

Definition 2.4: A nonlinear dynamical system is a system of the form

$$\dot{x} = F(x) \quad or \quad x_{n+1} = M(x_n)$$

where the vector field F or the map M is nonlinear, i.e., it contains:

- products of variables (e.g., xy),
- higher-order powers (e.g., x^2),
- nonlinear functions (e.g., $\sin x, e^x$),
- or does not satisfy superposition.

Nonlinearity is the fundamental source of complex phenomena such as:

- bifurcations,
- multiple equilibria,
- limit cycles,
- chaotic attractors,
- sensitive dependence on initial conditions. [2,9]

Example 2.1: Logistic Map (Discrete)

$$x_{n+1} = rx_n(1 - x_n)$$

which is nonlinear due to the quadratic term x_n^2 For certain parameter values r the system exhibits chaotic behavior [2,9,12].

Example 2.2: Lorenz System (Continuous)

$$\begin{cases} \dot{x} = \delta(y - x) \\ \dot{y} = x(\rho - z) - y \\ \dot{z} = xy - \beta z \end{cases}$$

The nonlinear interaction term xy generates chaotic attractors when parameters are appropriately chosen [4].

Example 2.3: Nonlinear Pendulum

$$\ddot{\theta} + \vartheta\dot{\theta} + \sin\theta = 0$$

where the sine term introduces nonlinearity.

Nonlinear dynamical systems form the mathematical framework for understanding complex behavior in natural and engineered systems. [2,10]. Their intrinsic nonlinearity allows the emergence of instability, bifurcation phenomena, and chaotic dynamics, which cannot occur in purely linear systems.

These theoretical foundations are essential before introducing data-driven and artificial intelligence approaches for modeling such systems.

3. Artificial Intelligence Approaches

The rapid development of artificial intelligence and machine learning has introduced powerful data-driven techniques for modeling nonlinear dynamical systems. Classical approaches rely on explicitly known governing equations, typically expressed as systems of differential equations. However, in many real-world problems the governing equations are partially known or completely unknown. In such situations, artificial intelligence provides an alternative framework capable of learning system dynamics directly from data [3].

Machine learning models approximate the evolution operator of a dynamical system. For a discrete dynamical system

$$x_{t+1} = F(x_t)$$

the objective of a learning algorithm is to construct an approximation

$$\hat{F}_\theta(x_t)$$

parameterized by θ, such that

$$\hat{x}_{t+1} \approx F(x_t)$$

Thus, artificial intelligence techniques can be interpreted as data-driven approximations of the dynamical evolution operator**.**

Recent research has demonstrated that machine learning methods are capable of reproducing complex nonlinear and chaotic dynamics with high accuracy, even when the underlying governing equations are unknown.

3.1. Neural Network Models for Dynamical Systems

Artificial neural networks represent one of the most widely used tools for modeling nonlinear dynamical systems due to their universal approximation capabilities [3].

Feedforward neural networks approximate nonlinear functions of the form

$$\hat{x}_{t+1} = NN(x_t;\theta)$$

where NN represents a multilayer neural network.

These models have been successfully applied to the prediction of chaotic time series generated by nonlinear systems such as the logistic map, Lorenz attractor, and intermittency maps. Studies have shown that neural networks such as multilayer perceptrons (MLP), residual networks, and long short-term memory networks (LSTM) can significantly reduce prediction errors compared with classical regression methods.

In particular, hierarchical neural architectures combining several specialized networks have been shown to improve prediction accuracy in chaotic time series modeling tasks.

3.2. Recurrent Neural Networks and Temporal Modeling

While feedforward neural networks approximate static nonlinear functions, recurrent neural networks (RNNs) are specifically designed to capture temporal dependencies in dynamical systems.

An RNN model evolves according to

$$h_{t+1} = \sigma(W_h h_t + W_x x_t)$$

$$\hat{x}_{t+1} = W_0 h_{t+1}$$

Here:

- h_t represents the hidden state
- $W_h, W_x, W_0,$ are weight matrices
- σ is a nonlinear activation function.

Long Short-Term Memory (LSTM) networks and gated recurrent units (GRU) are improved recurrent architectures capable of learning long-term temporal dependencies [5].

Applications include:

- climate system modeling
- fluid turbulence
- nonlinear oscillators
- chaotic attractors

Recent studies have demonstrated that LSTM networks can successfully learn the dynamics of coupled nonlinear systems and predict chaotic trajectories without explicit knowledge of the governing differential equations.

However, classical RNN architectures often suffer from instability when predicting chaotic systems over long time horizons due to error accumulation.

3.3. Reservoir Computing and Chaotic System Prediction

Reservoir computing is another powerful machine learning framework designed specifically for modeling dynamical systems [6].

Reservoir computing networks consist of:

1. a fixed high – dimensional dynamical reservoir
2. a trained linear readout layer

The system evolves according to

$$r_{t+1} = \tanh(W_r r_t + W_{in} x_t)$$

$$\hat{x}_{t+1} = W_{out} r_{t+1}$$

Unlike traditional neural networks, only the output layer is trained, which greatly simplifies the training process.

Reservoir computing models have demonstrated remarkable performance in predicting chaotic dynamical systems, including high-dimensional models such as the Lorenz-96 system. In many numerical experiments, reservoir computing has been shown to outperform traditional neural networks in short-term trajectory prediction and statistical reconstruction of chaotic attractors.

3.4. Koopman Operator and Data-Driven Linearization

A modern approach to modeling nonlinear dynamical systems is based on Koopman operator theory**.**

The Koopman operator K acts on observable functions $g(x)$ instead of the state variables themselves:

$$K(g(x)) = g(F(x))$$

Although the underlying system is nonlinear, the Koopman operator is linear but infinite-dimensional.

Data-driven techniques such as:

- Dynamic Mode Decomposition (DMD)
- Extended Dynamic Mode Decomposition (EDMD)

allow finite-dimensional approximations of the Koopman operator.

This approach transforms nonlinear dynamics into approximately linear systems in a lifted feature space. Numerical studies have demonstrated that Koopman-based models can accurately reproduce the statistical properties and temporal evolution of chaotic systems while remaining computationally efficient.

Koopman-based methods therefore provide a bridge between classical dynamical systems theory and modern machine learning [7].

3.5. Hybrid Physics–AI Models

A promising direction in recent research is the development of hybrid models combining physics-based equations with machine learning components**.**

In these models:

$$\dot{x} = F_{physics}(x) + F_{ML}(x)$$

where

- $F_{physics}(x)$ represents known physical laws
- $F_{ML}(x)$ approximates unknown or unresolved dynamics.

Hybrid neural-physics models have been successfully applied to systems such as the Lorenz equations, where neural networks are used to model missing physical processes [8]. These models significantly improve predictive accuracy when only partial knowledge of the governing equations is available.

3.6. Performance and Limitations

Despite their success, AI-based approaches face several challenges when applied to nonlinear dynamical systems:

1. ***Sensitivity to initial conditions***
   Chaotic systems amplify small prediction errors over time.
2. ***Long-term prediction instability***
   Neural networks often provide accurate short-term predictions but fail to reproduce long-term trajectories.
3. ***Interpretability issues***
   Deep learning models are often difficult to interpret in terms of physical laws.

Recent research attempts to address these issues by incorporating dynamical systems theory into machine learning architectures, including physics-informed neural networks and Koopman-based representations.

Artificial intelligence provides powerful tools for modeling nonlinear dynamical systems, particularly when governing equations are unknown or difficult to solve analytically. Neural networks, recurrent architectures, reservoir computing, and Koopman-based frameworks have all demonstrated the ability to approximate complex dynamical behaviors and chaotic systems.

The integration of machine learning with classical dynamical systems theory represents a promising research direction, enabling the development of hybrid models capable of combining data-driven learning with mathematical structure.

The artificial intelligence methods discussed above differ in their modeling philosophy, predictive capability, and interpretability when applied to nonlinear dynamical systems. In order to highlight the main characteristics of these approaches, Table 1 provides a comparative overview of the principal AI-based techniques used for modeling nonlinear dynamical systems and their relevance to the present study.

Table 1: *Comperative overview of AI-based techniques*

Method	Modeling Principle	Strengths in Nonlinear Dynamical Systems	Limitations	Relevance to This Study
Feedforward Neural Networks (MLP)	Approximate nonlinear mappings of the form $(x_{t+1} = F(x_t))$ using multilayer networks	Universal approximation capability; simple architecture; effective for nonlinear function approximation	Limited capability for capturing long-term temporal dependencies	Used to approximate the nonlinear mapping of the logistic map
Recurrent Neural Networks (RNN) / LSTM	Learn temporal dependencies through recurrent hidden states	Suitable for modeling sequential and time-series data; capable of capturing system dynamics	Training instability and error accumulation in chaotic systems	Applied to model Lorenz system trajectories
Reservoir Computing	Uses a fixed high-dimensional dynamical reservoir with a trained linear readout layer	Efficient training; strong performance for short-term prediction of chaotic systems	Limited interpretability and sensitivity to reservoir design	Discussed as an effective method for chaotic system forecasting
Koopman Operator Methods	Represent nonlinear dynamics through linear operators in a lifted feature space	Theoretically interpretable; bridges dynamical systems theory and data-driven modeling	Requires feature engineering and approximation of infinite-dimensional operators	Presented as a data-driven framework connecting AI and dynamical systems theory
Hybrid Physics–AI Models	Combine governing physical equations with machine learning components	Integrate prior physical knowledge with data-driven learning; improved prediction accuracy	Increased computational complexity and model design challenges	Proposed as a future direction for modeling complex nonlinear systems

4. Numerical Simulations

To evaluate the effectiveness of artificial intelligence approaches in modeling nonlinear dynamical systems, numerical experiments were conducted on two representative case studies:

1. the **logistic map**, representing a nonlinear discrete dynamical system,
2. the **Lorenz system**, representing a continuous chaotic dynamical system.

These examples are widely used benchmarks in nonlinear dynamics because they exhibit sensitive dependence on initial conditions and complex attractor structures [2, 4, 9].

These examples are widely used benchmarks in nonlinear dynamics because they exhibit sensitive dependence on initial conditions and complex attractor structures [2,4,9].

The objective of the numerical experiments is to investigate the ability of machine learning models to approximate the underlying evolution operator of these systems and to compare the predictive performance with classical numerical integration.

4.1. Case study 1

The logistic map is one of the simplest nonlinear dynamical systems exhibiting chaotic behavior and is defined as

$$x_{n+1} = rx_n(1-x_n),$$

where

$$0 \le x_n \le 1, \quad 0 \le r \le 4.$$

The parameter r controls the qualitative behavior of the system.

Different dynamical regimes occur depending on the value of r :

Parameter Range	Dynamical Behavior
$0 < r < 1$	Convergence to zero
$1 < r < 3$	Stable fixed point
$3 < r < 3.57$	Period doubling
$r > 3.57$	Chaos

For the simulations in this study we choose $r = 3.57$ which lies deep inside the chaotic regime.

4.1.1. Dataset Generation

The dataset used for training the AI model is generated through iterative simulation of the logistic map.

Initial condition:

$$x_0 = 0.2$$

The system is iterated for

$$N = 5000$$

time steps.

The resulting time series

$$\{x_0, x_1, ..., x_N\}$$

is then split into

- training data (70%)
- validation data (15%)
- testing data (15%)

The learning task consists of approximating the mapping

$$x_{n+1} = F(x_n)$$

using a neural network model.

4.1.2 Neural Network Architecture

A feedforward neural network is used with the following structure:

- Input layer: 1 neuron (x_n)
- Hidden layer 1: 32 neurons
- Hidden layer 2: 32 neurons
- Activation function: ReLU
- Output layer: 1 neuron (x_{n+1})

The loss function used for training is the mean squared error (MSE)

$$L = \frac{1}{N}\sum_{i=1}^{N}(x_{i+1} - \hat{x}_{i+1})^2.$$

Training is performed using the Adam optimization algorithm**.**

4.1.3 Results

The trained neural network successfully approximates the nonlinear mapping of the logistic system. Short-term prediction accuracy is very high, with prediction error remaining below 10^{-4} for the first 50 prediction steps.

However, due to the chaotic nature of the system, prediction errors grow exponentially over time. This behavior is expected because chaotic systems exhibit positive Lyapunov exponents:

$$\sigma x(t) = \sigma x(0)e^{\lambda t}.$$

Despite this limitation, the neural network accurately reconstructs the statistical properties of the attractor and reproduces the invariant distribution of the logistic map.

4.2 Case study 2

The Lorenz system is a classical example of a nonlinear continuous dynamical system exhibiting deterministic chaos.

The system is defined by

$$\begin{cases} \dot{x} = \delta(y - x) \\ \dot{y} = x(\rho - z) - y \\ \dot{z} = xy - \beta z \end{cases}$$

For the classical parameter values

$$\delta = 10, \ \rho = 28, \ \beta = \frac{8}{3}$$

the system exhibits the famous Lorenz chaotic attractor.

4.2.1 Numerical Integration

The Lorenz equations are solved numerically using the fourth-order Runge–Kutta method (RK4) with step size

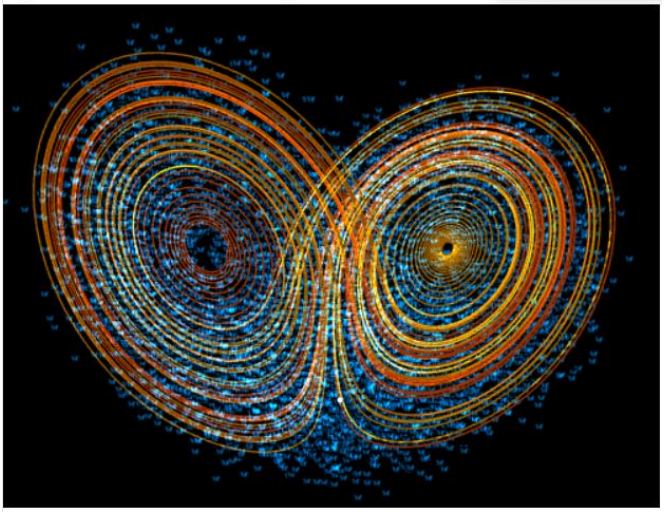
$$\Delta t = 0.01.$$

Initial condition:

$$(x_0, y_0, z_0) = (1, 1, 1).$$

A trajectory of 10000 time steps is generated. This dataset provides the training data for the AI model.

Figure 1: **Phase Portrait of the Lorenz System**



4.2.2 Machine Learning Model

To capture the temporal structure of the Lorenz system, a Long Short-Term Memory (LSTM) neural network is used.

Model structure:

- Input dimension: 3 variables (x, y, z).
- LSTM layer: 64 units
- Dense output layer: 3 neurons

The network predicts $(x_{t+1}, y_{t+1}, z_{t+1})$ given the previous system state.

The loss function is again the mean squared error:

$$L = \frac{1}{N} \sum \left\| x_{true} - x_{pred} \right\|^2.$$

4.2.3 Results

The trained LSTM model demonstrates strong predictive capability for short-term system evolution. The model accurately predicts the trajectory for approximately 5-7 Lyapunov time units before divergence becomes significant. More importantly, the AI model successfully reconstructs the geometry of the Lorenz attractor. The predicted trajectories preserve the characteristic butterfly-shaped structure of the attractor, indicating that the learned model captures the underlying nonlinear dynamics. Additionally, the probability distribution of states generated by the AI model closely matches that of the true Lorenz system.

4.3 Comparative Discussion

The numerical experiments highlight several important observations:

1. Artificial intelligence models can accurately learn the nonlinear evolution operators of dynamical systems from data.
2. Short-term predictions are highly accurate even for strongly chaotic systems.
3. Long-term prediction remains challenging due to exponential error growth.
4. AI models can successfully reproduce invariant statistical properties of chaotic attractors.

These results confirm that machine learning methods are powerful tools for modeling nonlinear dynamical systems, particularly when analytical models are unavailable or incomplete [3, 6].

4.4 Implications for Data-Driven Dynamical Modeling

The case studies demonstrate that artificial intelligence methods can approximate complex nonlinear dynamical behaviors with high accuracy.

This capability opens new possibilities for

- system identification,
- prediction of chaotic systems,
- modeling complex physical processes.

In particular, hybrid approaches combining machine learning with classical dynamical systems theory offer a promising direction for future research.

5. Conclusions

In this study, we investigated the application of artificial intelligence techniques for modeling nonlinear dynamical systems. The research combined the mathematical foundations of dynamical systems theory with modern machine learning methods in order to explore the capability of data-driven approaches to approximate complex nonlinear dynamics.

First, the theoretical framework of dynamical systems was presented, including the definitions of continuous and discrete dynamical systems and the role of nonlinearity in generating

complex behaviors such as bifurcations and chaos. These concepts form the mathematical basis for understanding the dynamics of many natural and engineered systems.

Next, several artificial intelligence approaches were discussed, including feedforward neural networks, recurrent neural networks, reservoir computing, and Koopman-based operator learning methods. These techniques allow the approximation of nonlinear evolution operators directly from data and provide powerful alternatives when the governing equations of a system are unknown or difficult to solve analytically.

To evaluate the effectiveness of these approaches, numerical simulations were conducted on two representative case studies: the logistic map and the Lorenz system. The results demonstrate that machine learning models can successfully learn the nonlinear evolution rules of dynamical systems and provide highly accurate short-term predictions. In addition, the trained models were able to reconstruct important statistical properties of chaotic attractors, indicating that they capture key structural features of the underlying dynamics.

However, the experiments also highlight inherent limitations when modeling chaotic systems. Due to the sensitive dependence on initial conditions, small prediction errors grow exponentially over time, which restricts long-term prediction accuracy. This phenomenon is consistent with the theoretical properties of chaotic systems characterized by positive Lyapunov exponents.

Despite these limitations, artificial intelligence provides a promising framework for the modeling and analysis of nonlinear dynamical systems. In particular, hybrid approaches that combine mathematical structure with machine learning techniques appear to be especially effective. Such methods may enable improved modeling, prediction, and control of complex systems in fields such as physics, engineering, climate science, and economics.

Future research may focus on the development of physics-informed machine learning models, Koopman operator learning techniques, and neural operator frameworks capable of capturing the underlying structure of nonlinear dynamical systems while preserving interpretability and stability.

Overall, the integration of dynamical systems theory and artificial intelligence represents an important and rapidly evolving research direction with significant potential for advancing the understanding of complex nonlinear phenomena. Several studies have also investigated dynamical system modeling, bifurcation analysis, and discrete dynamical systems in applied contexts [10–13].

References

[1] J. Guckenheimer, P. Holmes,Nonlinear Oscillations, Dynamical Systems and Bifurcations of Vector Fields, Springer, 2013.

[2] S. H. Strogatz, Nonlinear Dynamics and Chaos, CRC Press, 2018.

[3] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, MIT Press, 2016.

[4] E. N. Lorenz, Deterministic nonperiodic flow, Journal of the Atmospheric Sciences, 1963.

[5] P. R. Vlachas et al., Backpropagation algorithms and reservoir computing in recurrent neural networks for forecasting complex dynamics, Neural Networks, 2018.

[6] J. Pathak et al., Model-free prediction of large spatiotemporally chaotic systems from data, Physical Review Letters, 2018.

[7] S. L. Brunton, J. L. Proctor, J. N. Kutz, Discovering governing equations from data by sparse identification of nonlinear dynamical systems, PNAS, 2016.

[8] M. Raissi, P. Perdikaris, G. E. Karniadakis, Physics-informed neural networks, Journal of Computational Physics, 2019.

[9] E. Ott, Chaos in Dynamical Systems, Cambridge University Press, 2002.

[10] A. Daci, A. Spaho, "Bifurcation in a dynamical system: harvest models. "In International Conference on Research and Engineering-Challenges Toward the Future", Albania, May, pp. 24-25, 2013.

[11] A. Daci, Dynamical Systems: Phase Plane Analysis – Linearization Method, Journal of Multidisciplinary Engineering Science and Technology (JMEST), Vol. 3 Issue 8, 2016.

[12] A. Daci, S. Tola, "Discrete Dynamical Systems, Exponential Growth", First International Conference, Research, Application and Educational Methods. Book of Proceedings. Vol. 45., 2019.

[13] A. Daci, S. Tola, R. DATJA; "Laplace Transform in Difference Equations and Differential – Difference Equations", Journal of Multidisciplinary Engineering Science Studies (JMESS), ISSN: 2458-925X, Vol. 8 Issue 1, 2022.

Imagined pedestrian encounters: analyzing world-model rollouts in a reinforcement learning driving agent

Stefan Zlatinov.[1], Ilija Mizhimakoski.[1], Gorjan Nadzinski. [1], Mile Stankovski.[1]
Faculty of Electrical Engineering and Information Technologies – Ss. Cyril and Methodius University in Skopje, Skopje, North Macedonia[1]
zlatinov@feit.ukim.edu.mk

Abstract: *World-model reinforcement learning agents plan by imagining future trajectories through a learned latent dynamics model. In safety-critical driving tasks, the quality of these imagined rollouts — particularly their representation of vulnerable road users — directly determines whether the agent can anticipate and avoid dangerous situations. We analyze the imagined rollouts of a DreamerV3 agent trained to navigate urban environments in CARLA using semantic segmentation and depth observations. By decoding the agent's latent imagination into observation space, we qualitatively and quantitatively examine how pedestrians are represented in imagined futures, comparing episodes that result in collision against those where the agent successfully avoids. We investigate whether the world model accurately predicts pedestrian presence and motion in its imagined horizon, and whether reconstruction fidelity of pedestrian regions correlates with avoidance outcomes. Our analysis provides insight into the internal representations that underlie emergent safety behaviors in model-based driving agents and highlights limitations of finite imagination horizons for pedestrian safety.*

Keywords: WORLD MODEL, REINFORCEMENT LEARNING, AUTONOMOUS DRIVING, DREAMERV3, PEDESTRIAN SAFETY, IMAGINATION ROLLOUT, LATENT DYNAMICS, CARLA, SAFETY-CRITICAL AI

1. Introduction

Model-based reinforcement learning (MBRL) agents learn a predictive model of the environment and use it to plan by simulating future trajectories in a learned latent space [1]. In autonomous driving, this planning-by-imagination paradigm is appealing because it allows the agent to evaluate potential actions before executing them [2; 3]. DreamerV3 [4] is a recent world-model agent that has demonstrated strong performance across diverse domains, including simulated driving in CARLA [5]. Unlike model-free approaches that map observations directly to actions [6; 7], world-model agents rely on the quality of imagined rollouts to evaluate candidate action sequences. This dependence raises a natural concern: if the learned dynamics model fails to preserve a hazard in its imagined future, the agent may plan as if the road ahead is clear.

A central question for safety-critical deployment is therefore whether the learned world model faithfully represents vulnerable road users (in particular pedestrians) in its imagined futures. If the model fails to maintain an accurate representation of a pedestrian blocking the lane during imagination, the agent may select actions that lead to collision even when avoidance is physically possible. Pedestrian behavior prediction has been studied extensively in the context of conventional perception pipelines [8; 9], and intent-aware planning has been explored in model-predictive settings [10; 11]. However, relatively little work has examined how learned world models in MBRL agents internally represent pedestrians during latent imagination, the setting where planning decisions are actually made.

Recent work has applied world-model RL to driving in CARLA with varying observation modalities. Think2Drive [12] uses a latent world model for quasi-realistic autonomous driving, while CarDreamer [13] provides an open-source learning platform for world-model-based driving. A recent survey [14] covers the broader landscape of world models for autonomous driving. These works, along with broader RL-for-driving surveys [15; 16; 17], focus primarily on driving performance metrics (success rate, infraction count, route completion) rather than on the internal quality of imagined rollouts with respect to specific road users. On the evaluation side, adversarial testing frameworks such as ANTI-CARLA [18] stress-test driving agents but do not analyze the imagination content of world-model agents. Safe RL approaches have addressed uncertainty-aware planning [19] and latent-space trajectory optimization [20], yet the question of what the world model actually imagines about a pedestrian near the conflict point remains largely unexamined. Our work addresses this gap by directly decoding imagined latent states into observation space and measuring how well the world model preserves pedestrian representations during planning.

In this work we study the imagined rollouts of a DreamerV3 agent in a controlled urban scenario where a pedestrian walks to the center of the lane and stands still, blocking the ego vehicle's path. We adopt an event-centered approach: we capture imagination at the moment the stationary pedestrian is visible ahead and examine the imagined future only over the short conflict-resolution period that follows. We decode the agent's latent imagination into observation space and compare imagined segmentation frames against ground truth, using a strict localized fidelity metric. Our analysis covers 565 pedestrian-encounter episodes and reveals that (1) the world model retains only weak pedestrian support under a strict exact-class metric, and (2) this weakness appears in both collision and avoidance episodes, suggesting that incorrect imagined lane clearance, rather than simple pedestrian disappearance, is the more informative failure mode.

2. Method

Agent and Environment

We use a DreamerV3 agent [4] trained in the CARLA simulator [5]. DreamerV3 learns a recurrent state-space model (RSSM) that encodes observations into a compact latent state and predicts future latent states conditioned on actions. An actor-critic algorithm optimizes the policy entirely within the learned latent space, using imagined trajectories rather than real environment interactions for policy improvement [21].

Perception. The agent receives two 128 x 128 image channels at each step: a single-channel semantic segmentation map where class labels are ordered by collision risk (pedestrians occupy the highest intensity, value 255; drivable road occupies the lowest), and a normalized inverse-depth image that concentrates representational precision in the near field. These visual inputs are complemented by a low-dimensional numerical vector containing the vehicle's speed, longitudinal acceleration, yaw rate, lateral and heading deviations from the planned route, the previous steering and throttle/brake command, and a sequence of ten ego-frame waypoints spaced 3 m apart along the route (30 m look-ahead). A shared convolutional encoder processes the image pair while a separate multilayer perceptron encodes the numerical channels; both streams are fused into the RSSM latent state.

Control outputs. The policy produces two continuous signals per step: a steering command and a signed longitudinal command, both in [-1, 1]. Positive longitudinal values map to throttle and negative values to braking. Actions are repeated for two consecutive simulation ticks (each 0.05 s), yielding an effective control frequency of 10 Hz.

Training reward. The agent was trained on 200 m routes in Town02 for 10 million environment steps. The reward signal consists of a dense route-progress component (virtual gates placed every 2 m along the route, each yielding a reward proportional to lateral proximity to the route center), combined with a quadratic overspeed penalty above 30 km/h and a set of terminal bonuses and penalties for reaching the goal, timing out, leaving the route, or colliding with a vehicle or pedestrian. Crucially, pedestrian safety is supervised only through a large terminal penalty upon collision, scaled quadratically with impact speed. No distance-based proximity costs, time-to-collision shaping, or explicit braking incentives are included; any anticipatory avoidance behavior is therefore emergent rather than directly rewarded. During training the environment contained non-adversarial crossing pedestrians but no deliberately adversarial scenarios; the protester archetype used in this study was never encountered during learning.

Scenario Design

Each episode uses a straight 60 m route in a CARLA urban map. A jaywalking pedestrian ("protester" archetype) is spawned on the sidewalk at approximately 45 m from the route start. When the ego vehicle approaches, the pedestrian walks from the sidewalk to a blocking point near the center of the ego's lane at 2.0 m/s, then stops and remains stationary for up to 10 s. The pedestrian does not react to the approaching vehicle and does not yield. By the time the agent's imagination is captured, the pedestrian has typically reached the blocking point and is standing still in the lane. The agent therefore faces a stationary obstacle that must be avoided by braking or steering. Because the pedestrian is not visible at route start, early episode time is irrelevant; the analysis begins only when the pedestrian enters the conflict zone. Out of 640 recorded episodes, 75 had no pedestrian encounter (the pedestrian was never visible). Of the remaining 565 encounter episodes, 475 resulted in successful avoidance (the agent slowed or steered around the pedestrian) and 90 resulted in collision.

Imagination Capture

At each episode step where the pedestrian is visible in the segmentation observation, the agent's current latent state is used to produce a 32-step imagined rollout. These imagined latent states are decoded back into observation space by the learned decoder, producing predicted segmentation and depth frames that can be compared against ground truth. We select one capture per episode, specifically the one whose actual pedestrian mean depth is closest to 10\,m, to ensure a consistent comparison distance across episodes. At this distance the pedestrian is typically already stationary at the blocking point.

Event-Centered Analysis

We restrict analysis to the conflict-resolution period: from the capture step until the episode terminates (collision) or the pedestrian leaves the camera field of view (avoidance). This yields a right-censored future for each capture. We do not analyze the full nominal 32-step imagination horizon, as most episodes terminate or resolve the conflict well before that point.

Strict Local Overlap Metric

For each valid future step k after capture, we define the ground-truth pedestrian mask G_k as the set of pixels with exact semantic class value 255. Let $K = |G_k|$ denote the number of pixels in this mask. We apply no dilation, so the support region is exactly the ground-truth mask. Within this region, we threshold imagined segmentation values at 250 and retain up to the top-K surviving pixels, forming the predicted set P_k. The strict local overlap is then the Intersection over Union (IoU) between P_k and G_k:

$$IoU_k = \frac{P_k \cap G_k}{P_k \cup G_k} \quad (1)$$

This metric is intentionally strict: it measures whether the imagined segmentation reproduces the pedestrian at the correct location with the correct class value.

3. Results

Qualitative Case Study

Fig.1 shows decoded segmentation frames from one representative collision episode (capture at 11.8 s) and one representative avoidance episode (capture at 10.8 s). In the actual future, the pedestrian (highlighted in green) remains present in the lane throughout the conflict period: it is stationary at its blocking point and grows larger as the ego vehicle approaches. In the imagined future, the pedestrian representation is faint and degrades rapidly. By k=6 (0.6 s into imagination), the pedestrian signal in the imagined frames is substantially weaker than in the corresponding actual frames. In several imagined frames the road region where the pedestrian should be standing is instead predicted as empty lane. The world model effectively imagines the obstacle clearing when in reality the pedestrian is still there. This pattern (the world model imagining premature lane clearance rather than maintaining the stationary pedestrian) is the central qualitative finding and is observed across multiple episodes beyond the two shown here.

Strict Local Fidelity Over Horizon

Fig.2 shows the mean strict local overlap as a function of future time after capture, separately for collision and avoidance episodes. Both curves start near 0.65-0.70 at k=0 (the capture step itself, where the imagined state is initialized from the encoder) and decline to approximately 0.25-0.28 within the first second. After 1.0 s, both curves continue a gradual decline but remain above zero, indicating that the pedestrian signal does not disappear entirely. The collision and avoidance curves overlap substantially throughout the observed horizon, indicating that strict pedestrian fidelity does not cleanly separate the two outcomes. Shaded bands indicate the standard error of the mean. The lower panel shows the number of valid episodes contributing to each point; collision episodes are censored earlier due to episode termination upon impact. Each point on the curve averages only over episodes that are still valid at that future step, so the curves are not affected by survivorship bias from early terminations.

Episode-Horizon Heatmap

Fig.3 displays the per-episode strict local overlap over the first 1.0 s after capture. Rows represent individual episodes sorted by mean early overlap; columns represent future time steps at 0.1 s intervals. The heatmap reveals substantial heterogeneity: some episodes maintain moderate-to-high fidelity across the early horizon, while others show near-zero overlap from the first imagined step. Crucially, high-fidelity and low-fidelity episodes appear in both outcome groups, further supporting the observation that strict fidelity alone does not predict whether the agent avoids or collides. The grey cells in the heatmap indicate time steps where the pedestrian was no longer visible in the ground-truth frame, so no overlap could be computed.

Summary Table

Table 1 summarizes the imagination analysis across all 565 encounter episodes. The mean early strict overlap (first 1.0 s) is 0.40 overall, with similar values for avoidance (0.40) and collision (0.37) subgroups. The median valid horizon is approximately 2.2–2.4 s for both groups. On average, fewer than half of the ground-truth pedestrian pixels are reproduced by the imagination under the strict metric (mean predicted pixels: 6.3 out of 17.0 ground-truth pixels in the early window).

