



TIEES 2025 Proceedings

Trends and Innovations in E-business, Education and Security

Peer-Reviewed Conference Proceedings

All papers underwent a double-blind peer-review process

*Accepted papers were selected based on originality,
scientific quality and relevance to the conference themes.*

Budapest – Bratislava

2025

Editors:

Peter Schmidt, Zoltán Rajnai, Pavol Jurík

ISSN: 2729-8493

DOI: 10.5281/zenodo.18903876

Imprint

Publisher:

ÓBUDA UNIVERSITY BUDAPEST
DONÁT BÁNKI FACULTY OF MECHANICAL AND SAFETY ENGINEERING
H-1081 Budapest, Népszínház street 8. Hungary

In cooperation with:

The Slovak Society for Economic Informatics
Dolnozemská cesta 1,
85235 Bratislava-Petržalka, Slovak Republic

ISSN: 2729-8493

Date: December 16, 2025

Committees

Conference Chairs

Schmidt Peter (University of Economics in Bratislava, Slovakia)

INTERNATIONAL SCIENTIFIC COMMITTEE

Aws Jaber (KTH Royal Institute of Technology, Sweden)

Brezina Ivan (University of Economics in Bratislava, Slovakia)

Doucek Petr (University of Economics Prague, Czech Republic)

Gontar Zbigniew (Lodz University, Poland)

Kerimbaev Nurasil (Al-Farabi Kazakh National University, Kazakhstan)

Karimov Diyor Mukhtarovich (National University of Uzbekistan named after Mirzo Ulugbek, Tashkent, Uzbekistan)

Karelkhan Nursuale (ENU Astana, Kazakhstan)

Kiss Gábor (Óbuda University, Budapest, Hungary)

Kultan Jaroslav (University of Economics in Bratislava, Slovakia)

Meruert Serik (Eurasian National University, Kazakhstan)

Mišút Martin (University of Economics in Bratislava, Slovakia)

Novotný Ota (University of Economics Prague, Czech Republic)

Nogueira Marques Nuno Alexandre (Oslo Metropolitan University, Norway)

Ognjanović Ivana (University of Donja Gorica, Montenegro)

Panatie Maria (University of Geneva, Switzerland)

Rajnai Zoltán (Óbuda University, Budapest, Hungary)

Rakovská Eva (University of Economics in Bratislava, Slovakia)

Reiff Marian (University of Economics in Bratislava, Slovakia)

Schmidt Peter (University of Economics in Bratislava, Slovakia)

Simonka Zsolt (University of Economics in Bratislava, Slovakia)

Szabó Lajos (Óbuda University, Budapest, Hungary)

Valášek Ján (DAMA Slovak Republic – Bratislava)

Zelinová Silvia (University of Economics in Bratislava, Slovakia)

Zuev Vladimir (Institute for Social Sciences and Humanities, Republic of Tatarstan, Russia)

Organizing Committee

Jurík Pavol (University of Economics in Bratislava, Slovakia)

Kovács Tibor (Óbuda University, Budapest, Hungary)

Kultan Jaroslav (University of Economics in Bratislava, Slovakia)

Szűcs Endre (Óbuda University, Budapest, Hungary)

Zelinová Silvia (University of Economics in Bratislava, Slovakia)

Table of Contents

An Intelligent Fuzzy Logic–Based Control System for Adaptive Electric Fan Temperature Regulation in Sleep Environments	5
Ahmed Al-Areqi, Amgad Naji Ali Ahmed, Róbert Szabolcsi, György Györök	
Model Transparency: Integrating XGBoost with SHAP for Explainable Machine Learning	25
Andrej Bednařík	
Students’ Perceptions of Generative Artificial Intelligence in Higher Education	36
Veronika Horniaková, Eva Rakovská	
Institutional Dependence and Systemic Risk in Modern Money Markets	45
Karimov Diyor Mukhtarovich, Peter Schmidt, Jamila Bobonazarova, Azamat Khidirnazarov	
The Virtual Accounting Assistant: A Humanized and Expanded Analysis	54
Árpád Lángyi	
Enhancing the reliability of LoRa-based sensor networks in groundwater monitoring	67
Gulrukh Memonova, Peter Schmidt	
Artificial Intelligence in Operations Scheduling: A Knowledge-Driven Architecture Integrated with a Digital Twin	78
Martin Mišút, Mária Mišútová	
Game Theory and Inter-Company Data Sharing for AI Model Training	88
Stefan Otto Novak	
Generative Graphics for Scientific Visualization	103
Peter Procházka	
Detection of Artifacts in AI-Generated Portraits Using Frequency and Texture Analysis (FFT, DCT, LBP/GLCM)	116
Peter Procházka	
Application generated by artificial intelligence	126
Pavol Sojka	
Anomaly identification, Digital environment, Informatics teacher education, Information Security, Machine Learning, Threat detection	134
Jaroslav Kultán, Meruyert Serik, Danara Tleumagambetova	
Proposal for an Information System for Insurance Contract Management	142
Silvia Zelinová, Tomáš Šuhaj	

An Intelligent Fuzzy Logic–Based Control System for Adaptive Electric Fan Temperature Regulation in Sleep Environments

Ahmed Al-Areqi¹ Amgad Naji Ali Ahmed² Róbert Szabolcsi³ György Györök⁴

Abstract

Maintaining thermal comfort during sleep while minimizing energy consumption has become increasingly important under rising summer temperatures. This study presents the design and implementation of an intelligent fuzzy logic–based control system for adaptive regulation of household electric fans in sleep environments. The proposed system integrates an infrared array sensor for simultaneous detection of body and room temperatures with a Mamdani-type fuzzy inference controller implemented via MATLAB–Arduino communication. The controller dynamically adjusts the fan speed according to nine linguistic rules derived from thermal comfort criteria. Experimental validation was conducted across multiple temperature scenarios ranging from 25 °C to 45 °C, with body temperatures varying between 36 °C and 39 °C. Results show stable adaptive behavior, with fan speed smoothly varying between 12% and 82% according to thermal conditions. The proposed approach offers greater flexibility than fixed-speed operation and model-based controllers, providing an energy-efficient, personalized solution for smart home thermal management.

Keywords

Intelligent Control System, Fuzzy Logic Controller, Infrared Array Sensor, Thermal Comfort, Sleep Environment, Temperature Regulation

1 Introduction

The summer of 2024 was the hottest on record across Europe. According to the Copernicus Climate Change Service (C3S), the average European land-surface air temperature during June–August 2024 was +1.54 °C above the 1991–2020 climatological baseline, marking the warmest summer ever observed in the Copernicus ERA5 dataset (Copernicus Climate Change Service, 2024). This period represents the peak of the European summer when air temperatures typically reach their annual maximum. Several regions—particularly southern and southeastern Europe—experienced extreme heat anomalies exceeding +9 °C, resulting in prolonged and intense heatwaves that affected over half of the continent’s population (Agencia Estatal de Meteorología, 2025).

A comprehensive attribution analysis conducted by the European Environment Agency (EEA) and the UK Met Office concluded that the record-breaking heat of 2024 was largely driven by human-induced climate change. Using climate model simulations, the researchers demonstrated that without anthropogenic greenhouse gas emissions, such extreme summer temperatures would have been virtually impossible. The estimated recurrence interval of comparable heat events

¹Óbuda University, Bánki Donát Faculty of Mechanical and Safety Engineering, Doctoral School on Safety and Security Sciences, Népszínház u. 8, 1081 Budapest, Hungary, ahmedalareqi@stud.uni-obuda.hu

²Óbuda University, Bánki Donát Faculty of Mechanical and Safety Engineering, Doctoral School on Safety and Security Sciences, Népszínház u. 8, 1081 Budapest, Hungary, ahmed.amgad@uni-obuda.hu

³Óbuda University, Kandó Kálmán Faculty of Electrical Engineering, Institute of Electronic and Communication Systems, Tavaszmező u. 17, 1084 Budapest, Hungary, szabolcsi.robert@kvk.uni-obuda.hu

⁴Óbuda University, Alba Regia Faculty, Budai út 45, H-8000 Székesfehérvár, Hungary, gyorok.gyorgy@amk.uni-obuda.hu

has shortened from several millennia in the pre-industrial era to approximately once every three years under current climatic conditions. Projections under a medium-emission scenario indicate that by the end of the century, such extreme heat events could become an annual occurrence (Christidis, 2024).

As stated by Dr. Nikos Christidis, senior climate attribution scientist at the UK Met Office, “This latest attribution study is another example of how climate change is already making our weather extremes more severe. Our analysis of the European summer of 2024 reveals that what was once a rare occurrence has now become a regular feature of our summers” (Copernicus Climate Change Service, 2024).

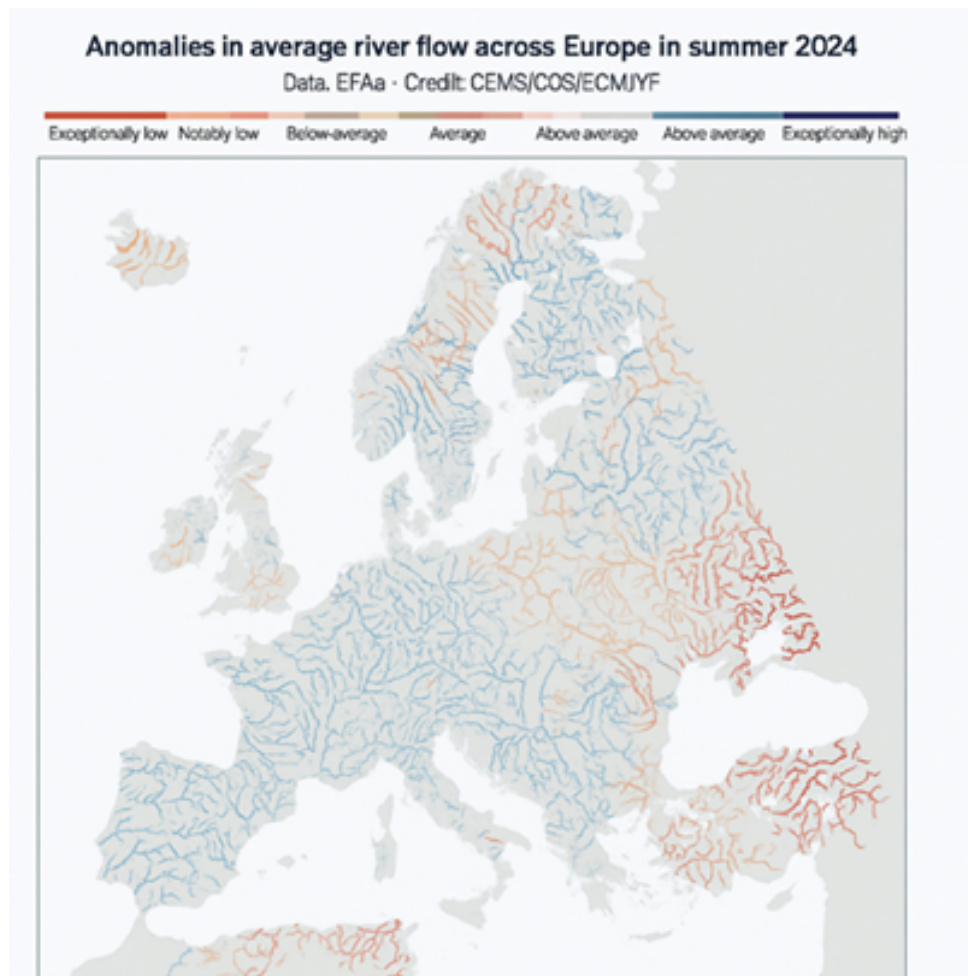


Figure 1. Average summer temperature anomalies in Europe for 2024 relative to 1991–2020 (Copernicus Climate Change Service, 2024)

The summer of 2025 maintained this warming trend, ranking among the warmest in Europe’s historical record. According to C3S seasonal data, mean temperatures were approximately 0.9 °C above the 1991–2020 average, with the most severe anomalies observed in southern and western Europe (Copernicus Climate Change Service, 2025). In Spain, the Agencia Estatal de Meteorología (AEMET) confirmed that 2025 was the country’s hottest summer on record, with a national mean temperature of 24.2 °C—about 2.1 °C above the recent climatological norm (Copernicus Climate Change Service, 2025). This exceptional heat contributed to widespread droughts, reduced river flows, and extensive wildfires across the Iberian Peninsula and the Mediterranean basin.

Preliminary projections for summer 2026 indicate the persistence of this warming pattern. While complete observational data are not yet available, the Copernicus Climate Outlook (issued late 2025) predicts that European summer surface temperatures will likely remain between

+0.9 °C and +1.3 °C above the 1991–2020 average, with an elevated likelihood of extended heatwaves in southern and eastern Europe (European Environment Agency, 2024). These findings reinforce the long-term trend of accelerated European warming—estimated at more than 2 °C above pre-industrial levels—making the continent one of the fastest-warming regions on Earth (Met Office, 2024). These escalating summer temperatures have significant implications for human health, energy demand, and indoor environmental comfort. Extended exposure to extreme heat can impair sleep quality, increase cardiovascular strain, and elevate mortality risks among vulnerable populations. Despite the growing impact of climate change on indoor comfort conditions, limited research has focused on low-cost adaptive cooling solutions specifically designed for sleep environments. Conventional fan systems typically operate at fixed speeds and fail to account for dynamic variations in human thermal perception. Therefore, this study proposes a closed-loop fuzzy logic-based control system that integrates physiological (body temperature) and environmental (room temperature) feedback to achieve adaptive fan speed regulation. The objective is to enhance thermal comfort, reduce unnecessary energy consumption, and provide a personalized cooling experience during sleep. The novelty of this study lies in integrating physiological thermal feedback and environmental sensing into a low-cost fuzzy-logic control architecture specifically designed for sleep environments. This approach has not been extensively addressed in previous fan-based cooling research.

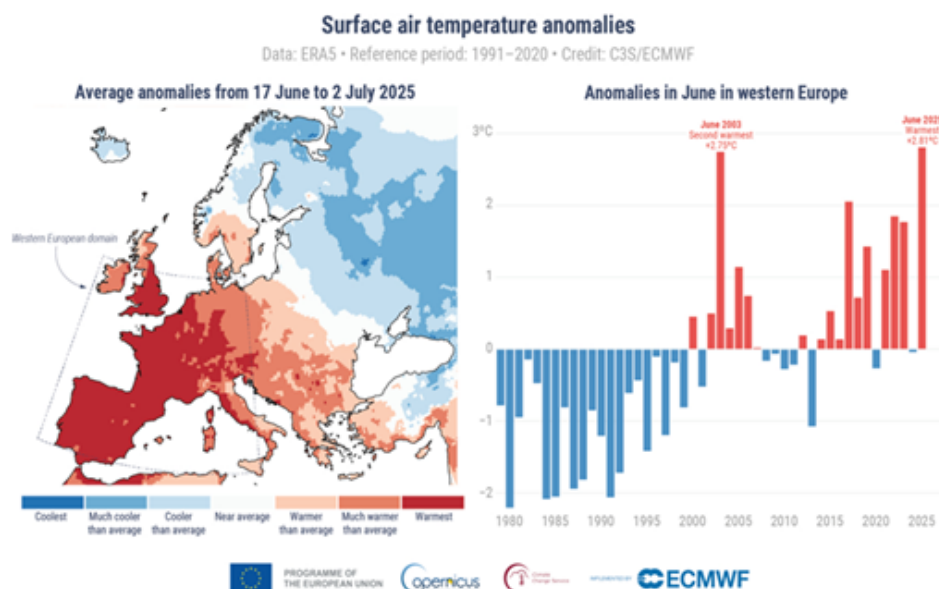


Figure 2. European surface temperature anomalies during 2024–2025[5,6]

2 Related work

A literature review was conducted on articles reporting the effects of the thermal environment on sleep quality, with the objectives of investigating how the thermal environment affects sleep quality and, thus, how to control the sleeping thermal environment more appropriately and energy efficiently. Previous studies have investigated the influence of thermal environments on sleep quality and physiological regulation. Experimental findings indicate that air temperature, airflow velocity, and sleep stage significantly influence sleep efficiency and thermal comfort perception. Although electric fans are commonly used for cooling, continuous operation without adaptive control may lead to discomfort or suboptimal energy use. These findings highlight the importance of intelligent control strategies that dynamically adjust airflow based on both environmental and physiological conditions, as reported by Sleep Advisor (Fig. 3). In almost all cases, except those

with an electric fan, the mean air temperature was roughly 26°C. The mean air temperature with the electric fan was 29.4°C, which might be attributed to faster air velocity, boosting convective heat loss, allowing for a higher temperature. This shows that there is a probable link between air temperature and airflow and that everyone responded to the environment through a range of adaptive behaviors. Sleep efficiency was marginally lower when the electric fan was used than when the other conditions were used (Akiyama et al., 2021).

Normal human sleep is divided into two stages: non-rapid eye movement (NREM) and rapid eye movement (RREM) (REM). NREM sleep is further classified as having three stages: N1, N2, and N3. The N1 and N2 phases are associated with light sleep, whereas the N3 stage is associated with slow-wave and deep sleep, as shown in Fig. 4. A person's sleep cycle is a series of events that progress from the N1 stage to the N3 stage and eventually to the REM state. Any disruption to these synchronized sleep periods can lead to poor sleep quality and the related detrimental mental and physical health repercussions (Ngarambe et al., 2019).

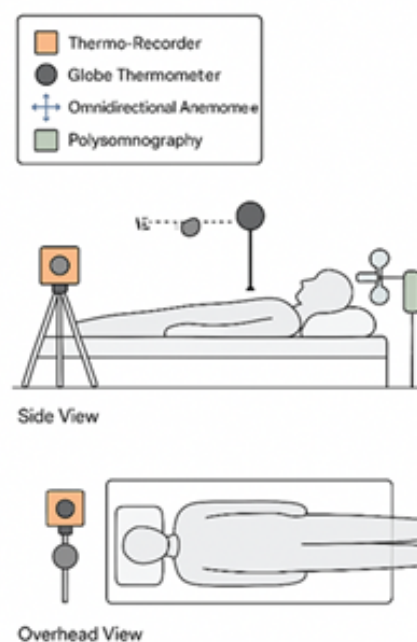


Figure 3. Measurement points of the body temperature (Akiyama et al., 2021)

Although the thermoregulatory and sleep regulatory systems are still poorly understood, there are some hints of how they might interact. Many studies have shown that the preoptic area/anterior hypothalamus (POAH), the key brain region that promotes heat loss, is also critical in sleep regulation (Economo, 1930). Even at the cellular level, neurons are sensitive to heat (warm-sensitive neurons, WSNs), and neurons that modify their firing patterns prior to and during sleep do so in concert. Von Economo (1930) argued that sleep is controlled by conflicting wake-promoting and sleep-promoting mechanisms in the hypothalamus (Candas et al., 1982). Later research has verified the existence of such a sleep-promoting mechanism in the POAH (Lan et al., 2017). High brain and skin temperatures have been linked to increased WSN activity and sleep induction (Economo, 1930). Thus, a thermally comfortable sleeping environment is important for sleep maintenance (Tateno et al., 2019).

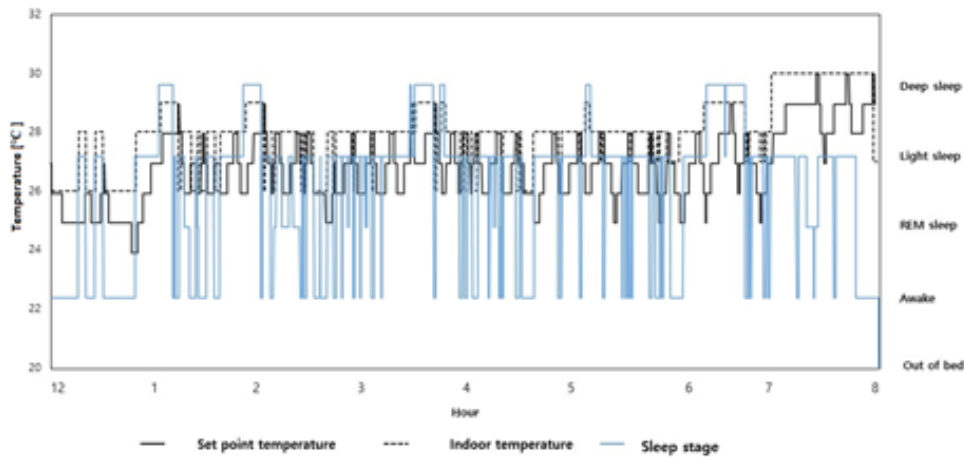


Figure 4. Setpoint temperature of the air conditioner in response to sleep stage (Ngarambe et al., 2019)

3 Methodology

In this section, the main components used in this paper are explained. The fuzzy controller will also be briefly explained. The main components are an Arduino UNO microcontroller, an Infrared Array Sensor, an LCD, and a Fuzzy controller.

3.1 Grid eye sensor

Compared with traditional optical sensors such as cameras, infrared-based sensors are less affected by environmental factors, making them more reliable and with a broader application range. Nowadays, thermal sensors are frequently applied in medical and military applications. However, this type of thermal sensor is much more expensive than cameras (Panasonic Industry, 2026).

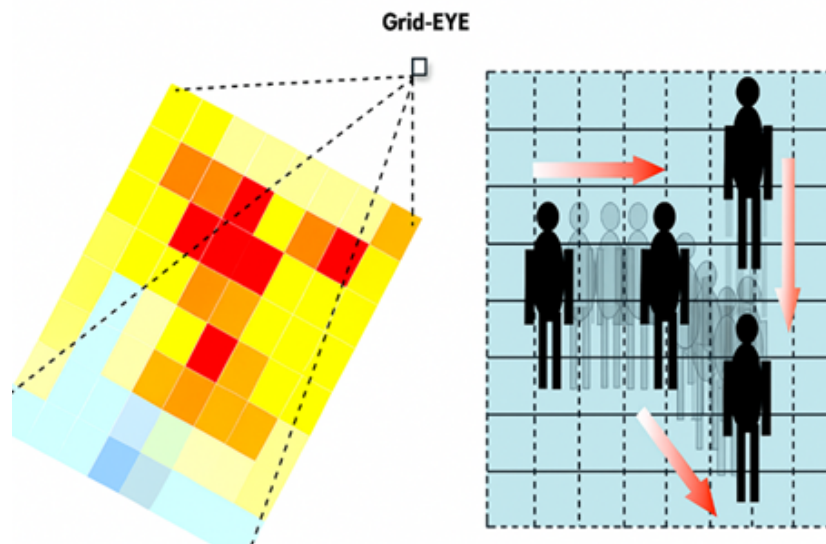


Figure 5. Temperature detection achieved on a two-dimensional area with 8×8 (64) pixels (Ngarambe et al., 2019).

The grid-eye sensor is based on MEMS (microelectromechanical systems). It comprises 64 MEMS thermopile elements arranged in an 8×8 grid on a single detector chip, as shown in Fig. 5. (Zhao et al., 2020) Each thermopile element delivers a single temperature value, yielding a total of 64 temperature outputs ranging from 0°C to 80°C . The Grid eye sensor can measure body and ambient temperatures contactlessly throughout the fully defined area, with a 60° viewing

angle in both horizontal and vertical directions, and a total range of 7 meters. Grid eye receives thermal energy (infrared radiation) that flows through the silicon lens to the thermopile sensor elements, which convert it into a proportional output. These analog signals are converted into digital temperature values via an integrated circuit I2C circuit and supplied to a CPU. A CPU creates a map of the individual thermopile temperature values to generate a binary picture or thermal image of these 8x8 pixels. The temperature distribution detected by the grid-eye sensor can be used to detect human presence. It can detect human presence in both stationary and moving postures. It also measures the surrounding temperature. Finally, this thermal map can be processed using a fuzzy controller to obtain higher-resolution measurements.

3.2 Fuzzy controller

Fuzzy logic control was selected for its ability to handle the nonlinear and uncertain relationships between human thermal perception and environmental temperature. Unlike classical control strategies that require accurate mathematical modeling, fuzzy logic enables rule-based decision-making using linguistic variables. In this study, two input variables (body temperature and room temperature) and one output variable (fan speed) were defined. A Mamdani-type inference system with centroid defuzzification was implemented using nine control rules to achieve adaptive fan speed regulation (Dernoncourt, 2013). Fuzzy logic is needed in control systems due to its ability to handle uncertain and imprecise data, as well as its flexibility, robustness, and user-friendliness. It excels at complex, dynamic systems and is adaptable to diverse objectives, making it a valuable tool for control and decision-making across a wide range of real-world applications. A fuzzy controller is a mathematical controller for dynamic systems. It consists of three parts: input, process, and output. Input parts, called fuzzification, read analog signals from sensors such as pressure, sound, temperature, etc., or digital signals from ON/OFF switches, and convert them into membership functions. The process part is the most important, and it can be designed through experimentation. In this part, the controller's rules are generated and translated to the output port of the fuzzy controller. The final part is the output, which consists of the defuzzification that converts the analyzed result into the specific output value (Al-Areqi and Szakács, 2021). A fuzzy logic system is shown in Fig. 6.

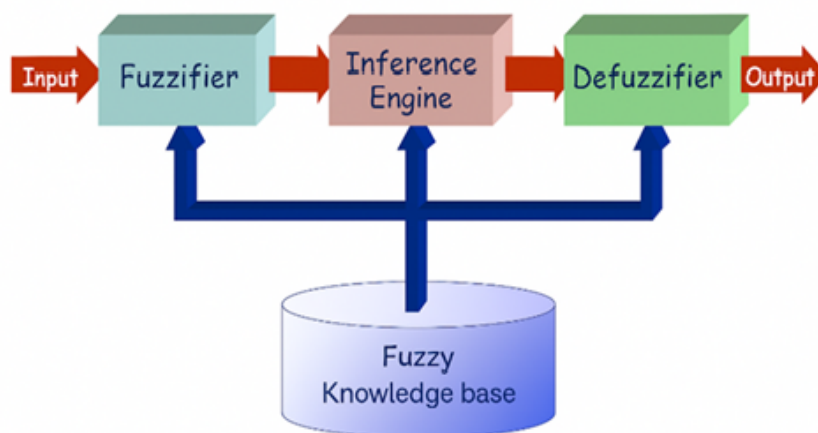


Figure 6. Overview diagram of a fuzzy system (and).

Fuzzification of input, which is a process of converting crisp (exact) input data into fuzzy values. Fuzzy logic systems operate with linguistic variables, which represent input and output values using linguistic terms like "low," "medium," and "high" rather than precise numerical values. Fuzzification is the initial step in the fuzzy logic control process. It involves mapping crisp input values to fuzzy sets or membership functions that represent the degree to which each linguistic term applies. Fuzzy rules are set up to take control of the output. The rules in fuzzy logic are

simple: an if-else condition has a condition and a target, as shown in Table 1. Defuzzification is the process of converting the fuzzy, linguistic output of a fuzzy logic controller into a specific numerical value that can be used for decision-making or for controlling a system. It makes the fuzzy results actionable in real-world applications.

Table 1. Rules of the fuzzy controller

Body temperature	Room temperature	Results
Low	Low	Very Slow speed fan
Low	Mid	Slow-speed fan
Low	High	Medium – Low speed fan
Mid	Low	Slow-speed fan
Mid	Mid	Medium – Low speed fan
Mid	High	Medium – High speed fan
High	Low	Medium – High speed fan
High	Mid	High-speed fan
High	High	Very high-speed fan

3.3 Arduino Uno microcontroller

A microcontroller is a tiny computer built on a single integrated circuit (IC). There are some things that all computers share in common. Arduino is one of the most popular microcontrollers, with an open-source platform for developing electronics projects. Uno is part of the Arduino series and is powered by an ATmega328 microcontroller. Because it includes an IDE (integrated development environment), this board is also excellent for educational and training reasons (Hossain, 2022).

3.4 MATLAB

MATLAB is a high-level programming language for scientific computing. It combines computing, visualization, and a programming environment. Furthermore, MATLAB is a modern programming language environment: it features sophisticated data structures, built-in editing and debugging tools, and enables object-oriented programming. These feature combinations make MATLAB a great tool for both teaching and research (Castrodes et al., 2020).

3.5 SSR

A solid-state relay (SSR) is a relay that does not have moving contacts. In general operation, SSRs are similar to mechanical relays with movable contacts. It depends on semiconductor switching elements such as thyristors, triacs, diodes, and transistors to operate. It also uses optical semiconductors, known as photocouplers, to separate the low-voltage side (control) and high-voltage side (load) signals. Photocouplers convert electrical signals from the microcontroller into optical signals and send them through space, effectively separating the input and output parts, enabling high-speed signal transfer (Teja, 2021). As seen in Fig. 7

4 System hardware design

The functionalities of the hardware components will be described in this section. The system's block diagram and the connections between its various components will be displayed. The system wiring and circuit diagram will be attached at the end of this section.

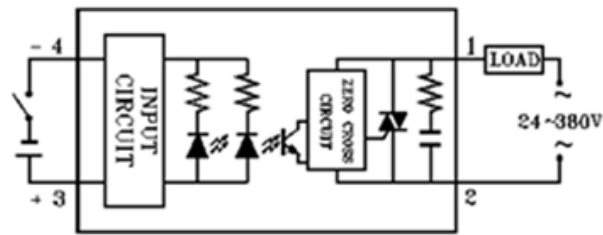


Figure 7. SSR Circuit Diagram (Teja, 2021).

4.1 Block diagram for hardware design

In general, any system is typically composed of three essential components: input, system block, and output. Such a system is referred to as an open-loop control system, as illustrated in Fig. 8, (Electronics Coach, 2026). However, in the case of a closed-loop control system, an additional sensor (feedback) is integrated into the system, resulting in the configuration of input, system block, output, and feedback, as demonstrated in Fig. 11. In a closed-loop control system, the inclusion of a controller, such as a PID or Fuzzy controller, becomes essential to manage the entire process in accordance with application parameters.

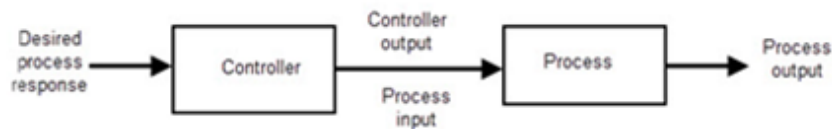


Figure 8. Open Loop System (end)

In this paper, a closed-loop control system is used to control the speed of the home fan based on the temperature of the human and the room. An infrared array sensor (AMG8833) grid eye sensor is integrated into the system. This sensor can measure both human and room temperature simultaneously. We use this sensor to provide feedback to the controller, which adjusts the system's output. As shown in Fig. 9, the system input is a PWM (Pulse-Width Modulation) voltage controlled by an Arduino microcontroller. The system's output is to adjust the fan speed.

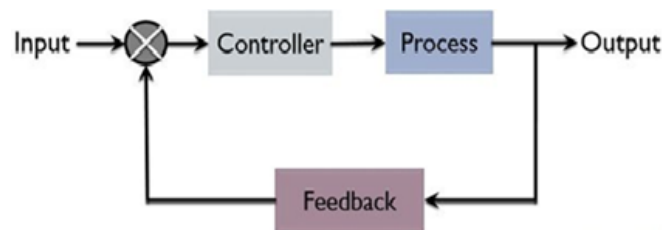


Figure 9. Closed-Loop Control System

4.2 System (black box)

Related to Fig.10, the connection of all system components will be explained. Initially, data from the (body-room) temperature sensor will be received by the Arduino MC. Subsequently, the MC will transmit this data to MATLAB through a serial connection. Within MATLAB, the input parameters of the system (body-room temperatures) will be read by the fuzzy controller, and, guided by the rules of the fuzzy controller, the output value will be adjusted by increasing or decreasing it. The output value will then be sent back to the MC from MATLAB to control the SSR voltage via a PWM signal. Ultimately, the electrical system will be controlled by the PWM signal, increasing or decreasing the fan's speed. The block diagram is depicted in Fig. 11.

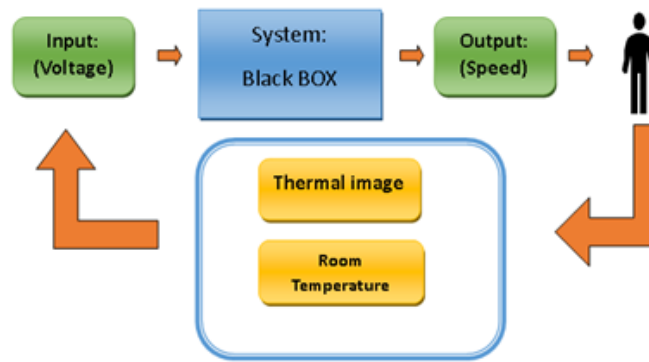


Figure 10. Diagram of our paperwork

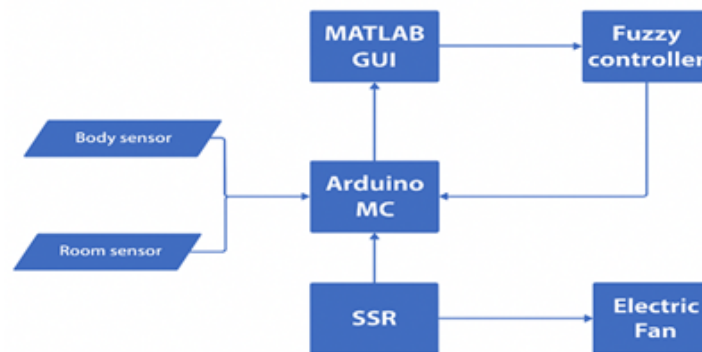


Figure 11. Block diagram of the system.

4.3 Circuit diagram of hardware design

In Fig. 12, the circuit diagram of our system is depicted. The circuit and wiring were designed and built using Fritzing software. This software serves as an open-source CAD tool for designing electronic hardware. It was explored at the University of Applied Sciences Potsdam. As illustrated in our circuit diagram, a DC-DC converter converts 12 VDC to 5 VDC, powering all components. The Arduino microcontroller is connected to MATLAB via a wired connection. Additionally, the LCD was used to display body temperature, room temperature, and fan speed.

5 System software design

In this section, the software of our system will be stated. Our system operates on three main parts: the system algorithm, MATLAB, and the fuzzy controller. Each part of our system has its own algorithm, which will be explained in this section.

5.1 System algorithm

The flowchart presented in Fig. 13 depicts the algorithm for the full operating system. From the starting point:

The Grid Eye sensor measures room temperature.

1. Subsequently, the Grid-EYE sensor scans the thermal image of a human body and determines whether the detected image corresponds to a human or another object.
2. The measurement data is transmitted to the MATLAB software by the Arduino microcontroller.

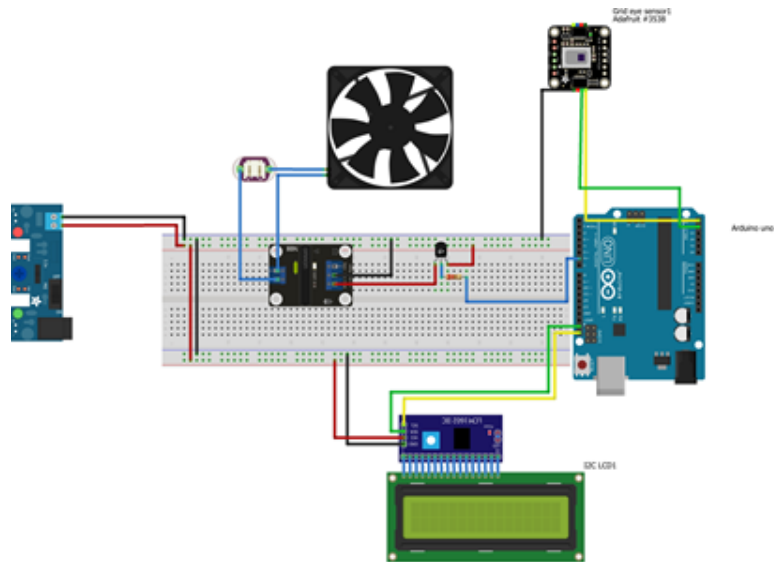


Figure 12. Hardware diagram of the system

3. Within MATLAB, the measurement is received by the fuzzy controller, which applies the rules detailed in the fuzzy controller section.
4. The fuzzy controller then sends the processed data back to the microcontroller (MC) via MATLAB.
5. The MC receives the data from the fuzzy controller and converts it into a PWM voltage signal using the DAC port.
6. Finally, all system data is displayed on the LCD by the MC.

5.2 MATLAB code implantation

A MATLAB script is utilized to establish a connection between the Arduino MC and the Fuzzy controller in Simulink mode within MATLAB, as shown in Figs. 14 and 15. The script receives data sent by the MC through the serial port (COM'6') with a baud rate of 115200. Subsequently, the script reads the measurements and sends them to the fuzzy controller in Simulink mode, as seen in Fig. 17. Within Simulink mode, there exist two inputs (Body_tem - Room_tem), the fuzzy controller, and the PWM output from the fuzzy controller. Once the fuzzy controller applies its rules, the output of the controller is sent to the MATLAB script as a variable named (PWM). Ultimately, the script transmits the PWM value to the Arduino MC, which converts this value into a PWM signal.