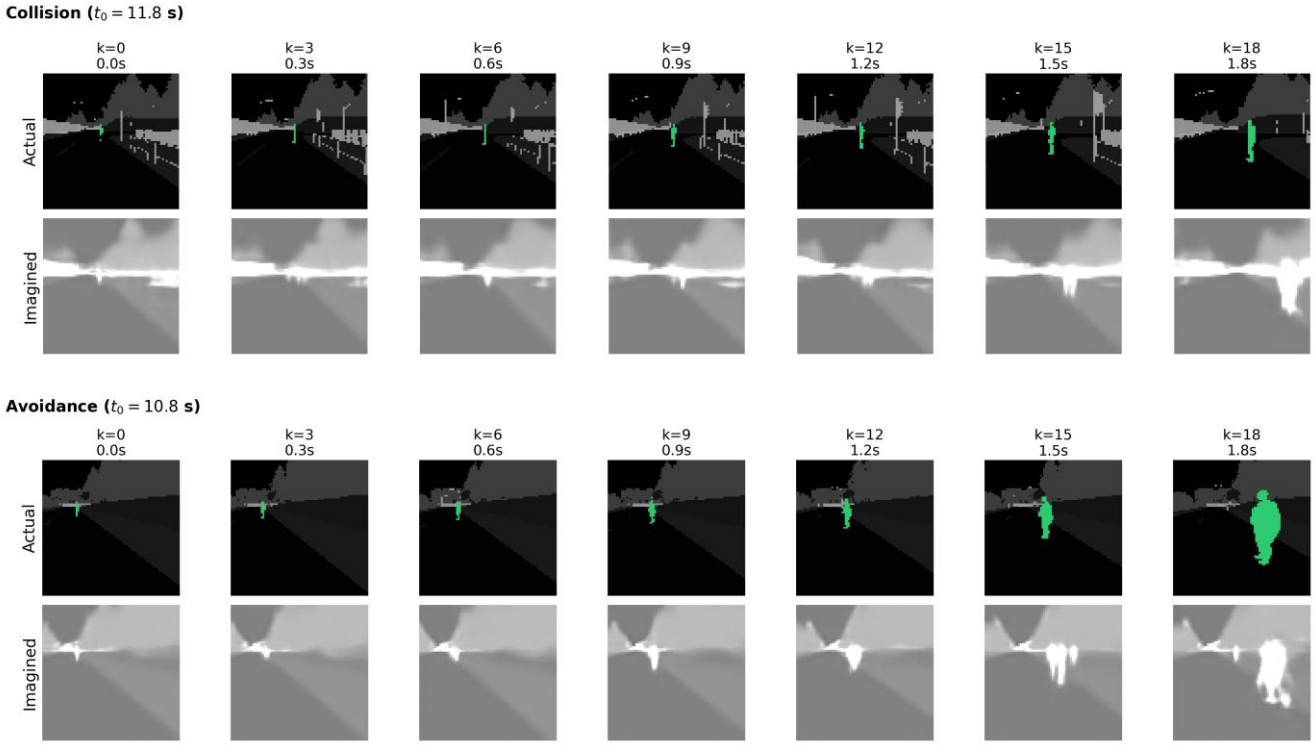


Fig. 1 *Actual versus imagined segmentation in a collision episode (top two rows) and an avoidance episode (bottom two rows). The pedestrian (green) remains stationary in the lane in actual frames. In imagined frames, the pedestrian signal is weak and fades rapidly, as if the lane were clearing.*

Table 1: *Protester mid-crossing imagination summary. One capture is selected per episode near 10 m depth; no-encounter episodes are excluded. Early overlap averages the first 1.0 s after capture (imagination threshold 250).*

	N	Med. Hor. [s]	Early IoU	Full IoU	GT px	Pred px
All	565	2.30	0.40	0.32	17.0	6.3
Avoidance	475	2.20	0.40	0.33	17.3	6.5
Collision	90	2.40	0.37	0.28	15.4	5.5

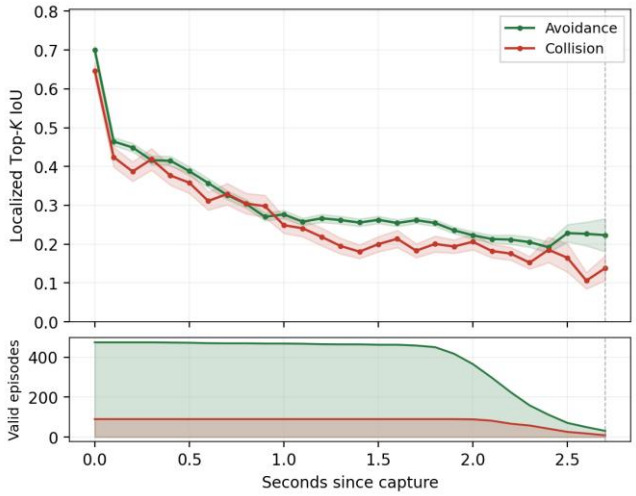


Fig. 2 *Strict local overlap versus future time after capture. Top: mean localized Top-K IoU for collision and avoidance episodes (shaded: 1 standard error of the mean). Bottom: number of valid episodes at each time step. Both outcome groups show similar fidelity decay; collision episodes are censored earlier.*

4. Discussion

The results show that the DreamerV3 world model retains some pedestrian signal in its imagined rollouts: the pedestrian does not vanish entirely at the capture step. However, under a strict exact-class metric with no dilation, this signal is weak: mean early strict overlap is at or below 0.40 in all subgroups, and fewer than half of the ground-truth pedestrian pixels survive the thresholded Top-K rule in imagination.

More importantly, the strict fidelity values for collision and avoidance episodes are similar (0.37 vs. 0.40). This means that pixel-level pedestrian reconstruction quality does not by itself explain why the agent avoids in some episodes and collides in

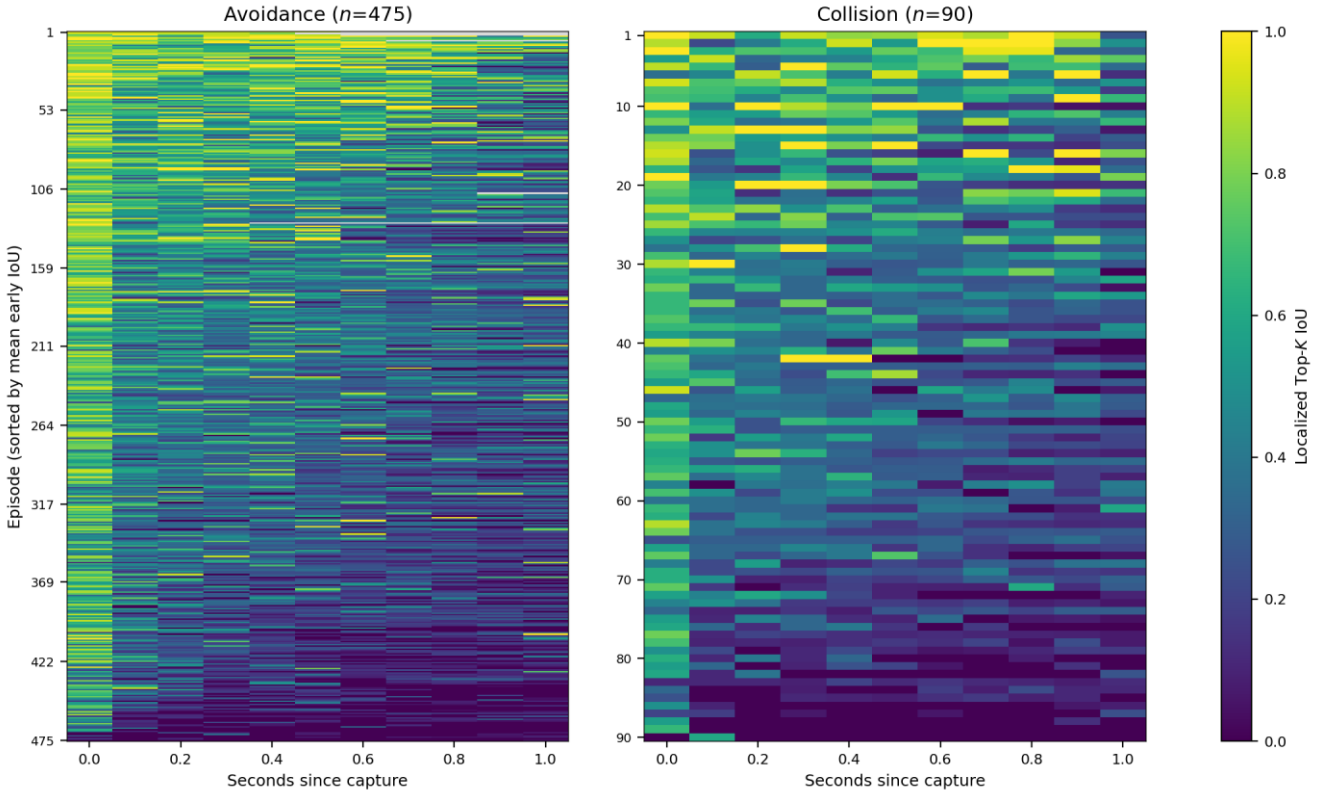


Fig. 3 *Per-episode strict local overlap (IoU) of imagined vs.\ actual pedestrian segmentation over the first 1.0\,s after capture. Each row is one episode; each column is one time step (0.1\,s intervals). Color encodes IoU from 0 (dark, world model misses the pedestrian entirely) to 1 (bright, world model reproduces the pedestrian pixel-for-pixel). Grey cells indicate steps where the pedestrian was no longer visible in ground truth, so no comparison is possible. Left panel: 475 avoidance episodes. Right panel: 90 collision episodes. Within each panel, episodes are sorted top-to-bottom by descending mean early IoU, so the best-imagining episodes appear at the top. High-fidelity and low-fidelity episodes are present in both outcome groups, showing that imagination quality alone does not predict whether the agent avoids or collides.*

others. The qualitative analysis (Fig.1) suggests a complementary explanation: the world model often imagines premature lane clearance. Although the pedestrian is stationary and should persist in the lane, the model's imagined future shows the pedestrian signal fading or disappearing, as if the obstacle were no longer there. Such incorrect predictions could lead the agent to select actions that do not account for the pedestrian's continued presence.

These findings are consistent with known limitations of learned world models in representing rare or out-of-distribution entities [22; 19]. The pedestrian occupies a small number of pixels and exhibits motion patterns that are uncommon in the training distribution, making faithful long-horizon prediction difficult. Our event-centered analysis complements full-episode evaluations by focusing on the critical conflict-resolution window.

From a practical perspective, these observations suggest that improving pedestrian safety in world-model agents may require more than increasing reconstruction fidelity. If the model can partially reconstruct the pedestrian at the capture step but imagines it disappearing shortly after, then the policy will plan around an incorrect future (one in which the lane is clear) regardless of initial pixel quality. Several directions could address this failure mode: (1) an *object-persistence prior* that biases latent dynamics toward maintaining detected static obstacles across the imagination horizon unless an explicit removal event is predicted; (2) *targeted data augmentation* during world-model training, such as oversampling pedestrian-blocking episodes, to better represent rare but safety-critical scenarios; (3) a *hazard-aware prediction head* trained alongside the decoder to explicitly forecast binary pedestrian occupancy at each imagined step, providing the actor with an auxiliary safety signal; and (4) *hybrid planning* that combines learned imagination with a lightweight rule-based safety layer (e.g., a time-to-collision check on the decoded depth channel) as a fallback when imagination is unreliable.

Several limitations apply. Our analysis uses a single scenario type (protester archetype) and one trained model. The sample size (565 encounters, 90 collisions) is substantial, but the strict metric is intentionally conservative. A larger-scale study with multiple pedestrian archetypes and relaxed metrics would provide a more complete picture. Additionally, the analysis is conducted in decoded observation space; the latent representations used for actual policy decisions may encode pedestrian information differently than what the decoder reveals.

5. Conlusion

We presented an event-centered analysis of imagination quality in a DreamerV3 driving agent facing a stationary pedestrian blocking the lane. Under a strict localized fidelity metric, the world model produces only weak pedestrian support in imagined futures, and this weakness does not clearly separate collision from avoidance outcomes. Qualitative inspection reveals that the more informative failure is imagined lane clearance: the model predicts that the pedestrian obstacle will disappear rather than persist. These findings highlight the importance of examining not only whether a world model retains object presence at the moment of capture, but whether it maintains that presence across the imagination horizon used for planning.

Acknowledgments

This work was funded by the COMET K2 - Competence Centers for Excellent Technologies Programme of the Austrian Federal Ministry for Transport, Innovation and Technology (bmvit), the Austrian Federal Ministry for Digital, Business and Enterprise

(bmdw), the Austrian Research Promotion Agency (FFG), the Province of Styria, and the Styrian Business Promotion Agency (SFG).

References

[1] Danijar Hafner, Timothy P. Lillicrap, Jimmy Ba, and Mohammad Norouzi. Dream to control: Learning behaviors by latent imagination. ArXiv, abs/1912.01603, 2020.

[2] B Ravi Kiran, Ibrahim Sobh, Victor Talpaert, Patrick Mannion, Ahmad A. Al Sallab, Senthil Kumar Yogamani, and Patrick Pérez. Deep reinforcement learning for autonomous driving: A survey. IEEE Transactions on Intelligent Transportation Systems, 23:4909–4926, 2021.

[3] Wilko Schwarting, Javier Alonso-Mora, and Daniela Rus. Planning and decision-making for autonomous vehicles. Annual Review of Control, Robotics, and Autonomous Systems, 1:187–210, 2018.

[4] Danijar Hafner, Jurgis Pašukonis, Jimmy Ba, and Timothy P. Lillicrap. Mastering diverse domains through world models. ArXiv, abs/2301.04104, 2023.

[5] Alexey Dosovitskiy, German Ros, Felipe Codevilla, Antonio Lopez, and Vladlen Koltun. CARLA: An open urban driving simulator. In Proceedings of the 1st Annual Conference on Robot Learning, pages 1–16, 2017.

[6] Marin Toromanoff, Emilie Wirbel, and Fabien Moutarde. End-to-end model-free reinforcement learning for urban driving using implicit affordances. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 7153–7162, 2020.

[7] Dian Chen, Brady Zhou, Vladlen Koltun, and Philipp Krähenbühl. Learning by cheating. In Proceedings of the Conference on Robot Learning, volume 100 of Proceedings of Machine Learning Research, pages 66–75. PMLR, 2020.

[8] Neha Sharma, Chhavi Dhiman, and S. Indu. Pedestrian intention prediction for autonomous vehicles: A comprehensive survey. Neurocomputing, 508:120–152, 2022.

[9] Amir Rasouli and John K. Tsotsos. Autonomous vehicles that interact with pedestrians: A survey of theory and practice. IEEE Transactions on Intelligent Transportation Systems, 21:900–918, 2018.

[10] Tirthankar Bandyopadhyay, Kok Sung Won, Emilio Frazzoli, David Hsu, Wee Sun Lee, and Daniela Rus. Intention-aware motion planning. In Workshop on the Algorithmic Foundations of Robotics, 2013.

[11] Fanta Camara, Nicola Bellotto, et al. Pedestrian models for autonomous driving part ii: High-level models of human behavior. IEEE Transactions on Intelligent Transportation Systems, 22:5453–5472, 2020.

[12] Qifeng Li, Xiaosong Jia, Shaobo Wang, and Junchi Yan. Think2drive: Efficient reinforcement learning by thinking in latent world model for quasi-realistic autonomous driving (in carla-v2). ArXiv, abs/2402.16720, 2024.

[13] Dechen Gao, Shuangyu Cai, Hanchu Zhou, Hang Wang, Iman Soltani, and Junshan Zhang. Cardreamer: Open-source learning platform for world-model-based autonomous driving. IEEE Internet of Things Journal, 12:2866–2875, 2024.

[14] Tuo Feng, Wenguan Wang, and Yi Yang. A survey of world models for autonomous driving. ArXiv, abs/2501.11260, 2025.

[15] Zeyu Zhu and Huijing Zhao. A survey of deep rl and il for autonomous driving policy learning. IEEE Transactions on Intelligent Transportation Systems, 23:14043–14065, 2021.

[16] Pranav Singh Chib and Pravendra Singh. Recent advancements in end-to-end autonomous driving using deep learning: A survey. IEEE Transactions on Intelligent Vehicles, 9:103–118, 2023.

[17] Elahe Delavari, Feeza Khan Khanzada, and Jaerock Kwon. A comprehensive review of reinforcement learning for autonomous driving in the carla simulator. ArXiv, abs/2509.08221, 2025.

[18] Shreyas Ramakrishna, Baiting Luo, Christopher B. Kuhn, Gabor Karsai, and Abhishek Dubey. Anti-carla: An adversarial testing framework for autonomous vehicles in carla. 2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC), pages 2620–2627, 2022.

[19] Christopher Diehl, Timo Sievernich, Martin Krüger, Frank Hoffmann, and Torsten Bertram. Uncertainty-aware model-based offline reinforcement learning for automated driving. IEEE Robotics and Automation Letters, 8:1167–1174, 2023.

[20] Oleh Rybkin, Chuning Zhu, Anusha Nagabandi, Kostas Daniilidis, Igor Mordatch, and Sergey Levine. Model-based reinforcement learning via latent-space collocation. ArXiv, abs/2106.13229, 2021.

[21] Danijar Hafner, Timothy P. Lillicrap, Mohammad Norouzi, and Jimmy Ba. Mastering atari with discrete world models. ArXiv, abs/2010.02193, 2021.

[22] Fan-Ming Luo, Tian Xu, Hang Lai, Xiong-Hui Chen, Weinan Zhang, and Yang Yu. A survey on model-based reinforcement learning. Science China Information Sciences, 67, 2022.

APPLICATION OF ARTIFICIAL INTELLIGENCE IN PRE-UNIVERSITY EDUCATION – CASE STUDY SCHOOLS IN THE MUNICIPALITY OF KAMENICA – KOSOVO

Selami Klaiqi [1], Alban Rakipi [2]
[1]PhD Candidate in Faculty of Information Technology, Polytechnic University of Tirana
[2]Associate Professor in Faculty of Information Technology, Polytechnic University of Tirana
selami.klaiqi@uni-gjilan.net

Abstract: *Pre-university education in general is facing various challenges, but in the schools of the Municipality of Kamenica the challenges are even greater in improving learning outcomes in teaching and learning, these challenges range from identifying individual student needs and optimizing learning resources.*

This paper examines the role and impact of artificial intelligence (AI) in improving and facilitating learning circumstances in schools in the Municipality of Kamenica in Kosovo. The main goal is to analyze how AI solutions can improve the level of learning, support effective teaching and help in assessing and attracting students by increasing their level of concentration.

In this paper, we have taken an empirical approach using real data from several pre-university education schools in Kamenica, which include test scores, attendance statistics, and information on learning activities. With the help of machine learning techniques, models have been built that identify student profiles with different performance and recommend personalized learning strategies for each profile. An AI-based recommender system has also been developed that suggests teaching materials and relevant exercises using virtual laboratories according to the needs and individual progress of students.

The results show that the use of AI tools helps in identifying student weaknesses more quickly, in creating personalized lesson plans and in facilitating the work of teachers for continuous monitoring and evaluation. The analysis also shows the perception of teachers and parents across the Municipality of Kamenica towards the integration of AI in teaching practice, highlighting the challenges and opportunities for wider implementation in pre-university education in Kosovo.

This case study that we have conducted in the Municipality of Kamenica offers a practical and strategic contribution to municipal education policies towards the effective use of advanced technologies in the teaching process not only in the Municipality of Kamenica but also beyond.

Keywords: Artificial intelligence, machine learning, pre-university education.

1. Introduction

As in other countries, technology has significantly transformed the education system, creating new opportunities for improving the teaching and learning process. In modern education, the use of technology is no longer considered just an auxiliary tool, but as an essential component for developing students' skills, developing critical thinking and increasing the quality of the education system.

The Municipality of Kamenica has a territory of 423 km/2. It consists of the town of Kamenica as an urban center and 56 villages, the Municipality of Kamenica also has 27,823 inhabitants, has 3 upper secondary schools in the Albanian language, 2 upper secondary schools in the Serbian language, 13 primary and lower secondary schools in the Albanian language and 3 primary and lower secondary schools in the Serbian language, where in total there are 5 upper secondary schools and 16 primary and lower secondary schools.

One of the most important aspects of using technology in education is facilitating access to information, facilitating the work of teachers and activities for students. Through the Internet and digital platforms, students and teachers have access to a large number of educational resources, such as electronic books, educational videos, scientific databases and other didactic materials. This helps to expand knowledge and develop independent learning. Technology also contributes to increasing interactivity in the learning process. The use of interactive whiteboards that the Municipality of Kamenica has invested heavily in, online platforms, educational applications and multimedia tools makes learning more attractive and understandable for students. In this way, students are more engaged in the learning process and develop critical and creative skills.

The integration of technology in education improves the quality of the learning process, increases student motivation and prepares them for the challenges of contemporary society, for this reason the Directorate of Education in Kamenica prioritizes the integration of technology in education, especially in lower secondary schools and primary schools. Therefore, it is very necessary for educational institutions to continuously invest in technological infrastructure and in the training of teachers for the effective use of technology in teaching.

2. Benefits of using AI in education

Kamenica, although a small city, the use of artificial intelligence in education has brought significant changes in the way the teaching and learning process takes place. AI offers new tools and methods that help increase the efficiency, innovation and quality of education.

One of the main benefits of artificial intelligence is the ability to provide coherent, innovative and current learning. AI-based systems can analyze student performance and adapt learning materials according to their individual level, pace and needs. This helps each student learn in the most effective way for them, and this helps create higher security among students also in relation to the assessment process which is defined in AI 06/2022.

AI can help teachers prepare teaching materials, design lesson plans and create exercises or tests. The use of AI helps reduce the administrative burden of teachers and gives them more time to focus on pedagogical aspects and direct work with students. The use of intelligent applications, digital platforms, virtual laboratories, interactive platforms and digital systems makes the learning process more interesting, interactive and attractive for students.

Through simulations, educational games and multimedia content, students are more actively involved in learning and this

increases student satisfaction and also allows parents to have easier access to their children.

The use of artificial intelligence in education contributes to improving the quality of learning, as quality is one of the main challenges in the entire educational process, increasing the efficiency of the learning process and developing the digital skills of students, making education more suitable for the demands of the parent community.

The municipality of Kamenica has a significant number of students who may have problems in their studies, in addition to the general difficulties that a municipality may have in education, Kamenica has also faced the closure of several schools for a while and a significant number of students have not attended school for several years, therefore AI can provide special tools for students with learning difficulties, such as voice recognition programs, automatic translation or applications that help with reading and writing.

The use of AI in the educational process is one of the advantages that the Municipality and the Directorate of Education can use to increase the level of satisfaction among parents, facilitate the work of teachers and increase the attractiveness among students.

Therefore, teaching AI in pre-university education is currently feasible, but the specific topics that should be taught at this age and to what extent these topics should be addressed are an open issue that needs to be discussed.

3. Challenges and risks of using AI in education

The use of AI in education is accompanied by challenges and eventual risks throughout the process, each municipality but especially the Municipality of Kamenica has challenges in implementing the use of AI in the educational process in primary schools, lower secondary schools and upper secondary schools.

If we start listing the challenges and risks in the use of AI in education, we should start from issues of privacy and data security, dependence on technology and the use of technology in an inappropriate way, plagiarism and credibility in teaching, lack of training for teachers, unequal access to technology, reliability and accuracy of information, etc.

3.1. Data privacy and security

The use of platforms and applications based on artificial intelligence requires the collection and processing of a large amount of data on students and teachers. If this data is not managed securely, there is a risk of privacy violations and misuse of personal information, therefore the use of platforms in the educational process also requires relevant legislation to prevent possible privacy violations.

The Directorate of Education has contracted a company through a public tender to create a unique platform for managing the education system within the municipality and access to this platform can be done through the link: https://www.kamenica-edu.net/ , but the issue of data privacy, their security was and remains one of the main challenges for management with this platform.

3.2. Technology addiction

Excessive use of technology by students, lack of control over the use of technology can create dependence on digital systems and reduce the active role of the teacher in the teaching process. The educational process should remain a balanced process where technology serves as an auxiliary tool and not as a complete substitute for the teacher, where the use of technology enables more effective learning and not degradation of the process.

The role of the teacher is key in preventing the creation of dependence on the use of technology, digital platforms or even virtual laboratories, the school should enable capacity building at the school level to increase control in preventing the creation of dependence on technology among students.

3.3. Plagiarism

Artificial intelligence tools, instruments and applications can be used by students to generate assignments, essays or projects without necessarily reflecting their real knowledge. This can negatively affect the development of critical thinking skills and the assessment of authorship.

Restricting the use of these AI-powered tools increases the level of assessment among students, puts into practice the legislation in force on the reliability of assessment and reduces the impact of AI on the authorship of papers, essays or projects by students.

3.4. Lack of training for teachers

Many teachers may not have the necessary knowledge and skills to effectively use artificial intelligence in the learning process, which makes teaching effectiveness difficult. Without proper training and professional support, the integration of AI in education may be limited, it may be an inefficient project and in order to have expectations of good results, the education directorate, together with the school management, must plan and organize school-based training, training that will help teachers increase their digital competence.

The Education Directorate in Kamenica has made and is making continuous efforts to increase the budget that will be allocated for the training of educational staff and this budget will be the weight of success in the future for a complete success.

3.5. Without equal access to technology

Not all schools in the Municipality of Kamenica and students have equal access to technological equipment and the internet. This can create a digital divide between different schools or regions, affecting equal opportunities for quality education.

Kamenica is a municipality bordering Serbia and has a large territory, a territory that does not allow equal access to technology, the internet and applications that are ready to provide resources for use in AI.

3.6 Reliability and accuracy of information

Artificial intelligence systems are not always accurate and can generate inaccurate or incomplete information. If students rely entirely on these systems without critical verification, this can affect the quality of learning.

Early childhood education, education on the use of technology and identification of safe sources of information makes the system more efficient and reliable by enabling students to identify fake news, identify safe sources and this helps us to have a stable and safe system.

4. Research methodology

During this work, we have taken into consideration the analysis of data obtained from student questionnaires, teacher questionnaires and interviews with school principals.

The sample was purposive, where we treated 2 upper secondary schools and 3 lower secondary schools, a total of 5 schools, where 37 teachers, 72 students and 5 school principals participated in this research, where we conducted a statistical and comparative analysis.

5. The current state of technology use in schools

Schools in the Municipality of Kamenica are generally supplied with technology equipment and this has led to better conditions for students and teachers. As a Municipality, it is a special specificity that the number of students in schools is not that large and the Directorate of Education has enabled each school at the level of lower secondary schools in each classroom to have a smart TV, laptop, computer, and the classrooms to be adapted in such a way that lighting is not an obstacle to the development of the learning process.

While the situation is different in upper secondary schools because there is a need for investments in creating favorable conditions for the learning process, supplying schools with technology equipment, increasing internet capacity throughout the school space, and training teachers in the use of technology and the use of AI platforms.

There has been and continues to be a problem in implementation with schools in the suburbs where there are problems with internet access, the need for teacher training, and also the need for students to have extra classes, supplementary lessons, or elective lessons for the subject of computer science.

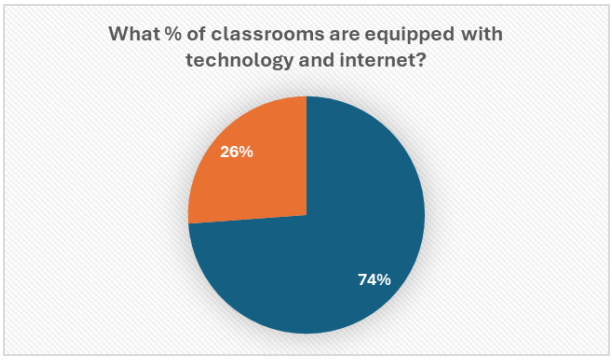


Figure 1. ***Infrastructure in schools regarding information technology***

6. Use of AI by students and teachers of the Municipality of Kamenica

The use of AI by students and teachers in the educational process is an indispensable need to enable increased quality and increased attractiveness to students. In addition to daily use in lessons, teachers have the opportunity to prepare their materials and then have them ready for their explanations to students.

During the implementation of the questionnaire with students, the question: Have you ever used artificial intelligence tools to assist in learning? In this case, 74% of students responded that they have AI tools to assist the learning process, while only 26% responded with No.

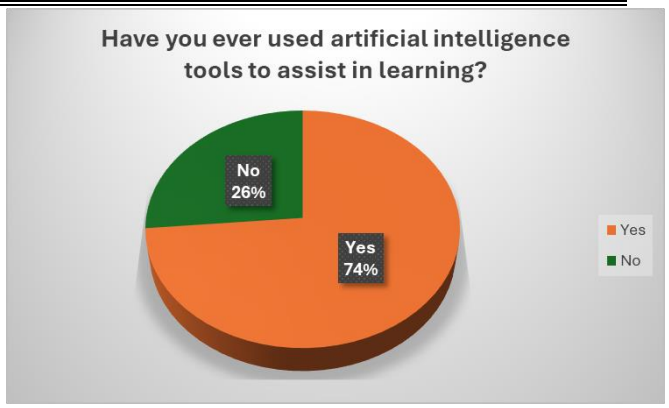


Figure 2. Have you ever used artificial intelligence tools to assist in learning?

Of the 74% of respondents who answered Yes, 64% of them said that they use AI more to understand a certain topic, 21% of them to do homework or write essays, and 15% of them to search for information.

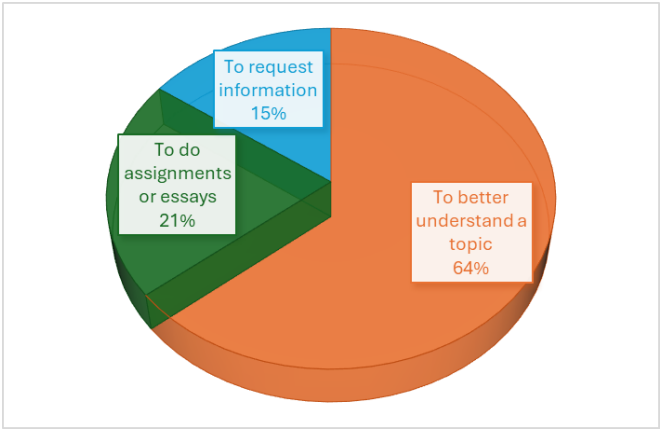


Figure 3. If so, what do you use them for most often?

During this study, we also addressed the impact of AI on lessons from both the student and teacher perspectives, asking: Do you think AI helps you understand lessons better?

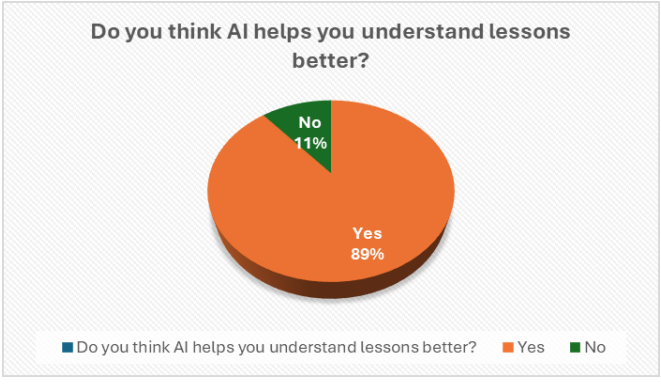


Figure 4. Do you think AI helps you understand lessons better?

From the diagram we can see that 89% of respondents answered Yes while only 11% of respondents answered No.

The treatment of this topic is very sensitive in terms of the possible impact of AI on the educational process and this situation is also presented to us in the following question: Do you think that the use of AI makes learning more interesting?

From here it can be seen that 62% of respondents answered that they think that the use of AI makes lessons more interesting, 31% of them think that the use of AI makes lessons interesting sometimes and only 7% think that the use of AI in lessons does not affect them to be more interesting and attractive.

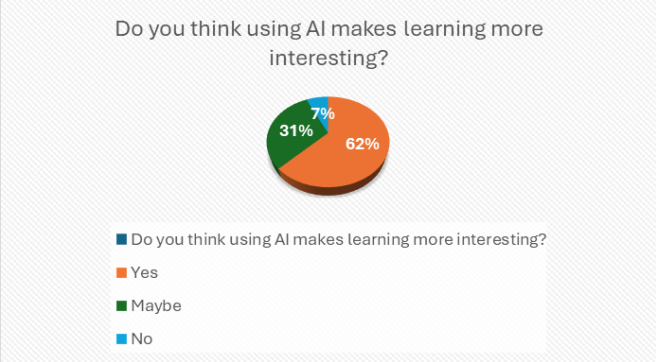


Figure 5. Do you think using AI makes learning more interesting?

The education system and the institutions that are responsible for improving the situation in education will increasingly focus on ensuring that the system in general is sustainable to ensure quality, to have security for students but also to be easily accessible for parents, to make the daily work of teachers easier by enabling them to make the learning and teaching process more attractive through the use of technology.

The next question that was asked to the students was: Do you think that the use of AI can cause students not to work enough on their own?

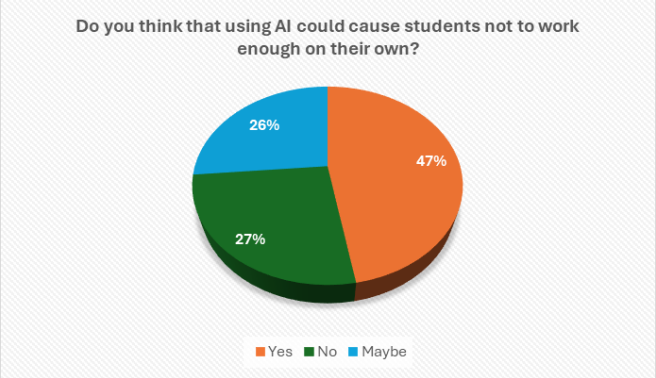


Figure 6. Do you think that using AI could cause students not to work enough on their own?

From the responses we have processed, we see that 47% of respondents think that the use of AI can cause students to not work independently and sufficiently, 26% think that this phenomenon sometimes occurs, while 27% of respondents think that AI does not affect students not working sufficiently.

One of the main challenges for the use of AI in the educational process is the preparation of educational staff, their readiness to use various platforms in the educational process.

We have conducted a questionnaire with 37 teachers from 5 schools in total that we conducted the research at the level of the Municipality of Kamenica.

In the question we conducted for teachers: Have you received any training on the use of digital technologies or AI in teaching?

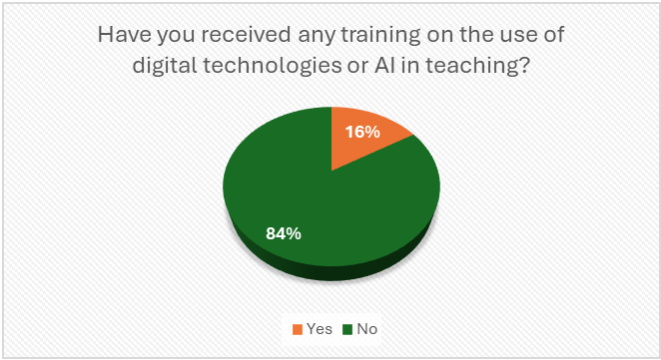


Figure 7. Have you received any training on the use of digital technologies or AI in teaching?

One of the main challenges to implementing and increasing the use of AI in the educational process is teacher training, the advanced age of teachers, and the preparation of legislation that would also include the use of AI in the teaching process.

In the next question, which deals with the use of AI tools in the teaching process, the question was: Do you use artificial intelligence tools in the preparation of teaching materials?

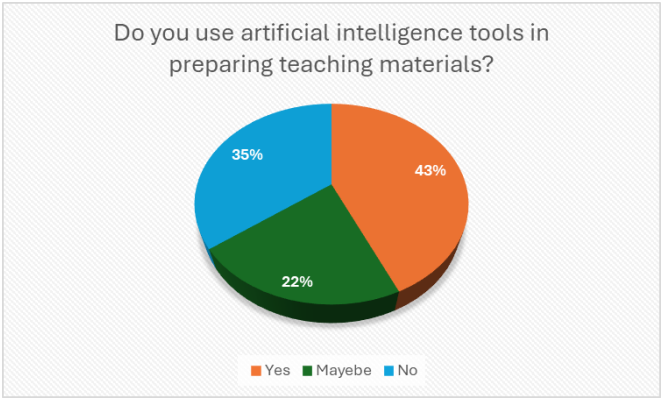


Figure 8. Do you use artificial intelligence tools in preparing teaching materials?

Of the teachers surveyed, we see that 43% of them use artificial intelligence tools to prepare teaching materials, 22% of them do not use these tools always, while 35% of teachers do not use artificial intelligence tools at all to prepare teaching materials.

A significant number of respondents, about 35% of whom stated that they do not use AI tools to prepare teaching materials, may result from the fact that teaching staff still use tools that are classical and are not yet ready for various reasons to use AI in the teaching process.

The entire research is about the state of the education system, the use of AI tools, and what impact it has on improving the state of the education system. The next question posed to teachers is: How much do you think the use of AI can improve the quality of teaching?

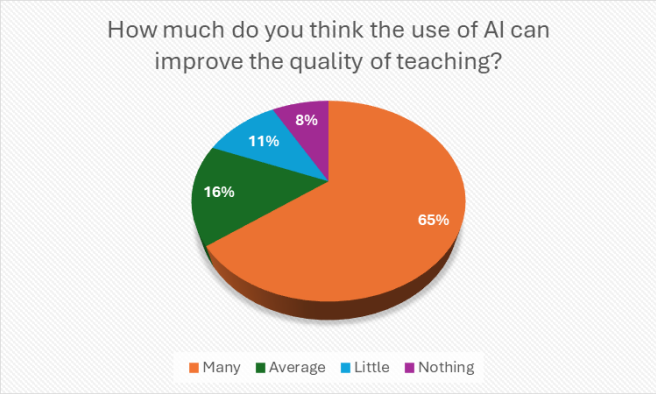


Figure 9. How much do you think the use of AI can improve the quality of teaching?

From the diagram we see the results that we have shown where 65% of teachers think that the use of AI tools can improve the quality in education, 16% of them think that the quality situation will improve on average if we use AI elements, 11% of teachers think that AI tools have little impact on improving the situation and quality in education while 8% of the surveyed teachers think that AI tools do not affect the improvement of the situation in education at all.

If the last question and result are carefully analyzed and compared with the previous question, we see that teachers are optimistic that the use of AI will positively affect the educational process, but we still have an unstable situation on the way and AI tools that teachers are using.

7. Challenges in implementing AI in schools

The integration of artificial intelligence tools into the education system in schools in the Municipality of Kamenica is an important step towards the modernization of the teaching and learning process. However, the implementation of these technologies in schools is accompanied by several challenges that need to be carefully addressed by educational institutions and specifically by the Directorate of Education.

The main challenges that may arise during the implementation of AI tools in schools in the Municipality of Kamenica are: lack of IT infrastructure, lack of training for teachers, resistance to change, lack of clear policies and guidelines, and the issue of creating a legal framework or regulations at the municipal or school level.

During the interview with the principals of the respective schools, they stated that one of the main challenges during the efforts to implement the use of AI tools in schools and their use is also the resistance from the staff, the resistance to switching from one system to another.

According to school principals, another challenge is the lack of supporting infrastructure and especially the lack of school inclusion in the internet or increasing internet capacity and inclusion of the entire school space.

One of the main concerns that should be resolved at the national level is for the Ministry of Education in Kosovo to draft Administrative Instructions that would enable this area to be regulated through legislation and for eventual actions to be based on the law, as this also avoids contestation by various actors who may be specific teachers, parents or other important factors.

Although artificial intelligence offers great opportunities for improving the quality of education, its implementation in schools, especially in a small municipality such as Kamenica, requires a budget to invest in infrastructure, professional development for teachers, and the creation of clear policies for the safe and ethical use of technology.

Only through a well-planned approach can the Directorate of Education successfully integrate artificial intelligence into the education system at both the upper secondary and lower secondary school levels.

8. Discussions and recommendations

During this study, we have managed to see the importance of using AI in the teaching process in both upper secondary and lower secondary schools, the impact of the use of AI by students during the teaching process, the challenges and advantages that we may have during this process.

If we carefully analyze the results of the responses from students and teachers, we see that we have very promising results and that they let us know that the use of AI in the educational process is the right alternative to have better results and to have a higher interest among students.

Recommendations for possible improvements during the process will include investing in infrastructure, increasing the level of awareness campaigns for students, increasing the budget for training to increase support for professional development for teachers, investing in human resources in the field of IT in schools that will help the process both from a hardware and software perspective.

Increasing the capacity to have the legal infrastructure at the appropriate level to be a legal support for all activity in schools related to the implementation and use of AI in educational institutions at the level of upper secondary schools and lower secondary schools.

Conclusions

During this study, we have managed to see the importance of using AI in the teaching process in both upper secondary and lower secondary schools, the impact of the use of AI by students during the teaching process, the challenges and advantages that we may have during this process.

If we carefully analyze the results of the responses from students and teachers, we see that we have very promising results and that they let us know that the use of AI in the educational process is the right alternative to have better results and to have a higher interest among students.

Recommendations for possible improvements during the process will include investing in infrastructure, increasing the level of awareness campaigns for students, increasing the budget for training to increase support for professional development for teachers, investing in human resources in the field of IT in schools that will help the process both from a hardware and software perspective.

Increasing the capacity to have the legal infrastructure at the appropriate level to be a legal support for all activity in schools related to the implementation and use of AI in educational institutions at the level of upper secondary schools and lower secondary schools.

References

[1] Thomas K.F. Chiu, Helen Meng, Ching-Sing Chai, Irwin King, Savio Wong, Yeung Yam, Creation and Evaluation of a Pre-tertiary Artificial Intelligence (AI) Curriculum

[2] Andrew Lee Hock Cheong Global College, Heriot-Watt University Malaysia, Use of Generative Artificial Intelligence among Selected Pre University Business Students at a Private University in Malaysia

[3] Serkan SAVAŞ, Artificial Intelligence and Innovative Applications in Education: The Case of Turkey

[4] Sara Guerreiro-Santalla *, Francisco Bellas and Richard J. Duro, Artificial Intelligence in Pre-University Education: What and How to Teach

[5] Sekvan Kuzu, Artificial intelligence in pre-service teacher education: Bibliometrics analysis

[6] Glenn W. Ellis, Katie L. Silva, Talia Epstein And Nicholas C. Giammaria, Artificial Intelligence in Pre-College Education: Learning within a Philosophy of the Mind Framework

Preparation and Optical Properties of Gelatin-SiO_2 Hybrid Coatings on Glasses for Photovoltaic Applications

Stancho Yordanov[1]*, Tatyana Ivanova[2], Shaban Uzun[1], Vladimir Petkov[1], Ina Gospodinova[1]
[1] Bulgarian Academy of Sciences, Institute of Metal Science, Equipment, and Technologies "Acad. A. Balevski" with Center for Hydro- and Aerodynamics, 67 "Shipchenski prohod" str., 1574 Sofia, Bulgaria
[2] Bulgarian Academy of Sciences, Central Laboratory of Solar Energy and New Energy Sources, Sofia, Bulgaria
e-mail: stancho14@abv.bg

Abstract: *In this study, gelatin-SiO_2 hybrid coatings were prepared on glass substrates using an acid-catalyzed sol-gel method. The deposited layers were subjected to thermal treatment at 50°C, 150°C, and 250°C to evaluate the effect of temperature on their structure and optical properties. Spectroscopic investigations in the UV-VIS-NIR range show that the hybrid coatings enhance optical transmittance and reduce reflective losses compared to uncoated glass. The best optical performance was observed for the sample thermally treated at 150 °C (GS 150), where an optimal balance between the organic (gelatin) and inorganic (SiO_2) phases was achieved, resulting in maximum transmittance and minimal light reflection. At lower treatment temperatures (50°C), the coating exhibited higher porosity but lower structural stability, whereas at higher temperatures (250°C), partial thermal degradation of the gelatin phase led to layer densification and a slight reduction in the antireflective effect. These results demonstrate that gelatin can be successfully used as a biopolymer component to control the microstructure and optical properties of sol-gel hybrid coatings. Improving optical transmittance and reducing reflective losses of glass surfaces is crucial for enhancing photovoltaic module efficiency, as it allows a greater amount of solar radiation to reach the active layer of solar cells.*

Keywords*: GELATIN-SIO_2 HYBRID, SOL-GEL METHOD, OPTICAL PROPERTIES, ANTIREFLECTIVE COATINGS*

1. Introduction

Photovoltaic (PV) systems are among the fastest-growing renewable energy sources and play a key role in the transition to sustainable energy technologies. The efficiency of photovoltaic panels depends not only on the characteristics of the semiconductor materials but also on the optical properties of the protective glass covering the active layer. Glass serves as mechanical and environmental protection but can cause optical losses due to light reflection at the air-glass interface. For standard glass, reflective losses can reach approximately 4 % per surface, reducing the amount of light reaching the photovoltaic cell and, consequently, lowering the module's energy efficiency [1, 2].

One of the most effective approaches to reduce these losses is the application of antireflective (AR) coatings on the glass surface. Such coatings are designed so that their effective refractive index lies between that of air and glass, allowing a reduction in reflection and an increase in light transmittance to the active layer of the photovoltaic cell [1]. In addition to optical performance, coatings for photovoltaic applications must have high transparency over a wide spectral range (UV-VIS-NIR), good mechanical stability, resistance to environmental conditions, and often self-cleaning properties that reduce dust and contamination accumulation on the panel surface [3].

Among the various methods for producing such functional coatings, the sol-gel technique is considered one of the most promising approaches. This low-temperature process allows the synthesis of inorganic and hybrid materials with controlled microstructure and optical properties [4,5]. The sol-gel method is compatible with industrial coating techniques for large surfaces, such as dip-coating, spin-coating, and spray-coating, making it particularly suitable for photovoltaic panel applications [6].

Silicon dioxide (SiO_2) is one of the most commonly used materials for sol-gel optical coatings due to its high transparency, chemical stability, and good adhesion to glass substrates. By controlling the porosity of the silica network, the effective refractive index of the coating can be tuned, enabling the formation of low-refractive-index layers with strong antireflective behavior [7]. Porous silica coatings with very low refractive indices ($n \approx 1.18$) have been reported, while maintaining nearly 100 % transmittance at certain wavelengths [8]. Furthermore, multilayer sol-gel silica coatings can provide broadband antireflective performance with transmittance above 98 % in the 350-1400 nm spectral range [9].

Despite these advantages, purely inorganic sol-gel silica films may exhibit limitations, such as cracking, internal stress, and limited mechanical durability during operation. One approach to overcome these issues is the development of organic-inorganic hybrid materials, where the organic component is integrated into the inorganic silica network. Such hybrid systems combine the mechanical and chemical stability of the silica phase with the flexibility and film-forming properties of the organic components [10, 11].

Biopolymers are an especially interesting class of organic additives for producing such hybrid coatings. Gelatin, obtained by partial hydrolysis of collagen, is widely used due to its biocompatibility, biodegradability, and excellent film-forming properties. The presence of functional groups such as amino (-NH_2), carboxyl (-COOH), and hydroxyl (-OH) allows interaction with hydrolyzed silicate species and participation in the formation of organic-inorganic hybrid structures [12]. Moreover, gelatin can influence the viscosity of the sol system, hydrolysis and condensation kinetics, and thin-film formation, facilitating the preparation of uniform coatings [13].

Previous studies have shown that silica-gelatin hybrid materials can form transparent coatings on glass substrates. For instance, Smitha et al. reported transparent SiO_2-gelatin coatings with a thickness of about 300 nm and optical transparency above 90 % [14]. Other studies have obtained silica-gelatin nanohybrid coatings with nearly 100 % optical transmittance, indicating that proper control of nanostructure and component dispersion can preserve high optical transparency [15].

Despite these promising results, the relationship between synthesis conditions, hybrid network structure, and optical properties of gelatin-SiO_2 coatings on glass is not yet fully understood. The optical characteristics of these materials depend on several factors, including porosity, film uniformity, particle size, and the degree of interaction between the organic and inorganic phases [16].

Therefore, the aim of this work is to prepare gelatin-SiO_2 hybrid coatings on glass substrates using the sol-gel method and to investigate their optical properties through UV-Vis spectroscopy. Particular attention is paid to the influence of synthesis parameters and thermal treatment on the transparency and antireflective performance of the coatings, with a view to their potential application in photovoltaic panels.

2. Experimental procedure

The gelatin-SiO_2 hybrid coatings were prepared via an acid-catalyzed sol-gel process using tetraethyl orthosilicate (TEOS, ≥98%), absolute ethanol, 0.1 M HCl, distilled water, and commercial gelatin, all used without further purification. Standard soda-lime glass slides were used as substrates.

The initial silica sol was prepared by mixing 19 mL TEOS (17.73 g, 0.085 mol), 30 mL ethanol (23.67 g, 0.514 mol), 2.7 mL water (2.70 g, 0.150 mol), and 0.5 mL of 0.1 M HCl. The resulting initial molar ratio of TEOS:H_2O:EtOH was 1:1.76:6.05. In a separate vessel, 2 g of gelatin was dissolved in 60 mL of distilled water at 45 °C under magnetic stirring until fully dissolved. The gelatin solution, containing an additional 60 g water (3.33 mol H_2O), was slowly added to the silica sol under continuous stirring. After addition, the total water content reached 62.7 g (3.48 mol), resulting in a final molar ratio of TEOS:H_2O:EtOH = 1:40.9:6.05 and a hydrolysis ratio R ≈ 41, characteristic of highly hydrolyzed silica systems.

For a more complete description of the composition, all components were also calculated in mass percentages. The final sol contained 59.1 wt% water, 22.3 wt% ethanol, 16.7 wt% TEOS, and 1.9 wt% gelatin. This high water fraction facilitates complete hydrolysis of TEOS and the formation of a hydroxyl-rich Si-O-Si network capable of interacting with the functional groups of gelatin. The resulting sol was stirred for 60 min and left to rest for 24 h at room temperature to stabilize the hydrolysis-condensation processes.

The glass substrates were cleaned ultrasonically in sequence with ethanol and distilled water and then dried at 100 °C. The coatings were applied via dip coating at a withdrawal speed of 72.9 mm/min under ambient conditions. After deposition, the films were allowed to dry for 2 h at 25°C and subsequently thermally treated in air at 50°C (GS 50), 150°C (GS 150), and 250°C (GS 250). These temperatures were selected to investigate the effect of partial or complete degradation of the organic phase on the optical properties and microstructure of the hybrid layers.

The optical properties of the coatings were examined using UV-VIS-NIR spectroscopy in the 300-1100 nm range in both transmittance and reflectance modes. Uncoated glass was used as a reference sample, allowing direct comparison of the effect of the hybrid layer on optical behavior.

3. Results and Discussion

The transmittance spectra of the hybrid coatings, measured with a Shimadzu UV 3600 UV-VIS-NIR spectrophotometer in the 240-1800 nm range, show a pronounced increase in optical transparency compared to uncoated glass. The spectra were recorded relative to air, with the spectrum of bare glass used as a reference. All samples exhibit higher transmittance in the 400-800 nm range, which can be attributed to the formation of a low-refractive-index porous silica layer modified by the organic gelatin phase. The highest transmittance is observed for the GS 150 sample, indicating optimal structuring of the SiO_2 network at this thermal treatment temperature. At 250°C, partial degradation of the gelatin leads to densification of the layer and a slight increase in the refractive index, reducing the antireflective effect (Figures 1, 2, 3, and 4).

Reflectance spectra, measured using an aluminum mirror as the reference standard, confirm a reduction in surface reflection for all hybrid coatings. The minimum reflectance is observed for the GS 150 sample, corresponding to the best match of the effective refractive index between air and the glass substrate. The reduced reflection indicates the presence of nanometer-scale porosity in the hybrid SiO_2 structure, induced by the organic component. At higher temperature (250°C), partial breakdown of the organic matrix results in a denser structure and, consequently, a slight increase in reflectance.

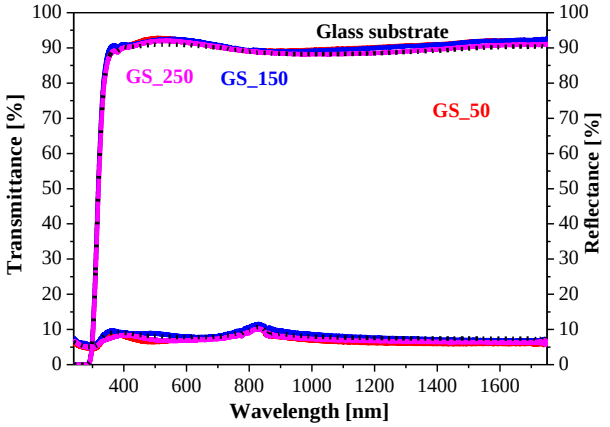


Figure 1. Transmittance and reflectance of gelatin-SiO_2 coatings after thermal treatment at 50°C, 150°C and 250°C in the region 240-1800 nm range.

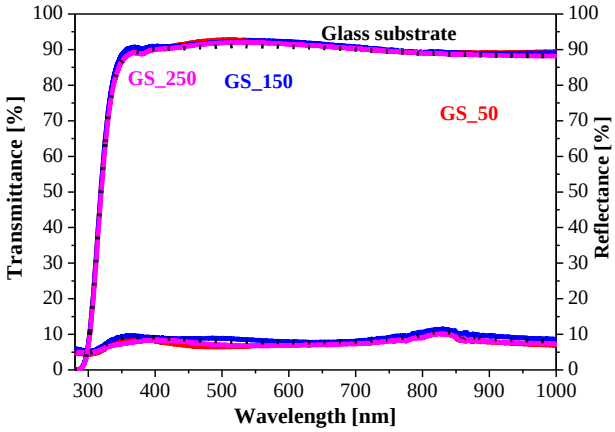


Figure 2. UV-VIS-NIR transmittance and reflectance of gelatin-SiO_2 coatings after thermal treatment at 50°C, 150°C and 250°C in the 240–1000 nm spectral range

The spectra are recorded against air, with bare glass used as reference. The GS 150 sample shows the highest transmittance due to optimal stabilization of the hybrid silica network.

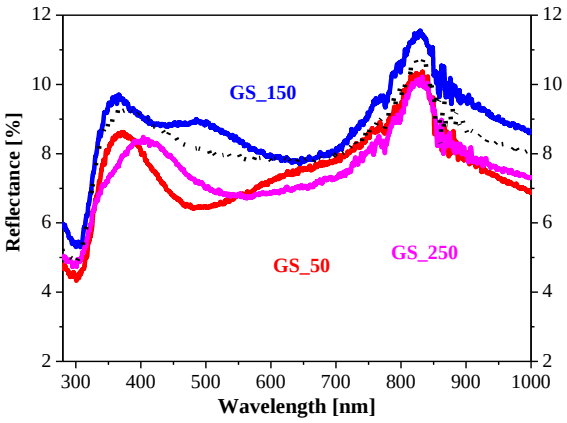


Figure 3. Reflectance in the range 2–12 % of gelatin-SiO_2 coatings after thermal treatment at 50°c, 150°c and 250°c in the region 240-1800 nm range.

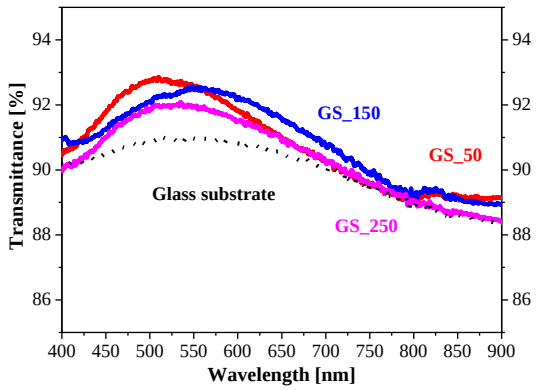


Figure 4. Transmittance in the range 86–94 % of gelatin-SiO_2 coatings after thermal treatment at 50°C, 150°C and 250°C in the region 240-1800 nm range.

Figure Summary

- GS 50: High organic phase → higher porosity, but lower structural stability.
- GS 150: Optimal balance between organic and inorganic phases → highest transmittance and lowest reflectance.
- GS 250: Partial gelatin degradation → densification of the layer, higher refractive index, reduced antireflective effect.

The results clearly demonstrate the influence of thermal treatment on the structure and optical properties of gelatin-SiO_2 hybrid coatings. At low temperature (50°C), the high organic content leads to increased porosity and lower structural stability, resulting in slightly reduced transmittance and higher reflectance. The optimal performance is observed for GS 150, where the interaction between gelatin and the SiO_2 network promotes the formation of a compact yet porous structure, which minimizes surface reflection and maximizes transmittance. This indicates that the organic phase not only facilitates the formation of a porous network but also stabilizes the hybrid structure, enhancing the optical properties.

At 250°C, partial degradation of gelatin results in densification of the layer, an increase in refractive index, and a slight rise in reflectance, confirming the role of the organic component in controlling porosity and the antireflective effect. These observations are consistent with previous studies on SiO_2-organic hybrid systems, where the balance between the organic and inorganic phases determines the optical characteristics and mechanical stability [9, 11].

Overall, the results demonstrate that by precisely controlling the thermal treatment temperature and sol composition, functional coatings with tailored optical properties can be obtained, suitable for applications such as antireflective layers on glass and potentially for photovoltaic panels.

4. Conclusion

1. Gelatin can be effectively integrated into the SiO_2 sol-gel network, providing structural adaptability and control over the porosity of the layer.
2. Gelatin-SiO_2 hybrid coatings exhibit enhanced transparency and reduced reflectance compared to uncoated glass, with optimal optical properties achieved at a thermal treatment of 150°C.
3. At lower temperature (50°C), the layer is porous but structurally less stable, whereas at higher temperature (250°C), partial gelatin degradation leads to densification and a slight reduction in the antireflective effect.
4. The results highlight the potential of gelatin-SiO_2 hybrid coatings for applications requiring transparent, antireflective, and functional glass layers, such as optical devices and photovoltaic panels.

References

1. Macleod, H. A. Thin-Film Optical Filters, CRC Press, 2010.
2. Green, M. A. Solar Cells: Operating Principles, Technology and System Applications, Prentice Hall, 1982.
3. Saravanan, S. et al. Optical investigation of sol-gel synthesized SiO_2/TiO_2 multilayers for antireflective coatings, Nanomaterials, 13 (2023) 389.
4. Brinker, C. J., Scherer, G. W. Sol-Gel Science: The Physics and Chemistry of Sol-Gel Processing, Academic Press, 1990.
5. Hench, L. L., West, J. K. The sol-gel process, Chemical Reviews, 90 (1990) 33-72.
6. Lu, S., Shao, J., Wu, F. Industrial applications of sol-gel derived coatings, Journal of Sol-Gel Science and Technology, 114 (2025) 87-97.
7. Guglielmi, M. Sol-gel coatings on metals, Journal of Sol-Gel Science and Technology, 1 (1990) 7-25.
8. Ye, L. et al. Ammonia treatment of sol-gel silica coatings to achieve a stable low refractive index, Ceramics International, 51 (2025) 25192-25197.
9. Chi, F., Zeng, Y., Liu, C. Tuning refractive indices of sol-gel silica coatings by ammonia treatment for broadband antireflection applications, Optik, 224 (2020) 165501.
10. Sanchez, C. et al. Applications of advanced hybrid organic-inorganic nanomaterials, Chemical Society Reviews, 40/2 (2011) 696-753.
11. Zvonkina, I. J., Soucek, M. D. Inorganic-organic hybrid coatings: common and new approaches, Current Opinion in Chemical Engineering, 11 (2016) 123-127.
12. Ramos, M. et al. Gelatin-based films and coatings for food packaging applications, Coatings, 6/4 (2016) 41.
13. Jitianu, A. et al. Hybrid silica-gelatin materials obtained by sol-gel method, Journal of Non-Crystalline Solids, 345-346 (2004) 596-600.
14. Smitha, S. et al. Silica-gelatin bio-hybrid and transparent nano-coatings through sol-gel technique, Materials Chemistry and Physics, 103 (2007) 318-322.
15. Smitha, S. et al. Antiwetting silica-gelatin nanohybrid coatings synthesised through an aqueous sol-gel process, Journal of Sol-Gel Science and Technology, 42 (2007) 157-163.
16. Xia, B. et al. Preparation of silica coatings with adjustable refractive indices via sol-gel method. *RSC* Advances, 8 (2018) 6091-6098.

Investigation of Positioning Accuracy Improvement Methods for Sigfox IoT Devices through WiFi MAC Address Analysis

Masaya Watanabe[1], Takuya Nanjo[1], Tomoo Sigehuzi[2], Satoru Tada[1], Masayuki Yamauchi[1],
Takeshi Nishiyone[3], Hironori Wakaoka[3], Shuji Oji[3], Naoki Kabemoto[4], Hiroki Mori[4] ,Takeshi Tanaka[1] and Katia Vutova[5]
Hiroshima Institute of Technology, Hiroshima, Japan[1]
National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Japan[2]
Chudenko Corporation, Hiroshima, Japan[3]
Interface Corporation, Hiroshima, Japan[4]
Institute of Electronics, Bulgarian Academy of Sciences, Sofia, Bulgaria[5]
tanaka@cc.it-hiroshima.ac.jp

Abstract**:** *This study presents and validates a method for achieving positioning accuracy comparable to standalone GPS (approximately 10 m) using a proprietary Sigfox WiFi positioning system that transmits only two MAC addresses per message. Due to Sigfox's severe payload limitations, conventional positioning attempts often fail or yield significant errors. Our approach introduces continuous scanning on the device to dynamically collect up to four MAC addresses, paired with meticulous backend analysis and filtering of MAC address quality using a custom Here API Validation Tool. The proposed method effectively addresses these limitations, demonstrating a 100% location acquisition rate and a median positional error of 19 meters, making it highly effective for both stationary and dynamic mobile tracking.*
Keywords*: SIGFOX, WIFI POSITIONING, MAC ADDRESS, HERE API, LPWA, IOT, ESP32-S3*

1. Introduction

Location information is crucial across various fields, including surveying, autonomous driving, construction, and disaster response. GPS positioning, a common method for location estimation, typically incurs an error of approximately 10–30 m[1]. With the widespread adoption of IoT devices, the demand for low-cost, low-power positioning methods that utilize Sigfox, a low-power wide area network, has increased[2].

Sigfox's official Atlas WiFi service enables positioning by obtaining the MAC addresses of nearby WiFi access points. However, within the proprietary mechanism developed in this study, the number of MAC addresses that can be transmitted in a single message is strictly limited to two due to Sigfox's payload constraint of 12 bytes. Attempting to perform positioning by directly sending this limited MAC address information to the Here Positioning API often results in complete failures or significant errors (ranging from tens to hundreds of meters).

To address these challenges, this study proposes and validates a method that achieves accuracy comparable to standalone GPS positioning (approximately 10 m) using a proprietary Sigfox WiFi positioning system that transmits only two MAC addresses per message. This is achieved through meticulous scanning of the WiFi environment using NetSpot, along with quality analysis and selection of MAC addresses via a custom-built Here API Validation Tool designed for this verification.

2. GPS Positioning (Baseline accuracy verification)

A. IoT Device Configuration

We first developed an IoT device equipped with a Sigfox network module and a GPS module. The configuration combines an Arduino Uno Rev3[3] with a Sigfox shield for Arduino V2S[4] and a GPS module[5]. Detailed construction methods and the modules utilized are outlined in the literature[6].



Fig. 1 *IoT device with Sigfox network and GPS modules.*

B. GPS Positioning Results (Accuracy: Approximately 10 m)

Latitude (multiplied by 1 M) and longitude (multiplied by 1 M) data acquired from the IoT device are transmitted via email through the Sigfox cloud.

An example of the received data is given as follows:

```
{
    "device" : "771E2D",
    "time" : 1771975311,
    "seqNumber" : 954,
    "data" : "020c7a2807e640d0",
    "slt" : 34372136,
    "slg" : 132530384
}
```

When the acquired location information was plotted on Google Maps, an error of approximately 10 m was observed between the GPS positioning (red dot) and the actual physical location (blue dot). In general, GPS positioning errors range from 10 to 30 m due to factors such as ionospheric delay and multipath effects, which align with the findings of this experiment. Therefore, this study establishes a GPS precision of approximately 10 m as the target accuracy for the Sigfox WiFi positioning system.



Fig. 2 *IoT GPS device displayed on Google Maps.*

3. Positioning using Sigfox: Mechanisms and Challenges

This section discusses the mechanisms of positioning using the Sigfox network, its accuracy, and the inherent challenges.