5.3 Fuzzy controller implantation

In the fuzzy controller shown in Fig. 16, the initial step is to define the controller's parameters, including inputs, outputs, and rules. Temperature ranges were established for body and room temperatures. As for body temperature, the range is maintained between [36 - 39] degrees Celsius, where 36°C represents the minimum and 39°C the maximum temperature. The body temperature membership function is shown in Fig. 17. Similarly, for room temperature, the range is set to [25 - 45] degrees Celsius, with 25°C as the minimum and 45°C as the maximum. The room-temperature membership function is shown in Fig. 18. Three conditions were defined for both body and room temperature: "Low," "Medium," and "High." Additionally, six conditions were defined for the output, including "Very Slow," "Slow," "Medium Low," "Medium High,"

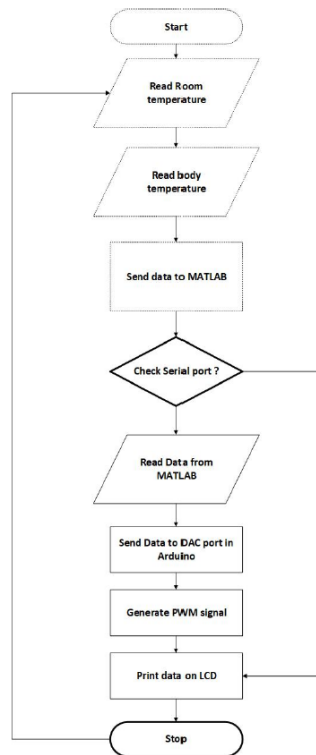


Figure 13. Flowchart of the system

"High," and "Very High" fan speeds. Subsequently, 9 rules were generated, as displayed in Table (Agencia Estatal de Meteorología, 2025). Finally, the speed fan membership function is shown in Fig. 19.

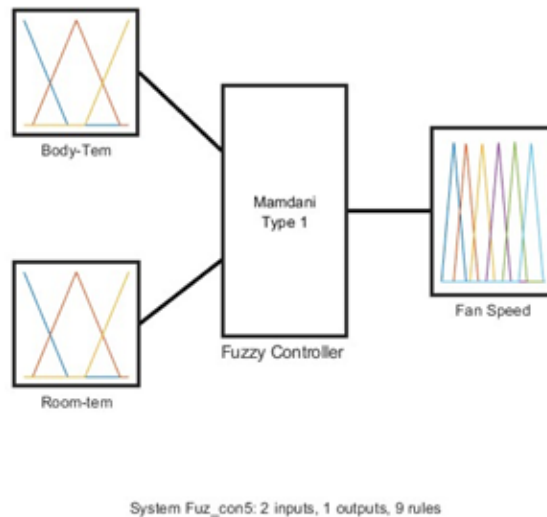


Figure 16. Design of fuzzy controller

```

s=serial('COM6','BaudRate',115200,'DataBits',8,'StopBits',1);
time=100;
i=1;
fopen(s);
while(i<time)

readData=fscanf(s); %reads "Ready"
if readData == 0
else
Temp1(i)=str2num(readData(1:2));
Body_Tem = Temp1(i);
Temp2(i)=str2num(readData(3:4));
Room_Tem =Temp2(i) ;
end

sim('controller_Fuz.slx');
data = PWM;
fprintf(s,'%s',char(data));

i=i+1;

```

Read From MC

Write to MC

Figure 14. MATLAB code for reading and writing from Arduino

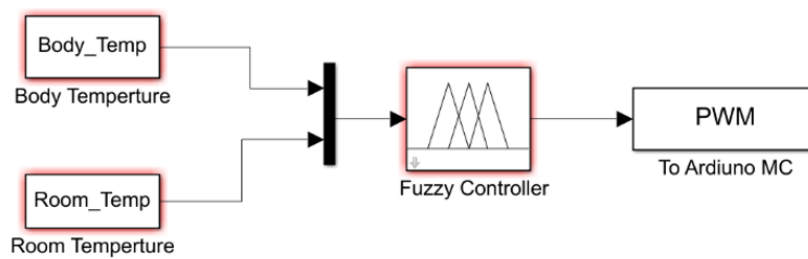


Figure 15. Simulink model of the fuzzy controller

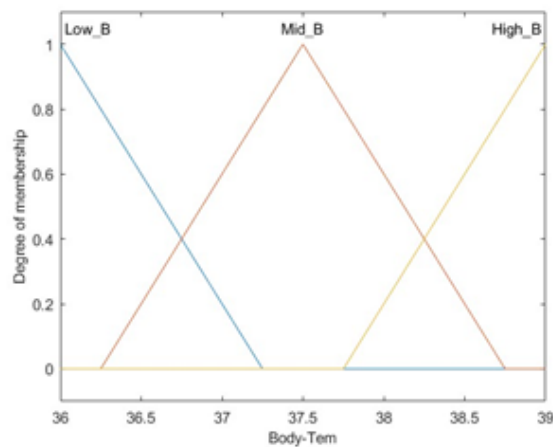


Figure 17. Human body temperature membership

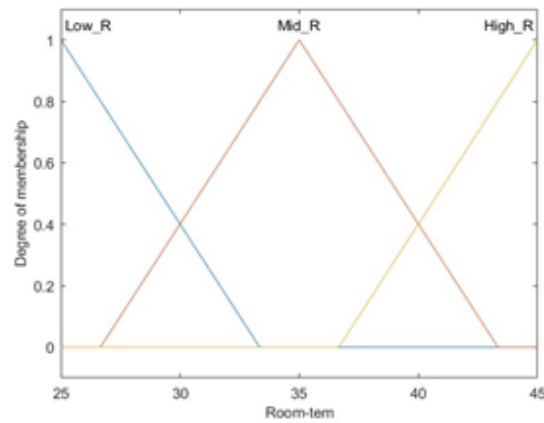


Figure 18. Human body temperature membership

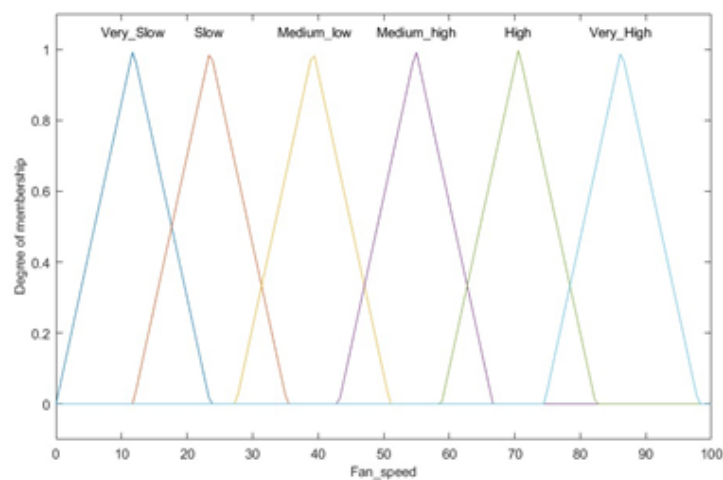


Figure 19. Output variable of the fuzzy controller

6 Main functionality of the system operation

The proposed system was experimentally validated in a controlled indoor environment. A total of 15 test scenarios were conducted, covering combinations of body temperature (36–39 °C) and room temperature (25–45 °C). Each experiment was maintained for approximately 20 minutes to ensure a steady-state system response. Performance evaluation focused on fan speed adaptation accuracy, response smoothness, and stability under dynamic temperature variations. The results confirm that the fuzzy controller provides continuous and proportional speed adjustment without oscillations or abrupt transitions. In this section, the overall system's function will be explained. The system has been put in a real environment to measure the room temperature.

Based on the data in Fig. 20, it appears that there is a relationship among body temperature, room temperature, and fan speed. Specifically, both body and room temperatures appear to affect fan speed.

6.1 Body Temperature Effect

As body temperature increases, there is a tendency for fan speed to increase. For example, when the body temperature is 36°C, the fan speed starts at 12%, and as the body temperature increases to 39°C, the fan speed reaches 82%.

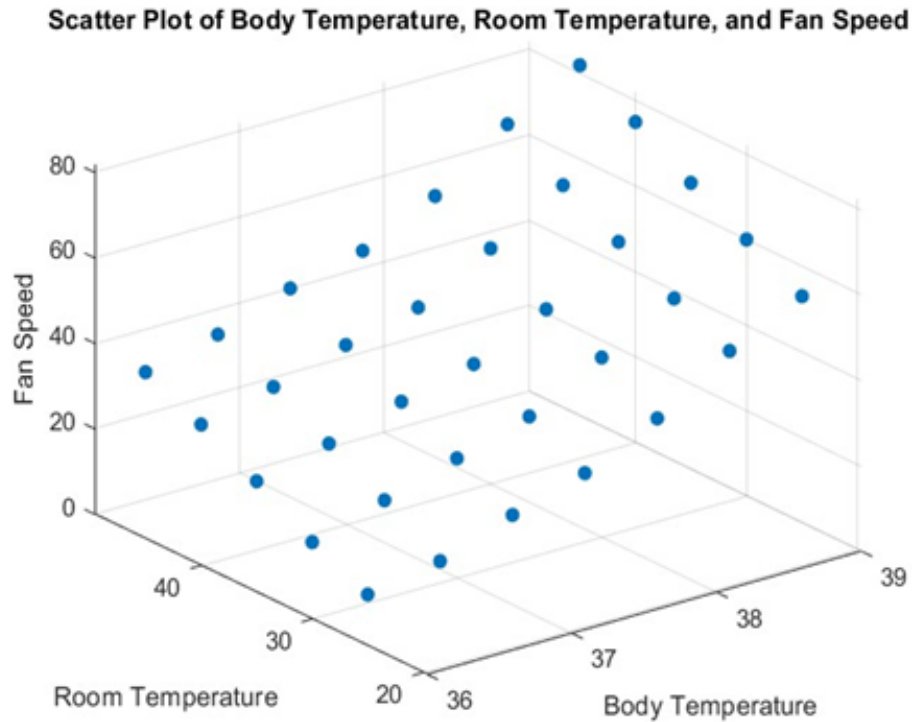


Figure 20. Relationship between (body Temperature and room Temperature) with fan speed

6.2 Room Temperature Effect

As room temperature increases, there is also a tendency for fan speed to increase. For example, when the room temperature is 25°C, the fan speed starts at 12%, and as the room temperature increases to 45°C, the fan speed reaches 82%.

6.3 Combined Effect

The combined effect of both body temperature and room temperature may be more complex, and it will be examined.

The first condition is a normal body temperature with room temperature as shown in Fig. 21.

Table 2. Result of the first condition

Body temperature	Room temperature	Fan speed
36°C	25°C	11.8%

From this single data point in Table 2, we can make some general observations:

- Effect of Body Temperature:** The body temperature remains within the normal range, indicating stable thermal conditions. The fan speed is set to 11.8%, reflecting a moderate airflow sufficient to maintain thermal comfort without excessive cooling.
- Room Temperature:** The room temperature is lower than the body temperature, suggesting a relatively cooler environment. Consequently, the control system maintains a low fan speed (11.8%), as only minimal airflow is required to preserve thermal balance.

The second condition is that while body temperature is increasing, as shown in Fig. 22.

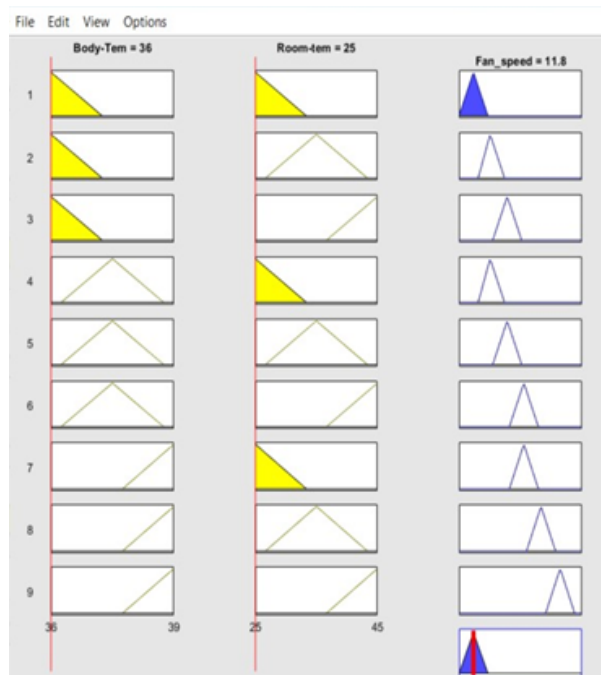


Figure 21. Fuzzy controller of 1st condition

Table 3. Result of the second condition

Body temperature	Room temperature	Fan speed
37.5°C	30.1°C	33.2%

From this single data point in Table 3, we can make some general observations:

- Effect of Body Temperature:** The body temperature was increased. This suggests that a person might have an elevated body temperature, possibly due to a fever or increased physical activity. In response to the elevated body temperature, the fan speed is set at 33.2%. This indicates that the fan speed is adjusted to cool down the person due to the elevated body temperature.
- Weather Temperature (Room Temperature):** The room temperature was also increased. The fan speed is likely set to provide additional cooling or comfort in a relatively warmer room.
- Relationship Between Body Temperature, Weather Temperature, and Fan Speed:** This data point suggests a positive correlation between body temperature and fan speed. As body temperature increases relative to room temperature, the fan speed increases. This implies that the fan speed is adjusted based on the temperature difference between the body and the room.

The third condition was with a hot object temperature as shown in Fig. 23.

Table 4. Result of the second condition

Body temperature	Room temperature	Fan speed
36 ° C	45 ° C	39.2 %

From this single data point in Table 4, we can make some general observations:

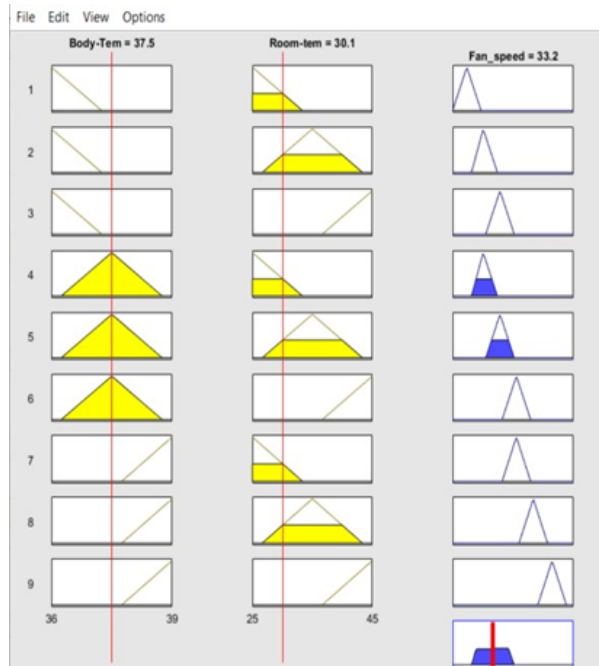


Figure 22. Fuzzy controller of the 2nd condition

- a. **Body Temperature Effect:** The body temperature (36°C) is lower than the room temperature (45°C), indicating a significant thermal imbalance. In response to this condition, the fan speed is set to 39.2% , providing active cooling to compensate for the elevated ambient temperature.
- b. **Room Temperature Effect:** The substantially higher room temperature suggests a warm indoor environment. The relatively high fan speed (39.2%) reflects the control system's effort to enhance airflow and maintain thermal comfort under elevated environmental conditions.

Furthermore, time-response analysis showed that the controller achieved steady-state within approximately 6 seconds after a step change in the temperature input. No oscillations or instability were observed during operation, indicating robust and smooth adaptive performance suitable for sleep environments.

7 Implementation system for testing

The test of the system before incorporating the SSR with a simple DC fan will be explained. Subsequently, an AC fan controlled by an SSR will be added.

7.1 DC fan with MOSFET

System with a DC fan which is controlled by a MOSFET, as seen in Fig.24. PWM controls the speed of the DC fan from the Arduino microcontroller through the MOSFET. The result was that the DC fan speed was adjusted according to the defined fuzzy controller rules.

7.2 AC Fan with SSR

After SSR was connected to the system, the AC fan could be added, as shown in Fig. 25. PWM was sent from the Arduino microcontroller to the SSR via the MOSFET to control the AC fan

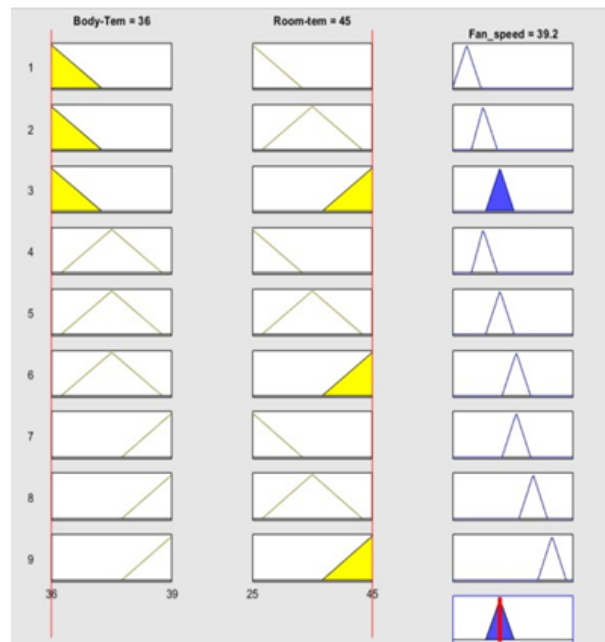


Figure 23. Fuzzy controller of the 3rd condition

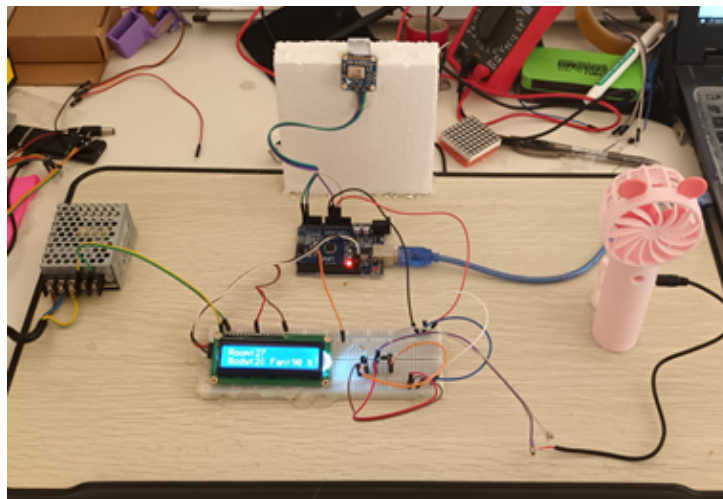


Figure 24. System with DC fan

speed. As a result, the AC fan speed increased and decreased according to the Fuzzy controller, and the results were displayed on the LCD.