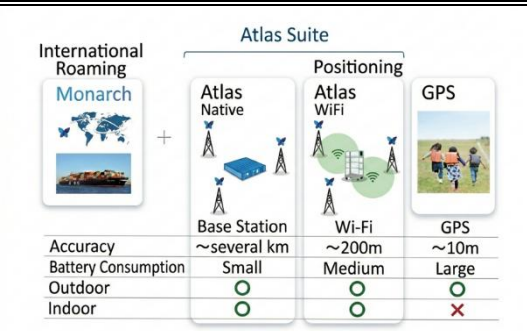


Fig. 3 Comparison of Sigfox Atlas Native and WiFi.

A. Atlas Native Positioning (Accuracy: Several Kilometers)

Atlas Native is a location measurement service that communicates with Sigfox base stations to provide geographical coordinates (latitude and longitude) and an estimated error range. This information is relayed via callbacks or the standard Sigfox API based on factors such as signal strength from transmitted messages. Although it is extremely low-power and does not require additional modules like GPS, its accuracy is limited to several kilometers, allowing only for rough area estimations. In this experiment, the indicated area spanned 90 km east-west and 110 km north-south.

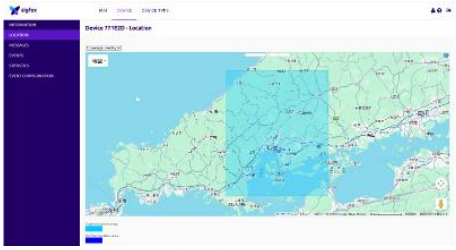


Fig. 4 *Mechanism of Sigfox Atlas location estimation.*

B. Atlas WiFi and Custom Sigfox WiFi Positioning (Here API Integration)

Sigfox's official Atlas WiFi service enables low-power acquisition of location information both indoors and outdoors by utilizing surrounding WiFi access points (BSSID/MAC addresses), provided the area has Sigfox coverage. It offers latitude, longitude, and an expected accuracy range based on MAC addresses, facilitating location identification even in environments where GPS signals are unavailable, such as indoors or in urban canyons.

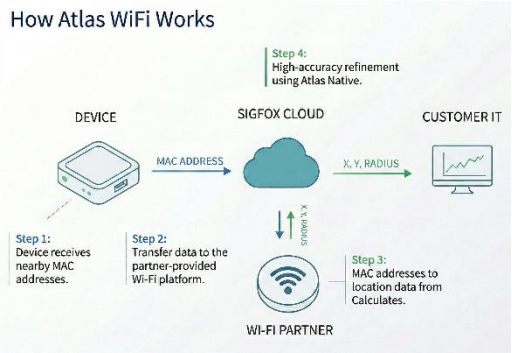


Fig. 5 *Mechanism of Sigfox Atlas WiFi.*

Instead of relying on the existing Atlas WiFi service, this research developed and validated a proprietary mechanism[7] that directly stores MAC addresses acquired from the device into the Sigfox message payload and queries the Here Positioning API through a custom backend server (e.g., AWS). Due to Sigfox communication specifications, the maximum payload for a single message is strictly limited to 12 bytes, allowing for the inclusion of only two MAC addresses (six bytes each). This limitation presents significant challenges:

- **Positioning Failure**: The Here API often cannot calculate a location based on only two MAC addresses, resulting in no coordinate data. An analysis of historical experimental data involving 5,431 unique transmissions revealed that approximately 75.3% of all transmissions failed to acquire even a single valid MAC address (resulting in no data), highlighting the extreme difficulty of data acquisition in sparsely populated areas.
- **Large Errors**: Even when a location is successfully identified, the accuracy can vary significantly, ranging from tens to hundreds of meters when relying on only two data points.
- **Effects of Inappropriate APs**: If mobile access points (APs), such as smartphone tethering devices, or APs with registered locations that differ from their actual physical locations are included in the transmission, the center of gravity for location estimation can be severely distorted.

The following visualization presents experimental data obtained using the conventional single-scan method (maximum of two MACs per location, no MAC filtering). As shown, the estimated coordinate points are widely dispersed across a vast area, visually confirming that practical positioning accuracy has not yet been achieved.

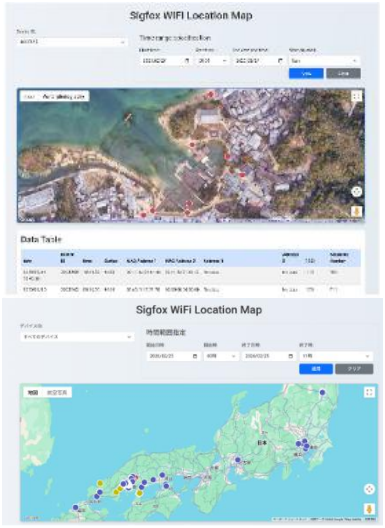


Fig. 6 *Data scattering by conventional methods (max two MACs per location, no MAC filter)*

C. Summary of Challenges

The previous analysis clearly shows that achieving practical accuracy (approximately 10 m, similar to GPS) in custom Sigfox WiFi positioning, while using a minimal number of MAC addresses (two), requires a mechanism to evaluate and prefilter the quality of MAC addresses before transmission.

4. Accuracy improvement method via WiFi MAC address analysis

A. Verification of Continuous Scanning via Walking (Device-side Enhancements)

In this verification, communication and positioning experiments were conducted using a dedicated edge device (shown in Fig. 7) that incorporates an ESP32-S3 chip for WiFi scanning, alongside the aforementioned GPS module.



Fig. 7 *ESP32-S3 edge device*.

Objective: To address payload constraints, the "Continuous Scan" approach, which acquires a maximum of four MAC addresses through overlapping scans in two separate transmissions, was tested for its effectiveness in improving positioning accuracy within a 110-m sectional environment.

Method: Data was collected in a 110-m section (11 points at 10-m intervals) under consistent conditions that simulated stopping every 10 m. Two methods were compared based on the minimum error reported by the Here API: "Single Scan (transmitting a maximum of two MACs per location)" and "Continuous Scan (scanning twice per location to transmit a maximum of four MACs)."



Fig. 8 *110-m section comparison experiment environment*

Summary of Results: In the Single Scan method (two MACs transmitted), the Here API lacked sufficient information to accurately determine the location, resulting in a failure to acquire data at eight of 11 points. At the remaining three points where positioning was successful, the average error was 26.5 m, with a median error of 25.3 m. By contrast, the Continuous Scan method (maximum of four MACs transmitted) successfully determined locations at all 11 points, achieving a 100% acquisition rate. Furthermore, the location information error improved, with an average of 25.8 m and a median of 19.2 m. These results indicate that an accuracy level comparable to standalone GPS can be achieved, and that the dynamic collection of multiple MACs in two stages is an effective strategy for significantly enhancing both the "positioning success rate" and "accuracy" within Sigfox's severe payload constraints.

Figure 9 illustrates the measurement results for the 110-m section at 10-m intervals. True coordinates are indicated by blue pins, single scan results are shown in orange, and continuous scan results are in red, accompanied by their respective accuracy radii. The single scan (orange) exhibited failures at numerous points, resulting in scattered data, whereas the continuous scan (red) is consistently plotted near the true coordinates (blue) at all points, visually demonstrating its superior tracing capabilities as a trajectory.

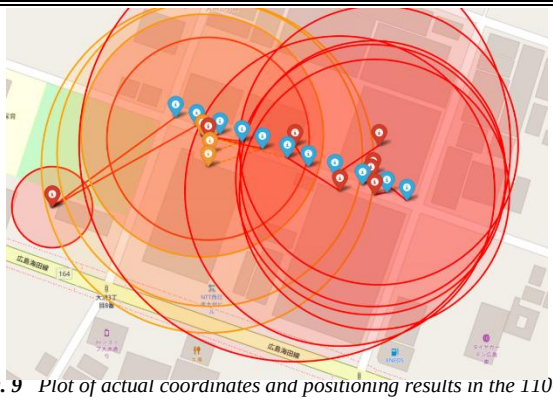
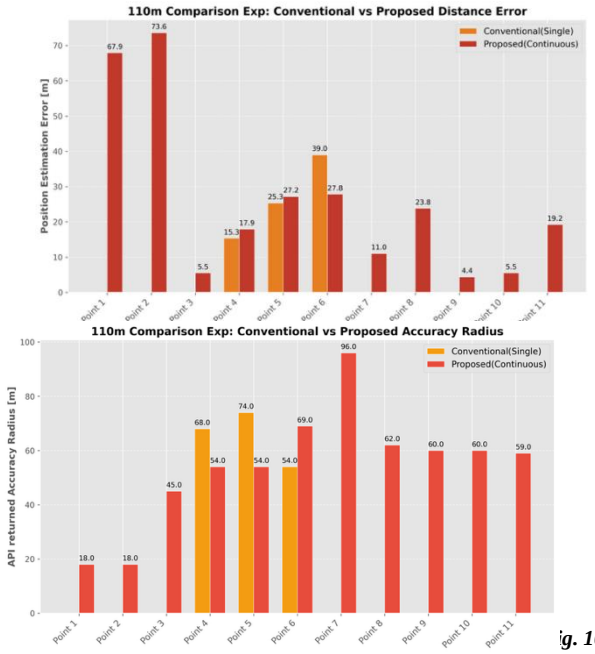


Fig. 9 *Plot of actual coordinates and positioning results in the 110-m section (single vs. continuous).*

Table 1: *presents the actual measured errors for each positioning method across the 110-m section (11 points in total). Entries marked with "-" in the Single Scan column denote points where positioning was not possible, as the Here API did not return coordinate data due to insufficient information.*

Point Name	**Single Scan Error (m)**	**Continuous Scan Error (m)**	**Accuracy Improvement (m) / Notes**
Point 1	-	67.9	Acquisition Improved
Point 2	-	73.6	Acquisition Improved
Point 3	-	5.5	Acquisition Improved
Point 4	15.3	17.9	−2.6
Point 5	25.3	27.2	−1.9
Point 6	39.0	27.8	11.2
Point 7	-	11.0	Acquisition Improved
Point 8	-	23.8	Acquisition Improved
Point 9	-	4.4	Acquisition Improved
Point 10	-	5.5	Acquisition Improved
Point 11	-	19.2	Acquisition Improved
Average	**26.5**	**25.8**	-
Median	**25.3**	**19.2**	-



***ig. 10** Comparison of error and accuracy radius under the positioning method*

Through the implementation of the aforementioned methodology, which relies on operational ingenuity at the Sigfox device end (Continuous Scanning combined with rough received signal strength indicator (RSSI) filtering), the acquisition rate significantly improved, resulting in a positioning accuracy refined to approximately 19 m.

B. Verification of Effectiveness during Continuous Walking (3 and 6 km/h)

Objective: To assess the feasibility of achieving highly accurate positioning by dynamically collecting and merging four MAC addresses while walking continuously at a constant speed.

Method: Positioning data was gathered within the same 110-m segment while moving at "Low-speed Walking (approximately 3 km/h)" and "High-speed Walking (approximately 6 km/h)." Two sequential messages, each transmitting two MAC addresses, were merged into a single group (with a maximum of four MACs at the backend). The estimated coordinates and accuracy radius were then calculated using the Here API.

Results and Discussion: Table 2 lists the positioning results for each group during continuous walking. When walking at a low speed of approximately 3 km/h, all three groups achieved successful positioning. Notably, Group 1 demonstrated a high precision, namely, a radius of 18 m, which is comparable to standalone GPS positioning. However, a significant issue was noted: when comparing the estimated coordinates (latitude and longitude) to the actual starting point of the walk, Group 1 exhibited a spatial offset exceeding 50 m. This discrepancy was not due to measurement errors from the device; rather, it indicated that the physical locations of the MAC addresses (access points) used by the Here Positioning API did not align with their true real-world positions. Thus, whereas the API reported a location with "high confidence" (an 18-m accuracy radius), inherent biases or deviations in its reference database remained. By contrast, during high-speed walking at approximately 6 km/h, the accuracy radius for Group 1 increased significantly to 89 m. At this speed, equivalent to 100 m per min (approximately 1.67 m/s), the subject covered the entire 110-m section in just 66 s. Given that the subject had already advanced nearly 100 m during the time spent transmitting multiple messages via Sigfox (including scanning intervals), achieving a location "within an 89-m radius" using dynamically acquired MAC addresses represents a sufficiently high and practically effective accuracy considering the speed of movement.

***Table 2:** Positioning results for each group during continuous walking.*

Walking Speed	Group	Transmission Count	Estimated Lat	Estimated Lng	Accuracy Radius(m)
Approx. 3 km/h	Slow ①	4 MACs	34.38839	132.49108	18
Approx. 3 km/h	Slow ②	4 MACs	34.38852	132.49209	65
Approx. 3 km/h	Slow ③	4 MACs	34.38868	132.49221	68
Approx. 6 km/h	Fast ①	4 MACs	34.38869	132.49235	89

Figure 11 plots the positioning data recorded at each walking speed while continuously scanning along the 110-m route. (The blue line represents the actual 110-m path; green pins indicate Slow Walking; purple pins denote the positioning results and accuracy radius for Fast Walking). The visual representation confirms that data acquisition remains consistently successful even during motion.



***Fig. 11** Positioning results plot during continuous walking (3 and 6 km/h)*

These results validate that the method of merging multiple scans operates effectively while in motion, achieving a reasonable positioning accuracy (ranging from tens of meters to approximately 10 m) relative to walking speed.

To address coordinate fluctuations (discrepancies in accuracy between groups) caused by walking, to achieve a more advanced and consistent improvement (ensuring errors remain under 10 m), and to cope with mobile APs (e.g., smartphone tethering), detailed data accumulation and thorough analysis on the backend (e.g., AWS) are essential. Therefore, to capture nuanced WiFi data that the device alone cannot interpret and to utilize this for constructing whitelists, the custom Here API Validation Tool was developed.

C. Detailed Investigation of the WiFi Environment using NetSpot

To enhance the accuracy of the Here Positioning API, NetSpot was employed to conduct thorough scans of the WiFi environment in the target area. This process effectively gathers the MAC addresses, SSIDs, and RSSI profiles of all nearby access points.

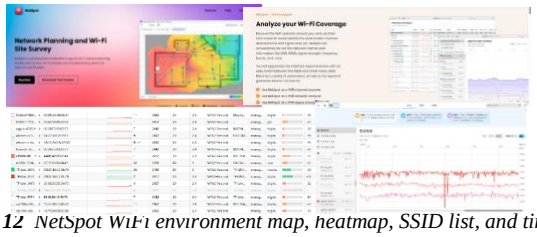


***Fig. 12** NetSpot WiFi environment map, heatmap, SSID list, and time-series graph*

From the data collected from NetSpot, such as the quantitative distribution of APs detected at each location, the localized signal

strength patterns, and the temporal stability of APs, the essential baseline data was established for the subsequent analytical phases.

D. Development of the Here API Validation Tool (WiFi Positioning Validator)

To support whitelist registration and advanced accuracy-enhancement measures in the upcoming backend (e.g., AWS), a robust validation tool (WiFi Positioning Validator) was autonomously developed. This tool automatically queries the Here Positioning API with the MAC address sets obtained from NetSpot and calculates the positional error relative to the actual coordinates. The tool includes the following key features:

- Queries the API with various combinations of MAC addresses to determine the coordinate error compared to the ground truth.
- Visualizes the "Contribution Parameter" for each MAC address, indicating how much each AP affects tracking accuracy.
- Analyzes systemic errors in orientation (directional bias vectors).
- Compiles error statistics from various waypoints into comparative analyses.

Through this validation tool, a thorough accuracy audit was conducted. The complete MAC address repository from the NetSpot data, mapped at the same 11 waypoints used in the Continuous Scan experiment, was input into the live Here API. The interface schematics and detailed analytical visualizations for Waypoint 1 are presented in Figs. 13–15.

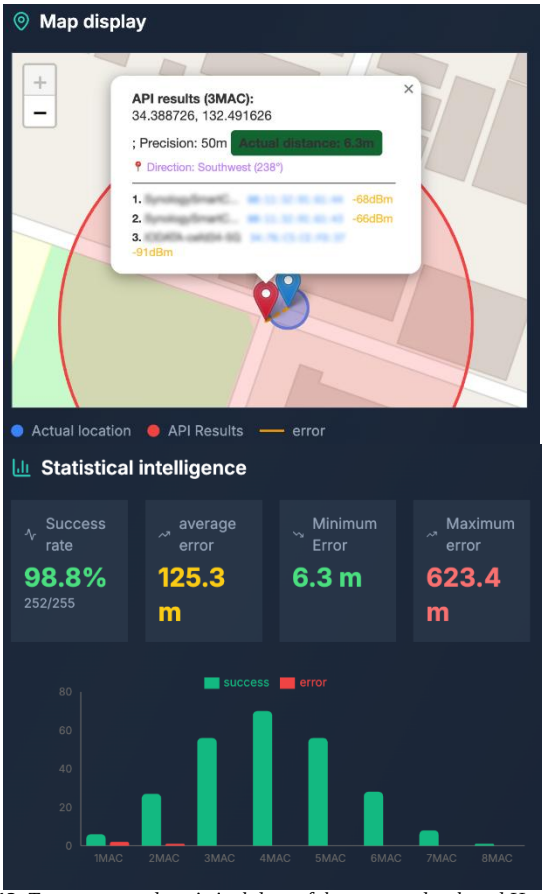


Fig. 13 *Top screen and statistical data of the custom-developed Here API Validation Tool*

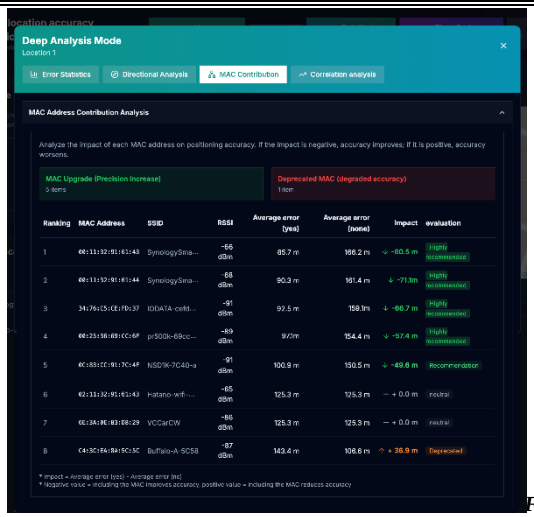


Fig. 14 *Analysis of the MAC address contribution to error at Point 1*



Fig. 15 *Bearing analysis and detailed statistical graphs at Point 1*

E. Implementation of Proactive Prefiltering on the Device (ESP32-S3)

To improve positioning accuracy at the device level, unique prefiltering and screening architectures were integrated and validated on the edge device (execution environment: ESP32-S3) prior to transmission over Sigfox.

Absolute Exclusion of LAA and Mobile OUIs

- **Mechanism**: During the ESP32-S3 WiFi scanning protocol, randomly generated MAC addresses (LAA: locally administered addresses) and OUIs (vendor codes) associated with mobile devices, which are prone to displacement via tethering, are strictly removed from the MAC cache.
- **Implementation Strategy**: Inspect the first octet of the MAC address to instantaneously identify and discard LAAs (where the second character is 2, 6, A, E). Maintain a dedicated blacklist of mobile OUIs within the ESP32 to continually eliminate MAC addresses originating from mobile terminals that match the blacklist.
- **Empirical Substantiation via the Validation Tool**: The Waypoint 1 audit, conducted with the designated validation tool, conclusively demonstrated that MAC addresses categorized as LAA have no effect on the Here API's positioning calculations (resulting in an error contribution of 0 m). In simpler terms, including these addresses in the limited Sigfox payloads (which allow two to four inputs) constitutes a complete waste of bandwidth. Therefore, a reliable local discard logic on the device side is essential.

Triage Optimization by Apex RSSI

- **Mechanism**: After passing through the filtration process, the most powerful APs, which emit the highest RSSI (signal strength), are prioritized for data transmission.
- **Implementation Evaluation**: Although the ESP32-S3 has sufficient computational bandwidth and memory (SRAM/Flash), the core protocol developed for this validation primarily focuses on applying a simple descending sort based on RSSI values to the filtered array (after LAA/OUI exclusion). Only the top-ranking APs are then used in the payload sequence.

F. Experiment 1: Correlation between MAC Address Volume and Positioning Accuracy

Objective: To quantitatively assess the extreme limits of positioning accuracy defined by the 2-unit Sigfox constraint and to identify the necessary threshold for MAC address inclusion.

Method: The validation tool was used to incrementally increase the volume of MAC addresses injected uniformly across all 11 waypoints, carefully recording and compiling the changes in positioning error.

Summary of Results: Data collected from all 11 waypoints revealed an average error of approximately 112 m when restricted to a 2-unit injection limit, significantly below the expected GPS accuracy of approximately 10 m. Notably, increasing the MAC address payload resulted in a significant improvement in accuracy, peaking at an average error of approximately 36 m, the most stable level of accuracy observed, with the injection of four units. However, raising the throughput threshold to eight units resulted in an average error of approximately 38 m, indicating a plateau (and slight regression) in accuracy improvement. This decline can be attributed to the introduction of distant, weak signal APs (ambient noise) that distort the API's estimation process. This finding underscores that optimal preprocessing logic is not achieved by simply increasing the volume of data but by methodically isolating and querying the optimal echelon of approximately four "superior-quality" MAC networks with maximum broadcast resilience.

G. Experiment 2: Accuracy Elevation through Rigorous MAC Quality Segmentation

Objective: This study aims to evaluate the exact kinetic effects of the "Qualitative Fidelity" of injected MAC addresses on terminal positioning accuracy and to develop optimized selection criteria.

Method: Implement and assess the following three-segment taxonomy to compare raw accuracy differentials.

Segmentation Protocol	Analytical Architecture	Formulated Expectation
RSS Filtration Pivot	Exclusive utilization of top N units displaying paramount RSSI vectors	Primacy applied to hyperstable adjacent arrays
Mobile AP Eradication	Systemic excision of mobile APs classified by chronological manifestation flux models	Eradication of unstable geographical transients
Contribution Diagnostics	Algorithmic calculation of individual MAC liability parameters to ostracize detrimental arrays	Radical deprecation of systemic bias

Summary of Results: Utilizing a foundational RSS filtration baseline (top eight units) combined with rigorous mobile AP excision led to a remarkable increase in accuracy, closely aligning with the 10-m radius target. In addition, applying the "Contribution Diagnostics" protocol from the validation tool to the Waypoint 1 datasets revealed a concerning finding: the integration of specific isolated APs (e.g., C4:3C:EA:...) increased the average coordinate error by a significant 37 m. This serious issue is identified with high certainty as a "corrosive rogue AP," incorrectly logged within the overarching database, and is directly responsible for the "50-m+ locational displacement" documented during the continuous traversal phase near Waypoint 1 in Chapter 4.2. By utilizing this tool, designed specifically to detect and blacklist rogue AP vectors from the operational backend, we can effectively eliminate large-scale systemic deviations from the positioning lifecycle.

H. Experiment 3: Calibration of Systemic Error (Azimuth Bias)

Objective: To create comprehensive error vectors based on azimuth orientation derived from successive positioning outputs and to develop correction matrices that optimize maximum theoretical precision.

Method: Calculate statistical azimuth error vectors from extended multi-point positional outputs to create a dynamic algorithmic bias-correction framework. Assess its overall effectiveness by comparing spatial dispersal matrices before and after calibration.

I. Experiment 4: Spatial Trajectory Filtration Using Chronological Data Clusters

Objective: To address dynamic transit-tracking paradigms by using chronologically ordered positional clusters to eliminate outlier anomalies and enhance tracking precision trajectories.

Method: Conduct iteratively staggered positioning bursts over the Sigfox substrate, then apply trajectory constraints (such as advanced Kalman filters). Analyze whether this temporal matrix yields significantly improved trajectory stabilization compared to isolated point-estimation sequences by incorporating logical movement constraints (e.g., pedestrian velocity limits of ≤ 5 km/h) to physically mitigate erratic positional spikes.

J. Comprehensive Synthesis of Precision Enhancement Frameworks

Based on the empirical findings of this research, the essential methodologies required to overcome the significant Sigfox 2-MAC payload limitations and achieve accuracy levels comparable to GPS paradigms (sub-10-m limits) are outlined as follows:

1. **Device-level Architecture Triage (Continuous Scan Implementation)**: Implement staggered, two-stage sequential scanning clusters combined with macroscopic RSSI truncation at static waypoints to significantly increase MAC throughput (up to four apex-tier inputs), thereby eliminating the high positioning attrition rate seen in single-scan applications and stabilizing coordinate isolation medians at approximately 19 m.
2. **Propulsive Dynamic Scan Integration**: Coupling messages to transmit up to four MAC arrays consistently maintains exceptionally sharp localized accuracy (with minimal extremes at approximately 18 m), matching static orientations, even during movement at speeds of up to 3 km/h.
3. **Backend Algorithmic Tier Segmentation**: Isolate erratic mobile arrays from the device using cloud-based (AWS) whitelisting mechanisms informed by comprehensive WiFi contextual analysis (Here API Validation Tool benchmarks).
4. **Calibration Protocol of Systemic Discrepancies**: Dynamically reduce macro-errors through the azimuth-oriented bias correction framework.
5. **Chronological Filtration Models**: Suppress outlier spikes using dynamic trajectory velocity matrix constraints.

Recently published studies demonstrate that machine learning strategies are improving the accuracy of localized indoor WiFi coordinate estimation to a mean absolute error of approximately 1 m[8]. The advanced filtration techniques and comprehensive modeling approaches discussed here are essential steps toward the next level of location estimation technology.

Integration Potential with the Atlas WiFi Private DB Initiative

In addition to proprietary network applications, a significant opportunity emerges when rigorously verified high-reliability APs are registered in the formal administrative registry of the Sigfox Atlas WiFi Private DB initiative. This could lead to significant improvements in both the overall success rate and accuracy of their official public platform infrastructure.

5. Conclusion

This research addressed the persistent challenges in positioning accuracy caused by the limitations of payload architecture (restricted to two MAC addresses) in custom Sigfox-based WiFi localization systems. The "Continuous Scan" algorithm was introduced as a corrective protocol for devices stationed at fixed waypoints every 10 m. By overlapping multiple transmissions, this protocol enables the dynamic use of up to four MAC addresses, which effectively eliminates the significant failure rate associated with single-scan attempts (which exhibited an approximately 70% failure margin under test conditions) resulting in precise coordinate isolation at each waypoint and stabilizing the overall positional median at approximately 19 m. In addition, empirical tests involving continuous walking at speeds of approximately 3 km/h confirmed that chaining as many as four MAC addresses consistently yields practical and efficient accuracy in dynamic mobile tracking scenarios. This validates the robust functional capability of the system. Critically, data accumulation combined with macro-analytical processing on a centralized backend is vital for achieving greater accuracy levels (maintaining within 10-m thresholds). To this end, we propose a strong MAC address segmentation strategy driven by extensive WiFi environmental analysis using NetSpot, in conjunction with our custom-built Here API Validation Tool. By incorporating the empirical filtering logic validated by this toolset, such as aggressive RSS filtering, elimination of mobile AP profiles, recalibration of systemic biases, and longitudinal tracking models, into the backend network implementations (e.g.,. AWS whitelists) and aligning it with the edge device's continuous scanning infrastructure, we aim to deliver highly stabilized coordinate projection operations that match or exceed the performance of traditional standalone GPS methodologies.

6. References

[1] B. W. Remondi, "Global positioning system carrier phase: Description and use," Bulletin Géodésique, vol. 59, no. 4, (1985) pp. 361-377.

[2] U. Raza, P. Kulkarni and M. Sooriyabandara, "Low power wide area networks: An overview," IEEE Communications Surveys & Tutorials, vol. 19, no. 2, (2017), pp. 855-873

[3] https://docs.arduino.cc/hardware/uno-rev3 (March 1, 2026)

[4] https://unabiz.github.io/unashield/hardware.html (March 1, 2026)

[5] http://akizukidenshi.com/catalog/g/gK-09991/ (March 1, 2026)

[6] T. Tanaka, N. Okamitsu, and K. Nishino, IoT device with Global Positioning System using Sigfox, ThingSpeak, and Anaconda platform, E+E, vol. 55, (2020)pp. 94-98.

[7] M. Watanabe, S. Ohji, T. Ishikawa, Y. Ueda, S. Tada, R. Masumoto, M. Yamauchi, K. Kakugawa, T. Tanaka, and K. Vutova, "Location estimation using Wi-Fi media access control address and device miniaturization", Industry 4.0, vol. 10, no. 2, (2025) pp. 43-46.

[8] H. Esmaeili Gorjan and V. P. Gil Jiménez, "Improving indoor WiFi localization by using machine learning techniques," Sensors, vol. 24, no. 19, (2024) p. 6293.

Innovations in industrial enterprises: the role of AI, IOT, and automation in enhancing competitiveness

Slava Petkova
Technical University of Sofia branch Plovdiv, Bulgaria
E-mail: spet1788@gmail.com

***Abstract**: Industrial enterprises are entering a phase in which competitiveness is increasingly shaped by advanced digital technologies and the organizational capabilities that transform them into measurable outcomes. This paper analyzes how Artificial Intelligence (AI), Internet of Things (IoT), and automation influence operational excellence and how their effects are mediated by leadership, knowledge management capabilities, digitalization maturity and e-HRM systems. Based on the resource-based view and dynamic capability theory, a capability-mediated framework is proposed. Evidence from Tata Steel, Freeport-McMoRan and Siemens demonstrates that technology alone does not ensure competitiveness; value emerges when digital systems are embedded into organizational routines, governance mechanisms and workforce development processes.*
***Keywords**: AI; IOT; AUTOMATION; INDUSTRIAL COMPETITIVENESS; DIGITAL TRANSFORMATION; KNOWLEDGE-ORIENTED LEADERSHIP; E-HRM; HR ANALYTICS; DYNAMIC CAPABILITIES*

1. Introduction

Industrial enterprises are undergoing structural transformation driven by AI, IoT, and automation—technologies increasingly embedded across the industrial value chain, from engineering and manufacturing to operations, supply chains, and workforce systems [32, 23]. While earlier digital initiatives often emphasized automation or IT modernization, the current wave is characterized by AI-enabled decision systems, real-time cyber-physical connectivity, and continuous optimization through learning loops [17, 23].

From a competitiveness standpoint, firms able to sense, interpret, and act on operational data faster can reduce downtime, increase throughput, improve yield, and shorten time-to-market—advantages that accumulate into sustained performance differences when digitalization is managed as an organizational capability rather than a one-time technology investment [1, 10, 24, 33]. Evidence consistently shows that digital transformation is rarely an automatic driver of performance; instead, benefits depend on complementary organizational capabilities, governance, and digitalization management maturity [10, 24, 33].

Research in SMEs and MSMEs emphasizes that digital technologies often work through mediating mechanisms: leadership orientation, knowledge sharing quality, organizational agility, and dynamic capabilities determine whether digital adoption becomes performance improvement [10, 19, 20, 21, 33]. Extending these insights to industrial enterprises, AI/IoT/automation should be treated as strategic resources whose performance impact is contingent on organizational readiness and human systems alignment [1, 27, 33].

A further context is Bulgaria's digital transformation trajectory. Despite relatively strong connectivity, Bulgarian industrial sectors remain below EU averages in digital maturity, with limited integration of Industry 4.0 technologies and persistent capability gaps [2, 3, 7]. This reinforces the need to analyze competitiveness not only through technology adoption but through systemic capability building, ecosystem development, and change management [4, 5, 6].

This paper contributes by integrating:

(1) operational impacts of AI/IoT/automation;

(2) capability-based theories (resource-based view; dynamic capabilities);

(3) leadership and HR digitalization mechanisms (KOL, e-HRM, HR analytics).

The argument is strengthened using industrial cases across steel, mining, and industrial software ecosystems [15, 8, 23].

2. Prerequisites and means for solving the problem

2.1. Conceptual framing: RBV and strategic digital assets

According to the resource-based view (RBV), sustained competitive advantage stems from resources that are valuable, rare, difficult to imitate, and non-substitutable [1]. In industrial enterprises, AI models, IoT sensor networks, industrial data platforms, and automation systems can meet these criteria when they are deeply integrated into production routines and continuously improved through learning cycles via digitalization management and structured capability ownership [33]. Importantly, the "resource" is not only the technology artifact but also embedded know-how: data pipelines, domain-specific feature engineering, model retraining, and decision governance [8, 15].

This explains why similar technology investments can yield different outcomes across firms: performance differences emerge from the capability to convert digital tools into repeatable routines under operational complexity (e.g., mining and steelmaking environments) [15, 8].

2.2. Dynamic capabilities: sensing, seizing, and reconfiguration

Dynamic capability logic complements RBV by emphasizing a firm's ability to sense opportunities, seize them through coordinated action, and reconfigure assets as environments shift [33]. In industrial contexts, this includes iterative deployment, operational experimentation within safe boundaries, and rapid process reconfiguration based on data feedback [8]. Evidence also indicates that agility and digital process maturity act as key enablers and barriers—particularly where digital processes are fragmented, readiness is uneven, and governance is weak [21, 27, 33].

2.3. Knowledge-oriented leadership and knowledge management capabilities

Technology-driven competitiveness depends heavily on leadership systems that cultivate knowledge flow, experimentation, and high-quality knowledge sharing. Knowledge-oriented leadership (KOL) is linked to performance through mechanisms such as knowledge sharing quality, service innovation, and ambidextrous innovation—especially under technological turbulence [20, 19]. In fast-evolving environments, leaders who institutionalize learning loops and protect constructive experimentation are more likely to extract value from digital investments [10, 19].

2.4. Human systems: e-HRM, HR analytics, and adoption readiness

Industrial digitalization requires workforce readiness: skills, engagement, role clarity, and behavioral adoption. Evidence shows that e-HRM tools can improve employee engagement and HRM system effectiveness, thereby supporting organizational effectiveness [25, 31]. HR analytics contributes to organizational effectiveness by enabling evidence-based workforce decisions [34, 29]. Adoption and scaling are shaped by contextual enablers, culture, and readiness [28, 30], and readiness assessment models emphasize that industrial digitalization requires multi-dimensional

preparation (process, people, technology, governance)—not merely installing tools [27].

2.5. Bulgarian industrial context: digital transformation and resistance

Studies on Bulgarian industrial enterprises emphasize gaps in readiness, ecosystem development, and change management [2, 3, 5, 7]. Resistance to change and insufficient capabilities for reengineering projects are reported barriers that can prevent industrial digital initiatives from generating measurable performance outcomes [4]. This supports the capability-mediated perspective adopted in this paper.

3. Solution of the examined problem

3.1. Capability-mediated model

This paper applies an integrative synthesis: peer-reviewed evidence on leadership, digital transformation, and HR digitalization is combined with practitioner case studies on AI deployment in industrial operations. The core proposition is:

AI, IoT, and automation improve competitiveness primarily through capability pathways—digital transformation maturity, leadership orientation, knowledge management capability, and HR digitalization readiness—rather than through adoption alone [10, 24, 33].

Key mediators and enablers include:

- KOL and knowledge sharing quality [20, 19];
- Digitalization management and agile/dynamic capabilities [21, 27, 33];
- HR digitalization readiness and analytics [29, 27, 34];
- Organizational culture and contextual barriers [28, 30].

Table 1: *Capability-mediated competitiveness mechanisms in industrial digital transformation*

No.	Technology	Capability mediator	Competitiveness outcome
1	AI	Data governance; model retraining routines; agile execution; KOL	Higher throughput; lower downtime; better yield and cost efficiency
2	IoT	Real-time monitoring capability; cybersecurity governance; data backbone maturity	Faster detection of drift; fewer failures; improved resilience
3	Automation	Standardized workflows; reskilling and adoption management; e-HRM support	Stable quality; higher consistency; reduced cost of poor quality
4	Digital twins	Simulation + real-time data integration; decision validation routines	Reduced CapEx risk; faster change validation; capacity optimization
5	e-HRM / HR analytics	Workforce readiness; engagement; evidence-based HR decisions	Sustained adoption; faster upskilling; higher organizational effectiveness

3.2. AI, IoT, and automation as operational competitiveness levers

3.2.1. AI: predictive and prescriptive value

AI creates value through prediction (failures, yield, energy consumption) and optimization (recommended control settings, scheduling, resource allocation). Predictive maintenance is a documented use case in industrial environments [14]. At scale, AI improves decision quality by detecting patterns and anomalies in complex process data [17]. The highest value often comes from prescriptive optimization combined with iterative improvement cycles and adoption routines embedded into daily work [8, 15].

3.2.2. IoT: real-time monitoring and risk governance

IoT enables continuous data flows across equipment, processes, and environments. In smart manufacturing, sensor infrastructures support monitoring of equipment condition, process drift, energy waste, and quality deviations [16]. Industry tracking also highlights that IoT adoption is accelerating across industrial operations, reinforcing its role as a baseline enabler for efficiency and data-driven management [11]. These data streams become the backbone for analytics, optimization, and digital twins [23]. IoT connectivity also expands cyber risk exposure, meaning cybersecurity capability is necessary for protecting competitiveness and operational integrity [12].

3.2.3. Automation: quality stability and throughput consistency

Automation improves consistency, throughput, and quality by reducing process variability and human error. In automotive manufacturing, robotics and automation are widely linked to throughput and operational performance improvement [13]. Automated inspection and feedback loops support quality performance and reduce cost of poor quality [18]. Yet automation also intensifies workforce transition requirements, making reskilling and adoption management central to sustained outcomes [22, 25].

3.2.4. Ecosystems and platformization

Competitiveness increasingly depends on participation in digital business ecosystems where platforms enable coordination, co-creation, and scaling across networks [5]. This aligns with Siemens' ecosystem framing of Industrial AI as an integrated operating layer across lifecycle functions [23].

4. Results and discussion

4.1. Case evidence: what works in practice

4.1.1. Tata Steel (Kalinganagar): analytics as a capability system

Tata Steel's Kalinganagar plant illustrates analytics transformation in a high-variance steel process. Analytics models were used to improve control of a critical thermal stage in production, and the transformation required not only model development but workforce capability building, adoption management, and continuous model updating as conditions shifted [15]. The case demonstrates a core lesson: early model success can degrade when input conditions change, requiring retraining, governance, and internal ownership [15]. The plant's capability development contributed to recognition as an advanced Industry 4.0 adopter [15, 32].

4.1.2. Freeport-McMoRan (Bagdad): AI optimization combined with agile execution

Freeport-McMoRan's Bagdad mine demonstrates how AI can unlock gains by challenging entrenched assumptions and turning prediction into optimization. The case highlights cross-functional

teamwork, iterative deployment (MVP logic), and operational experimentation within safety boundaries—an organizational design aligned with dynamic capabilities and agility [8]. The competitiveness mechanism is not the algorithm alone, but the ability to embed model recommendations into daily operational routines and continuously improve the system [8]. This aligns with the principle that organizations should prioritize measurable business improvements over technology novelty, supported by pragmatic change and governance practices [9].

4.1.3. Siemens at CES 2026: Industrial AI operating system, digital twins, and copilots

Siemens' CES 2026 announcements illustrate an ecosystem-scale strategy: Industrial AI as a lifecycle-integrated operating layer spanning design, simulation, manufacturing, and supply chain coordination [23]. Digital Twin Composer is positioned as infrastructure for connecting comprehensive digital twins with simulation and real-time engineering data, enabling virtual decision-making and validation at scale [23]. Siemens also announced industrial copilots embedded across engineering and manufacturing software to reduce workflow friction and accelerate adoption [23]. This supports the argument that competitiveness is reinforced when AI is operationalized into routines and human systems, consistent with evidence on engagement and HR digitalization mechanisms [25, 29].

4.2. Bulgaria: why capability mediation matters even more

Bulgarian enterprise studies highlight that connectivity alone is insufficient for industrial digital maturity. Transformation requires ecosystem strengthening, platform development, and structured change management [2, 3, 5, 7]. Resistance to change and lack of reengineering skills are reported barriers, reinforcing the need for KOL, learning routines, and HR digitalization as adoption infrastructure [4, 19, 20, 27].

4.3. Synthesis: what actually creates competitive advantage

Across literature and cases, three recurring patterns emerge:

•Technology is necessary but insufficient: adoption alone does not guarantee competitiveness [10, 24, 33].

•Embedding beats pilots: value compounds when AI/IoT/automation are institutionalized into daily routines, with governance and iterative improvement [8, 15].

•Leadership and human systems are multipliers: KOL supports innovation under turbulence through knowledge sharing quality and KM capability [20, 19], while e-HRM and HR analytics strengthen adoption, engagement, and workforce readiness [31, 25, 29, 34].

Additionally, the regulatory environment can influence digital transformation performance, particularly for SMEs and ecosystem-level digitalization [26].

5. Conclusion

Industrial competitiveness in the AI era is not determined by technology adoption alone but by the ability to build an integrated capability system: IoT-enabled data infrastructure, AI decision intelligence, automation-based stability and precision, and continuous improvement enabled by agile execution and workforce readiness. Evidence from steel and mining shows that measurable operational gains depend on internal capability ownership, adoption routines, and iterative model improvement—not one-time deployments [15, 8]. Siemens' ecosystem strategy further indicates that Industrial AI is moving toward lifecycle-integrated architectures where digital twins and copilots embed intelligence into daily workflows [23].

The reviewed literature reinforces that knowledge-oriented leadership and knowledge management capabilities enable organizations to convert digital tools into innovation and performance under technological turbulence [19, 20]. Digitalization management and agile/dynamic capabilities mediate outcomes [21, 27, 33], while e-HRM and HR analytics provide the human systems infrastructure required for scaling and sustaining advantage [31, 25, 29, 34]. Regulatory frameworks can further shape digital transformation capacity and outcomes, especially for smaller firms [26].

In sum, industrial enterprises that treat AI/IoT/automation as a strategic capability program—not a technology project—are more likely to achieve durable competitiveness in global markets [1, 10, 33].

References

[1] Barney J. Firm resources and sustained competitive advantage – Journal of Management, 17, 1991, 99–120.

[2] Gigova T., Nikolova-Alexieva V., Valeva K., in: Proc. Int. Conf. Creative Business for Smart and Sustainable Growth (CREBUS), 2019 (IEEE, 2019).

[3] Gigova T., Valeva K., Nikolova-Alexieva V., in: Proc. Int. Conf. Creative Business for Smart and Sustainable Growth (CREBUS), 2019 (IEEE, 2019).

[4] Gigova T., Nikolova-Alexieva V., Valeva K. Resistance against changes in Bulgarian industrial enterprises – reasons and methods for overcoming – AIP Conference Proceedings, 2449, 2022, 070014.

[5] Gigova T., Valeva K., Nikolova-Alexieva V. Development of digital business ecosystems through digital transformation of industrial enterprises – AIP Conference Proceedings, 2980, 2024, 070002.

[6] Nikolova-Alexieva V., Valeva K., Gigova T. Eco-innovation activity of industrial enterprises in Bulgaria – AIP Conference Proceedings, 3274, 2025, 080013.

[7] Nikolova-Alexieva V., Valeva K., Gigova T. State of the digital economy and the place of Bulgaria – AIP Conference Proceedings, 3274, 2025, 080012.

[8] Conger R., Robinson H., et al. Inside a mining company's AI transformation – McKinsey & Company, 2020.

[9] Deloitte. Navigating the Future of Work: Insights on AI and Automation – Deloitte, 2023.

[10] Fahrudi A.N.L.I., et al. Strategic business research evidence on digital transformation and competitiveness mechanisms – Strategic Business Research, 2, 2026, 100036.

[11] Gartner. Top 10 IoT Technologies – Gartner, 2023.

[12] IFC. Cybersecurity for Smart Manufacturing: Protecting Your Investment – International Finance Corporation, 2022.

[13] Industry Week. Toyota drives production with robotics and automation – IndustryWeek, 2023.

[14] Ivanov I., Petrov P. Artificial intelligence for predictive maintenance applications – Applied Sciences, 10, 2020, 7446.

[15] Kumar R., Dubashi N., Gupta R., Singh K. How a steel plant in India tapped the value of data—and won global acclaim – McKinsey & Company, 2019.

[16] Lee J., Bagchi A. Internet of Things (IoT) in smart manufacturing – International Journal of Production Research, 57, 2019, 1–15.

[17] McKinsey & Company. AI in Manufacturing and Operations: Patterns and value creation – McKinsey & Company, 2020.

[18] Nestlé. Nestlé quality assurance in automated processes – Nestlé Corporate Reports, 2022.

[19] Riaz M., Jie W., Ali Z., Sherani M., Yutong L. Do knowledge-oriented leadership and knowledge management capabilities help firms to stimulate ambidextrous innovation: moderating role of technological turbulence – European Journal of Innovation Management, 28, 2025, 235–270.

[20] Sahibzada U.F., Janjua N.A., Muavia M., Aamir S. Knowledge-oriented leadership and organizational performance: mediating role of service innovation, knowledge sharing quality – Journal of Organizational Effectiveness: People and Performance, 11, 2024, 69–89.

[21] Schmid D., Büechl J., Härting R.C., Beldarrain Y., Schwörer R., Naeve J. Opportunities and barriers of agility and digital processes in HR in SMEs – Procedia Computer Science, 246, 2024, 3605–3613.

[22] Schwab K. The Fourth Industrial Revolution, Crown Business, 2016.

[23] Siemens. Siemens unveils technologies to accelerate the industrial AI revolution at CES 2026 – Siemens Newsroom, 2026.

[24] Skare M., de Obesso M., Ribeiro-Navarrete S. Digital transformation and European SMEs: a comparative study using DESI data – International Journal of Information Management, 68, 2023, 102594.

[25] Stachová K., Stacho Z., Šamalík P., Sekan F. The impact of e-HRM tools on employee engagement – Administrative Sciences, 14, 2024, 303.

[26] Succurro M., Donati C. The role of the regulatory framework in enhancing SMEs' digital transformation – International Review of Law and Economics, 83, 2025, 106263.

[27] Tavana M., Sorooshian S., Ashrafidehkordi R., Mina H. A readiness assessment model for human systems management digitalization in industrial organizations – Journal of Industrial Production Engineering, 42, 2025, 47–64.

[28] Theotokas I.N., Lagoudis I.N., Syntychaki A., Prosilias J., et al. Factors affecting e-HRM practices: the role of organizational culture, cultural intelligence, and innovation – Journal of Shipping and Trade, 9, 2024, 13.

[29] Verma S., Rana N., Meher J.R. Identifying the enablers of HR digitalization and HR analytics using ISM and MICMAC analysis – International Journal of Organizational Analysis, 32, 2024, 504–521.

[30] Waheed A., Xiaoming M., Waheed S., Ahmad N., Tian-tian S. E-HRM implementation, adoption and its predictors: a case of SMEs of Pakistan – International Journal of Information Technology Management, 19, 2020, 162–180.

[31] William A.A., Singh K. The assessment of e-HRM tools and their impact on HRM system effectiveness and organizational effectiveness – Electronic Journal of Information Systems in Developing Countries, 89, 2023, e12267.

[32] World Economic Forum. The Future of Jobs Report 2021 – World Economic Forum, 2021.

[33] Zaman S.A.A., Vilkas M., Zaman S.I., Jamil S. Digital technologies and digitalization performance: the mediating role of digitalization management – Journal of Manufacturing Technology Management, 36, 2025, 307–333.

[34] Zeidan S., Itani N. HR analytics and organizational effectiveness – International Journal of Emerging Technologies, 11, 2020, 683–688.

AI ethics education as a tool to minimize risks in medicine

Nikoleta Leventi [1], Alexandrina Vodenicharova [2], Vidin Kirkov [1]
Department of Social and Preventive Medicine and Disaster medicine, Faculty of Public Health „Prof. Dr. Tzekomir Vodenitcharov, PhD", Medical University of Sofia, Sofia, Bulgaria [1]
Department of Bioethics, Faculty of Public Health „Prof. Dr. Tzekomir Vodenitcharov, PhD", Medical University of Sofia, Sofia, Bulgaria [2]

n.leventi@foz.mu-sofia.bg

Abstract: *The rapid integration of artificial intelligence (AI) into medical practice has created unprecedented opportunities for clinical decision support, diagnostics, and patient communication. At the same time, it has introduced new categories of risk, including hallucinated outputs, biased recommendations, privacy breaches, and overreliance by inexperienced practitioners. These risks are amplified in high- stakes environments such as medicine, where erroneous AI- generated information can directly impact patient safety. This paper argues that robust AI ethics education is an essential risks mitigation strategy for future healthcare professionals. Drawing on the results of a survey conducted among medical students, the study examines perceptions of AI use in medicine, the perceived importance of AI ethics education, and expectations regarding its content. Respondents showed clear optimism about the role of generative artificial intelligence (GenAI) in medicine, with strong majorities agreeing that it can positively change clinical practice, improve patient care, and find useful applications. At the same time, the high number of undecided responses (especially regarding patient- care quality) reflects ongoing uncertainty about reliability and safety. Students also strongly emphasized the need for structured AI ethics education, as most agreed that it raises awareness of ethical issues and should involve multidisciplinary expertise. Across all proposed topics, including data privacy, bias, explainability, safety, fairness, and autonomy, students rated the relevance of ethics-related content as high, indicating a clear expectation that medical curricula must prepare them to navigate the respective risks and limitations. By synthesizing these findings with current debates on AI governance and medical safety, the paper positions ethics- driven education as a foundational component of responsible AI deployment in healthcare. The paper argues for embedding AI ethics directly into medical curricula as a proactive risk- mitigation tool. The contribution of this work lies in demonstrating, through empirical evidence, that future clinicians recognize both the promise and the dangers of AI in medicine.*

Keywords: Artificial intelligence (AI), generative artificial intelligence (GenAI), medical ethics education, ethical risks in medicine, AI in healthcare.

1. Introduction

The rapid emergence of artificial intelligence has begun to reshape the landscape of modern medicine. Once limited to rule- based decision support systems and narrow diagnostic tools, AI in healthcare has evolved into a new generation of models capable of producing clinical text, summarizing patient histories, generating differential diagnoses, drafting treatment plans, and even simulating patient interactions [1]. These capabilities have made GenAI attractive to clinicians, educators, and healthcare institutions seeking to improve efficiency, reduce administrative burden, and expand access to medical expertise. As a result, AI systems are increasingly integrated into everyday clinical workflows, from triage chatbots and radiology reporting assistants to tools that support medical students in learning complex subjects [2]. This rapid adoption signals a profound shift: generative models are no longer peripheral technologies but emerging collaborators in the clinical environment [3].

Yet the same qualities that make AI powerful also introduce significant risks when applied in high- stakes medical contexts. Unlike traditional medical software, generative models do not guarantee factual accuracy, consistency, or transparency. They may produce hallucinated clinical information, fabricate references, or generate plausible- sounding but unsafe recommendations. Biases embedded in training data can lead to inequitable treatment suggestions, while privacy concerns arise when models inadvertently reproduce sensitive patient information [4]. Overreliance on AI outputs (particularly among inexperienced practitioners) can further amplify these risks, potentially undermining clinical judgment and patient safety. In medicine, where decisions carry immediate and sometimes irreversible consequences, such failures are not merely technical issues but ethical and professional challenges. The growing presence of AI in clinical settings therefore demands a deeper understanding of how future healthcare professionals perceive these risks and how prepared they feel to navigate them.

Despite the increasing visibility of AI in healthcare, a critical gap persists in medical education, namely, the absence of structured, focused training on AI ethics for minimizing risks. Existing curricula often emphasize technological enthusiasm rather than critical evaluation, leaving students without the conceptual tools needed to assess AI- generated information, identify potential harms, or understand the boundaries of safe use. As AI becomes more embedded in clinical practice, this educational gap becomes increasingly consequential. Without targeted ethics education, future clinicians may lack the ability to recognize when AI outputs conflict with established medical standards, when biases may distort recommendations, or when human oversight is essential [5].

This paper addresses this gap by examining the role of AI ethics- driven education as a mitigation strategy for the safe use of AI in medicine. Drawing on survey data from medical students, it explores how emerging professionals perceive AI's benefits and risks, how they evaluate the importance of AI ethics education, and what content they believe such education should include. By situating these findings within broader debates on trustworthy AI and medical safety, the paper argues that ethics education is not an optional supplement but a foundational requirement for responsible AI integration in healthcare.

2. AI in Medicine: Opportunities and Risks

The emergence of artificial intelligence has opened a new chapter in the digital transformation of healthcare. Unlike earlier AI systems that relied on structured inputs and narrow predictive models, AI can produce coherent clinical text, synthesize complex information, and simulate reasoning processes in ways that resemble human communication. These capabilities have accelerated its adoption across medical education, clinical practice, and administrative workflows. Yet the same generative flexibility that makes these systems powerful also introduces new categories of risk that are particularly consequential in medicine. Understanding both the opportunities and the vulnerabilities is essential for developing responsible strategies for integration [6, 7].

2.1. Clinical applications

AI is increasingly used to support a wide range of clinical tasks. In documentation, models assist with drafting patient histories, summarizing clinical notes, and generating discharge instructions. Help reducing administrative burden and freeing clinicians to focus

on patient care. In diagnostics, AI tools can help generate differential diagnoses, interpret imaging findings, or highlight potential red flags in laboratory results. Medical education has also been transformed, as students use AI to explain complex concepts, simulate patient interviews, and practice clinical reasoning in low- risk environments. These applications demonstrate the potential of AI to enhance efficiency, support decision- making, and expand access to medical expertise, especially in resource- constrained settings [6, 7].

2.2. Documented risks

Despite these benefits, AI introduces significant risks that differ in nature and magnitude from those associated with traditional clinical technologies. One of the most widely recognized issues is hallucination, where the generation of plausible but factually incorrect or clinically unsafe information. In a medical context, hallucinations can lead to inaccurate diagnoses, inappropriate treatment suggestions, or fabricated references that mislead clinicians and students. Bias is another critical concern with models trained on unrepresentative or historically biased datasets that may produce inequitable recommendations, disproportionately affecting marginalized patient groups. Privacy risks arise when models reproduce sensitive patient information embedded in training data or when users input identifiable data into unsecured systems [8].

A further challenge is the lack of transparency of AI models. Their internal reasoning processes are not easily interpretable, making it difficult for clinicians to understand why a particular recommendation was generated or to assess its reliability. This lack of transparency complicates accountability, when an AI- generated suggestion contributes to patient harm.

3. Risks in Medicine and AI Ethics Education as a Mitigation Strategy

3.1. Why medicine is uniquely vulnerable

Medicine is particularly sensitive to the risks of AI for several reasons. First, clinical decisions often involve high stakes, where errors can lead to delayed diagnoses, inappropriate treatments, or irreversible harm. Unlike domains where AI mistakes may be inconvenient or costly, in healthcare they can be life- threatening. Second, medical information is inherently complex, context- dependent, and nuanced. AI models, which rely on statistical patterns rather than true understanding, may struggle with rare conditions or atypical specific health practices. Third, the clinician- patient relationship is built on trust, confidentiality, and professional accountability. Introducing AI into this relationship raises concerns about transparency, informed consent, and the preservation of human oversight [2].

Additionally, medical training emphasizes evidence- based practice, yet AI systems do not always align with established clinical guidelines. Their outputs may blend accurate information with inaccuracies that are difficult to detect, especially for learners. The rapid speed of AI development further complicates, leaving institutions without clear frameworks for safe deployment. Finally, the ethical diversity of global healthcare systems means that AI must operate within varied norms, and expectations [6].

3.2. The ethical foundations

The integration of AI into clinical environments introduces a set of risks that extend beyond technical malfunction and reach deeply into ethical, professional, and societal domains [2]. Because medical decisions directly affect human health and well- being, the ethical foundations guiding AI use in medicine must be exceptionally robust. Ethical principles traditionally applied to healthcare, such as beneficence, non- maleficence, autonomy, justice, and accountability, take on new significance when clinicians interact with generative models capable of producing authoritative- sounding but potentially flawed outputs. These principles provide a framework for evaluating when and how AI systems should be used, what safeguards are necessary, and how responsibility should be distributed between human practitioners and algorithmic tools.

Beneficence and non- maleficence require that AI systems contribute to patient welfare while minimizing the risk of harm. Yet AI models may hallucinate clinical facts, misinterpret symptoms, or produce biased recommendations, challenging clinicians to critically assess AI- generated content rather than accept it at face value.

Autonomy demands that patients remain informed decision- makers, but opaque AI systems complicate transparency and informed consent. Justice requires equitable treatment across diverse populations, yet biases embedded in training data can lead to unequal outcomes. Accountability becomes especially complex when AI tools influence clinical decisions, and even determining who is responsible for errors, as the clinician, the institution, or the model's developers, remains an evolving ethical question. These principles highlight the need for clinicians to understand not only how AI works but also how its limitations intersect with core medical values.

AI ethics education emerges as a practical and necessary mitigation strategy precisely because it equips future healthcare professionals with the skills to navigate these risks. In this way, ethics education does more than raise awareness, it actively strengthens clinical safety. It empowers future clinicians to use generative AI responsibly, critically, and in alignment with the ethical commitments that define medical practice.

4. Survey Methodology

4.1. Participants

The cross-sectional study targeted first- year medical students enrolled in the Faculty of Medicine at the Medical University – Sofia. All students were invited to participate through an online announcement distributed between October 15 and November 30, 2025. Participation was entirely voluntary and anonymous, with no incentives offered and no academic consequences for declining. This approach ensured that students could respond freely and without pressure, thereby enhancing the reliability of the collected data. The sample represents individuals at the beginning of their medical training, a group for whom perceptions of AI and its ethical implications are particularly relevant as they form foundational professional attitudes.

4.2. Instruments

Data were collected using a structured web- based survey adapted from a previously validated instrument developed by Lukas Weidener and Michael Fischer [9, 10, 11]. The survey consisted of 53 items organized into six thematic sections, combining both closed- ended questions and Likert- scale statements.

- Part 1 gathered demographic and educational background information, enabling contextual interpretation of responses.
- Part 2 assessed students' prior exposure to AI- based (chat) applications, particularly conversational or genAI tools.
- Part 3 included statements evaluating students' perceptions of AI use in medicine, covering aspects such as usefulness, reliability, and potential risks.
- Part 4 focused on students' views regarding the teaching of AI in medical education.
- Part 5 explored attitudes toward AI ethics, including the importance of ethical principles and responsible use.

- Part 6 examined the perceived relevance of specific topics that could be included in an AI ethics curriculum.

Items in Parts 3 through 6 were rated on a 5- point Likert scale, allowing respondents to express varying degrees of agreement or perceived importance. The structure of the instrument ensured comprehensive coverage of both experiential and attitudinal dimensions related to AI in medicine.

4.3. Data Collection and Analysis

The survey was administered online to facilitate broad accessibility and to accommodate students' schedules. In total 144 students took participation, of 350 first academic year students of medicine.

5. Survey Results

The results reflect both enthusiasm and caution as students recognize the transformative potential of AI while simultaneously expressing concerns about reliability, safety, and the ethical challenges it introduces. The following subsections present the key findings across three major themes, namely, perceptions of AI in medicine, the perceived importance of AI ethics education, and the relevance of specific topics that should be included in such training.

5.1. Perceptions of AI in Medicine

The survey results show that first-year medical students hold a predominantly positive view of the potential role of artificial intelligence in healthcare. When asked whether AI can *positively change medicine*, *find useful applications*, and *improve the quality of patient care*, the majority of respondents expressed agreement.

Across the three statements, between 49 and 85 students selected "agree," and an additional 17 to 36 selected "strongly agree," indicating a strong overall endorsement of AI's transformative potential. Only a small minority expressed disagreement. A notable proportion remained undecided (particularly regarding AI's ability to improve patient care, where 60 students selected "undecided") suggesting that while enthusiasm is high, some students still lack sufficient exposure or confidence to fully evaluate AI's clinical impact. Overall, the data reflect a student cohort that is optimistic about AI's usefulness in medicine but still cautious about its reliability and real-world implications (see Table 1.).

Table 1: *AI contribution, perceptions of AI in medicine.*

Please indicate your degree of agreement with the following statements: AI in Medicine can...	[1. positively change medicine.]	[2. find useful applications in medicine.]	[5. improve the quality of patient care.]
1: I strongly disagree	4	2	3
2: I disagree	15	6	15
3: Undecided	36	15	60
4: I agree	66	85	49
5: I strongly agree	23	36	17

5.2. Importance of AI Ethics Education

Students demonstrated a clear recognition of the importance of AI ethics education within medical training. When evaluating whether AI ethics education *raises awareness of ethical issues in clinical practice* and *should involve experts from multiple disciplines*, the majority responded favorably. Between 59 and 64 students agreed with the statements, and an additional 30 to 42 strongly agreed, indicating broad support for integrating ethics instruction into the curriculum. Only a small number of respondents disagreed (15–14 students) or strongly disagreed (3–4 students), while a moderate group remained undecided (32–25 students). These results suggest that students not only value ethics education but also expect it to be interdisciplinary, reflecting the complex nature of AI-related challenges in healthcare. The findings highlight a strong student-driven mandate for structured, comprehensive ethics training as a necessary component of preparing future clinicians for responsible AI use (see Table 2.).

Table 2:*Importance of AI ethics education.*

Please indicate your degree of agreement with the following statements: AI ethics education…	[4. contributes to raising awareness for ethical issues in clinical everyday life.]	[7. should involve experts from various fields (e.g., medicine, computer science, philosophy) to ensure a multidisciplinary perspective on AI ethics.]
1: I strongly disagree	3	4
2: I disagree	15	14
3: Undecided	32	25
4: I agree	64	59
5: I strongly agree	30	42

5.3. Expected Content of AI Ethics Education

When asked to assess the relevance of specific topics for inclusion in AI ethics education, students rated nearly all proposed themes as important. Topics such as *data privacy*, *bias*, *explainability*, *safety of AI-based applications*, and *fairness* received particularly high relevance scores, with the majority of responses falling into the "quite relevant" or "very relevant" categories. For example, 47 students rated *data privacy* as "very relevant," and 40 to 46 students rated *safety*, *fairness*, and *explainability* as "quite relevant." Even more conceptually complex topics (such as *informed consent* in the context of AI and *autonomy*) were viewed as relevant by most respondents, though with slightly more variability. Only a small minority rated any topic as "not relevant," and moderate relevance ratings were common across all categories, indicating thoughtful engagement rather than uniform endorsement. Overall, the results show that students expect AI ethics education to address a broad spectrum of issues, combining foundational ethical principles with practical considerations related to risk, transparency, and patient rights (see Table 3.).

6. Discussion with Interpretation of Findings

The findings of this study reveal a student cohort that is both optimistic about the potential of AI in medicine and acutely aware of the risks associated with its clinical use. The strong agreement that AI can positively transform healthcare, improve patient care, and find meaningful applications suggests that students recognize the growing relevance of AI-driven tools in future medical practice. At the same time, the substantial number of undecided responses (particularly regarding AI's impact on patient care) indicates uncertainty, likely coming from limited hands-on experience with AI systems or concerns about their reliability. This underscores the need for structured educational interventions that help students critically evaluate AI outputs rather than rely on intuition or external narratives.

The results also demonstrate a clear demand for AI ethics education. Students agreed that such education is essential for raising awareness of ethical issues and should involve multidisciplinary expertise. This aligns with broader discussions in the literature emphasizing that ethical challenges in AI cannot be addressed solely from a technical or medical perspective. The strong support of diverse curricular topics (including bias, data privacy, explainability, safety, fairness, and autonomy) further illustrates that students expect ethics education to be both conceptually rich and practically oriented.