7.3 Comparison with Conventional Control Methods

To evaluate the effectiveness of the proposed fuzzy logic controller, a conceptual comparison was performed with conventional fixed-speed operation and classical PID control, as illustrated in Figure 26. In fixed-speed systems, airflow remains constant regardless of thermal variations, which may lead to insufficient cooling or unnecessary energy consumption. PID control offers a dynamic response but requires precise mathematical modeling and may exhibit overshoot under nonlinear conditions. In contrast, the fuzzy logic controller dynamically adjusts fan speed based on real-time body and room temperature inputs, providing smooth, proportional, and adaptive regulation. Due to its model-free design and robustness against nonlinear thermal perception, the proposed system is better suited for personalized sleep environments.

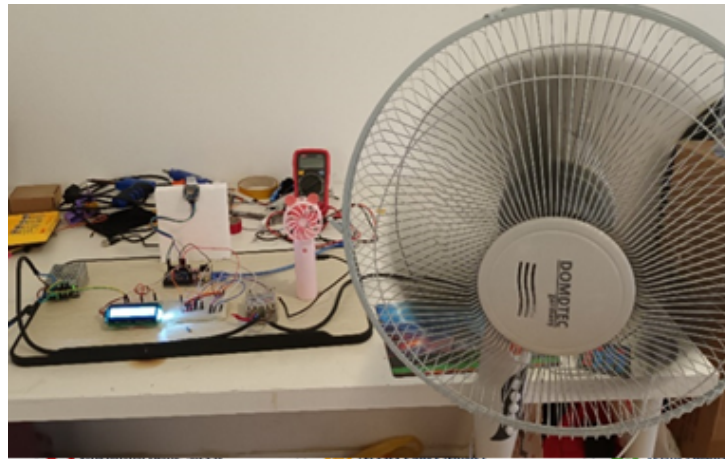


Figure 25. Applied system in a real environment

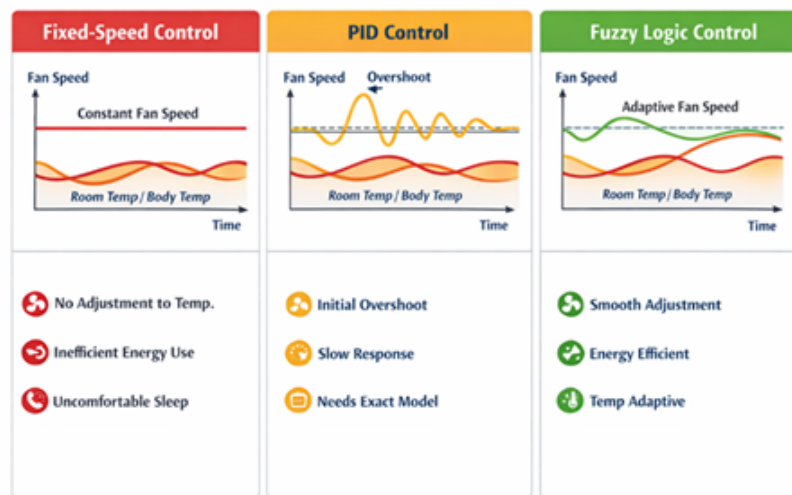


Figure 26. Comparative Evaluation of Control Strategies for Adaptive Thermal Regulation

8 Conclusion

In summary, this study successfully developed and validated an intelligent fuzzy logic-based control system for adaptive electric fan regulation in sleep environments. The system demonstrated high effectiveness in maintaining thermal comfort and potentially improving energy efficiency through adaptive operation, dynamically adjusting fan speed in response to real-time variations in body and room temperature. Experimental results confirmed that fan speed increased proportionally from 12% to 82% as temperature rose from 25 °C to 45 °C, ensuring stable and comfortable sleep conditions while reducing unnecessary energy use. These outcomes clearly highlight the capability of fuzzy logic control to manage the uncertainty and nonlinearity in the relationship between human thermal perception and environmental conditions—an advantage that conventional PID controllers cannot achieve with equal precision. The integration of an infrared array sensor allowed accurate, contactless detection of both body and ambient temperature, improving system responsiveness and personalization. The developed model not only enhanced user comfort but also established a foundation for intelligent, human-centered environmental control. Furthermore, the successful MATLAB–Arduino interface and the inclusion of a solid-state relay (SSR) for AC fan speed modulation confirmed the feasibility and robustness of the proposed design. This study presented the design and implementation of an adaptive fuzzy logic-based fan

control system for sleep environments. Experimental results demonstrated stable, proportional fan speed regulation across varying body and room temperatures. The integration of infrared sensing and fuzzy inference enabled real-time, contactless, and personalized thermal management. The findings confirm the practical feasibility and energy-aware potential of intelligent fan control in smart home applications.

9 Future work

Although the proposed intelligent fuzzy-logic-based control system has shown promising results in enhancing thermal comfort and energy efficiency during sleep, several aspects remain open to further development. Future research should aim to improve both hardware and software components to achieve greater precision, reliability, and adaptability under varying environmental conditions. Advancements could include integrating high-performance microcontrollers or embedded processors, such as the Raspberry Pi, Nvidia Jetson Nano, or STM32 series, to increase computational capability and response time. Incorporating machine learning and artificial intelligence techniques would further enhance the controller's adaptability by enabling it to learn user preferences and dynamically adjust thermal regulation patterns. In addition, the use of multi-sensor fusion combining infrared, humidity, and airflow sensors could provide a more comprehensive understanding of the sleeping environment. Wireless communication and Internet of Things (IoT) integration would also allow real-time monitoring and remote control through smart home platforms. On the hardware side, future prototypes may utilize compact printed circuit boards (PCBs) for direct integration into fan systems, supporting miniaturization and improved reliability. Extended testing under different climatic conditions and user profiles would help validate system durability, energy efficiency, and comfort performance. Furthermore, the fuzzy logic control concept can be adapted for broader applications in household or healthcare devices that require adaptive thermal management. Pursuing these research directions will enable the system to evolve into a fully autonomous, intelligent, and user-centered solution for next-generation smart environments.

Resources

- Agencia Estatal de Meteorología (2025). *Spain's 2025 Summer Temperature Report*. Madrid, Spain.
- Akiyama, Y., E. Miyake, R. Matsuzaki, M. Ogata, K. Tsuzuki, and S. Tanabe (2021). "Effect of thermal environment on sleep quality in actual bedroom in summer by sleep stages analysis". In: 4.1, pp. 211–221.
- Al-Areqi, A. and T. Szakács (2021). "Can Bus Communication Demonstration Tool for Education". In: *2021 IEEE 15th International Symposium on Applied Computational Intelligence and Informatics*.
- Candas, V., J. Libert, and A. Muzet (1982). "Heating and cooling stimulations during SWS and REM sleep in man". In: 7.3, pp. 155–158.
- Castrodes, R. Q., E. J. J. Funa, H. N. G. Lim, H. L. P. Angelia, and N. B. Linsangan (2020). "Android Application-Based Controller for Air Conditioning Units". In: *2020 IEEE 12th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management*.
- Christidis, N. (2024). *Met Office Climate Attribution Analysis for Europe 2024*. UK Met Office.
- Copernicus Climate Change Service (2024). *European Summer 2024 – Hottest on Record*. European State of the Climate Report 2024. European Centre for Medium-Range Weather Forecasts (ECMWF).
- (2025). *European Summer 2025 – Hot in the West and South, Dry in the Southeast*. ECMWF.
- Deroncourt, F. (2013). *Introduction to fuzzy logic*. Lecture notes / online material.

- Economo, C. von (1930). "Sleep as a problem of localization". In: 71.3, pp. 249–259.
- Electronics Coach (2026). *Closed-Loop Control System*. URL: <https://electronicscoach.com/closed-loop-control-system.html> (visited on 02/18/2026).
- European Environment Agency (2024). *Attribution of European Heat Events to Anthropogenic Climate Change*. EEA Technical Report No. 18/2024.
- Hossain, E. (2022). "Introduction to MATLAB". In: *MATLAB and Simulink Crash Course for Engineers*. Springer, pp. 1–18.
- Lan, L., K. Tsuzuki, Y. Liu, and Z. Lian (2017). "Thermal environment and sleep quality: A review". In: *Energy and Buildings* 149, pp. 101–113.
- Met Office (2024). *Press Release: European Heat Events and Climate Change Attribution*. London, UK.
- Ngarambe, J., G. Y. Yun, K. Lee, and Y. Hwang (2019). "Effects of changing air temperature at different sleep stages on the subjective evaluation of sleep quality". In: 11.5, p. 1417.
- Panasonic Industry (2026). *Infrared Array Sensor Grid-EYE*. URL: <https://industrial.panasonic.com/ww/products/pt/grid-eye> (visited on 02/18/2026).
- Tateno, S., Y. Zhu, and F. Meng (2019). "Hand gesture recognition system for in-car device control based on infrared array sensor". In: *2019 58th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE)*.
- Teja, R. (2021). *Open Loop System*. URL: <https://www.electronicshub.org/open-loop-system> (visited on 02/18/2026).
- Zhao, S., F. Blaabjerg, and H. Wang (2020). "An overview of artificial intelligence applications for power electronics". In: *IEEE Transactions on Power Electronics* 36.4, pp. 4633–4658.

Model Transparency: Integrating XGBoost with SHAP for Explainable Machine Learning

Andrej Bednařík¹

Abstract

This paper examines integrating the XGBoost algorithm with SHAP values to balance predictive performance and model interpretability. XGBoost is widely recognized for its high accuracy and efficiency, yet its ensemble structure makes the internal decision-making process difficult to interpret. SHAP offers a theoretically grounded framework based on Shapley values that enables both global and local explanations of model behavior. The paper focuses on identifying key predictors using SHAP summary analysis, exploring variable interactions, and providing detailed explanations of individual predictions through local SHAP visualizations. The results show that combining XGBoost with SHAP creates a robust and transparent modeling framework suitable for domains where explainability is essential. Moreover, SHAP uncovers complex feature relationships that traditional feature-importance methods miss, thereby improving the overall interpretive value of the model.

Keywords

XGBoost; SHAP; model interpretability; feature importance; machine learning

1 Introduction

In recent years, the field of machine learning has increasingly emphasized not only achieving strong predictive performance but also ensuring that models are interpretable and transparent. In application domains such as banking, insurance, and healthcare, understanding why a model makes a particular decision is critical for accountability, regulatory compliance, and user trust. However, the highest predictive accuracy is often achieved by complex models, most notably tree-based ensemble methods, which can obscure the underlying decision mechanism and create a black-box effect. One of the most widely used tools in practice is XGBoost, a powerful gradient-boosting framework capable of handling large-scale data and delivering excellent predictive performance (Chen and Guestrin, 2016). At the same time, XGBoost typically relies on an ensemble of hundreds or even thousands of decision trees, which makes direct interpretation of decision paths difficult. For this reason, in recent years, methods and tools have been developed to make such high-performing models more explainable. A prominent approach is SHapley Additive exPlanations, or SHAP, which is grounded in cooperative game theory and assigns each input feature a contribution to a given model output. SHAP provides a unified framework for explaining predictions, satisfying desirable properties such as local accuracy and symmetry, and enabling consistent comparison across explanation techniques (Lundberg and Lee, 2017). In practice, simpler classical feature importance measures are still commonly used for XGBoost, such as Gain, Cover, or Weight, which produce global rankings of feature relevance. These measures, however, may overlook feature interactions, nonlinear relationships, and the distinction between global and local importance. In contrast, SHAP supports both global and instance-level explanations and can reveal patterns that classical importance metrics may fail to capture (Ponce-Bobadilla et al., 2024). Accordingly, the focus of this article, *Model Transparency: Integrating*

¹Department of Mathematics and Actuarial Science, Faculty of Economic Informatics, Bratislava University of Economics and Business, Slovakia. Email: andrej.bednarik@euba.sk

XGBoost with SHAP for Explainable Machine Learning, is to integrate predictive performance with interpretability by combining XGBoost with SHAP-based explanations. The goal is to evaluate how SHAP complements or differs from traditional XGBoost feature importance analysis.

2 A search for sources and the current state of the problem

To position this article within the current research landscape, the literature search focused on peer-reviewed and widely cited work on explainable machine learning for tree ensembles, with emphasis on XGBoost and SHAP (Barredo Arrieta et al., 2020). The search strategy combined queries in Scopus, Web of Science, ACM Digital Library, IEEE Xplore, and Google Scholar using keyword sets such as XGBoost, gradient boosted trees, explainable machine learning, explainable AI, SHAP, Shapley values, TreeSHAP, feature importance, and post hoc explanations (Angelov et al., 2021). Priority was given to foundational methodological papers, systematic reviews, and domain-focused surveys in high-impact application areas where transparency is required, especially finance and healthcare (Weber et al., 2024). Across the broader explainability literature, there is strong agreement that predictive performance alone is insufficient in many real-world deployments, because stakeholders often need to understand and justify model behavior (Doshi-Velez and Kim, 2017). This need is especially visible in high-stakes domains, where explanations support accountability, auditing, and error analysis (Barredo Arrieta et al., 2020). Survey papers provide structured taxonomies of explainability and clarify common distinctions such as global versus local explanations and model-specific versus model-agnostic methods (Guidotti et al., 2019). At the same time, critical perspectives argue that explanations do not automatically guarantee safety or validity, and in some settings, inherently interpretable models can be preferable to explaining complex black box systems (Rudin, 2019). This tension motivates the practical focus of this article, namely how to combine high-performing tree boosting with explanation methods in a way that is useful and methodologically defensible (Doshi-Velez and Kim, 2017).

Within this space, SHAP is one of the most influential frameworks for feature attribution because it is grounded in cooperative game theory and uses a consistent additive form to explain individual predictions (Lundberg and Lee, 2017). For tree-based models, later work introduced efficient algorithms and tooling that make Shapley-style explanations feasible at scale and enable moving from local explanations to global understanding (Lundberg, Erion, et al., 2020). These contributions are particularly relevant for XGBoost, a widely used gradient boosting implementation known for its strong accuracy and scalability (Chen and Guestrin, 2016). In practice, however, many projects still rely on built-in XGBoost importance measures such as gain, cover, or split counts (Chen and Guestrin, 2016). Such summaries can miss interactions, nonlinearities, and differences across instances, leading to a partial or misleading narrative about what drives model outputs (Guidotti et al., 2019). This creates an applied gap because users may obtain different conclusions depending on whether they interpret the model with built-in importance or with SHAP-based attributions (Barredo Arrieta et al., 2020).

Recent research also clarifies limitations that shape the current state of the problem. SHAP attributions can be sensitive to feature dependence because common implementations may rely on assumptions that are violated when predictors are correlated, thereby altering how contributions are allocated (Aas et al., 2021). Other studies question whether Shapley-based importance should be treated as a human-centered explanation rather than a mathematical attribution, noting that stronger explanatory claims often require additional assumptions and, in some cases, causal framing (Kumar et al., 2020). There is also evidence that post hoc explanation methods can be manipulated under certain threat models, which is important when explanations are used for compliance or trust rather than primarily for debugging (Slack et al., 2020). Together, these findings support the view that explainability should be treated as a careful methodological layer rather than a simple visualization add-on (Guidotti et al., 2019).

From an application perspective, systematic reviews and bibliometric studies show rapid

growth of explainable machine learning in regulated sectors (Sharma et al., 2024). In finance, a systematic review documents diverse explainability goals and methods and links adoption to increasing expectations for transparency in risk-sensitive decision support (Weber et al., 2024). In healthcare, bibliometric analysis reports a rapid expansion of empirical work on explainable AI and highlights a concentration of output in leading research countries and its frequent use in prediction and diagnostic settings (Dhiman et al., 2023). These trends indicate that integrating high-performance models, such as XGBoost, with explanation frameworks, such as SHAP, aligns with current research priorities and real-world deployment needs worldwide (Barredo Arrieta et al., 2020).

3 The main findings of the article

The task was formulated as supervised regression, where the target variable was the insurance premium (`poistne_pzp`). The dataset contained approximately 124,000 records and combined numeric predictors, categorical variables, and date or time information. Before model training, the dataset underwent a structured data processing workflow to improve data quality, reduce noise, and ensure that all variables were represented in a model-compatible form. This pre-processing step included basic consistency checks, handling of missing or invalid entries, and verification of ranges for key numeric variables to prevent extreme values from dominating the learning process. Categorical predictors were standardized to remove label inconsistencies and encoded in a format suitable for gradient-boosted trees, while numeric variables were retained at their original scales because tree-based methods do not require normalization. Date and time variables were transformed into usable representations, such as extracted years or derived duration-based quantities, enabling temporal signals to be incorporated without relying on raw timestamps. Following preprocessing, the final feature set was constructed from a combination of contract-related factors, vehicle attributes, and regional or segment descriptors, namely `Datum_pociatku`, `Datum_storna`, `Druh_auta`, `Frekvencia_platenia`, `Hmotnost`, `Kanal`, `Kategoria_vozidla_popis`, `Objem`, `Sposob_pouzitia_auta`, `Vykon`, `Znacka_vozidla`, `bonus_malus`, `okres`, `vek`, and `vek_vozidla`.

The goal was not only to achieve strong predictive performance in ex post analysis, but also to ensure that the resulting model could be interpreted transparently. For this reason, the modeling workflow combined performance evaluation with a systematic interpretability layer. After training, predictive accuracy was assessed using standard regression metrics, including RMSE, MAE, and R-squared, to capture both typical prediction error and sensitivity to large deviations. In parallel, model explanations were generated using SHAP values, which decompose each prediction into an additive combination of feature contributions relative to a baseline expectation. This enabled two complementary perspectives on the model: a global view describing which variables tend to matter most across the portfolio, and a local view explaining why a particular policy receives a higher or lower predicted premium. In practical terms, this integration supports model validation and communication by enabling analysts to compare classical XGBoost importance rankings with SHAP-based attributions and to identify nonlinear effects, thresholds, and segment-specific behavior that are common in insurance pricing data.