Overall, the findings suggest that students are not passive recipients of technological change; they are actively seeking the knowledge and ethical grounding needed to navigate AI responsibly. This positions ethics- driven AI education as a crucial component of preparing future clinicians for safe and accountable AI issue

Table 3: *AI contribution, perceptions of AI in medicine.*

Please estimate the relevance of the following topics for education within your medical university studies:	[Informed Consent: Given the complexity of AI, it is questionable whether doctors will be able to understand the technology itself in the clinical	[Bias: The use of AI in medicine can lead to discrimination if the data used for training or programming the AI lack representative ness.]	[Data Privacy: As the use of AI in medicine involves highly sensitive patient data, security gaps or data misuse can have far-reaching consequences	[Explainability: Decisions made by AI-based applications cannot always be traced by the users due to the technical structure and complexity.]	[Safety of AI-based Applications: If AI-based applications are used for medical purposes, such as in diagnosis or treatment decision-making, faulty]	[Fairness: In addition to fairness in terms of equal treatment by the AI-based applications used (e.g., risk of bias and discrimination), access to]	[Autonomy: The use of AI in medicine can limit the autonomy of patients (e.g., regarding the use of AI in their own treatment) and doctors (e.g., i
1: Not relevant	9	9	4	5	3	6	9
2: Slightly relevant	18	28	19	18	11	21	26
3: Moderately relevant	44	38	27	50	38	46	37
4: Quite relevant	53	40	47	40	44	46	52
5: Very relevant	20	29	47	31	48	25	20

Conclusions

This study highlights the importance of integrating ethics- driven AI education into medical education as GenAI becomes increasingly embedded in clinical practice. The survey results show that first- year medical students view AI as a promising tool capable of improving efficiency and patient care, yet they remain cautious about its limitations. Their responses reflect a balanced understanding, with enthusiasm for innovation paired with recognition of the risks posed by hallucinations, bias, opacity, and overreliance. This dual perspective reinforces the need for educational frameworks that equip students to critically assess AI- generated information and apply it responsibly in clinical contexts.

A key insight from the study is the strong student support of AI ethics education. Participants not only affirmed its importance but also articulated clear expectations regarding its content. Topics such as data privacy, fairness, explainability, and safety were consistently rated as highly relevant, indicating that students understand the multifaceted nature of AI- related risks. Their emphasis on multidisciplinary teaching further suggests that effective AI ethics education must bridge medicine with computer science to provide a comprehensive foundation for responsible practice. The findings position ethics education not as an optional supplement but as a necessary mitigation strategy for reducing AI- related risks in medicine.

Future research should explore how AI ethics education can be effectively implemented and evaluated within medical curricula. Further studies could examine how students' perceptions evolve as they gain clinical experience and interact with AI tools in real- world settings. Additionally, comparative studies across institutions and countries would provide insight into how cultural, regulatory, and educational contexts shape attitudes toward AI and its ethical challenges.

Acknowledgement

This publication is part of a project on the topic "Trustworthy and Ethical AI in Healthcare – the Opinion of Future Medical Professionals", funded by the Medical Science Council of the Medical University - Sofia, GRANT-2025, Contract No. D-181/04.06.2025.

References

[1] Leventi, N. (2023). The Adoption of Information Technologies in Health Promotion. ISESIA, 18. Accessed at: https://isesia.fmi.uni-sofia.bg/documents/ISESIA_2023_e-book.pdf#page=24

[2] Leventi, N. (2025). "The Digital Doctor: Navigating the Ethical Landscape for Artificial Intelligence in Healthcare", ISBN: 978-619-7452-35-8, Publisher: Faculty of Public Health "Prof. Dr. Tsekomir Vodenicharov, MD", in Bulgarian.

[3] Rabbani, S. A., El-Tanani, M., Sharma, S., Rabbani, S. S., El-Tanani, Y., Kumar, R., & Saini, M. (2025). Generative artificial intelligence in healthcare: applications, implementation challenges, and future directions. BioMedInformatics, 5(3), 37.

[4] Patias, I. (2026). GPT AS GPT Generative Pre-trained Transformer as General-Purpose Technology, ISBN:978-954-07-6293-7, St. Kliment Ohridski University Press, Sofia, Bulgaria

[5] Leventi, N., Vodenitcharova, A., & Popova, K. (2020). Ethical aspects of the use of innovative information technologies in clinical trials. Proceedings of CBU in Medicine and Pharmacy..., 1, 66.

[6] Vodenitcharova, A., Leventi, N., & Popova, K. (2022). Innovative Information Technologies in Medicine, the Ethical Aspects–Medical Students' Opinion. In ISGT2022 conference" (CEUR Volume 3191, ISSN: 16130073- http://ceur-ws.org/Vol-3191/paper08.Pdf.

[6] Ogut, E. (2025). AI in clinical medicine: challenges across diagnostic imaging, clinical decision support, surgery, pathology, and drug discovery. Clinics and practice, 15(9), 169.

[7] Patias, I., & Georgiev, V. (2020). The Use of Big Data in Medicine and Public Health Policy-Making: Opportunities and Challenges. In Proceedings of the thirteenth International Conference on ISGT'2020, Sofia, Bulgaria (pp. 7-13).

[8] Nastoska, A., Jancheska, B., Rizinski, M., & Trajanov, D. (2025). Evaluating trustworthiness in AI: Risks, metrics, and applications across industries. Electronics, 14(13), 2717.

[9] Weidener, L., & Fischer, M. (2024). Artificial intelligence in medicine: cross-sectional study among medical students on application, education, and ethical aspects. JMIR medical education, 10(1), e51247.

[10] Weidener, L., & Fischer, M. (2023). Teaching AI ethics in medical education: a scoping review of current literature and practices. Perspectives on medical education, 12(1), 399.

[11] Weidener, L., & Fischer, M. (2023). Artificial intelligence teaching as part of medical education: qualitative analysis of expert interviews. JMIR Medical Education, 9(1), e46428.

AI and digital ethics in the age of generative systems-principles, standards and accountability across cultures

Ognen Firfov
University American College Skopje, North Macedonia; e-mail: ognen.firfov@uacs.edu.mk

Abstract: *Generative AI intensifies ethical risks around opacity, bias, responsibility gaps, and cross-cultural legitimacy. This paper synthesizes contemporary literature and proposes a layered governance model with three layers that translates universal ethical principles into culturally adaptive and sector-specific controls. The contribution is a practical pathway from principles to standards, enabling transparency, accountability, and contestability across the AI lifecycle.*
KEY WORDS*: GENERATIVE AI, DIGITAL ETHICS, TRANSPARENCY, ACCOUNTABILITY, STANDARDS, GOVERNANCE*

1. Introduction

Generative artificial intelligence (AI) systems are more and more included in important areas with possible high impact such as public administration, telecommunications, healthcare, education, and finance. Their outputs which are probability based, scale of coverage, and limited interpretability introduce ethical challenges that existing governance regimes struggle to manage [1,2]. A central problem is in fact operational: how to move from widely recomomended ethical principles to implementable operational standards that remain legitimate across different cultures and various contexts.

2. Prerequisites and means for solving the problem

Ethical AI governance draws on wide known recurring principles — transparency, accountability, fairness, privacy, and human autonomy — yet generative AI complicates implementation because harms can be diffuse, emergent, and value-laden [4,3]. Further in this section, the state of the art is reviewed and gaps that motivate the proposed governance model are identified.

2.1 Literature review: principles, practice, and socio-technical governance

Early AI ethics work emphasized that ethical quality is not only a property of algorithms but of socio-technical systems — institutions, incentives, and accountability processes surrounding them [4]. Floridi et al. propose a principled foundation [beneficence, non-maleficence, autonomy, justice, explicability] that has influenced subsequent debates, while also highlighting explicability as a distinct requirement for AI-mediated societies [1].

However, principle proliferation has raised concerns about 'ethics washing' and non-actionability, things would be done on paper and formally but then no effect could be seen later . Mittelstadt argues that principles are too indeterminate without enforcement, metrics, and institutional mechanisms [2]. This critique is reinforced by the observation that different stakeholders require different forms of transparency and justification, making single, universal transparency mandates inadequate [5].

Opacity has been conceptualized and demonstrated in multiple ways. Burrell distinguishes intentional secrecy, technical complexity, and inherent model inscrutability, noting that modern machine learning can be opaque even absent malintent [6]. Interpretable machine learning research frames explainability as a scientific challenge and calls for evaluation criteria to avoid misleading explanations [7]. In governance contexts, the emphasis shifts from full interpretability to auditability and procedural transparency that can support oversight and contestability [5].

Accountability research similarly stresses socio-technical embedding. Binns connects algorithmic fairness to political philosophy and argues that normative commitments (e.g., equality) require explicit choices and trade-offs rather than purely technical fixes [8]. Selbst et al. warn that abstraction can sever fairness methods from social context, producing 'false precision' and governance blind spots [9].

Operational accountability mechanisms are more and more intensively discussed in the context of audits and documentation related to AI related ethics. Raji et al. propose structured auditing processes to close the 'accountability gap' between stated principles and deployed systems, emphasizing organizational controls and independent review [10]. Yeung's concept of algorithmic regulation highlights the institutional design challenges of delegating governance to computational systems and the need for oversight that preserves democratic accountability [11].

Cultural diversity complicates any move toward universal standards which can be applied across the board, also due to the reason that this diversity is multilayered and different in different societies or groups. Jobin et al. find convergence at high-level values across guidelines but significant divergence in interpretation, scope, and enforcement proposals [3]. Ess argues that digital media ethics must take plural moral traditions seriously, not merely as 'local noise' but as sources of legitimacy and societal stability [12]. Hagerty further cautions that global AI ethics can reproduce power asymmetries unless standard-setting includes diverse voices and socio-economic realities [13].

Together, it is quite obvious from the mentioned above that this literature indicates a persistent gap: [i] principles are widely accepted, [ii] implementation mechanisms remain uneven, and [iii] universalization efforts risk cultural illegitimacy. The next section responds by proposing a layered governance model that connects ethical baselines to context-sensitive standards and sector controls.

2.2 Gap statement and design requirements

From the reviewed literature, the governance solution should satisfy four requirements: [R1] enable accountability across the AI lifecycle (design–deployment–monitoring), [R2] provide procedural transparency sufficient for audit and contestability, [R3] maintain interoperability through a minimal universal baseline, and [R4] allow cultural and sectoral adaptation without ethical relativism [2,10,12].

3. Solution of the examined problem

To meet these requirements, we propose a layered governance model for generative AI that translates universal principles into implementable standards and controls. The model is designed to bridge the 'principles-to-practice' gap identified in the literature [2].

The model consists of three layers: [L1] a universal ethical baseline (e.g., human dignity, fairness, accountability, safety); [L2] cultural and regional interpretation [legal traditions, ethical norms, risk tolerances, and institutional capacity]; and [L3] sector-specific operationalization (policies, technical controls, documentation, audits, and redress). Figure 1 presents the already mentioned model.

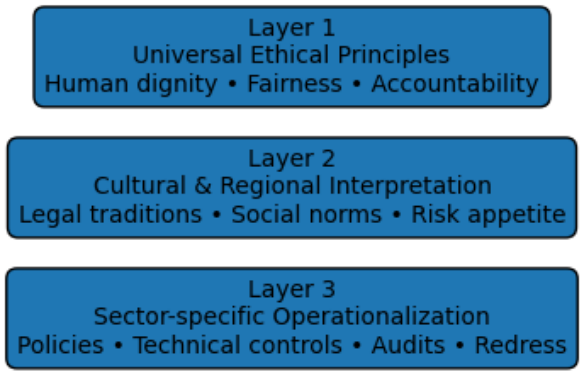


fig. 1. *Layered governance model for ethical generative AI systems*

In practice, L1 functions as a minimum normative floor (the basis) that prevents harmful 'race-to-the-bottom' dynamics, L2 provides legitimacy and interpretive alignment across jurisdictions, and L3 delivers measurable controls aligned with concrete risks (e.g., discrimination, privacy violations, misinformation). This structure supports procedural transparency and auditable accountability without assuming full model explainability [6,5].

4. Results and discussion

Because this study is a conceptual governance analysis, the 'results' are presented as analytical findings derived from synthesis of the literature and design requirements. The core proposed outcome is that a layered approach presented on Figure 1, can operationalize ethics while still keeping alive cultural legitimacy and interoperability.

4.1 Interpretation: transparency as procedural, not total disclosure

A key implication from opacity research is that transparency should be specified by purpose and audience. Rather than total disclosure of model internals in all cases, governance should require procedural transparency: documentation of intended use, training data provenance at an appropriate granularity, known limitations, and evaluation protocols. This aligns with critiques of the 'transparency ideal' and supports contestability and oversight [5,7].

Within the layered model, transparency obligations are distributed: L1 defines the transparency minimum (e.g., explainability proportional to risk), L2 maps this to local regulatory and cultural expectations, and L3 prescribes specific artifacts (e.g., model cards, audit logs, incident reporting). The result is a feasible pathway to oversight even for complex models where interpretability remains limited [6].

4.2 Interpretation: accountability as lifecycle and institutional control

The accountability gap is appearing when responsibility is diffused across creators, developers, deployers, and downstream integrators. Prior work emphasizes organizational governance mechanisms such as independent audits and structured risk assessments [10]. In the layered model, L3 enforces lifecycle controls, including pre-deployment impact assessments, ongoing monitoring, and corrective actions when harms occur. This institutional framing aligns with socio-technical perspectives on machine behavior and governance [14,11].

A practical implication is that accountability should be clearly documented, visible and assignable to a person, position or unit: who(or which unit) approves deployment, who monitors performance drift, and who is responsible for user-facing remedies. These controls support traceability and provide the basis for redress (remedy to set right), an often under-specified element in principle-only frameworks [2].

4.3 Cultural diversity: avoiding both relativism and ethical imperialism

The literature shows that universal principles can be agreed at a high level, yet their operational meaning differs across cultures [3]. L2 explicitly accommodates pluralism by allowing interpretive translation (e.g., how autonomy, consent, and harm are socially understood) while preserving the normative floor defined by L1. This approach aims to avoid ethical imperialism while still preventing relativistic 'anything goes' deployment [12,13].

4.4 Implications and future research directions

Future research should test and refine the layered model via empirical studies: [i] comparative case analyses of generative AI governance across jurisdictions, [ii] evaluation of auditing artifacts and their effectiveness in detecting harm, and [iii] development of metrics for 'procedural transparency' and 'contestability' that can be standardized without forcing uniform moral assumptions. Work is also needed on governance for model supply chains (foundation models, fine-tuning, and downstream deployment) and on culturally legitimate stakeholder participation in standard-setting [10,11].

5. Conclusion

This paper argues that ethical governance of generative AI requires a shift from abstract principles to implementable standards embedded in institutional processes. By integrating universal ethical baselines with culturally adaptive interpretation and sector-specific controls as shown on Figure 1, the layered governance model offers a feasible and desirable direction toward transparency, accountability, and legitimacy.

5.1 Tangible hypotheses

H1: Generative AI deployments that implement lifecycle auditing and documentation [L3] will show lower rates of recurring ethical incidents than deployments relying only on principle statements.
H2: Governance regimes that explicitly include cultural/regional interpretation [L2] will have higher perceived legitimacy among affected stakeholders than regimes that apply uniform standards without adaptation.
H3: A minimal universal baseline [L1] combined with sector controls [L3] will improve cross-border interoperability without increasing ethical risks, compared to purely localized governance.
H4: Procedural transparency artifacts (documentation + audit logs) will enable more effective redress and contestation than technical 'black-box' explanations alone.

These hypotheses can be tested through mixed-method research combining incident datasets, operational audits, stakeholder surveys, and comparative regulatory analysis. If supported, they would substantiate the model's practical impact and provide evidence-based foundations for universal standards that remain culturally legitimate.

6. Synthesis and practical applications

Reviewing the study's insights and view, the layered governance model can be applied as a design template for organizations and regulators. At L1, governmental institutions and commercial entities should define a non-negotiable baseline: commitments to non-discrimination, safety, accountability, and meaningful transparency proportional to risk which is present. At L2, these commitments should be converted into locally legitimate requirements (e.g., consent expectations, data protection norms, acceptable uses). At L3, organizations should implement operational controls: documentation of the model, assessment impact before deployment, monitoring for drift and typical misuse, incident response plans, and user-facing remedy directions.

For high-impact areas (e.g., critical infrastructure, high hazard applications and telecommunications), practical controls include access controls for model outputs, robust logging and hazard logging and retention policies, third-party auditing, and governance committees with cross-functional representation. When it comes to public generative AI systems, practical protection points should contain public disclosure of AI-generated content, abuse monitoring, and typical escalation procedures. The broader impact is a governance architecture that can quite easy scale up and grow with innovation while maintaining ethical accountability and societal trust.

7. References

1. Floridi L. AI4People – An ethical framework for a good AI society – Minds and Machines, 28, 2018, 689–707.

2. Mittelstadt B. Principles alone cannot guarantee ethical AI – Nature Machine Intelligence, 1, 2019, 501–507.

3. Jobin A. The global landscape of AI ethics guidelines – Nature Machine Intelligence, 1, 2019, 389–399.

4. Mittelstadt B. The ethics of algorithms: Mapping the debate – Big Data & Society, 3, 2016.

5. Ananny M. Seeing without knowing: Limitations of the transparency ideal – New Media & Society, 20, 2018, 973–989.

6. Burrell J. How the machine thinks: Understanding opacity in machine learning algorithms – Big Data & Society, 3, 2016.

7. Doshi-Velez F. Towards a rigorous science of interpretable machine learning – arXiv, 2017.

8. Binns R. Fairness in machine learning: Lessons from political philosophy – Proceedings of FAT* (FAccT), 2018, 149–159.

9. Selbst A. Fairness and abstraction in sociotechnical systems, in: Proceedings of FAT* (FAccT), 2019, 59–68.

10. Raji I. Closing the AI accountability gap, in: Proceedings of FAT* (FAccT), 2020.

11. Yeung K. Algorithmic regulation: A critical interrogation – Regulation & Governance, 12, 2018, 505–523.

12. Ess C. Digital Media Ethics, Cambridge, Polity Press, 2020.

13. Hagerty A. Global AI ethics: A review of the social impacts of artificial intelligence – AI & Society, 34, 2019, 497–512.

14. Rahwan I. Machine behaviour – Nature, 568, 2019, 477–486.

Legal regulation problems in transport logistics: challenges and prospects

Irakli Kochadze[1], Ana Mgeladze[2].
Georgian Technical University[1] - Tbilisi, Georgia
Ivane Javakhishvili Tbilisi State University[2] – Tbilisi, Georgia
E-mail: i.kochadze@gtu.ge, ani.mgeladze@yahoo.com.

Abstract: *Transport logistics plays a decisive role in supporting global trade and economic development. The effectiveness of logistics systems largely depends on the quality, coherence, and adaptability of their legal regulation. In recent decades, profound changes in supply chains, technological innovation, and sustainability requirements have revealed significant shortcomings in existing legal frameworks governing transport logistics. This article analyzes the key problems of legal regulation in transport logistics, including fragmentation of legal regimes, challenges of multimodal transport, contractual and liability issues, digitalization, environmental regulation, and enforcement mechanisms. Using doctrinal legal analysis and comparative perspectives, the study identifies systemic weaknesses and proposes directions for regulatory improvement. The findings contribute to ongoing academic and policy debates on the modernization and harmonization of transport logistics law.*

KEY WORDS *TRANSPORT LOGISTICS; LEGAL REGULATION; MULTIMODAL TRANSPORT; LIABILITY DIGITALIZATION; SUSTINABILITY*

1. Introduction

Transport logistics constitutes the backbone of modern supply chains, enabling the movement of goods across domestic and international markets. The growing integration of national economies and the expansion of global trade have increased the importance of efficient and predictable logistics systems. Legal regulation is a fundamental element of such systems, as it defines the rights and obligations of logistics actors, allocates risks, and ensures safety, security, and environmental protection [3].

Despite its central role, the legal framework governing transport logistics remains fragmented and, in many respects, outdated. Many regulatory instruments were developed for traditional, mode-specific transport operations and do not adequately address contemporary logistics practices such as multimodal transport, digital platforms, and integrated supply chain management [1]. As a result, legal uncertainty and regulatory inefficiencies persist.

This article aims to examine the principal legal regulation problems in transport logistics and to assess their implications for market participants and public authorities. It further explores potential directions for legal reform and harmonization.

2. Methodological and theorical framework

The study employs a doctrinal legal research methodology, focusing on the analysis of international conventions, national legislation, and academic literature on transport and logistics law. The doctrinal approach is complemented by comparative legal analysis, which allows for the identification of similarities and differences between regulatory frameworks across jurisdictions. This combination makes it possible to assess the coherence, consistency, and practical effectiveness of existing legal regimes, particularly in the context of cross-border and multimodal transport operations.

The theoretical framework of the research is grounded in the principles of legal certainty and economic efficiency. Legal certainty is regarded as a key prerequisite for predictable contractual relations and effective risk allocation in transport logistics, while economic efficiency emphasizes the role of legal regulation in reducing transaction costs and facilitating trade. The theoretical framework is grounded in the concept of legal certainty and economic efficiency, emphasizing the role of law in reducing transaction costs and facilitating trade [6].

3. Functions of Legal Regulation in Transport Logistics

Legal regulation in transport logistics fulfills several interconnected functions. First, it establishes market access rules, including licensing and authorization of carriers and logistics service providers. Second, it governs contractual relations among shippers, carriers, freight forwarders, and intermediaries. Third, it determines liability for loss, damage, or delay of goods, thereby allocating risks among parties [8].

In addition, transport logistics law addresses public interests such as safety, security, environmental protection, and fair competition. Effective regulation should strike a balance between facilitating efficient transport services and safeguarding societal objectives. However, achieving this balance has proven increasingly difficult in the context of globalized logistics networks.

4. Fragmentation of legal frameworks

4.1 Diversity of National Regulations

One of the most persistent and structurally significant problems in the legal regulation of transport logistics is the diversity of national legal systems. Each state maintains its own set of transport, commercial, labor, tax, and administrative regulations, all of which directly influence the organization and execution of logistics operations. In the context of cross-border transport, this regulatory diversity results in increased compliance costs, legal uncertainty, and administrative complexity for logistics operators, carriers, and freight forwarders [10]. Market participants are often required to navigate multiple legal regimes simultaneously, which complicates planning and risk management.

Differences in technical and operational requirements, such as vehicle standards, cabotage regulations, driver working conditions, licensing systems, and mandatory insurance coverage, further complicate international logistics operations. These disparities create barriers to market access and limit operational flexibility, particularly for small and medium-sized enterprises that lack the resources to adapt to diverse regulatory environments. Moreover, the absence or limited application of mutual recognition mechanisms between states exacerbates regulatory fragmentation, undermining legal predictability and impeding the development of integrated and efficient international logistics networks.

4.2 Limitations of International Legal Instruments

International legal regulation of transport logistics is primarily based on mode-specific conventions governing road, rail, air, and maritime transport. These instruments aim to promote legal harmonization by establishing uniform rules on contractual relations, carrier obligations, and liability regimes. However, their practical effectiveness remains limited due to incomplete ratification by states and significant differences in national implementation and interpretation [1]. As a result, the intended uniformity of international transport law is often undermined at the domestic level, leading to inconsistent legal outcomes.

Furthermore, the coexistence of multiple international conventions with distinct scopes, liability standards, limitation amounts, and defenses creates substantial legal complexity, particularly in cases involving multimodal or combined transport operations. When a single logistics chain involves several transport modes, determining the applicable legal regime becomes problematic, especially if the location of loss or damage cannot be clearly identified. This fragmented legal landscape, often described as a "legal patchwork," undermines predictability, increases the risk of disputes, and weakens legal certainty for transport and logistics actors, ultimately limiting the efficiency and reliability of international logistics systems.

5. Multimodal Transport and Regulatory Gaps

5.1 Absence of a Comprehensive Legal Regime

Multimodal transport has become a central component of modern logistics systems, allowing the coordinated use of different transport modes under a single contractual framework and contributing to greater efficiency, flexibility, and reliability in global supply chains. Through the integration of various modes of transport, logistics operators are able to optimize routes, reduce costs, and respond more effectively to operational and market demands.

Despite its growing importance, multimodal transport remains inadequately regulated at the international level, as there is no universally accepted and binding legal regime governing such operations [7]. Most legal systems continue to rely on mode-specific rules derived from existing international conventions, which creates legal uncertainty when determining the applicable liability regime, particularly in cases where the location of loss, damage, or delay cannot be clearly identified. This regulatory gap complicates claims handling and dispute resolution, increases transaction costs, and undermines legal certainty for logistics actors, thereby limiting the full potential of multimodal transport solutions.

5.2 Liability Allocation Issues

A central legal challenge in multimodal transport concerns the allocation of liability among carriers, freight forwarders, and other logistics service providers involved in a single transport operation. When loss, damage, or delay occurs, identifying the responsible party and determining the applicable legal rules is often complex, particularly in situations where the exact location or stage of transport at which the damage occurred cannot be clearly established [8].

The absence of clear and uniform liability standards across different transport modes leads to legal uncertainty and inconsistent outcomes in claims handling and dispute resolution. This uncertainty increases transactional risks and costs for market participants and weakens effective risk allocation. As a result, the lack of harmonized liability rules discourages the wider adoption of multimodal transport solutions, despite their potential economic and operational advantages.

6. Contractual Regulation and Liability

6.1 Standard Form Contracts and Inequality of Bargaining Power

Transport logistics contracts frequently rely on standard terms and conditions developed by carriers or large logistics service providers. Such standardization plays an important role in enhancing efficiency, reducing transaction costs, and facilitating the rapid conclusion of contracts in complex logistics operations. However, the widespread use of standard form contracts may also reinforce structural imbalances of bargaining power between contractual parties, particularly where smaller shippers or subcontractors are involved [1].

In practice, shippers and subcontractors often have limited opportunities to negotiate key contractual terms, including liability limitations, exclusions of responsibility, choice of law, and jurisdiction clauses. These standardized provisions are typically imposed on a "take-it-or-leave-it" basis, which may lead to unequal risk distribution. Legal systems differ significantly in their willingness to intervene in such contractual arrangements, with some jurisdictions offering stronger protective mechanisms than others. This divergence results in uneven levels of legal protection and contributes to further fragmentation of contractual regulation in transport logistics.

6.2 Adequacy of Liability Regimes

Existing liability regimes in transport logistics differ considerably with regard to liability limits, available defenses, and the allocation of the burden of proof. These differences stem from the mode-specific nature of international conventions and their varied implementation at the national level. In many cases, the established liability limits are outdated and no longer correspond to the value of modern, high-value cargo or to the significant economic losses that may result from delay or disruption in contemporary supply chains [7].

This misalignment between legal liability standards and the economic realities of transport logistics undermines effective risk allocation and weakens the protective function of liability regimes. As a consequence, market participants may face inadequate compensation for losses, which can reduce trust in transport services and discourage investment in efficient and innovative logistics solutions.

7. Digitalization and Legal Challenges

7.1 Electronic Transport Documents

Digitalization has significantly transformed transport logistics through the increasing use of electronic contracts, electronic bills of lading, and real-time tracking and information systems. These technologies enhance transparency, speed, and efficiency within logistics operations and contribute to better coordination across supply chains. However, despite technological progress, many legal frameworks continue to prioritize or formally require paper-based transport documentation, which limits the full realization of digital benefits [9].

The lack of uniform legal recognition and acceptance of electronic transport documents creates barriers to interoperability and legal certainty, particularly in cross-border logistics operations involving multiple jurisdictions. Divergent national approaches to electronic documentation result in inconsistent legal treatment, increased administrative burdens, and reduced efficiency gains. This regulatory uncertainty discourages wider adoption of digital solutions and highlights the need for harmonized legal standards that recognize electronic transport documents as functionally equivalent to their paper counterparts.

7.2 Data Protection and Cybersecurity

Logistics operations increasingly depend on extensive data exchange involving both commercial and personal information, including shipment data, customer details, contractual records, and real-time tracking information. This growing reliance on digital technologies exposes logistics actors to significant legal, financial, and operational risks, particularly in the context of inconsistent data protection frameworks and insufficiently developed cybersecurity regulation across jurisdictions [5].

Divergent national approaches to data protection and cybersecurity compliance create uncertainty for cross-border logistics operations and increase vulnerability to data breaches, cyberattacks, and unauthorized access to sensitive information. Ensuring robust data security while simultaneously enabling efficient information sharing and interoperability between logistics stakeholders remains a key regulatory challenge, requiring coordinated legal standards and effective enforcement mechanisms.

8. Environmental Regulation and Sustainability

Transport logistics is a significant contributor to greenhouse gas emissions, air pollution, and other forms of environmental degradation, making it a key focus of contemporary environmental policy. In response to these challenges, states and regional organizations have introduced increasingly strict environmental regulations, including emission standards, carbon reduction targets, and requirements aimed at promoting energy efficiency and the use of cleaner transport technologies [4].

While such regulatory measures are essential for achieving sustainable development goals, divergent national and regional approaches create substantial compliance challenges and legal uncertainty for logistics operators, particularly those engaged in cross-border transport. Differences in environmental standards, reporting obligations, and enforcement practices may distort competition and increase operational costs. Moreover, the absence of harmonized environmental regulations can hinder long-term investment in green technologies and sustainable logistics solutions, underscoring the need for greater coordination and legal convergence in environmental regulation within the transport logistics sector.

9. Safety, Security, and Customs Regulation

Safety and security constitute core objectives of transport regulation, aimed at protecting human life, property, and the integrity of supply chains. However, significant differences in safety standards, control procedures, and enforcement practices across jurisdictions reduce the overall effectiveness of regulatory frameworks. In the context of international logistics, the duplication of inspections, security checks, and administrative procedures often leads to increased delays, higher operational costs, and reduced reliability of transport services [10].

Customs regulation remains a critical bottleneck in global supply chains, particularly in cross-border transport operations. Despite international initiatives to promote trade facilitation and streamline customs procedures, many customs regimes continue to rely on outdated legal frameworks and paper-based processes. Limited coordination and information sharing among customs authorities further exacerbate inefficiencies, impeding the smooth flow of goods and undermining the effectiveness of modern transport logistics systems.

10. Dispute Resolution and Enforcement

Disputes arising in transport logistics frequently involve multiple jurisdictions, diverse contractual arrangements, and the application of different national and international legal regimes. Such complexity is particularly evident in cross-border and multimodal transport operations, where conflicting laws and divergent procedural rules may apply simultaneously. Differences in judicial procedures, evidentiary standards, and enforcement mechanisms significantly complicate dispute resolution and undermine legal certainty for logistics actors [2].

Arbitration and other alternative dispute resolution mechanisms are often regarded as effective tools for resolving transport logistics disputes, due to their flexibility, expertise, and international enforceability. However, the effectiveness of these mechanisms largely depends on their consistent legal recognition and enforcement at the national level. Variations in domestic arbitration laws, court practices, and enforcement standards may limit the practical benefits of alternative dispute resolution, highlighting the importance of harmonized legal frameworks and effective international cooperation in this area.

11. Directions for Legal Reform

Addressing the identified problems requires a coordinated and forward-looking approach. Key reform directions include:

- Harmonization of international transport and logistics law
- Development of a comprehensive legal framework for multimodal transport
- Modernization of liability regimes
- Legal recognition of electronic documents and platforms
- Greater coordination of environmental and safety regulations

Such reforms require cooperation between states, international organizations, and industry stakeholders.

12. Conclusions

The legal regulation of transport logistics faces significant challenges arising from fragmentation, technological change, and sustainability pressures. Existing legal frameworks often fail to provide sufficient certainty, efficiency, and adaptability.

Modernizing and harmonizing transport logistics law is essential to support global trade, promote sustainable development, and ensure fair and efficient logistics markets. Continued academic research and international cooperation will play a crucial role in achieving these objectives.

13. References

[1] Baughen, S. (2015). Shipping Law (6th ed.). London: Routledge.
[2] Born, G. B. (2021). International Commercial Arbitration (3rd ed.). Alphen aan den Rijn: Kluwer Law International.
[3] Branch, A. E. (2009). Global Supply Chain Management and International Logistics. London: Routledge.
[4] European Commission. (2020). Sustainable and Smart Mobility Strategy. Brussels.
[5] OECD. (2020). Enhancing Cybersecurity in the Transport Sector. Paris: OECD Publishing.
[6] Posner, R. A. (2014). Economic Analysis of Law (9th ed.). New York: Wolters Kluwer.
[7] Ramberg, J. (2011). ICC Guide to Incoterms 2010. Paris: ICC Publishing.
[8] Tetley, W. (2008). Marine Cargo Claims (4th ed.). Montreal: International Shipping Publications.
[9] UNCITRAL. (2017). Model Law on Electronic Transferable Records. New York: United Nations.
[10] UNCTAD. (2021). Review of Maritime Transport. Geneva: United Nations.