3.1 What is SHAP

To overcome the limitations of classical feature importance measures in XGBoost-type models, such as weight and gain, SHAP has emerged as an advanced approach. SHAP is grounded in cooperative game theory and assigns each input feature an individual share of the model prediction for a specific observation. As a result, it provides not only a global overview of importance but also the ability to explain predictions locally. A key advantage of SHAP is that it satisfies properties such as symmetry, efficiency, and additivity, which makes it a theoretically

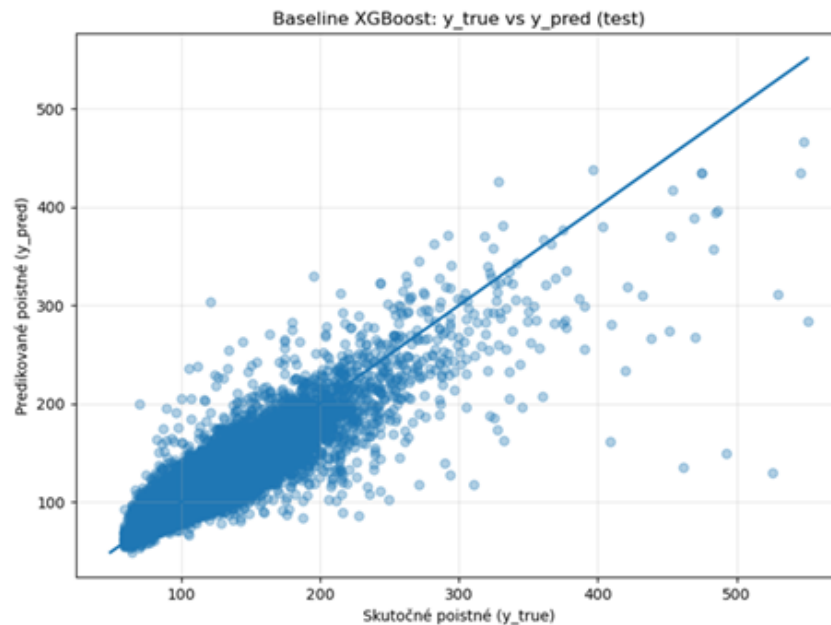


Figure 1. Gain-based feature importance ranking (schematic).

well-founded framework for determining feature importance (Lundberg and Lee, 2017). On a practical level, this means that for every instance in the data, we can obtain a set of SHAP values showing how each variable contributed to shifting the prediction from the baseline value to the final output. SHAP therefore enables the identification of high-impact variables, the determination of whether their contributions are positive or negative, and the detection of interactions or nonlinear effects that may be masked by simpler metrics. In the context of XGBoost, this means that while metrics such as gain or weight provide a certain perspective, SHAP offers a deeper, more detailed interpretation of model behavior across data segments and at the level of individual predictions. From a technical standpoint, SHAP values rely on evaluating all possible combinations of input variables, or an approximation of these combinations, in order to estimate the marginal contribution of each feature, which supports consistent and fair attribution (Lundberg and Lee, 2017). However, using SHAP also incurs higher computational costs and requires careful interpretation. In particular, large SHAP values for a feature do not automatically imply causality; rather, they indicate a strong associative relationship within the trained model. For this reason, integrating SHAP-based interpretations should be accompanied by a thorough understanding of the data, domain requirements, and the broader modeling context.

3.2 Predictive performance of the XGBoost regression model

Using the selected feature set, an XGBoost regression model was trained to predict premium values. Performance was assessed on a held-out test set using standard regression metrics. RMSE captures sensitivity to larger errors, MAE provides an interpretable measure of typical absolute deviation, and R-squared summarizes the proportion of variance explained. The final test performance achieved $RMSE = 16.63$, $MAE = 10.58$, and $R^2 = 0.805$. These results confirm that gradient boosted tree ensembles are well-suited for premium prediction in heterogeneous tabular insurance data and can model nonlinear relationships between risk drivers and premium levels.

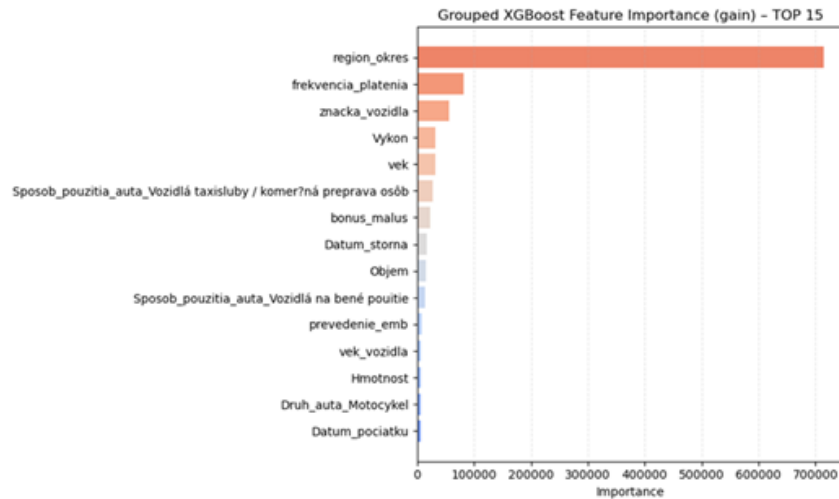


Figure 2. SHAP global importance ranking (mean absolute SHAP) (schematic).

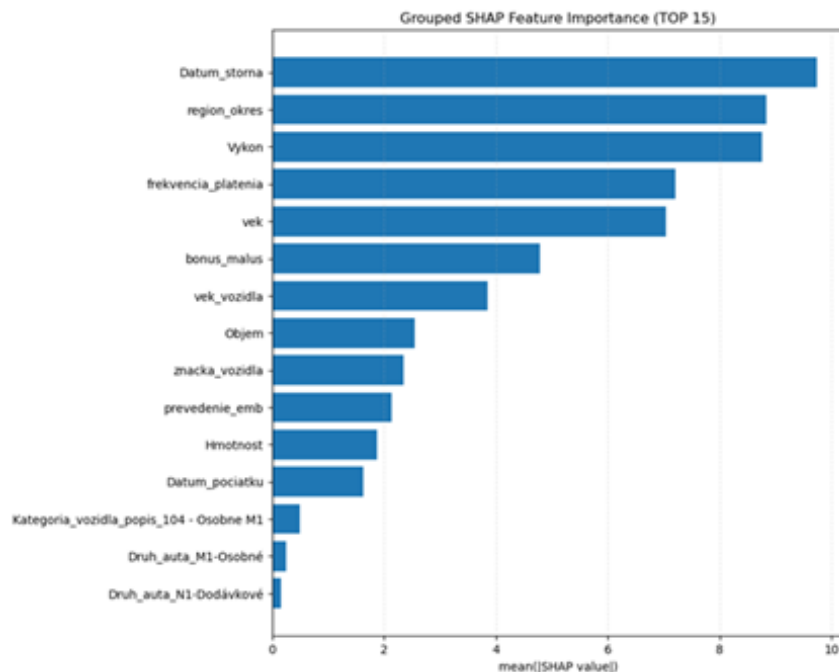


Figure 3. SHAP beeswarm plot (schematic).

3.3 Global importance differences: gain vs SHAP (full text from the paper)

The SHAP-based ranking can differ substantially from the gain-based feature importance because the two metrics capture different concepts: gain reflects how much a feature improves the objective when used in splits during training, whereas mean absolute SHAP reflects how much the feature actually shifts predictions across all observations. As a result, a variable can have high gain due to a small number of very strong splits yet show lower SHAP importance if it affects only a narrow subgroup. Conversely, a variable can have a moderate gain but high SHAP importance if it consistently shifts across many policies. In the context of this study, which is explicitly framed as an ex post analysis, the dominance of Datum_storna is therefore interpreted as evidence that the model captures a strong policy lifecycle and temporal signal present in the historical data-generating process, and that it meaningfully differentiates premium outcomes across the portfolio.



Figure 4. SHAP dependence: bonus_malus vs Vykon (schematic).

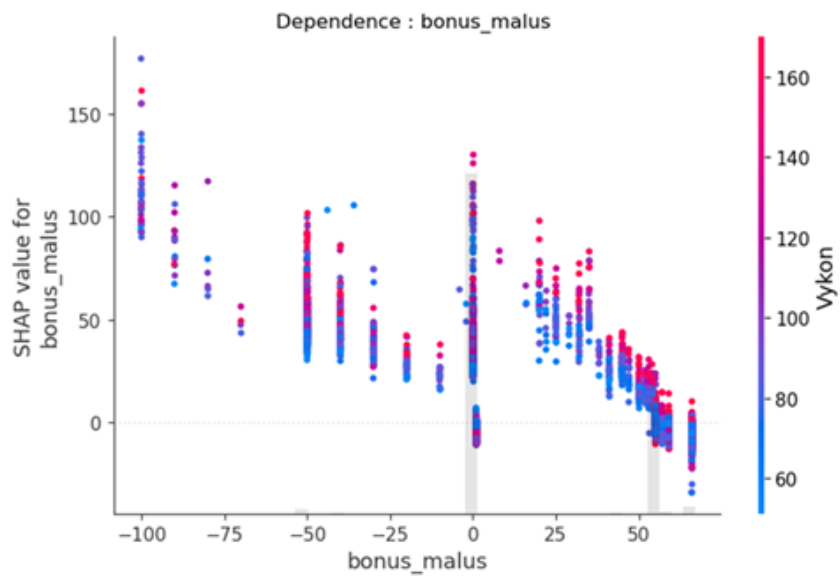


Figure 5. SHAP dependence: vek vs bonus_malus (schematic).

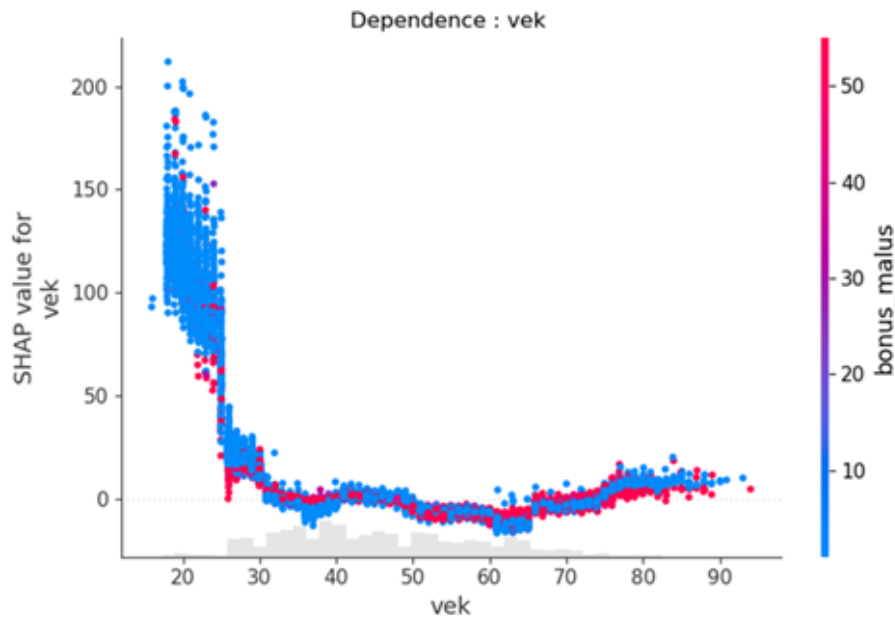


Figure 6. Local interpretability: SHAP waterfall plot (schematic).

3.4 SHAP beeswarm plot (full text from the paper)

Figure 3 presents a SHAP beeswarm plot providing a global summary of feature effects by displaying the distribution of SHAP values for each variable across all observations. Features are ordered from top to bottom by overall importance, typically measured by the mean absolute SHAP value, so variables at the top have the largest average impact on the predicted premium. Each point represents one policy in the dataset, positioned on the x-axis according to its SHAP value, which indicates how much that feature pushes the prediction above or below the model baseline. The color encodes the actual feature value, with blue indicating low values and red indicating high values, allowing the direction of the relationship to be interpreted visually.

The plot confirms that the strongest global drivers are Datum_storna, Vykon, and vek. For Datum_storna, higher values are associated with positive SHAP contributions, meaning that later cancellation-related values tend to increase predicted premiums relative to the baseline, while lower values tend to have smaller or negative effects. For Vykon, the color pattern shows that higher engine power values are concentrated on the positive SHAP side, indicating that more powerful vehicles systematically increase predicted premiums, whereas lower power values are associated with negative or near-zero contributions. For vek, the spread is asymmetric, with younger ages predominantly on the positive SHAP side, indicating that younger customers receive higher predicted premiums, while older ages tend to cluster near zero or slightly negative values, consistent with the dependence plot, which shows the age effect becoming milder after early adulthood.

For bonus_malus, the beeswarm shows a wide spread of SHAP contributions, indicating that this variable can strongly shift predictions in both directions depending on its level. The color distribution suggests that different bonus malus values correspond to distinct premium adjustments rather than a single smooth trend, which aligns with the discrete bands seen in the dependence plot and reflects stepwise segmentation learned by the tree ensemble. The feature vek_vozidla also shows a clear directional pattern, where higher vehicle age tends to be associated with negative SHAP values, indicating that older vehicles reduce the predicted premium relative to the baseline in this dataset, while newer vehicles contribute positively. Objem and Hmotnost contribute with smaller but still visible effects, and their spreads indicate that these technical vehicle characteristics matter primarily in certain ranges rather than uniformly

across all observations.

Several one-hot-encoded regional indicators, such as `okres_Bratislava`, `okres_Komárno`, `okres_Nové Zámky`, and `okres_Žilina`, are identified as important contributors, confirming that geographic segmentation is embedded in the model and that specific districts systematically shift predicted premiums. Because these are binary indicators, the color typically distinguishes between the presence and absence of the category, and the horizontal spread reflects the magnitude of the premium shift when the policy belongs to that district. Brand indicators such as `Znacka_vozidla_BMW`, `Znacka_vozidla_AUDI`, `Znacka_vozidla_CITROEN`, and `Znacka_vozidla_KIA` appear lower in the ranking, with narrower SHAP spreads, suggesting that brand effects are present but less pronounced than the dominant drivers such as cancellation timing, power, age, and region. Finally, variables such as `Duplicita`, `Kanal`, `limit_plnenia`, and `Fyz_osoba` show SHAP values clustered tightly around zero, indicating minimal average contribution to the model output, either because their effect is weak in this dataset or because their information is largely captured by other correlated predictors.

3.5 Nonlinear effects and segment-specific behavior revealed by SHAP

A key advantage of SHAP is that it supports deeper diagnostics beyond a single ranking. Dependence plots can be used to examine how predicted premiums change across the range of a variable and to identify thresholds or nonlinearities typical for tree-based models. For example, continuous vehicle attributes such as `Objem`, `Hmotnost`, and `Vykon` often exhibit piecewise patterns, with their impact increasing after certain breakpoints. Similarly, contract-related variables such as `Frekvencia_platenia` can shift predicted premiums in a structured way that reflects the learned segmentation of payment patterns. Categorical variables such as `Znacka_vozidla`, `Druh_auta`, `Kategoria_vozidla_popis`, or `Kanal` can be investigated through the distribution of their SHAP contributions, which highlights whether certain categories systematically increase or decrease premiums.

3.6 `Bonus_malus` vs `Vykon` (full text from the paper)

Figure 4 presents a SHAP dependence plot illustrating how `bonus_malus` affects the model prediction through its SHAP contribution. The x-axis shows the bonus malus value, while the y-axis shows the SHAP value for `bonus_malus`, interpreted as the marginal impact of bonus malus on the predicted premium relative to the model baseline. Positive SHAP values indicate that the given bonus malus level increases the predicted premium, whereas values near zero or negative indicate little effect or a decrease relative to the baseline. Points are colored by `Vykon` (engine power), which helps reveal whether the effect of bonus malus changes systematically with vehicle power.

A clear overall pattern is visible: as `bonus_malus` increases into higher positive values, the SHAP contribution generally declines toward zero and can even become negative for the highest levels. This means the model associates higher bonus malus values in this dataset with smaller premium increases, and in some cases with a reduction relative to the baseline. In contrast, strongly negative bonus malus values are associated with large positive SHAP values, indicating substantial premium increases for those policies. The plot also shows distinct vertical bands, suggesting that `bonus_malus` takes on discrete or heavily clustered values in the dataset, so the model learns stepwise adjustments rather than a smooth, continuous relationship.

The coloring by `Vykon` indicates that engine power contributes additional differentiation within the same bonus malus levels. For many bonus malus values, observations with higher `Vykon` tend to have higher SHAP values than those with lower power, suggesting that the model combines risk experience captured by bonus malus with vehicle performance when setting premiums. This interaction is particularly visible in the mid-range of bonus malus values, where there is noticeable spread in SHAP contributions at the same x value, and the color gradient

indicates that part of this spread is explained by differences in engine power. Overall, the figure demonstrates that the model treats bonus malus as a strong, structured risk signal, but its effect is not purely linear and is moderated by other risk-related variables, such as vehicle power.

3.7 Vek vs Bonus_malus (full text from the paper)

Figure 5 presents a SHAP dependence plot showing how the feature vek (customer age) influences the model prediction through its SHAP contribution. The x-axis represents age, while the y-axis shows the SHAP value for vek, interpreted as the marginal effect of age on the predicted premium relative to the model baseline: positive SHAP values indicate that age increases the predicted premium, while negative SHAP values indicate that age decreases it. The points are colored by bonus_malus, which allows the plot to reveal potential interaction effects between age and the bonus malus level.

The most prominent pattern is a very large positive SHAP contribution at young ages, roughly 28-30, where SHAP values are strongly positive and highly dispersed, in some cases exceeding 200. This indicates that the model assigns substantially higher predicted premiums to younger policyholders and that the effect varies across individuals, suggesting the presence of additional interacting factors. Immediately after this region, there is a sharp drop, and from approximately the early 30s onward, the SHAP values cluster close to zero, indicating that, for the majority of middle-aged policyholders, age alone contributes only modestly to premium adjustments relative to other drivers.

A second, weaker pattern appears at higher ages, where SHAP values become slightly positive again in older segments, indicating that the model begins to increase predicted premiums for elderly policyholders, although the magnitude is far smaller than for very young customers. Between roughly 40 and 65, many points fall slightly below zero, suggesting a mild discount effect for these ages relative to the baseline, consistent with a lower-risk segment identified by the model. The color gradient indicates how bonus_malus interacts with age. While the overall shape of the age effect is present across the full range of bonus malus values, the dispersion and extremes in the youngest segment show that bonus malus levels contribute to additional differentiation inside that group. In practice, this means the model does not treat young age as a single category; instead, it combines age with a bonus-malus to produce more granular premium adjustments, a typical behavior of tree-based ensembles that capture nonlinearities and interactions.

3.8 Local explanations of individual premium predictions

To support transparency on the level of single policies, local explanations were produced for selected observations using SHAP waterfall-style decompositions. Each local explanation breaks down a predicted premium into a baseline expected value and feature-specific contributions that push the prediction upward or downward. This makes it possible to explain why two policies with similar characteristics can still receive different premiums, for example, due to interactions between bonus_malus and vehicle attributes, or because certain categorical segments, such as okres or Znacka_vozidla, shift the model output even when numeric risk factors are comparable. Such local explanations are useful for case-based validation, error analysis, and communication with non-technical stakeholders, because they translate a complex ensemble decision into an additive narrative while remaining faithful to the trained model (Lundberg and Lee, 2017).

3.9 Local interpretability (full text from the paper)

Figure 6 presents a SHAP waterfall plot for a single observation. It decomposes the model's prediction into a baseline value and a sequence of feature contributions that push the prediction up or down. The starting point on the x-axis is the expected model output, $E[f(x)] = 109.547$,

and the final predicted premium for this particular policy is $f(x) = 152.678$. Red bars indicate features that increase the prediction, while blue bars indicate features that decrease it, and the length of each bar corresponds to the magnitude of the contribution. In this case, the strongest positive driver is `frekvencia_platenia`, which increases the prediction by approximately +23.08. It is important to note that payment frequency was encoded as a grouped categorical feature, combining the original four categories into a single grouping. Because of this grouping, the SHAP contribution shown here reflects the cumulative effect of belonging to the specific grouped payment frequency category for this policy. In other words, the model does not attribute the effect to separate dummy levels in this visualization; instead, the displayed impact aggregates the contribution of the relevant category level, which is why `frekvencia_platenia` appears with a large single contribution.

Additional positive contributions come from `Vykon` (+9.7), `Datum_storna` (+9.11), `vek_vozidla` (+7.42), and `znacka_vozidla` (+6.2), indicating that the vehicle's power, cancellation-related timing, vehicle age, and brand segment all increase the predicted premium relative to the baseline. Smaller positive effects are also visible for `Objem` (+1.78), `Hmotnost` (+1.57), and `prevedenie_emb` (+0.89). On the negative side, `vek` decreases the prediction by about -12.98, suggesting that the customer's age places this observation in a segment with lower premiums than the average baseline. Further modest downward adjustments are attributed to `bonus_malus` (-1.67), `Datum_pociatku` (-1.02), `zmluva` (-0.64), and several minor categorical indicators, while the remaining features collectively contribute only a small additional negative shift. Overall, the plot provides a transparent case-level explanation: the predicted premium is substantially above the baseline primarily because the grouped payment frequency category contributes a strong upward adjustment, reinforced by power and temporal lifecycle-related features, while age offsets part of the increase with a notable negative contribution.

4 Conclusion

SHAP adds a substantial interpretability layer beyond standard XGBoost feature importance, explaining not only which variables matter but also how they influence predictions. Unlike gain-based importance, which provides a global ranking without indicating whether a feature increases or decreases the model's output, SHAP quantifies both the direction and magnitude of each feature's impact on the model's output. This enables clear local explanations at the level of individual policies, where each prediction can be decomposed into a baseline value and feature contributions that push the premium upward or downward. When aggregated across the dataset, SHAP also provides more reliable global insights because it reflects how features actually change predictions across the portfolio, rather than how often they are used in splits during training. In addition, SHAP reveals nonlinear relationships, threshold effects, and interactions between predictors, which are common in insurance pricing data and may remain hidden under classical importance metrics. Overall, integrating SHAP with XGBoost improves model transparency and supports more defensible interpretation and communication of model behavior in domains where explainability is critical.

Resources

- Aas, K., M. Jullum, and A. Løland (2021). "Explaining individual predictions when features are dependent: More accurate approximations to Shapley values". In: *Artificial Intelligence* 298, p. 103502. DOI: 10.1016/j.artint.2021.103502.
- Angelov, P. P., E. A. Soares, R. Jiang, N. I. Arnold, and P. M. Atkinson (2021). "Explainable artificial intelligence: An analytical review". In: *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 11.5, e1424. DOI: 10.1002/widm.1424.

- Barredo Arrieta, A., N. Diaz Rodriguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil Lopez, D. Molina, R. Benjamins, R. Chatila, and F. Herrera (2020). “Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI”. In: *Information Fusion* 58, pp. 82–115. DOI: 10.1016/j.inffus.2019.12.012.
- Chen, Tianqi and Carlos Guestrin (2016). “XGBoost: A scalable tree boosting system”. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Association for Computing Machinery, pp. 785–794. DOI: 10.1145/2939672.2939785.
- Dhiman, P., A. Bonkra, A. Kaur, Y. Gulzar, Y. Hamid, M. S. Mir, and A. B. Soomro (2023). “Healthcare trust evolution with explainable artificial intelligence: Bibliometric analysis”. In: *Information* 14.10, p. 541. DOI: 10.3390/info14100541.
- Doshi-Velez, Finale and Been Kim (2017). *Towards a rigorous science of interpretable machine learning*. arXiv: 1702.08608. URL: <https://arxiv.org/abs/1702.08608>.
- Guidotti, Riccardo, Anna Monreale, Salvatore Ruggieri, Franco Turini, Fosca Giannotti, and Dino Pedreschi (2019). “A survey of methods for explaining black box models”. In: *ACM Computing Surveys* 51.5, p. 93. DOI: 10.1145/3236009.
- Kumar, I. E., S. Venkatasubramanian, C. Scheidegger, and S. Friedler (2020). “Problems with Shapley value based explanations as feature importance measures”. In: *Proceedings of the 37th International Conference on Machine Learning*. PMLR, pp. 5491–5500.
- Lundberg, Scott M., Gabriel Erion, Hugh Chen, Alex DeGrave, Jordan M. Prutkin, Bala Nair, Ronit Katz, Jonathan Himmelfarb, Nisha Bansal, and Su-In Lee (2020). “From local explanations to global understanding with explainable AI for trees”. In: *Nature Machine Intelligence* 2.1, pp. 56–67. DOI: 10.1038/s42256-019-0138-9.
- Lundberg, Scott M. and Su-In Lee (2017). “A unified approach to interpreting model predictions”. In: *Advances in Neural Information Processing Systems*. Vol. 30. Curran Associates, pp. 4768–4777.
- Ponce-Bobadilla, A. V., V. Schmitt, C. S. Maier, S. Mensing, and S. Stodtmann (2024). “Practical guide to SHAP analysis: Explaining supervised machine learning model predictions in drug development”. In: *Clinical and Translational Science* 17.11, e70056. DOI: 10.1111/cts.70056.
- Rudin, Cynthia (2019). “Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead”. In: *Nature Machine Intelligence* 1.5, pp. 206–215. DOI: 10.1038/s42256-019-0048-x.
- Sharma, C., S. Sharma, K. Sharma, and G. K. Sethi (2024). “Exploring explainable AI: A bibliometric analysis”. In: *Discover Applied Sciences* 6, p. 615. DOI: 10.1007/s42452-024-06324-z.
- Slack, Dylan, Sophie Hilgard, Emily Jia, Sameer Singh, and Himabindu Lakkaraju (2020). “Fooling LIME and SHAP: Adversarial attacks on post hoc explanation methods”. In: *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*. Association for Computing Machinery, pp. 180–186. DOI: 10.1145/3375627.3375830.
- Weber, P., K. V. Carl, and O. Hinz (2024). “Applications of explainable artificial intelligence in finance: A systematic review of finance, information systems, and computer science literature”. In: *Management Review Quarterly* 74, pp. 867–907. DOI: 10.1007/s11301-023-00320-0.

Students' Perceptions of Generative Artificial Intelligence in Higher Education

Veronika Horniaková¹ Eva Rakovská²

Abstract

The rapid development of information technologies, accelerated by the pandemic, has significantly increased the presence of Artificial Intelligence (AI) in higher education. Although AI has long been embedded in many educational tools, our surveys show that students, including those in informatics programs, often lack a clear understanding of its mechanisms, possibilities, and limitations. Through two questionnaire-based studies, we examined the extent to which students adopt AI for learning, writing academic assignments, and programming, and their expectations regarding the proper use of these technologies. The findings reveal that, despite their technical proficiency, students' adoption of AI is uneven and largely intuitive, often without awareness of underlying principles or ethical considerations. At the same time, students express a strong need for formal education that would enable them to use AI safely, effectively, and responsibly in their academic work. The results confirm that universities should play an active role in guiding AI usage and building competencies that support its responsible integration into the educational process. Based on survey results, we identify key advantages and disadvantages of AI adoption in universities and provide recommendations for its effective implementation in academic environments.

Keywords

AI Adoption; Higher Education; AI in Learning; Educational Technology; Responsible AI Use; Digital Literacy

1 Introduction

The release of large language models such as ChatGPT, Gemini, or Claude significantly accelerated the integration of AI tools into higher education. Although AI technologies have long been embedded in search engines, recommendation systems, and learning platforms, the public accessibility of generative AI (GenAI) has fundamentally transformed students' interaction with digital tools.

Empirical evidence suggests that student adoption of generative AI is both rapid and widespread. According to the Higher Education Policy Institute (Freeman, 2025) Student Generative AI Survey, approximately 88% of UK students report using generative AI tools in some form to support their studies. Importantly, the survey highlights a significant discrepancy between the speed of student adoption and the level of formal institutional guidance. While students perceive AI as increasingly important for academic success, structured training and clear policies often remain limited.

Similarly, research conducted by Jisc (Attewell, 2025) confirms that AI tools have become deeply integrated into students' everyday academic routines. However, alongside enthusiasm, students also express concerns regarding misinformation, declining quality of work, overreliance

¹Bratislava University of Economics and Business, Faculty of Economic Informatics, Bratislava, Slovakia, veronika.horniakova@euba.sk

²Bratislava University of Economics and Business, Faculty of Economic Informatics, Bratislava, Slovakia, eva.rakovska@euba.sk

on automation, privacy risks, and long-term employability implications. These findings indicate that AI adoption is not merely a technological shift but also a pedagogical and ethical challenge.

A global perspective is provided by the Digital Education Council (Digital Education Council, 2024) Global AI Student Survey, which reports consistently high adoption rates across multiple countries. The survey emphasizes that students expect universities to play an active role in defining acceptable use, offering training, and ensuring responsible implementation. Thus, international evidence suggests that while AI usage is largely student-driven, governance and strategic direction are expected from institutions.

While existing international surveys and policy frameworks provide valuable macro-level insights into the adoption of generative AI in higher education, empirical evidence from Central and Eastern European contexts remains limited. Little attention has been paid to informatics-oriented students, who constitute a distinctive group due to their advanced digital literacy and frequent use of AI tools. This study contributes to the current discourse by contextualizing global findings within the Slovak higher education environment and by offering discipline-specific insight into AI usage patterns, ethical awareness, and expectations regarding institutional guidance. Rather than proposing a new theoretical adoption model, the paper provides empirically grounded evidence that refines and situates international trends within a specific national and disciplinary setting.

1.1 Students' Perceptions: Benefits, Risks, and Usage Patterns

Beyond adoption rates, recent studies have examined how students perceive and practically use generative AI tools.

Chan and Hu, in their study *Students' Voices on Generative AI*, surveyed 399 university students in Hong Kong and identified several perceived benefits of GenAI, including (Chan and Hu, 2023):

- personalized academic support,
- brainstorming assistance,
- explanation of complex concepts,
- support in research and writing tasks.

However, the study also highlighted concerns about accuracy, ethical misuse, academic integrity, and long-term societal impacts. Students recognized both the utility and potential risks of AI, demonstrating a nuanced perception rather than unconditional acceptance.

Mohammad et al. provide further insight into practical usage patterns, identifying common applications such as drafting assignments (Mohammad et al., 2025), summarizing texts, generating ideas, debugging code, and clarifying theoretical concepts. These usage patterns align closely with findings from our survey, in which students predominantly reported using AI for programming, web research, and software-related tasks.

A broader synthesis is offered by Campbell University's *AI in Higher Education: A Meta Summary of Recent Surveys of Students and Faculty*, which aggregates multiple large-scale surveys (including HEPI and institutional studies). The meta-summary indicates that roughly 80% of students globally use generative AI to support learning. Notably, it also identifies a gap between students' generally positive attitudes and faculty members' more cautious perspectives.

Further evidence from *Inside Higher Ed* (Flaherty, 2025) suggests that students do not perceive AI as diminishing the value of higher education. Instead, AI changes how they conceptualize learning and academic work. While some express concern about critical thinking skills, many report that AI can enhance understanding when used appropriately.

Collectively, these studies indicate that AI adoption among students is neither purely instrumental nor purely problematic. Rather, it is characterized by pragmatic usage, perceived efficiency gains, and ongoing ethical uncertainty.

While studies on perception are abundant, the question of whether AI positively or negatively affects learning outcomes remains central. A comprehensive meta-analysis conducted by Wang et al., published within *Nature Humanities and Social Sciences Communications*, examined 51 empirical studies from 2022–2024 (H. Wang et al., 2024). The analysis found:

- a large positive effect on learning performance,
- a moderate positive effect on learning perception,
- a moderate positive effect on higher-order thinking skills.

Importantly, the study emphasizes that the educational impact of ChatGPT and similar tools depends heavily on instructional design, duration of use, and pedagogical integration. AI demonstrates the strongest positive effects when embedded in structured learning environments rather than used as an unsupervised shortcut.

This finding is particularly relevant to our research. Although our survey indicates widespread AI use among IT students, it also reveals an uneven understanding of the underlying principles and limited awareness of ethical considerations. The literature suggests that without pedagogical framing, the potential cognitive benefits of AI may not be fully realized.

1.2 Empirical Study: AI Usage Among IT Students at the University of Economics in Bratislava

To complement international findings on AI adoption in higher education, we conducted an empirical study focused on students enrolled in informatics-oriented programs at the University of Economics in Bratislava, Slovakia. The objective was to examine students' awareness of AI technologies, their actual usage patterns in academic contexts, and their perceptions regarding potential AI applications within university educational and managerial processes.

1.2.1 Research Design and Methodology

The study was conducted between September and November 2024 using a structured questionnaire distributed electronically to students of:

- Economic Informatics (Bachelor's level), and
- Information Management (Master's level).

Although more than 300 students were invited to participate, we obtained 48 valid responses. The questionnaire consisted of 12 questions combining single-choice, multiple-choice, and open-ended items. The questions addressed:

- general attitudes toward AI usage at the university,
- perceived usefulness of AI tools,
- specific areas where AI could improve university processes,
- personal usage of AI tools in academic work,
- perceived advantages and risks of ChatGPT and similar models.

A primary limitation of this study is the relatively small sample size ($n = 48$), which limits statistical generalizability. The findings should therefore be interpreted as exploratory rather than representative of the broader higher education population.

However, the focused and homogeneous nature of the sample informatics-oriented students enhances contextual relevance. As future IT professionals with high exposure to digital technologies, this group provides meaningful insight into AI adoption patterns within a clearly defined disciplinary environment. The study thus prioritizes contextual depth over broad generalization.

Additionally, the cross-sectional design captures perceptions and usage patterns at a single point in time amid rapid technological development. Future research will expand the dataset and incorporate a longitudinal follow-up to strengthen generalizability and track changes in AI adoption over time.

1.2.2 General Attitudes Toward AI in University Processes

Most respondents indicated that AI could improve university processes (Figure 1). However, several students emphasized that the benefits would depend heavily on the quality of implementation. This conditional support reflects a pragmatic rather than idealistic view of technological integration.

When asked which technologies would be appropriate for university use, students most frequently suggested:

- chatbots,
- machine learning methods,
- rule-based systems,
- augmented and virtual reality solutions.

Chatbots were perceived as particularly suitable for administrative support functions, especially in the study office context.

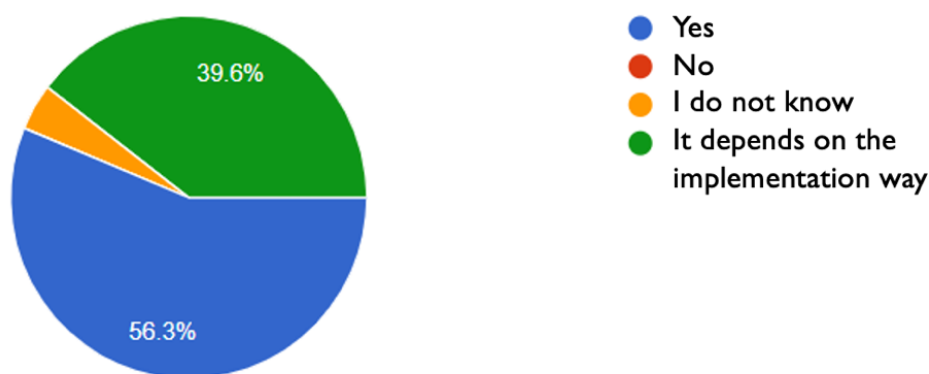


Figure 1. Process improvement by AI at the university (Own processing)

1.2.3 Perceived Applications of Chatbots

Among the 48 respondents, chatbot usage was most frequently suggested for:

- answering frequently asked questions (79.2%),
- providing general study information (77.1%),

- supplying information about subjects and syllabi (72.9%),
- consultation hours and contact details (60.4%),
- academic calendar information (56.3%).

These responses indicate that students associate AI primarily with operational efficiency and information accessibility rather than with strategic or pedagogical transformation.

Interestingly, one respondent explicitly suggested that AI could support “more effective learning with critical evaluation of gained information,” demonstrating awareness of AI’s potential educational role beyond administrative support. Below are students’ responses to the suggested tools (Figure 2).

■ What technologies and tools would you suggest?

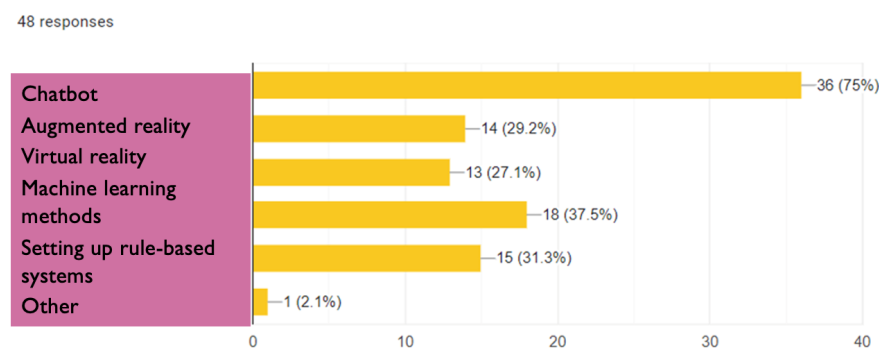


Figure 2. AI tools suggested by the students (Own processing)

Most respondents reported using some form of AI in their studies. ChatGPT was by far the most frequently mentioned tool. Other tools included:

- Bing AI,
- Bard,
- GitHub Copilot,
- PhotoMath,
- DeepL,
- automated syntax correction systems in programming environments.