Integrating ai into insolvency procedures, challenges and opportunities in Albania

Teuta Hoxha
1. "Faculty of Social Sciences" Albanian University of Tirana
2. Faculty of Law and Human Sciences Mediterranenan University of Tirana
Albania .hoxha_t@yahoo.com

Abstract: *Bankruptcy procedures are complex and often slow, requiring extensive documentation, strict deadlines, multiple creditors, significant human resources, and numerous judicial decisions In Albania, these procedures remain predominantly manual; therefore, the use of Artificial Intelligence opens new opportunities to enhance efficiency and transparency in these procedures. The main benefits include reducing the duration and improving the effectiveness of actions, increasing the accuracy and reliability of the process, and generating useful statistics for the formulation of public policies. Nevertheless, challenges remain related to the adoption of an appropriate legal framework, the protection of sensitive data, the transparency of algorithms, and the determination of legal responsibility. AI has the potential to transform bankruptcy procedures in Albania, but it requires a cautious approach that balances technological innovation with the protection of the legal rights of the parties involved. In this regard, harmonization of legislation with European Union standards and its proper implementation represent the most critical issues in this domain.*

Keywords: ARTIFICIAL INTELLIGENCE, BANKRUPTCY PROCEDURES, ALBANIA, LEGAL TECHNOLOGY, DOCUMENT ANALYSIS, DATA PROTECTION, EUROPEAN UNION STANDARDS.

1. Introduction

Bankruptcy proceedings constitute a cornerstone of market economies by providing mechanisms for debt restructuring, liquidation, and creditor protection. Their complexity stems from procedural formalities, documentation requirements, and multi-stakeholder coordination. Empirical and doctrinal literature consistently highlights procedural delays and administrative burdens as systemic challenges across jurisdictions.

In Albania, bankruptcy procedures are still largely manual, reflecting limited digital integration in judicial administration. This context presents an opportunity for technological innovation. Artificial intelligence particularly machine learning and natural language processing is increasingly used in legal contexts to analyze documentation, identify patterns, and support decision-making processes [1-5].

The potential integration of AI into bankruptcy proceedings raises fundamental questions about efficiency, accountability, and legal safeguards [6-10]. This paper evaluates these dimensions through a qualitative doctrinal approach and considers alignment with European regulatory developments.

2. Literature Context

Recent studies highlight the growing intersection between artificial intelligence and legal procedural systems, emphasizing its potential to enhance efficiency, analytical capacity, and transparency [11-13]. Studies examining legal natural language processing demonstrate that AI tools can automate document classification, extract relevant information, and assist in legal reasoning, thereby reducing administrative burdens in complex proceedings. At the same time, research focused on insolvency and bankruptcy systems indicates that machine-learning models are increasingly capable of identifying financially distressed entities and supporting case prioritization, improving resource allocation within courts [14-16]. However, the literature also stresses that technological implementation must address governance concerns, including algorithmic explainability, accountability, and protection of sensitive data [17, 18]. These findings suggest that while AI integration offers meaningful procedural benefits, its adoption must be accompanied by regulatory safeguards and institutional preparedness to ensure fairness and legitimacy in legal decision-making.

2.1 AI and Legal Digitalization

AI's capacity to process extensive legal and financial datasets has driven its adoption in judicial analytics, contract review, and document classification [2]. Large language models and related technologies are transforming legal research and procedural management [1].

2.2 AI in Insolvency and Bankruptcy Systems

Recent research demonstrates AI's ability to predict financial distress and assist courts in prioritizing cases [3]. Digital technologies also contribute to procedural transparency, improved communication, and reduced administrative burden in insolvency systems [4].

2.3 Governance and Accountability Concerns

Despite efficiency gains, legal scholars emphasize challenges related to explainability and accountability in automated systems [5]. Algorithmic opacity may undermine procedural fairness if safeguards are not implemented.

3. Methodology

This study adopts a qualitative doctrinal and analytical research design complemented by conceptual modelling. The objective is not to develop or test a computational AI system but to evaluate the feasibility and implications of integrating artificial intelligence into bankruptcy procedures in Albania from a legal-institutional perspective.

3.1 Research Design

The research is structured around three methodological components:

a. Relevant academic literature, regulatory discussions, and comparative studies on artificial intelligence in legal and insolvency contexts were reviewed. This analysis focused on identifying recurring themes regarding efficiency, transparency, accountability, and procedural safeguards. The findings were interpreted within the context of Albania's procedural environment.

b. Developments in technologically advanced insolvency frameworks were examined to identify best practices and transferable insights. This approach enables evaluation of potential alignment with European regulatory standards and digital governance initiatives.

c. Because empirical procedural datasets are not publicly available, analytical graphs were constructed to illustrate theoretical relationships between AI integration and procedural outcomes. These visualizations serve as heuristic tools to support interpretation rather than empirical validation.

3.2 Conceptual Variables

The analytical modelling considered relationships among the following conceptual variables:

- *Level of AI Integration* - extent of automation in document processing, prediction tools, and decision support
- *Processing Time Index* - relative measure of procedural duration
- *Dataset Volume* - availability of digitized legal and financial records
- *Prediction Accuracy* - reliability of AI-supported analytical outputs

These variables reflect commonly discussed performance indicators in legal technology research.

3.3 Analytical Visualization Approach

Figure 1 and 2 shows the analytical visualization approach which are base respectively on AI integration vs Procedural duration and dataset volume in relation to AI accuracy. The Figures are illustrative simulations intended to support conceptual discussion and policy interpretation rather than represent empirical measurement.

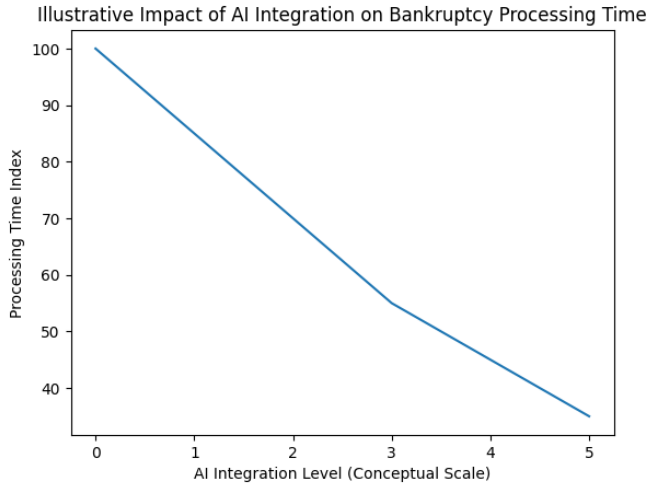


Fig. 1 *AI Integration vs Procedural Duration*

From above figure it has been seen the hypothesized inverse relationship between the level of AI implementation and bankruptcy processing time. Increasing automation is expected to reduce administrative workload and case duration.

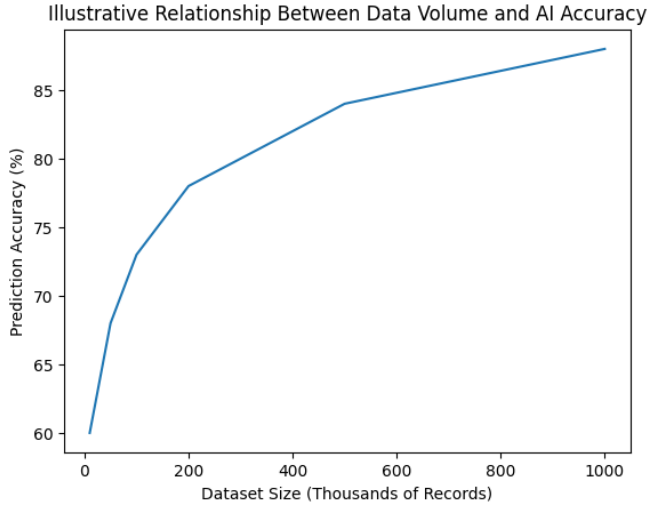


Fig. 2 *Dataset Volume vs AI Accuracy*

It has been demonstrated the expected positive relationship between data availability and predictive reliability of AI tools, emphasizing the importance of judicial digitization.

3.4 Limitations

This methodology presents several limitations:

- Absence of empirical court datasets restricts quantitative validation
- Conceptual modelling cannot capture institutional variability
- Findings depend on interpretation of comparative literature

Despite these constraints, the approach provides valuable theoretical insights relevant for early-stage policy and academic discussion.

3.5 Future Research Directions

Subsequent research could strengthen findings through:

- Collection of Albanian court procedural data
- Empirical testing of machine-learning models
- Surveys of judicial and practitioner readiness
- Pilot program evaluation

4. Opportunities of AI Integration in Bankruptcy Procedures

4.1 Efficiency and Time Reduction

Automation of document classification and information extraction can significantly reduce administrative workloads. NLP tools enable rapid review of filings and financial records, decreasing processing time [2].

4.2 Accuracy and Reliability

Machine-learning models assist in detecting financial irregularities and predicting insolvency risk, improving reliability in procedural evaluation [3]. These tools complement professional judgment rather than replacing it.

4.3 Statistical Insight for Policy Development

Data-driven analytics enable aggregation of procedural statistics, supporting evidence-based policymaking and monitoring of systemic trends. Such insights are valuable for economic governance and judicial reform [4].

4.4 Transparency Enhancement

Digital platforms improve accessibility and traceability of procedural actions, strengthening stakeholder trust and reducing informational asymmetry.

5. Challenges and Risks

5.1 Legal Framework Adaptation

Existing bankruptcy and procedural laws provide limited guidance on AI-assisted decision-making. Clarifying liability and evidentiary status of algorithmic outputs is essential.

5.2 Data Protection

Digitalization requires processing sensitive personal and financial data. Strengthening safeguards aligned with European standards remains a priority for Albania's regulatory environment.

5.3 Algorithmic Transparency

Ensuring explainability of automated recommendations is critical for maintaining procedural fairness and trust [5].

5.4 Accountability and Responsibility

Determining responsibility for AI-supported decisions developers, administrators, or judicial actors remains an unresolved legal challenge.

6. Summary and Conclusion

This study has examined the potential role of artificial intelligence in transforming bankruptcy procedures in Albania, highlighting both opportunities and structural limitations. Bankruptcy processes are inherently complex, involving intensive documentation, strict timelines, coordination among multiple creditors, and substantial judicial oversight. The predominantly manual nature of these procedures within the Albanian context contributes to administrative delays and limits the ability to generate structured data for policy evaluation.

The analysis demonstrates that AI integration could substantially improve procedural efficiency by automating document review, assisting in financial risk assessment, and enabling data-driven monitoring of case progression. Such tools have the potential to reduce case duration, enhance consistency in analytical tasks, and provide statistical insights supporting institutional reform and economic policymaking. Moreover, digital transparency mechanisms could strengthen stakeholder trust and contribute to a more predictable legal environment, which is essential for investment confidence and market stability.

However, the findings also underscore that technological adoption is not merely a technical matter but a governance challenge. Effective implementation requires addressing legal and ethical concerns related to data protection, algorithmic transparency, and accountability. Without a clear regulatory framework defining responsibility for AI-assisted outputs and ensuring procedural safeguards, technological innovation may risk undermining fairness and legal certainty rather than improving it. These considerations are particularly significant in bankruptcy proceedings where the rights of debtors, creditors, and employees must be carefully balanced.

Another central issue concerns institutional readiness. AI systems depend on digitized and standardized data infrastructures, yet gaps remain in electronic record management and data accessibility. Furthermore, adoption requires capacity building among judges, administrators, and practitioners to ensure meaningful interpretation and oversight of AI-generated insights. Investment in technical training and interdisciplinary collaboration will therefore be as critical as legislative reform.

Harmonization with European regulatory developments represents both an obligation and an opportunity. Aligning national legislation with emerging European standards on digital governance and data protection can facilitate responsible adoption while strengthening Albania's integration trajectory. Nonetheless, harmonization must extend beyond formal legal alignment to include effective implementation, monitoring, and institutional adaptation.

In conclusion, artificial intelligence presents a viable pathway for modernizing bankruptcy procedures in Albania, offering measurable gains in efficiency, transparency, and analytical capability. Yet, its successful integration depends on a balanced strategy that combines technological innovation with strong legal safeguards and institutional preparedness. Rather than replacing human expertise, AI should function as a complementary tool that enhances professional judgment and procedural integrity.

Future research should move beyond conceptual evaluation toward empirical pilot testing and stakeholder-based assessments, enabling evidence-driven policy development in this evolving field.

7. References

1. Lai, J.; Gan, W.; Wu, J.; Qi, Z.; Yu, P. S. Large Language Models in Law: A Survey. *arXiv* **2023**.
2. Zhong, H.; Xiao, C.; Tu, C.; Zhang, T.; Liu, Z.; Sun, M. How Does NLP Benefit Legal System: A Summary of Legal Artificial Intelligence. *arXiv* **2020**.
3. Baeck, J.; Arno, H.; Van Ruymbeke, S.; Audenaert, A.; Habils, T.; Mulier, K.; Demeester, T. Developing an AI Model for Detection of Financially Distressed Companies by Courts. *Eur. Insolvency Restructuring J.* **2025**.
4. Mishra, P.; Dhall, A. Investigating the Potential of Digital Technologies in Transforming Insolvency Procedures. *Library Progress Int.* **2024**, *44*, No. 3.
5. Doshi-Velez, F.; Kortz, M.; Budish, R.; Bavitz, C.; Gershman, S.; O'Brien, D.; Scott, K.; Schieber, S.; Waldo, J.; Weinberger, D.; Weller, A.; Wood, A. Accountability of AI Under the Law: The Role of Explanation. *arXiv* **2017**.
6. Hoxha, T.; Olldashi, E.; AL-Thabhawee, G.; Alkattan, H.; & Abotaleb, M. The Role of the Court in the Bankruptcy Process: A Statistical and Comparative Case Study. *International Journal of Innovative Technology and Interdisciplinary Sciences*, **2025**, *8*(4), 1039–1081.
7. Almeida,L. Risk and Bankruptcy Research: Mapping the State of the Art. *Journal of Risk and Financial Management*, **2023**, 16(8), 361.
8. Fu, J.; Yang, X.; Li, Y. Productive Entrepreneurship and the Effectiveness of Insolvency Legislation: A Cross-Country Study. *Small Bus. Econ.*, **2020**, 54, 383–404
9. Gajdosikova, D.; Michulek, J. Artificial Intelligence Models for Bankruptcy Prediction in Agriculture: Comparing the Performance of Artificial Neural Networks and Decision Trees. *Agriculture* **2025**, *15*, 1077.
10. Hoxha, T.; Olldashi, E.; Almahmodi, B. H.; Annuk, A.; Alkattan, H.; & Abotaleb, M. Data-Driven Regression Modelling of Insolvency Outcomes: Judicial Efficiency, Foreign Participation, and Recovery Trends. *International Journal of Innovative Technology and Interdisciplinary Sciences*, **2026**, *9*(1), 172–209.
11. Vásquez-Serpa, L.-J.; Rodríguez, C.; Pérez-Núñez, J.-R.; Navarro, C. Challenges of Artificial Intelligence for the Prevention and Identification of Bankruptcy Risk in Financial Institutions: A Systematic Review. *J. Risk Financial Manag.* **2025**, *18*, 26.
12. Samara, K.; Shinde, A. Bankruptcy Prediction Using Machine Learning and Data Preprocessing Techniques. *Analytics* **2025**, *4*, 22.
13. D'Ercole, A.; Me, G. A Novel Approach to Company Bankruptcy Prediction Using Convolutional Neural Networks and Generative Adversarial Networks. *Mach. Learn. Knowl. Extr.* **2025**, *7*, 63.
14. Matsumaru, M.; Katagiri, H. A Two-Stage Machine Learning Approach to Bankruptcy Prediction: Integrating Full-Feature Modeling and Optimized Feature Selection. *J. Risk Financial Manag.* **2025**, *18*, 662.
15. Omar, M.A.; Gomaa, I.I.; Sabry, S.H.; Moubarak, H. Artificial Intelligence's Role in Predicting Corporate Financial Performance: Evidence from the MENA Region. *J. Risk Financial Manag.* **2026**, *19*, 51.
16. Silva, A.F.d.; Brito, J.H.; Lourenço, M.; Pereira, J.M. Sustainability of Transport Sector Companies: Bankruptcy Prediction Based on Artificial Intelligence. *Sustainability* **2023**, *15*, 16482.
17. Muñoz-Izquierdo, N.; Camacho-Miñano, M.-d.-M.; Segovia-Vargas, M.-J.; Pascual-Ezama, D. Is the External Audit Report Useful for Bankruptcy Prediction? Evidence Using Artificial Intelligence. *Int. J. Financial Stud.* **2019**, *7*, 20.
18. Shetty, S.; Musa, M.; Brédart, X. Bankruptcy Prediction Using Machine Learning Techniques. *J. Risk Financial Manag.* **2022**, *15*, 35.

Review of the Development of Modern Green Energy Generators and Systems

Raul Turmanidze[1], Giorgi Popkhadze[2]

Georgian Technical University (GTU), Faculty of Transportation and Mechanical Engineering, 0171 Tbilisi – Georgia[1,2]
inform@gtu.ge[1], popxadzegiorgi@gmail.com[3]

Abstract: *Green energy (GE) is a type of renewable energy (RE) that is obtained from natural sources that are constantly renewed and do not pollute the environment or do so minimally. The main green energy sources (GES), which are considered the cleanest forms of energy or types of renewable energy sources (RES), are wind, water, sun and earth. While the world, particularly in the most developed countries, has made significant progress in adopting and applying the various forms of green energy (GE), in Georgia this field is in its initial stage and it is not possible to predict when the first major positive developments will be made. Currently, more than 10% of the world's primary energy consumption comes from green energy (GE) technologies, namely: hydropower with approximately 5.1%, followed by wind power with 2.5%, solar power with 1.5% and other types of green energy with 0.9%. The development of cheap, dominant and fast-growing technologies and the development of cheap, intelligent and smart green energy generators and systems will ensure the increase of all types of green energy (GE) in the future. The paper provides a brief review of the historical development of different types of green energy systems, (GES) with a special focus on the historical development of wind turbines (WT). It also provides an review and analysis of several types of intelligent and smart green energy systems (GES), which include: Internet of Things (IoT), intelligent and smart sensors and other components of systems.*

KEYWORDS: *ENERGY, RENEWABLE ENERGY (RE), RENEWABLE ENERGY SOURCES (RES), GREEN ENERGY (GE), GREEN ENERGY GENERATORS (GEG), GREEN ENERGY SYSTEMS (GES), SUSTAINABLE ENERGY (SE), CLEAN ENERGY (CE).*

1. Introduction

Energy is the capacity of a body or system to accomplish work. There are different forms of energy, namely: kinetic, potential, thermal (also known as heat), internal, electrical, chemical, elastic potential energy and many others.

Electricity, as a fundamental aspect of nature, is among the most widely utilized forms of energy. Considering the time possibility of exhaustion, natural or. primary sources (forms) of energy can be divided into [9, 28, 47, 49, 63-66]:

1. Non-renewable energy sources (NRES) and
2. Renewable energy sources (RES).

Non-renewable energy sources (NRES) are sources that are consumed much faster than they can be naturally renewed. This means that they are depleted over time and cannot be regenerated in the short term. Their quantity is limited, so they are considered unsustainable in the long term. These are: Fossil fuels (coal, oil, natural gas, oil shale), Nuclear fuels, etc.

Renewable energy sources (RES) are natural resources (sources) of energy that are constantly replenished and cannot be depleted through use within a reasonable period of time. Their key advantage is that they come from processes in nature that occur continuously, and therefore represent a sustainable and environmentally friendly way of obtaining energy. Unlike non-renewables (oil, coal, gas), they are environmentally friendly and contribute to reducing pollution and CO_2 emissions. These are [9, 28, 37, 47, 49]: Water power or Hydropower energy; Biomass and biogas (including wood and waste); Solar energy; Solar radiation energy; Wind energy; Earth's internal heat (geothermal energy); Tidal energy; Wave energy, etc.

Renewable energy sources (RES) in the last thirty years represent a new trend in the use and production of various forms of energy in all areas.

The largest portion of global energy consumption currently goes to industry, accounting for approximately 50%. Transport follows with approximately 25%, while households and the commercial sector together consume the remaining 25%.

The consumption and production of energy worldwide are steadily increasing each year, closely linked to the continual growth of the global human population. Currently, the overwhelming majority of global energy production, (exceeding 80%) comes from non-renewable sources, leading to adverse environmental effects. To address this, sustainable development (SD) principles advocate reducing reliance on fossil fuels in energy production and consumption. Consequently, there is a persistent need to substantially increase the use of renewable energy sources (RES) in the future. This shift is imperative due to the diminishing reserves of non-renewable energy sources (NRES) and their escalating environmental impact observed in recent years.

This means that, globally, there is a growing trend towards generating electricity from renewable energy sources (RES) due to environmental considerations.

In addition to RE, the following terms are often used in the literature: green energy (GE), sustainable energy (SE) and clean energy (CE) [9, 20, 22, 28, 31, 37, 47, 49]. The question arises as to whether there is a difference between: GE, SE, CE and RE energy and what is the difference.

The analysis and application of energy and the energy system in Georgia is presented in the papers [13, 15-17, 23, 46, 57, 64], while the works describe the current state and future prospects for the development and application of various types of renewable energy sources (RES) in Georgia [14, 29, 48, 52, 63, 65-66].

Due to the increasing use of green energy, the aim of this paper is to describe the variety of forms and types of generators and systems for green energy.

2. Green Energy and Green Energy Sources

Green energy (GE) is a type of renewable energy (RE) that is obtained from natural sources that are constantly renewed and do not pollute the environment or do so minimally [8, 18, 21, 52, 65, 66, 68, 71]. It is actually clean (CE), renewable (RE) and sustainable energy (SE).

The main types of green energy (GE) are:

- Solar energy (SE) [25, 51, 63], is a type of energy that uses the energy of the Sun or the sun's rays to obtain useful forms of energy (most often electrical or thermal energy), through solar panels and collectors.
- Wind energy (WE) [4, 27, 55, 56], is a type of renewable or green energy that is obtained by exploiting the movement of air (wind). It occurs as a result of uneven heating of the Earth's surface by the Sun, which leads to differences in pressure and movement of air masses. This process is based on converting the kinetic energy of the wind into electrical energy, most often by means of wind turbines (WT). It is realized in such a way that: The wind moves the blades of the turbine; The blades turn the rotor; The rotor drives the generator and The generator produces electricity.
- Hydropower energy (HPE) [63-66], is a type of energy that is obtained by exploiting the movement or fall of water (rivers, streams, water is stored or accumulated by means of dams and used as needed) in hydropower plants. It is realized in such a way that: Water from the river or reservoir (dam) is directed through pipes; The flow of water drives the turbine; The turbine drives the generator and the Generator produces electricity. With the HPE form of electricity, the production is completely stable and reliable.
- Energy from biomass [63-66], is a type of renewable (RE) or green energy (GE) obtained from organic materials (wood, plants, agricultural waste, etc.).
- Geothermal energy [63-66], is a type of energy that represents heat from the Earth's interior, etc.

The advantages of using green energy (GE) are multiple, compared to other types of energy, such as:

- There is no direct air pollution and carbon dioxide (CO_2) emissions are reduced;
- Reduces the amount of waste;
- It is a type of renewable energy (RE), which means that natural resources are not depleted, unlike oil, coal, gas, etc.;
- Helps fight climate change etc.

Green energy's history spans from ancient civilizations utilizing wind and water power to modern solar and wind technologies. Key milestones include 19th-century solar cell invention, early 20th-century hydroelectricity growth, and accelerating 21st-century adoption driven by climate action, resulting in today's cheap, dominant, and rapidly expanding renewable technologies [7, 60].

On Fig. 1 shows the timeline evolution of wind turbines (WT) for the period 19th century to 2025, with the characteristics and increase power of wind turbines (WT) on a time scale [7].

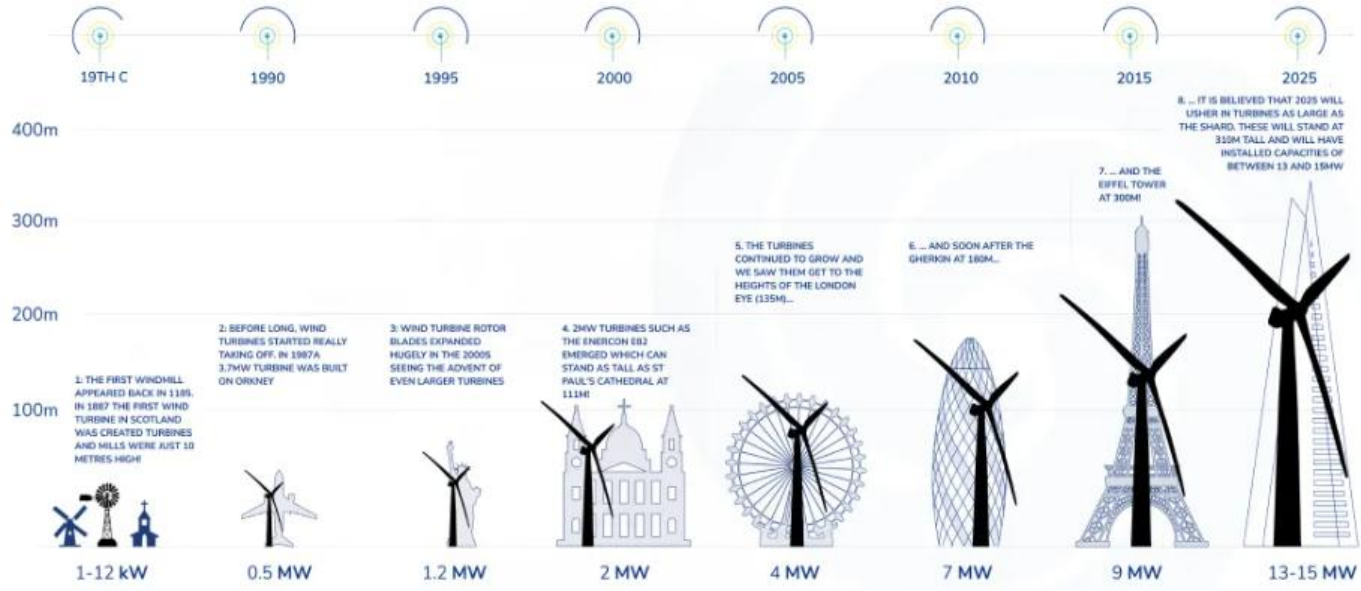


Fig. 1: *The history of Wind Turbines (WT) [7]*

Currently, more than 10% of the world's primary energy consumption for heating, transport and electricity comes from green energy (GE) technologies. Hydropower energy (HPE) accounts for the majority of green electricity production with approximately 51%, followed by wind energy with 25% and solar energy with 15% (Fig. 2).

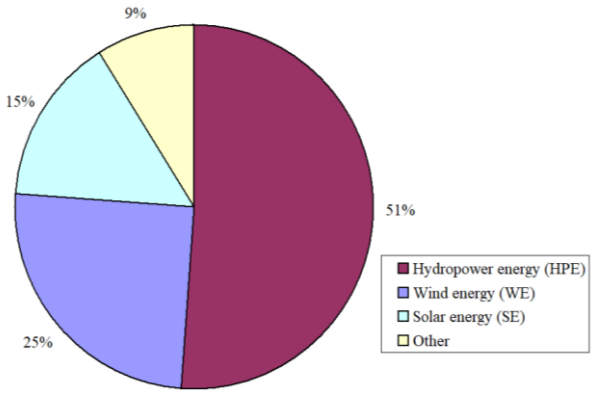


Fig. 2: *Graphics representation the world's primary energy consumption by major types of green energy (GE) technologies*

In the last few decades, the global demand for renewable (RE) and green energy (GE) has been continuously growing, which is mainly driven by environmental protection and sustainable energy development [14, 43].

Currently, many of the world's most developed countries have significantly advanced the adoption of various forms of green energy (GE).

A large number of research and studies describe in detail the current state and future prospects of the development and implementation of various types of renewable energy sources (RES) and green energy (GE) in different countries around the world [9, 11, 33, 36-37, 41, 45, 52-55, 62, 65-66, 69].

The current global political, economic, and energy crisis has led to increased energy prices, worsening economic recession, triggering geopolitical tensions, and worsening environmental problems, with a serious impact on the global economy. This situation in the world encourages scientists, manufacturers, and business people to turn and direct their thoughts to the use of renewable (RE) and green energy (GE).

3. Green Energy Generators and Systems

The energy system (ES) is the entire process of transforming or converting natural energy resources into effective energy required for human social production and life.

Similar to the energy system (ES), for the process of transforming or converting natural green energy resources into efficient green energy (GE) needed for human social production and life uses various types of green energy generators (GEG) and systems (GES) [44, 59].

There are two types of green energy generators (GEG) and systems (GES) and they are:

- Different types of traditional and modern green energy generators (GEG) and systems (GES) that directly depend on the type of green energy (GE) they convert. They all work on the principle of converting natural (renewable or green) energy from the environment into electricity without significant pollution. The most common traditional and modern green energy generators (GEG) and systems (GES) are for:
  - Solar energy (SE): Solar energy generators (SEG) and systems (SES); Solar energy generation systems (SEGS); Solar PV (Photovoltaic) Power Generators (SPPG); Solar PV (Photovoltaic) Systems (SPVS); Solar PV (Photovoltaic) panels (SPVP); Solar thermal power plants (STPP) etc.;
  - Wind energy (WE): Wind turbines (WT); Wind energy generators (WEG) and systems (WES); Wind power generation systems (WPGS) etc.;
  - Hydropower energy (HPE): Hydropower systems (HPS); Hydropower plants (HPP) etc.;
  - Energy from biomass: Biomass systems (BMS); Biogas plants (BGP); Cogeneration power plants (CPP) etc.;
  - Geothermal energy (GTE): Geothermal systems (GTS) Geothermal power systems (GTPS) etc.
- Hybrid green energy generators (HGEG) and systems (HGES).

Recently, a large number of different types of modern solar energy generators and systems have been developed [2-3, 10, 19, 30, 32, 34-35, 38-40, 43, 70], such as:

- Smart solar PV (photovoltaic) generator (SSPG) or Smart solar PV (photovoltaic) system (SSPS), is an advanced version of the classic solar system for electricity production, which uses modern technologies (sensors, software, automation, IoT, AI etc.) to work more efficiently, reliably and economically.
- Sustainable solar PV (photovoltaic) generator (SSPG) or Sustainable solar PV (photovoltaic) system (SSPS), is an electricity generation system that uses solar energy, but is designed to be long-term efficient, environmentally friendly and economically profitable.
- Smart solar PV (photovoltaic) power generator (SSPPG), is a system for the production of electricity or thermal energy from solar energy which integrate Internet of Things (IoT) technologies, Artificial intelligence (AI) and advanced inverters to maximize efficiency, automate maintenance and improve grid stability. These systems allow for real-time monitoring of energy production, weather forecasting, and automatic monitoring to increase power output, often increasing efficiency by over 20%.
- Smart solar energy generator (SSEG) or Smart solar energy system (SSES), is an advanced system for generating and managing electricity from the sun that uses digital technologies, automation and often artificial intelligence to operate more efficiently and economically.

- Sustainable solar energy generator (SSEG) or Sustainable solar energy system (SSES), is a system for the production of electricity or thermal energy from solar energy that is designed to be environmentally friendly, economically profitable and socially beneficial in the long term etc.

Recently, a large number of different types of modern Wind turbines (WT), Wind turbine generators (WTG), wind energy generators (WEG) and systems (WES) have been developed [34, 38, 61], such as:

- Horizontal axis wind turbine (HAWT), is a type of wind turbine (WT) in which the axis of rotation is parallel to the ground (horizontal), and the blades rotate around that axis;
- Vertical axis wind turbine (VAWT), is a type of wind turbine where the axis of rotation is placed upright (vertical) in relation to the ground.;
- Intelligent wind turbine (IWT), is a modern type of wind farm that uses advanced digital technologies to automatically optimize electricity production and system operation.
- Smart wind turbine (SWT), is an advanced wind turbine (WT) that uses modern technologies such as sensors, automation, artificial intelligence (AI) etc. to optimize its operation and electricity production by itself.
- Sustainable wind turbine (SWT), is a wind turbine (WT) that is designed, built and used in a way that minimizes the negative impact on the environment, while ensuring long-term and efficient production of energy from the wind etc.

Hybrid green energy generators (HGEG) and systems (HGES) are a type of green energy systems (GES) that combines multiple sources (e.g. solar + wind + batteries). HGES have gained attention as an effective approach to address the variability and intermittency issues of standalone renewable energy technologies, while increasing the stability and reliability of energy production. In papers is given analysis and applications hybrid green energy systems (HGES) and hybrid renewable energy systems (HRES) [1, 6, 12, 26, 42-45, 58, 67, 71-75].

Several hybrid green energy systems (HGES) will be presented and described below.

The **RES-GES** (RES (renewable energy sources) GES (green energy system)) represents a system of green energy (GE) based on the complementary integration of renewable energy sources (RES) (Fig. 3) [72], which can form different types of integrated energy systems based on different forms of drive devices, input and output energy. The RES-GES system contains a variety of green energy and integrates multiple renewable energy sources (RES): solar, wind and biomass, which contribute to minimizing fossil fuel consumption, improve the richness of energy structure, and promote green environmental protection [71].

In paper [44] is prezentation of study that presents a model and optimization approach for a green energy system through the complementary integration of renewable energy sources (RES-GES). The energy system model incorporates solar, wind and biomass energy, along with distributed functional subsystems. The optimization objective is formulated using environmental sustainability, economic profitability and efficient use of energy as key evaluation criteria.

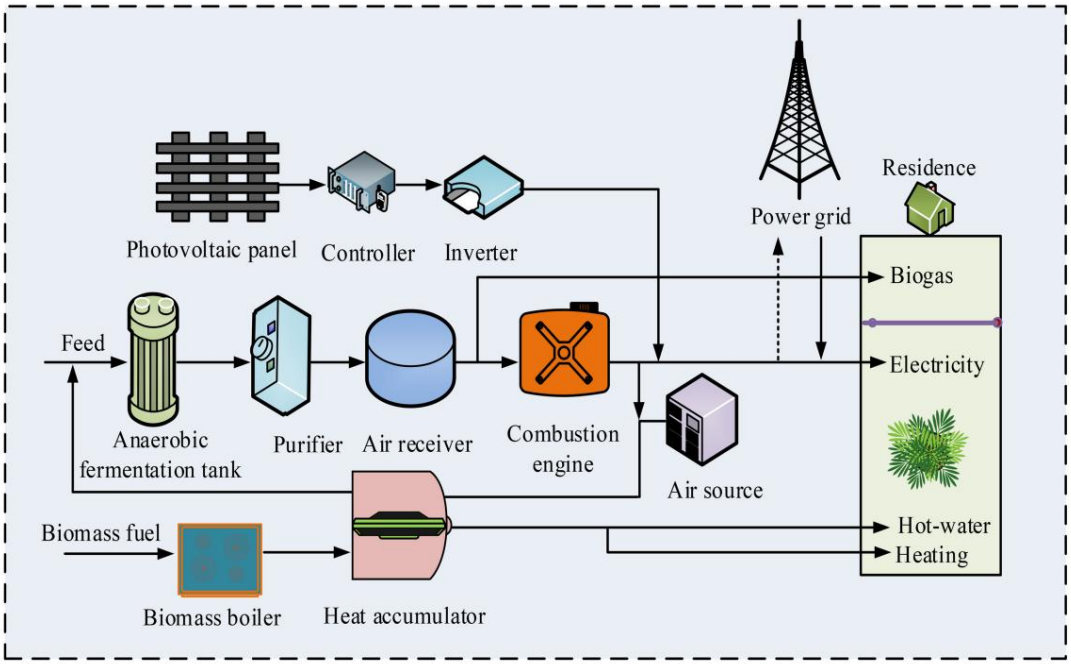


Fig. 3: *A green energy system (GES) model based on complementary integration of renewable energy sources (RES) [44, 72]*

The **HRES-MESS** (HRES (hybrid renewable energy sources) with MESS (mechanical energy storage system)) represents an integral variant of a green energy system (GES) which coupled HRES with the MESS system, thus enhances energy reliability and long-term sustainability [42]. In paper [42] was given of special focus to the primary technical, economic and environmental challenges associated with MESS, including high capital costs, conversion inefficiencies, environmental impacts and site-specific constraints.

The **HRES-PHES** (HRES (Hybrid renewable energy sources) with PHES (pumped hydro energy storage)) represents an integral variant of a green energy system (GES) for improve energy reliability, stability and sustainability, especially in off-grid and remote regions [74]. By utilizing the complementary characteristics of solar and wind energy, together with the large storage capacity of PHES, these systems can effectively reduce fluctuations in renewable energy generation, minimize fossil fuel use, and improve grid independence. An example of an integrated HRES-PHES system is shown in Fig. 4 [42, 74].

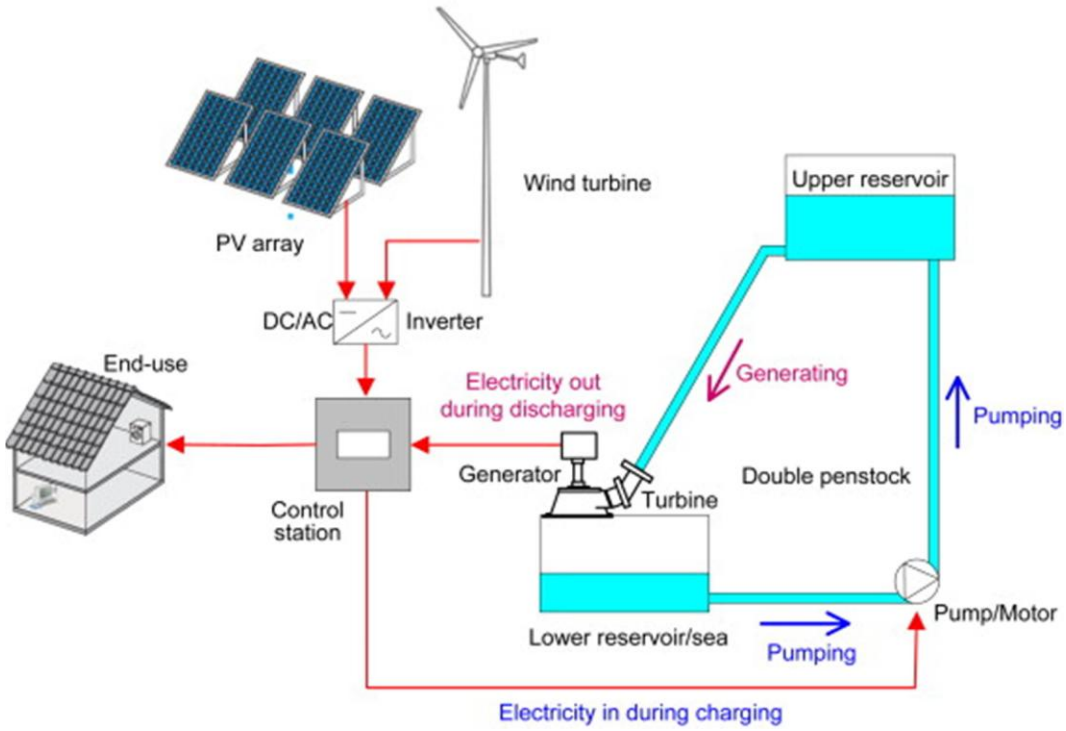


Fig. 4: *Schematic diagram of a photovoltaic wind pumped hydro energy storage system [42, 74]*

Due to the development and application of various variants of electricity generation systems, many authors use various multi-objective optimization (MOO) methods for the selection of green energy sources (GES) and green energy generators (GEG) and systems (GES) [21, 24, 38, 44, 58, 67, 73, 75]. Energy optimisation refers to the optimal allocation of energy structure to determine the optimal energy supply system solution [44].

The development and application of modern smart remote monitoring systems (SRMS) for renewable energy systems (RES) using the Internet of Things (IoT) are presented in [5, 50].

4. Conclusion

The development of modern green energy generators (GEG) and systems (GES) represents one of the key directions of modern technological and social progress. During the last decades, significant progress has been made in the efficiency, reliability and availability of green energy sources (GES), as a part renewable energy sources (RES), such as solar, wind, hydropower, biomass systems etc. The integration of smart technologies, automation, artificial intelligence (AI), Internet of Things (IoT) etc., has further improved the management of energy production and consumption, making these systems more flexible and adaptable to modern needs.

In addition to technical development, economic and ecological aspects also play a significant role, because green energy contributes to reducing the emission of harmful gases, preserving natural resources and mitigating climate change. Although challenges remain, such as energy storage and initial investment, trends clearly indicate that green energy systems (GES) will play an increasing role in the future energy mix.

In this context, further development of innovation, investment in infrastructure and support through regulation and education will be crucial for a complete transition towards sustainable and environmentally friendly energy solutions.

References

[1] Abdellatif, H.; Syed, M.N.; Hossain, M.I. & Abido, M.A.: Standalone hybrid renewable energy system optimization using linear programming. *Arabian Journal for Science and Engineering*, Vol. 48, Issue 5 (May 2023), pp. 6361–6376. ISSN 2193-567X. http://dx.doi.org/10.1007/s13369-022-07363-7.

[2] Aboltins, R. & Blumberga, D.: In search for market-based energy efficiency investment in households: Smart home solutions as an option for optimized use of energy and reduction of costs for energy. *Energy Procedia*, Vol. 147 (2018) pp. 1–6. ISSN 1876-6102. http://dx.doi.org/10.1016/j.egypro.2018.07.026.

[3] Acar, C.; Erturk, E. & Firtina-Ertis, I.: Performance analysis of a stand-alone integrated solar hydrogen energy system for zero energy buildings. *International Journal of Hydrogen Energy*, Vol. 48, Issue 5 (January 2023), pp. 1664–1684. ISSN 0360-3199. https://doi.org/10.1016/j.ijhydene.2022.10.051.

[4] Ackermann, T. & Söder, L.: Wind energy technology and current status: A review. *Renewable and Sustainable Energy Reviews*, Vol. 4, Issue 4 (December 2000), pp. 315–374. ISSN 1364-0321. http://dx.doi.org/10.1016/S1364-0321(00)00004-6.

[5] Adhya, S.; Saha, D.; Das, A.; Jana, J. & Saha, H.: An IoT based smart solar photovoltaic remote monitoring and control unit. In: *Proceedings of the 2nd International Conference on Control, Instrumentation, Energy & Communication (CIEC-2016)*; Kolkata, India; 28-30 January 2016. Piscataway (New Jersey – USA): Institute of Electrical and Electronics Engineers (IEEE), 2016, pp. 432–436. ISBN 978-1-5090-0036-4. https://doi.org/10.1109/CIEC.2016.7513793.

[6] Agarkar, B.D. & Barve, S.B.: A review on hybrid solar/wind/hydro power generation system. *International Journal of Current Engineering and Technology (IJCET)*, Vol. 4, Special Issue 4 (March 2016), pp. 188–191. ISSN 2347-5161. https://doi.org/10.14741/Ijcet/22774106/Spl.4.2016.38.

[7] *A History of Renewable Energy in the UK (2026 Guide)*. By Travis Benn, 30th May 2025. https://lumifyenergy.com/blog/renewable-energy-how-did-we-get-here/.

[8] Anandhi, R.J.; Singh, N.; Palakurthy, D.; Yadav, D.K.; Albawi, A.; Parashar, A.K. & Jithendar, P.: Hybrid solar-hydro-power systems for green energy production: A comprehensive analysis. *E3S Web of Conferences*, Vol. 552 (July 2024), Article no. 01123: pp. 1–11. eISSN 2267-1242. https://doi.org/10.1051/e3sconf/202455201123.

[9] Balat, M.: The use of renewable energy sources for energy in Turkey and potential trends. *Energy Exploration & Exploitation*, Vol. 22, Issue 4 (August 2004), pp. 241–257. ISSN 0144-5987. http://dx.doi.org/10.1260/0144598042886317.

[10] Bhukya, L.; Kedika, N.R. & Salkuti, S.R.: Enhanced maximum power point techniques for solar photovoltaic system under uniform insolation and partial shading conditions: A review. *Algorithms*, Vol. 15, Issue 10 (October 2022), Article no. 365: pp. 1–29. eISSN 1999-4893. https://doi.org/10.3390/a15100365.

[11] Blumberga, A.; Lauka, D.; Barisa, A. & Blumberga, D.: Modelling the Baltic power system till 2050. *Energy Conversion and Management*, Vol. 107 (January 2016), pp. 67–75. ISSN 0196-8904. http://dx.doi.org/10.1016/j.enconman.2015.09.005.

[12] Chen, X.-Y.; Leung, K.C. & Lam, A.Y.S.: Power output smoothing for renewable energy system: Planning, algorithms and analysis. *IEEE Systems Journal*, Vol. 14, Issue 1 (March 2020), pp. 1034–1045. ISSN 1932-8184. https://doi.org/10.1109/JSYST.2019.2958176.

[13] Chomakhidze, D.: Energy balance of Georgia. *Annals of Agrarian Science*, Vol. 14, Issue 3 (2016), pp. 196–200. ISSN 1512-1887. http://dx.doi.org/10.1016/j.aasci.2016.08.002.

[14] Chomakhidze, D.: *The regulation principles of sustainable energy development* (in Georgian). Tbilisi (Georgia): Georgian Technical University (GTU), 2012.

[15] Chomakhidze, D. & Shengelia, G.: *Energy complex of Georgia*. Saarbrücken (Germany): Lambert Academic Publishing (LAP), 2017. – 72 pp. ISBN 978-3-330-04864-5.

[16] Chomakhidze, D.; Tskhakaia, K. & Shamaevi, D.: Twenty years' experience of the regulation of energy in Georgia. *Energy Procedia*, Vol. 128 (2017), pp. 130–135. ISSN 1876-6102. http://dx.doi.org/10.1016/j.egypro.2017.09.027.

[17] Chomakhidze, D.; Tskhakaia, K.; Zivzivazde, L.; Moseshvili, T. & Shamaevi, D.: Electricity balance of Georgia: Trends and prospects. *Energy Procedia*, Vol. 147 (August 2018), pp. 581–587. ISSN 1876-6102. http://dx.doi.org/10.1016/j.egypro.2018.07.074.

[18] Conroy, G.: Data are key to proving green-energy benefits. *Nature*, Vol. 609, Issue 7926 (September 2022), pp. S2–S4. ISSN 0028-0836. http://dx.doi.org/10.1038/d41586-022-02831-4.

[19] Dasi, S.; Kuchibhatla, S.M.; Ravindra, M.; Kumar, K.S.; Chekuri, S.S. & Kavuru, A.K.: IoT-based smart energy management system to meet the requirements of EV charging stations. *Journal of Theoretical and Applied Information Technology (JTAIT)*, Vol. 102, Issue 5 (March 2024), pp. 2116–2127. ISSN 1992-8645.

[20] Dašić, P.: *Scientific and technological trends: Selected scientific-professional papers* (in Serbian). Vrnjačka Banja: SaTCIP Publesher Ltd., 2020. – 305 pp. ISBN 978-86-6075-072-5.

[21] Datta, A.; Ray, A.; Bhattacharya, G. & Saha, H.: Green energy sources (GES) selection based on multi- criteria decision analysis (MCDA). *International Journal of Energy Sector Management*, Vol. 5, Issue 2 (June 2011), pp. 271–286. ISSN 1750-6220. https://doi.org/10.1108/17506221111146020.

[22] Drago, C. & Gatto, A.: Policy, regulation effectiveness and sustainability in the energy sector: A worldwide interval-based composite indicator. *Energy Policy*. Vol. 167 (August 2022), Article no. 112889: pp. 1–13. ISSN 0301-4215. http://dx.doi.org/10.1016/j.enpol.2022.112889.

[23] *Energy system of Georgia*. Available on the Web site: https://www.iea.org/countries/georgia/energy-mix.

[24] Fadaeenejad, M.; Radzi, M.A.M.; Fadaeenejad, M.; Zarif, M. & Gandomi, Z.: Optimization and comparison analysis for application of PV panels in three villages. *Energy Science & Engineering*, Vol. 3, Issue 2 (March 2015), pp. 145–152. ISSN2050-0505. http://dx.doi.org/10.1002/ese3.52.

[25] Goswami, D.Y.; Vijayaraghavan, S.; Lu, S. & Tamm, G.: New and emerging developments in solar energy. *Solar Energy*, Vol. 76, Issue 1-3 (January-March 2004), pp. 33–43. ISSN 0038-092X. http://dx.doi.org/10.1016/S0038-092X(03)00103-8.

[26] Heiskanen, E & Matschoss, K.: Understanding the uneven diffusion of building-scale renewable energy systems: A review of household, local and country level factors in diverse European countries. *Renewable and Sustainable Energy Reviews*, Vol. 75 (August 2017), pp. 580–591. ISSN 1364-0321. http://dx.doi.org/10.1016/j.rser.2016.11.027.

[27] Herbert, G.M.J.; Iniyan, S.; Sreevalsan, E. & Rajapandian, S.: A review of wind energy technologies. *Renewable and Sustainable Energy Reviews*, Vol. 11, Issue 6 (August 2007), pp. 1117–1145. ISSN 1364-0321. http://dx.doi.org/10.1016/j.rser.2005.08.004.

[28] Hoogwijk, M.M.: *On the global and regional potential of renewable energy sources*. Ph.D. Thesis. Utrecht

(Nederlands): Utrecht University, Faculty of Chemistry, 2004. – 256 pp. ISBN 978-90-393-3640-3.
[29] *Industrial production of wind power plants to begin in Georgia following signing of memorandum*. Tbilisi (Georgia): Agenda.ge, 23 March 2022. Available on the Web page: https://agenda.ge/en/news/2022/849.
[30] Jayashri, P. & Raza, F.: Solar smart power generation. *International Journal of Research and Analytical Reviews (IJRAR)*, Vol. 11, Issue 3 (July 2024), pp. 5–9. ISSN 2349-5138.
[31] Jiang, L.; Zou, F.; Qiao, Y. & Huang, Y.: Patent analysis for generating the technology landscape and competition situation of renewable energy. *Journal of Cleaner Production*, Vol. 378 (December 2022), Article no. 134264: pp. 1–21. ISSN 0959-6526. http://dx.doi.org/10.1016/j.jclepro.2022.134264.
[32] Jurasz, J. & Ciapała, B.: Solar–hydro hybrid power station as a way to smooth power output and increase water retention. *Solar Energy*, Vol. 173 (October 2018), pp. 675–690. ISSN 0038-092X. https://doi.org/10.1016/j.solener.2018.07.087.
[33] Khabdullin, A.; Khabdullina, Z.; Khabdullin, A.; Khabdullina, G.; Lauka, D. & Blumberga, D.: Analysis of industrial electricity consumption flexibility: Assessment of saving potential in Latvia and Kazakhstan. *Energy Procedia*, Vol. 113 (2017), pp. 450–453. ISSN 1876-6102. http://dx.doi.org/10.1016/j.egypro.2017.04.037.
[34] Khan, F.A.; Pal, N. & Saeed, S.H.: Review of solar photovoltaic and wind hybrid energy systems for sizing strategies optimization techniques and cost analysis methodologies. *Renewable and Sustainable Energy Reviews*, Vol. 92 (September 2018), pp. 937–947. ISSN 1364-0321. https://doi.org/10.1016/j.rser.2018.04.107.
[35] Kumar, D.: Urban energy system management for enhanced energy potential for upcoming smart cities. *Energy Exploration & Exploitation*, Vol. 38, Issue 5 (September 2020), pp. 1968–1982. ISSN 0144-5987. https://doi.org/10.1177/0144598720937529.
[36] Lazdovska, A. & Jaunzems, D.: Case analysis in Latvia on involvement of end users in energy system. *Energy Procedia*, Vol. 128 (2017), pp. 423–430. ISSN 1876-6102. http://dx.doi.org/10.1016/j.egypro.2017.09.049.
[37] Levitt, C.J. & Lovell, S.: Market share, plant ownership and the merit-order effect of renewable resources: Evidence from Australia's national electricity market. *Utilities Policy*, Vol. 96 (October 2025), Article no. 101990: pp. 1–11. ISSN 0957-1787. http://dx.doi.org/10.1016/j.jup.2025.101990.
[38] Li, Z.; Liu, Y.; Du, M.-P.; Cheng, Y.-H. & Shi, L.: Modeling and multi-objective optimization of a stand-alone photovoltaic wind turbine-hydrogen-battery hybrid energy system based on hysteresis band. *International Journal of Hydrogen Energy*, Vol. 48, Issue 22 (March 2023), pp. 7959–7974. ISSN 0360-3199. https://doi.org/10.1016/j.ijhydene.2022.11.196.
[39] Liu, J.; Chen, X.; Cao, S.-L. & Yang, H.-X.: Overview on hybrid solar photovoltaic-electrical energy storage technologies for power supply to buildings. *Energy Conversion and Management*, Vol. 187 (May 2019), pp. 103–121. ISSN 0196-8904. https://doi.org/10.1016/j.enconman.2019.02.080.
[40] Liu, Y.; Zhao, J.; Wang, L.-Q. & Wang, W.: Unified modeling for multiple-energy coupling device of industrial integrated energy system. *IEEE Transactions on Industrial Electronics*, Vol. 70, Issue 1 (January 2022), pp. 1005–1015. ISSN 0278-0046. https://doi.org/10.1109/TIE.2022.3152019.
[41] Locmelis, K.; Bariss, U. & Blumberga, D.: Latvian energy policy on energy intensive industries. *Energy Procedia*, Vol. 113 (2017), pp. 362–368. ISSN 1876-6102. http://dx.doi.org/10.1016/j.egypro.2017.04.008.
[42] Mahmoud, M.;. Adhari, O.H.K ; Sayed, E.T.; Abdelkareem, M.A. & Olabi, A.G.: Advances in hybrid renewable energy systems coupled with mechanical energy storage: Progress and challenges. *Results in Engineering*, Vol. 29 (March 2026), Article no. 108994: 1–22. ISSN 2590-1230. https://doi.org/10.1016/j.rineng.2026.108994.
[43] Mamodiya, U.; Indra Kishor, Ramakrishna Garine, Priyam Ganguly & Nithesh Naik, : Artificial intelligence based hybrid solar energy systems with smart materials and adaptive photovoltaics for sustainable power generation. *Scientific Reports*, Vol. 15, Issue 1 (May 2025), Article no. 17370: pp. 1–29. eISSN 2045-2322. https://doi.org/10.1038/s41598-025-01788-4.
[44] Mao, W.-W.; Yang, S.-R. & Gao, X.-L.: Modelling and optimisation of green energy systems based on complementary integration of renewable energy. *International Journal of Power and Energy Systems*, Vol. 44, Issue 10 (October 2024), pp. 1–9. ISSN 1078-3466. http://dx.doi.org/10.2316/J.2024.203-0524.
[45] Mulumba, A.N. & Farzaneh, H.: Techno-economic analysis and dynamic power simulation of a hybrid solar-wind-battery-flywheel system for off-grid power supply in remote areas in Kenya. *Energy Conversion and Management*: *X*, Vol. 18 (April 2023), Article no. 100381: pp. 1–21. eISSN 2590-1745. https://doi.org/10.1016/j.ecmx.2023.100381.
[46] Narmania, D.; Chokheli, E.; Kharkheli, M.; Davitaia, S.; Vardiashvili, N.; Makasarashvili, M. & Morchiladze, G.: Energy potential and opportunities in Georgia: The context of making Georgia a regional energy hub. *E3S Web of Conferences*, Vol. 635 (2025), Article no. 01001: pp. 1–7. eISSN 2267-1242. http://dx.doi.org/10.1051/e3sconf/202563501001.
[47] Panwar, N.; Kaushik, S. & Kothari, S.: Role of renewable energy sources in environmental protection: A review. *Renewable and Sustainable Energy Reviews*, Vol. 15, Issue 3 (April 2011), pp. 1513–1524. ISSN 1364-0321. http://dx.doi.org/10.1016/j.rser.2010.11.037.
[48] Patarkalashvili, T. & Mirianashvili, N.: Prospects and challenges of alternative renewable energy resources development in Georgia. *Journal of Earth and Environmental Science Research*, Vol. 4, Issue 1 (2022), pp. 1–6. ISSN 2634-8845.
[49] Pavlović, T.; Milosavljević, D. & Mirjanić, D.: *Renewable energy sources* (in Serbian). Banja Luka (Bosnia and Herzegovina): Academy of Sciences and Arts of the Republic of Srpska (ANURS), 2013. – 365 pp. ISBN 978-99976-42-41-0.
[50] Pimpalkar, R.; Sahu, A. & Roy, A.: A smart solar PV monitoring system using Internet of Things (IoT). *Concurrent Engineering*, Vol. 33, Issue 1-4 (March-December 2025), pp. 50–60. ISSN 1063-293X. https://doi.org/10.1177/1063293X251325155.
[51] Reddy, V.S.; Kaushik, S.C.; Ranjan, K.R. & Tyagi, S.K.: State-of-the-art of solar thermal power plants – A review. *Renewable and Sustainable Energy Reviews*, Vol. 27 (November 2013), pp. 258–273. ISSN 1364-0321. http://dx.doi.org/10.1016/j.rser.2013.06.037.
[52] Ristau, O.: *Georgia*: *A source of green energy for Europe*? DW.com, 25 June 2023. Available on the Web page: https://www.dw.com/en/georgia-a-source-of-green-energy-for-europe/a-66010720.
[53] Rybak, A. & Rybak, A.: Analysis of the strategy for the energy policy of Poland until 2030 implementation effects in the aspect of environmental protection taking into account the energy security of Poland. *IOP Conference Series*: *Earth and Environmental Science*, Vol. 261 (April 2019), Article no. 012044: pp. 1–12. ISSN 1755-1307. http://dx.doi.org/10.1088/1755-1315/261/1/012044.
[54] Rybak, A. & Włodarczyk, E.: Energy policy until 2040 and the future of hard coal in Poland. *IOP Conference Series*: *Earth and Environmental Science*, Vol. 609 (December

2020), Article no. 012009: pp. 1–9. ISSN 1755-1307. http://dx.doi.org/10.1088/1755-1315/609/1/012009.

[55] Sahu, B.K.: Wind energy developments and policies in China: A short review. *Renewable and Sustainable Energy Reviews*, Vol. 81 (January 2018), pp. 1393–1405. ISSN 1364-0321. http://dx.doi.org/10.1016/j.rser.2017.05.183.

[56] Saidur, R.; Islam, M.R.; Rahim, N.A. & Solangi, K.H.: A review on global wind energy policy. *Renewable and Sustainable Energy Reviews*, Vol. 14 (2010), pp. 1744–1762. ISSN 1364-0321. http://dx.doi.org/10.1016/j.rser.2010.03.007.

[57] Samkharadze, I.; Gomiashvili, Z. & Pkhakadze, I.: *Energy law in Georgia*. In: *International Encyclopaedia of Laws (IEL): Energy Law*. Alphen on the Rhine (Netherlands): Kluwer Law International B.V., 2024. – 120 pp. ISBN 978-94-035-4721-3.

[58] Sharma, B.; Rizwan, M. & Anand, P.: A new intelligent approach for size optimization of a renewable energy based grid connected hybrid energy system. *International Journal of Numerical Modelling: Electronic Networks, Devices and Fields*, Vol. 36, Issue 2 (March-April 2023), Article no. e3050. ISSN 0894-3370. https://doi.org/10.1002/jnm.3050.

[59] Singh, V.K.; Bangari, N.; Tiwari, R.; Dubey, V.; Bhoi, A.K. & Babu, T.S. (editors): *Green energy systems*: *Design, modelling, synthesis and applications*. London (United Kingdom): Academic Press Ltd., 2023. - 255 pp. ISBN 978-0-323-95108-1. http://dx.doi.org/10.1016/C2021-0-00244-3.

[60] Sørensen, B.: A history of renewable energy technology. *Energy Policy*, Vol. 19, Issue 1 (January-February 1991), pp. 8–12. ISSN 0301-4215. https://doi.org/10.1016/0301-4215(91)90072-V.

[61] Tercan, S.M.; Elma, O.; Gokalp, E. & Cali, U.: An expansion planning method for extending distributed energy system lifespan with energy storage systems. *Energy Exploration & Exploitation*, Vol. 40, Issue 2 (March 2022), pp. 599–618. ISSN 0144-5987. https://doi.org/10.1177/01445987211058304.

[62] Topçu, Y.I. & Ulengin, F.: Energy for the future: An integrated decision aid for the case of Turkey. *Energy*, Vol. 29, Issue 1 (January 2004), pp. 137–54. ISSN 0360-5442. http://dx.doi.org/10.1016/S0360-5442(03)00160-9.

[63] Turmanidze, R.; Gomidze, N. & Popkhadze, G.: Analysis of using of solar energies in Georgia. *International Scientific Journal "Science, Business, Society"*, Vol. 10, Issue 2 (2025), pp. 60–67. ISSN 2367-8380.

[64] Turmanidze, R.; Inasharidze, K. & Popkhadze, G.: Analysis of the utilization of various energy sources in Georgia. *International Scientific Journal "Innovations"*, Vol. 12, Issue 2 (June 2024), pp. 78–82. ISSN 2603-3763.

[65] Turmanidze, R. & Popkhadze, G.: Classification of green energies and possibilities of their effective use.*"Industry 4.0"*, Vol. 7, Issue 6 (2022), pp. 254–258. ISSN 2534-8582.

[66] Turmanidze, R. & Popkhadze, G.: Trend analysis of using of green energies in Georgia until 2040. *International Scientific Journals "Industry 4.0"*, Vol. 8, Issue 5 (2023), pp. 193–197. ISSN 2534-8582.

[67] Wei, H.-L.; Chen, S.; Pan, T.-H.; Tao, J. & Zhu, M.-X.: Capacity configuration optimisation of hybrid renewable energy system using improved grey wolf optimizer. *International Journal of Computer Applications in Technology*, Vol. 68, Issue 1 (January 2022), pp. 1–11. ISSN 0952-8091. https://doi.org/10.1504/IJCAT.2022.123234.

[68] *What is green energy*? (*Definition, types and examples*). Available on the Web site: https://www.twi-global.com/technical-knowledge/faqs/what-is-green-energy.

[69] *Wind energy generation by region*. Our World in Data & Energy Institute Statistical Review of World Energy, 2023. Available on the Web site: https://ourworldindata.org/grapher/wind-energy-consumption-by-region/.

[70] Zainol Abidin, M.A.; Mahyuddin, M.N. & Mohd Zainuri, M.A.A.: Solar photovoltaic architecture and agronomic management in agrivoltaic system: A review. *Sustainability*, Vol. 13, Issue 14 (July 2021), Article no. 7846: pp. 1–27. eISSN 2071-1050. http://dx.doi.org/10.3390/su13147846.

[71] Zakari, A.; Oryani, B.; Alvarado, R. & Mumini, K.: Assessing the impact of green energy and finance on environmental performance in China and Japan. *Economic Change and Restructuring*, Vol. 56, Issue 2 (April 2023), pp. 1185–1199. ISSN 1573-9414. https://doi.org/10.1007/s10644-022-09469-2.

[72] Zeng, H.-Q.: Multi-load prediction modelling for combined energy system of intelligent city buildings. *International Journal of Global Energy Issues*, Vol. 42, Issue 3-4 (August 2020), pp. 198–211. ISSN 0954-7118. https://doi.org/10.1504/IJGEI.2020.108957.

[73] Zhang, Y.-S.; Chen, K.; Jiang, J.-J. & Sun, L.: Two-stage stochastic programming of a hybrid SOFC-renewable integrated energy system considering time sequence correlation. *International Journal of Green Energy*, Vol. 20, Issue 1 (2023), pp. 1–14. ISSN 1543-5075. https://doi.org/10.1080/15435075.2021.2021416.

[74] Zhao, P.; Xu, W.-P.; Zhang, S.Q.; Wang, & Dai, Y.-S.: Technical feasibility assessment of a standalone photovoltaic/wind/adiabatic compressed air energy storage based hybrid energy supply system for rural mobile base station. *Energy Conversion and Management*, Vol. 206 (February 2020), Article no. 112486. ISSN 0196-8904. https://doi.org/10.1016/j.enconman.2020.112486.

[75] Zhen, J.-L.; Zhang, Y.; Liu, X.-R.; Huang, G. & Li, J.-B.: Optimization modeling for regional energy system management coupled with energy-water nexus and carbon emission reduction: A case study. *Environmental Engineering Science*, Vol. 40, Issue 3 (March 2023), pp. 114–126. ISSN 1092-8758. https://doi.org/10.1089/ees.2022.0204.