Students most commonly used AI for:

- web research,
- programming tasks,
- working with various software tools (over 55% in each category).

A smaller proportion reported using AI for:

- writing assignments,
- verifying information,



Figure 3. Usage of AI tools (Own processing)

- mathematical problem solving.

Only a very small number of respondents reported not using AI at all. We can see the findings in Figure 3.

These findings align with international evidence indicating that AI adoption is particularly strong in technical and programming-related contexts (Mohammad et al., 2025; Chan and Hu, 2023).

1.2.4 Perception of ChatGPT as a Learning Tool

Approximately two-thirds of respondents considered ChatGPT to be a suitable learning tool. However, students also demonstrated awareness of important limitations. Key concerns included:

1. The need to verify information due to possible inaccuracies.
2. Limited contextual understanding of the model.
3. The importance of not replacing human interaction and traditional learning resources.

This nuanced perception suggests that students do not view AI as a substitute for education, but rather as a supplementary tool; this is consistent with broader international findings (J. Wang and Fan, 2025).

2 Conclusion

The rapid diffusion of AI tools in higher education has prompted international organizations to develop governance frameworks.

The UNESCO (Miao, Holmes, et al., 2021) publication, *AI and Education: Guidance for Policy-makers*, emphasizes the inclusive, human-centered integration of AI into educational systems. Key principles include:

- development of AI literacy,
- teacher training,
- ethical safeguards,
- protection of privacy and data,
- equitable access to technology.

More specifically, UNESCO's *Guidance for Generative AI in Education and Research* (Miao and Holmes, 2026) addresses generative AI tools directly, highlighting the need for ethical

validation, transparency, human oversight, and protection against overreliance on automated systems.

Complementing this perspective, the OECD AI Principles (Russo and Oder, 2023) define international standards for trustworthy AI, emphasizing human-centered values, robustness, transparency, accountability, and safety. These principles provide a normative framework for institutional AI governance.

Institutional policy analysis by Wang (H. Wang et al., 2024) further shows that universities worldwide are still developing coherent AI policies. Many institutions adopt reactive strategies, issuing guidelines after widespread student adoption has already occurred. This mirrors findings from our own survey context, where AI usage appears to have outpaced formal institutional regulation.

Key Findings and Identified Gaps

Despite being informatics students, respondents revealed several knowledge gaps:

- Many were unable to clearly distinguish between augmented and virtual reality.
- A significant number did not fully understand that they already use AI daily (e.g., search engines, MS Office tools, developer software, translation systems).
- Students often associate AI exclusively with specific subjects such as “Artificial Intelligence” or “Machine Learning,” rather than recognizing its broader systemic integration.

Furthermore, when asked about AI’s role in managerial university processes (e.g., enrollment, Erasmus mobility, financial procedures), many students demonstrated limited awareness of institutional workflows. This suggests that students’ AI-related thinking is primarily tool-oriented rather than process-oriented.

Our findings indicate that AI adoption among IT students is:

- widespread but largely intuitive,
- operational rather than strategic,
- tool-focused rather than principle-based,
- accompanied by limited awareness of governance and ethical implications.

Even within a technically proficient population, AI literacy regarding underlying mechanisms, risks, and institutional processes remains incomplete.

At the same time, students expressed openness to formal AI-related education and recognized the potential benefits of structured implementation. This confirms international calls for universities to assume an active role in developing AI competencies (Miao, Holmes, et al., 2021; Miao and Holmes, 2026; Russo and Oder, 2023).

The present study provides an initial empirical insight into AI adoption among informatics-oriented students within the Slovak higher education context. Future research should focus on expanding the sample size and including additional cohorts of IT students to strengthen statistical robustness and improve the stability of the findings.

A key direction for further investigation involves the planned longitudinal follow-up. Repeated data collection over time would enable examination of evolving AI usage patterns, shifts in ethical awareness, and changes in students’ expectations regarding institutional regulation and guidance. Considering the rapid development of generative AI technologies and the gradual establishment of university-level policies, a longitudinal design would provide valuable insight into the dynamics of AI adoption in higher education.

Future studies may also explore relationships between AI usage intensity, academic performance, digital competencies, and attitudes toward responsible AI use. Integrating qualitative methods, such as interviews or focus groups, could further enrich the understanding of students' motivations and decision-making processes when using AI tools.

Implications for Higher Education Institutions

Drawing on both our empirical findings and international frameworks, universities should:

- integrate AI literacy into curricula across disciplines,
- establish transparent institutional policies on acceptable AI use,
- provide systematic training for academic staff,
- implement pilot AI-supported learning initiatives with ongoing evaluation,
- maintain human oversight in assessment and decision-making processes.

The literature consistently indicates that AI is not inherently detrimental to learning. Rather, its impact depends on governance, pedagogical design, and responsible integration. Institutions that proactively address these dimensions are more likely to harness AI's educational benefits while mitigating associated risks.

Acknowledgement

This paper has been supported by the ESG grant A-25-103/3020-17 “Improving students' knowledge and skills in project management using the IPMA methodology”.

Resources

- Attewell, Sarah (May 2025). *Student Perceptions of AI 2025 – Artificial Intelligence*. Student Perceptions of AI 2025. URL: <https://nationalcentreforai.jiscinvolve.org/wp/2025/05/21/student-perceptions-of-ai-2025/>.
- Chan, Cecilia K. Y. and Wen Hu (2023). “Students' Voices on Generative AI: Perceptions, Benefits, and Challenges in Higher Education”. In: *arXiv*. DOI: 10.48550/arXiv.2305.00290. URL: <https://doi.org/10.48550/arXiv.2305.00290>.
- Digital Education Council (2024). *Digital Education Council Global AI Student Survey 2024*. URL: <https://www.digitaleducationcouncil.com/post/digital-education-council-global-ai-student-survey-2024>.
- Flaherty, Colleen (Aug. 2025). *Survey: College Students' Views on AI*. Inside Higher Ed | Higher Education News, Events and Jobs. URL: <https://www.insidehighered.com/news/students/academics/2025/08/29/survey-college-students-views-ai>.
- Freeman, Josh (Feb. 2025). *Student Generative AI Survey 2025*. HEPI Policy Note 61. URL: <https://www.hepi.ac.uk/reports/student-generative-ai-survey-2025/>.
- Miao, Fengchun and Wayne Holmes (Jan. 2026). *Guidance for Generative AI in Education and Research*. UNESCO.org. URL: <https://www.unesco.org/en/articles/guidance-generative-ai-education-and-research>.
- Miao, Fengchun, Wayne Holmes, Ronghuai Huang, and Hui Zhang (2021). *AI and Education: Guidance for Policy-Makers*. UNESCO. URL: <https://unesdoc.unesco.org/ark:/48223/pf0000376709>.
- Mohammad, N. M., M. Demers, E. McCubbin, J. Mitchell, and S. M. Fulmer (2025). “How College Students Use ChatGPT”. In: *Pedagogical Research* 10.4, em0250. DOI: 10.29333/pr/17428. URL: <https://doi.org/10.29333/pr/17428>.

- Russo, Laura and Norman Oder (Oct. 2023). *How Countries Are Implementing the OECD Principles for Trustworthy AI*. OECD.AI. URL: <https://oecd.ai/en/wonk/national-policies-2>.
- Wang, Hui, Anh Dang, Zihao Wu, and Son Mac (2024). “Generative AI in Higher Education: Seeing ChatGPT Through Universities’ Policies, Resources, and Guidelines”. In: *Computers and Education: Artificial Intelligence* 7, p. 100326. DOI: 10.1016/j.caeai.2024.100326. URL: <https://doi.org/10.1016/j.caeai.2024.100326>.
- Wang, Jun and Wei Fan (May 2025). *The Effect of ChatGPT on Students’ Learning Performance, Learning Perception, and Higher-Order Thinking: Insights from a Meta-Analysis*. Nature News. URL: <https://www.nature.com/articles/s41599-025-04787-y>.

Institutional Dependence and Systemic Risk in Modern Money Markets

Karimov Diyor Mukhtarovich¹ Peter Schmidt² Jamila Bobonazarova³ Azamat Khidirnazarov⁴

Abstract

The money market represents a key segment of the financial system, ensuring short-term capital allocation, liquidity, and the transmission mechanism of monetary policy. Although it is traditionally perceived as an example of the “free market,” empirical evidence and current trends demonstrate that regulatory frameworks, institutional safeguards, and trust in state and banking institutions condition its stability. This article analyzes the main money market instruments: Treasury bills, commercial paper, repurchase agreements, banker’s acceptances, and Eurodollars, and highlights their dependence on regulation and the solvency of issuers. Particular attention is devoted to systemic risk, moral hazard, and the concentration of power (“too big to fail”), which challenge the sustainability of the self-regulating market ideal in a globalized economy. In addition to global analysis, regional perspectives from Central Europe (Slovakia) and Central Asia (Uzbekistan) are considered, illustrating how institutional frameworks shape money-market stability across diverse contexts. An empirical illustration of Treasury bill yields confirms that prices and interest rates are shaped by central bank monetary policy rather than spontaneous market equilibrium. The conclusion emphasizes the need for a macro-prudential regulatory framework that fosters transparency, protects smaller participants, and minimizes systemic risks. Regulation thus appears not as a restriction, but as a necessary precondition for the stability and efficiency of the money market in the 21st century.

Keywords

Financial Stability; Monetary Policy Transmission; Shadow Banking; Market Regulation; Global Liquidity

1 Introduction

The money market constitutes one of the most critical segments of the financial system, where short-term, highly liquid financial instruments with maturities usually up to one year are traded. These instruments include Treasury bills, commercial paper, certificates of deposit, and interbank market instruments (Fabozzi and Modigliani, 2009). A key feature of the money market is its ability to efficiently redistribute temporarily available financial resources among economic entities, thereby maintaining liquidity and the stability of the entire financial system (Howells and Bain, 2008; Cecchetti and Schoenholtz, 2021).

At the same time, the money market plays a significant role in monetary policy, as it is through operations in this market that central banks influence short-term interest rates and regulate the amount of money in circulation. This transmission mechanism is crucial for managing aggregate demand, price stability, and macroeconomic equilibrium (Mishkin, 2019; European

¹National University of Uzbekistan named after Mirzo Ulugbek, Tashkent, Uzbekistan

²Faculty of Economic Informatics, Bratislava University of Economics and Business, Slovakia

³National University of Uzbekistan named after Mirzo Ulugbek, Tashkent, Uzbekistan

⁴National University of Uzbekistan named after Mirzo Ulugbek, Tashkent, Uzbekistan

Central Bank, 2020). Effective functioning of the money market is therefore essential not only for the stability of the banking sector but also for overall economic growth.

In economic theory, the money market is often considered an illustration of the functioning of the “free market,” in which the supply and demand for capital naturally determine equilibrium prices and yields. However, in practice, this ideal encounters several limitations. Current trends, such as extensive regulation, central bank interventions, moral hazard, and the concentration of market share in the hands of a few large financial institutions, suggest that the reality of the money market increasingly diverges from the classical liberal model (Bank for International Settlements, 2022; Pilbeam, 2018). These challenges raise questions about the boundaries and sustainability of the “free market” ideal in today’s globalized economy.

The aim of this article is therefore to analyze the current development of the money market, focusing on trends that challenge its traditional model, and to critically assess the extent to which it is still possible to speak of an autonomous, self-regulating financial environment.

2 Analysis of Individual Instruments

2.1 Treasury Bills

Treasury bills (T-bills) are short-term government securities issued to finance public expenditure. They are typically issued with maturities of up to one year and sold at a discount to their face value. T-bills are widely regarded as one of the safest financial instruments due to the backing of sovereign governments.

However, their safety is not purely market-driven. The demand for Treasury bills depends largely on trust in government solvency and macroeconomic stability. Central bank monetary policy significantly influences T-bill yields through policy rate adjustments and open market operations. Thus, even this seemingly free-market instrument operates within a framework of state-backed guarantees and institutional oversight.

2.2 Repurchase Agreements (Repos)

Repurchase agreements (repos) are structured as secured short-term loans. In a repo transaction, one party sells securities to another with an agreement to repurchase them at a specified date and price. Repos play a central role in liquidity provision for banks and financial institutions.

Despite their technical simplicity, repos are highly dependent on regulatory standards, collateral valuation practices, and central bank interventions. During periods of financial stress, repo markets may experience liquidity freezes, as observed during the 2008 financial crisis. Central banks frequently act as lenders of last resort to stabilize repo markets, underscoring their dependence on institutional support.

2.3 Commercial Paper

Commercial paper (CP) is an unsecured short-term debt instrument issued by corporations to finance working capital. While it appears to function within competitive market conditions, its viability depends strongly on credit ratings, investor confidence, and regulatory oversight.

The collapse of the commercial paper market during financial crises illustrates its vulnerability to systemic risk. When investor confidence deteriorates, corporations may face immediate refinancing challenges. Consequently, central banks often introduce emergency facilities to support commercial paper markets during crises.

2.4 Banker’s Acceptances

Banker’s acceptances are time drafts guaranteed by banks and commonly used in international trade. Their credibility rests on the financial strength of the guaranteeing bank. The instrument

demonstrates how private market transactions rely on institutional trust and regulatory capital requirements.

2.5 Eurodollars

Eurodollars refer to U.S. dollar-denominated deposits held outside the United States. Although they operate beyond direct U.S. jurisdiction, their stability depends on global banking regulations and international financial cooperation. The Eurodollar market highlights the globalized nature of money markets and their interconnected risks.

3 Limitations of the Free Market Ideal in the Era of Corporate Globalization

In the classical theoretical model of the “free market,” it is assumed that a large number of relatively small, mutually competing entities operate under conditions of symmetric information, high transparency, and equal opportunities. Such an environment is expected to allow market forces to spontaneously achieve efficient capital allocation without the need for external intervention. This ideal proved relatively functional in less concentrated economies with simpler financial market structures.

However, in today’s global economy and highly interconnected financial systems, it has become increasingly evident that this idealized model does not correspond to reality. As Schwarcz notes, modern financial markets function as complex and interdependent systems, where the failure of a single major institution or the uncoordinated behavior of several actors can trigger a domino effect with the potential to cause a systemic crisis. This phenomenon is known as systemic risk (Schwarcz, 2008). Regulation, therefore, cannot remain solely at the micro level (focused on individual institutions). However, it must also reflect macroprudential aspects of the financial system: relationships among institutions, external risks, and potential systemic shocks.

Another key issue is the concentration of economic and informational power. Large multinational corporations and financial institutions operate on such a scale that they can influence asset prices, capital flows, and the information environment, thereby creating unequal conditions for smaller market participants. As Stiglitz emphasizes, such an environment leads to systemic externalities, where the decisions of individual banks or investment funds may have negative consequences for the entire financial system—consequences that the market itself cannot effectively correct (Stiglitz, 2010).

3.1 Concentrated Risks and the “Too Big to Fail” Phenomenon

Institutions considered “too big to fail” implicitly expect state support in times of crisis. These expectations weaken market discipline—large institutions may engage in riskier operations because profits remain private, while potential losses are socialized. This distorts risk-taking behavior and leads to moral hazard (Acharya et al., 2010; Federal Reserve, 2020).

3.2 Interconnectedness, Complexity, and the Domino Effect

The high degree of interconnectedness among financial institutions means that problems in one bank or market segment can quickly spread throughout the system. A classic example is a liquidity crisis, where a lack of confidence and fears of default paralyze the market (European Central Bank, 2020). Such cascading effects are typical signs of the failure of the “free market” in complex systems.

3.3 Financial Innovation vs. Information Asymmetry

The emergence of complex and difficult-to-monitor financial products, including derivatives and securitized assets, has led to a situation that some authors liken to “financial alchemy” (European Corporate Governance Institute, 2022). These instruments can obscure real risks, create hidden leverage, and deepen information asymmetry. Without adequate regulation, the market’s ability to assess risks effectively is undermined.

3.4 Failure of Market Self-Regulation

In theory, the market should be able to “regulate itself” through competition and the rational behavior of actors. In reality, however, investors and creditors often lack sufficient information, motivation, or capacity to monitor risky behavior by large institutions. Moreover, when major entities are systemically important, market forces often fail to prevent destabilizing practices that benefit individuals but harm the system as a whole (Acharya et al., 2010; Federal Reserve, 2020).

These factors challenge the notion that the market mechanism in modern financial environments can ensure stability, efficiency, and fairness without significant state or regulatory intervention. The following sections of this article therefore analyze specific case studies (e.g., the 2008 crisis, the COVID-19 pandemic, volatility in repo and CP markets) and assess the extent to which the concept of the “free market” remains sustainable in the context of global corporate capitalism.

4 Regulatory Framework and Its Role in the Modern Financial System

The preceding analysis shows that relying solely on spontaneous market forces—the “invisible hand”—is insufficient in the context of globally interconnected, complex financial systems. As Harmon, Bar-Yam, and colleagues demonstrate in their study, *Networks of Economic Market Interdependence and Systemic Risk*, the modern global economy represents a network of complex interactions in which a shock in one part of the system often triggers a domino effect with consequences for the entire market (Harmon, Y. Bar-Yam, et al., 2011). Such systemic interconnections weaken markets’ ability to respond through self-regulation, thereby increasing the need for external regulatory and supervisory frameworks.

According to Acharya et al. (Acharya et al., 2010), several serious market failures, including excessive risk-taking, low transparency, and weak protection of small participants, contributed to the emergence of financial crises. They conclude that the market mechanism alone cannot ensure systemic stability if regulation focuses solely on individual risks while ignoring systemic risks.

For this reason, regulation should be viewed not as an obstacle to the market, but as its necessary precondition, a tool for creating an environment that fosters trust, responsibility, and resilience of the system as a whole. As the International Monetary Fund (International Monetary Fund, 2021) notes, effective regulation serves not only to protect small market participants but also to reduce systemic risks and ensure fair access to capital.

4.1 Key Functions of Regulation

a) Prevention of Systemic Collapses Regulation reduces the probability of cascading failures arising from strong interconnections among institutions, market segments, and countries. For example, requirements to hold adequate capital reserves lower the likelihood that a single liquidity shortfall will trigger a chain reaction across the system (Federal Reserve, 2020; International Monetary Fund, 2021).

b) Transparency and Information Disclosure Effective regulation promotes transparency in financial operations, disclosure of risk exposures, structures of derivative instruments, and

accounting standards. This helps reduce information asymmetry, which is a frequent cause of inefficient investor decisions (Acharya et al., 2010; ResearchGate, 2021).

c) Protection of Small Participants. In markets characterized by high concentration of power and informational inequality, regulation plays a crucial role in protecting retail investors, small banks, and non-professional participants who would otherwise lack the capacity to defend themselves against dominant actors (Stiglitz, 2010; Financial Regulation Courses, 2022).

d) Capital and Liquidity Requirements One of the most fundamental forms of regulation is capital adequacy requirements (e.g., Basel III), which oblige banks and other financial institutions to maintain sufficient reserves to absorb losses and shocks. These requirements also limit incentives for excessive risk-taking that could jeopardize system stability (Federal Reserve, 2020).

e) Rules of Conduct and Accountability The complex structure of modern financial instruments enables practices that maximize short-term profits for a small group of actors but may have adverse consequences for the system as a whole. Regulation can establish behavioral frameworks that account for long-term effects, externalities, and ethical standards (ScienceDirect, 2022; JSTOR, 2021).

Regulation, therefore, does not represent a denial of market principles, but rather their institutional complement in conditions where self-regulatory mechanisms cease to be effective. In global, complex, and technologically advanced markets, an active regulatory framework is an indispensable condition for stability, credibility, and sustainable development of the financial system.

5 Empirical Illustration of the Dependence of the Money Market on Regulation and Monetary Policy

Empirical analysis of key money market instruments, Treasury bills, repurchase agreements, commercial paper, and banker's acceptances shows that none of these segments operate based on pure market equilibrium. Treasury bill yields reflect central bank policy; repo operations depend on trust and legal certainty; the commercial paper market is subject to confidence volatility; and BA instruments rely on the reputation and solvency of banks.

These observations confirm that regulatory frameworks and institutional trust systemically condition the money market. Its stability and functionality do not arise from "self-regulation," but from formal mechanisms of supervision, rules, and the state's capacity to stabilize the financial environment. This supports the conclusion that, in a globalized economy, the free market alone is insufficient without a macroprudential framework (Acharya et al., 2010; Gorton and Metrick, 2012; Mishkin, 2019).

The money market is often perceived as the "purest" example of free market functioning—with high liquidity, low risk, and efficient allocation of short-term capital. International practice defines its main instruments as short-term government securities (Treasury bills), repurchase agreements (repos), certificates of deposit, commercial paper, banker's acceptances, and Eurodollars. These instruments are generally considered low risk and highly liquid, making them fundamental tools for liquidity management and short-term investment (Yaneer Bar-Yam et al., 2022).

However, their reliability does not arise spontaneously from market forces. On the contrary, it is closely tied to institutional guarantees, state credibility, central bank oversight, and firmly established regulatory frameworks. This observation directly supports the central thesis of this article: that the ideal of a self-regulating free market is not sustainable in modern, interconnected financial systems without active institutional control (Yaneer Bar-Yam et al., 2022; Stiglitz, 2010).

5.1 Treasury Bills as a “Conservative” Instrument and Their Institutional Conditioning

Short-term government securities, known as Treasury bills (T-Bills), are among the most conservative money market instruments. They are issued to cover short-term budget deficits and typically have maturities of 28, 91, or 182 days. Their yield does not arise from interest payments but from the difference between the purchase price (discount) and the nominal value at maturity. Since they represent obligations of the state, which possesses monetary sovereignty and the ability to issue new funds, T-Bills are traditionally regarded as “risk-free” (Federal Reserve, 2020). It must be emphasized, however, that this “risk-free” status does not originate from the market itself, but from trust in the state and its central institutions. The stability of these instruments is therefore the result of organized financial governance rather than spontaneous market development (Yaneer Bar-Yam et al., 2022; International Monetary Fund, 2021).

5.2 Development of T-Bill Yields as Evidence of Market Management by the Central Bank

The following table illustrates the development of yields (discount rates) of selected U.S. Treasury bills in 2025:

Table 1. Discount rates of selected U.S. Treasury bills in 2025

Type of security	Yield (approx., % p.a.)
4 week (\approx 28 days) T Bill	\sim 3.72%
91-day (\approx 13-week) T Bill	\sim 3.72–3.78%
182-day (\approx 26-week) T Bill	\sim 3.63–3.70%
1 year (\approx 52 weeks) T Bill	\sim 3.61–3.63%

Source: (A Red Ventures Company, n.d.).

This sharp increase in yields within a single year is not the result of organic movements in demand and supply for short-term investments, but rather a direct reflection of the Federal Reserve’s decision to raise benchmark interest rates (Federal Reserve, 2020). It is therefore a case of macroeconomic regulation of credit conditions, which is immediately transmitted into the prices and yields of money market instruments.

As the data clearly show, T-bill yields are not determined by the market but by monetary policy. This dependence demonstrates that even the most conservative and liquid money market instruments do not operate in free competition without external intervention (Acharya et al., 2010; International Monetary Fund, 2021). The market is not autonomous—the central bank structurally and price-wise shapes it.

T-Bills, often regarded as “zero risk” instruments, do not arise from pure market equilibrium but from state credibility, regulatory rules, and monetary policy. Their example shows that even in parts of the financial system considered closest to the free market ideal, institutional shaping of behavior and prices dominates. This once again confirms that, in a globalized and complex economy, regulation is an indispensable component of the money market, not its disruption.

5.3 Regional Perspectives: Central Europe and Central Asia

Although the fundamental instruments of the money market are broadly similar across jurisdictions, their functioning is strongly conditioned by local regulatory frameworks, institutional structures, and historical development. A comparison of Slovakia (as part of the euro area) and Uzbekistan (as a transforming Central Asian economy) illustrates the diversity of approaches to money market regulation and stability.

Slovakia (Central Europe)

- Integrated into the European Union and the euro area, Slovakia's money market is subject to the rules of the European Central Bank (ECB) and European legislation.
- The National Bank of Slovakia (NBS) supervises the banking sector and implements the ECB's monetary policy domestically.
- Short-term government securities and interbank operations are closely tied to euro area liquidity management.
- Challenges include oversight of smaller banks (LSIs) and fiscal deficits, which influence the issuance of short-term government securities.
- The Slovak money market is stable but has limited autonomy, as interest rates and liquidity are determined mainly by ECB policy.

Uzbekistan (Central Asia)

- Uzbekistan is undergoing a transition from a state-dominated banking sector to a more market-oriented system.
- Since 2019, reforms have strengthened the independence of the Central Bank of Uzbekistan (CBU), with a mandate to ensure price stability and financial resilience.
- The government has launched a strategy to privatize state-owned commercial banks and liberalize the financial sector.
- The interbank market has grown rapidly, with transaction volumes exceeding 100 trillion UZS in 2023, reflecting increasing liquidity and market activity.
- Despite progress, the Uzbek money market remains heavily dependent on state oversight, currency regulation, and institutional reforms.

United States (Global Benchmark)

- The U.S. money market is the largest and most liquid globally, serving as a benchmark for international practice.
- Instruments such as Treasury bills, repos, and commercial paper are widely traded, with yields directly influenced by Federal Reserve monetary policy.
- Challenges include shadow banking, systemic risk, and the "too big to fail" phenomenon, which highlight the need for robust macro prudential regulation.

6 Conclusion

The money market plays an indispensable role in modern economies by providing liquidity, facilitating short-term capital allocation, and transmitting monetary policy. Through instruments such as Treasury bills, commercial paper, and repo operations, the financial system can function effectively both in everyday operations and during periods of crisis. However, their apparent simplicity and low risk often conceal the fact that their stability and reliability are not the result of spontaneous market mechanisms, but instead of institutional oversight, trust in the state, and regulatory frameworks.

As demonstrated by the empirical analysis of Treasury bill yields in 2024–2025, money market interest rates are not solely determined by supply and demand. They are actively influenced by central bank monetary policy, confirming that even the most conservative segments of the market do not function autonomously.

Table 2. Discount rates of selected U.S. Treasury bills in 2025

Criterion	Slovakia	Uzbekistan	United States
Regulatory framework	EU/ECB integration, NBS oversight	Reforming the system, privatization, and the new CBU law	Federal Reserve + SEC, highly developed
Main authority	Slovak National Bank	Central Bank of Uzbekistan	Federal Reserve System
Market conditions	Stable, ECB-driven	Liberalizing, growing interbank market	Deep, liquid, global benchmark
Challenges	Oversight of small banks, fiscal deficits	Privatization, building robust regulation	Shadow banking, systemic risk
Market character	Stable, regulated, low autonomy	Transforming, state-influenced	Global, highly liquid, systemic

In theoretical terms, the free market model rests on assumptions of decentralization, information symmetry, and equal opportunity. In the reality of a globalized economy, however, high interconnectedness, power concentration, and systemic risks dominate, factors that the market alone cannot manage. Regional perspectives reinforce this conclusion: Slovakia’s money market is stable but highly dependent on ECB policy; Uzbekistan’s market is expanding yet remains strongly shaped by state oversight and ongoing reforms; and the United States, while serving as a global benchmark, faces challenges of shadow banking and systemic risk. These examples demonstrate that across diverse contexts, the money market cannot sustain itself without institutional frameworks.

- The money market is essential for maintaining liquidity and financial stability, but its functioning is unsustainable without a regulatory framework that accounts for systemic risks and macroeconomic factors.
- The free market ideal, based on the invisible hand and self-regulation, proves insufficient in the context of global, complex, and technologically advanced economies.
- It is therefore necessary to build and maintain a macro prudential framework that not only responds to crises but also proactively shapes market behavior, enhances transparency, protects smaller participants, and reduces systemic risk.

In conclusion, regulation should not be seen as a restriction on market freedom, but rather as its prerequisite in the 21st century. A truly effective and stable money market emerges only where market dynamics are supported by a responsible, transparent, and credible institutional environment—whether in Central Europe, Central Asia, or the global benchmark of the United States.

Resources

- A Red Ventures Company (n.d.). *91 Day T-bill treasury rate*. Bankrate. URL: <https://www.bankrate.com/rates/interest-rates/91-day-treasury-bill/>.
- Acharya, V. V., L. H. Pedersen, T. Philippon, and M. Richardson (2010). “Measuring systemic risk”. In: *Federal Reserve Bank of Cleveland Working Paper Series*. URL: <https://pages.stern.nyu.edu>.
- Bank for International Settlements (2022). *Money markets and central bank liquidity management*. BIS.

- Bar-Yam, Yaneer, Duncan Harmon, and al. (2022). *Networks of Economic Market Interdependence and Systemic Risk*. Updated version of systemic risk network research. URL: <https://arxiv.org> (visited on 01/10/2025).
- Cecchetti, S. G. and K. L. Schoenholtz (2021). *Money, banking, and financial markets*. 6th ed. McGraw-Hill Education.
- European Central Bank (2020). *The implementation of monetary policy in the euro area: General documentation*. Frankfurt am Main: European Central Bank. URL: <https://www.ecb.europa.eu>.
- European Corporate Governance Institute (2022). *Financial Innovation and Risk Governance*. URL: <https://www.ecgi.global> (visited on 01/10/2025).
- Fabozzi, F. J. and F. Modigliani (2009). *Capital markets: Institutions and instruments*. 4th ed. Pearson Education.
- Federal Reserve (2020). *Financial Stability Report*. URL: <https://www.federalreserve.gov>.
- Financial Regulation Courses (2022). *Financial Regulation and Protection of Retail Investors*. URL: <https://www.financialregulationcourses.com>.
- Gorton, G. and A. Metrick (2012). “Regulating the shadow banking system”. In: *Brookings Papers on Economic Activity* 2, pp. 261–312.
- Harmon, D., Y. Bar-Yam, et al. (2011). *Networks of Economic Market Interdependence and Systemic Risk*. arXiv preprint. URL: <https://arxiv.org>.
- Howells, P. and K. Bain (2008). *The economics of money, banking and finance: A European text*. 4th ed. Pearson Education.
- International Monetary Fund (2021). *The IMF and Financial Sector Regulation*. URL: <https://www.imf.org>.
- JSTOR (2021). *Selected articles on financial ethics and regulatory accountability*. URL: <https://www.jstor.org>.
- Mishkin, F. S. (2019). *The economics of money, banking, and financial markets*. 12th ed. Pearson.
- Pilbeam, K. (2018). *Finance and financial markets*. 4th ed. Palgrave Macmillan.
- ResearchGate (2021). *Selected research articles on transparency and information asymmetry in financial markets*. URL: <https://www.researchgate.net>.
- Schwarz, Steven L. (2008). “Systemic Risk”. In: *Georgetown Law Journal* 97.1, pp. 193–249.
- ScienceDirect (2022). *Selected articles on financial ethics and regulation*. URL: <https://www.sciencedirect.com>.
- Stiglitz, J. E. (2010). *Risk and global economic architecture: Why full financial integration may be undesirable*. Columbia Business School Discussion Papers.

The Virtual Accounting Assistant: A Humanized and Expanded Analysis

Árpád Lángyi¹

Abstract

In recent years, the adoption of remote and cloud-based services has enabled a new professional class: the virtual accounting assistant (VAA). While bookkeeping and financial support have long been core functions of business operations, the Virtual Accounting Assistant (VAA) leverages digital platforms, automation, and distributed workforce models to deliver accounting services with greater flexibility, cost-efficiency, and scalability. This paper explores the definition and roles of the virtual accounting assistant, outlines the primary drivers of adoption, analyses key benefits and challenges, and presents concluding reflections on future developments and regulatory considerations.

Keywords

virtual assistant; accounting; bookkeeping; remote work; cloud services; digital transformation; scalability; automation; Artificial Intelligence (AI); data security; compliance; regulation.

1 Introduction

As companies increasingly embrace digital transformation and remote collaboration, the role of traditional in-house accounting and bookkeeping is evolving. Historically, accounting work required on-site staff handling ledgers, invoices, reconciliations, and compliance. Today, the emergence of the virtual accounting assistant (VAA) model offers an alternative: remote professionals or, increasingly, Artificial Intelligence (AI)-enhanced digital assistants performing accounting tasks via cloud platforms, automation tools, and global networks.

Beyond convenience, this shift represents a significant optimisation of human resources. Many organisations face chronic shortages of skilled accountants, rising labour costs, and high turnover driven by repetitive, time-consuming administrative tasks. The Virtual Accounting Assistant model reduces dependency on large in-house teams by automating or outsourcing routine work, allowing existing employees to focus on analysis, strategy, and client relations. In practice, this translates into measurable savings in recruitment, onboarding, and long-term salary expenditure. A particularly important feature of modern virtual accounting assistants is their knowledge base, which may include continuously updated legal and regulatory frameworks. By integrating structured data sources such as national tax codes, financial reporting standards, and compliance directives VAAs can automatically adjust workflows to reflect the latest legislative changes. This capability significantly reduces the risk of human error and ensures that reporting and documentation remain compliant without requiring constant manual intervention. At the same time, the introduction of VAAs prompts a redefinition of staff development and training. Rather than replacing human accountants entirely, VAAs complement them by taking over routine processes, freeing up time for higher-value tasks. This shift necessitates continuous professional development: employees must acquire new skills in digital collaboration, data interpretation, automation management, and the ethical use of Artificial Intelligence (AI)-based accounting systems. Organisations that integrate VAAs successfully tend to invest in upskilling programs

¹Department of Applied Informatics, Faculty of Economic Informatics, Bratislava University of Economics and Business, Slovakia. Email: arpad.langyi@euba.sk

that blend financial expertise with technological literacy, ensuring both efficiency and long-term employability.

Thus, the rise of the virtual accounting assistant is not only a technological innovation but also an organisational transformation one that combines automation, legal intelligence, and human learning to redefine how financial operations are executed in the digital era.

2 Defining the Virtual Accounting Assistant

A virtual accounting assistant is a professional or system that performs accounting and bookkeeping tasks remotely through digital and cloud-based tools. It assists with invoice processing, bank reconciliation, tax preparation, document management, and financial reporting. However, it is not an autonomous system; rather, it operates under human supervision, supporting accountants and financial officers by combining automation with expert oversight. A key differentiator of the VAA model is its dynamic knowledge base, which is continuously updated (Ossisto, 2022) with the latest accounting standards, tax regulations, and legal requirements relevant to the jurisdiction in which the business operates. This ensures that all recommendations, calculations, and document templates are fully compliant and reflect the most recent legislative changes. The knowledge base functions as a living repository that evolves in real time as rules and practices change.

In addition to performing clerical or transactional work, modern VAAs can provide intelligent guidance to human accountants (Wise, 2025). They can suggest optimal accounting treatments, flag potential compliance risks, and recommend procedural improvements. By analysing financial data and regulations, the VAA can generate draft documents and templates, for example, financial statement proposals, expense reports, or tax declarations, ready for human review and finalisation. Another essential capability is quality assurance: the virtual assistant can verify the consistency and correctness of bookkeeping tasks performed by human accountants. By cross-checking entries, comparing them with regulatory thresholds, and identifying anomalies, the VAA acts as an additional control layer that enhances accuracy and reliability. This hybrid collaboration between human expertise and intelligent automation leads to a more efficient, error-resistant, and compliant accounting process. Thus, the virtual accounting assistant should be understood as an augmented intelligence system, one that empowers professionals rather than replaces them. It combines human judgment with computational precision, integrating an adaptive knowledge base, compliance awareness, and collaborative interaction to create a next-generation accounting environment.

3 Drivers of Adoption

The adoption of virtual accounting assistants is accelerating across the entire financial ecosystem (IBS for You, 2025), with a particularly pronounced impact on small and medium-sized enterprises (SMEs), accounting professionals, and regulatory institutions. This section analyses the main drivers of adoption and the broader systemic benefits observed in both the private and public sectors.

3.1 Structural Pressures in SMEs

SMEs typically operate with limited staff and capital, forcing owners to focus primarily on revenue generation, client acquisition, and operational continuity. Consequently, many lack the time or expertise to explore the full range of financial advantages and incentives available under national law. In practice, this leads to under-utilisation of tax benefits, deductions, or government grants that could improve profitability. Because of these constraints, SME owners often delegate financial and compliance tasks entirely to external accountants (Dong et al., 2024). While this

delegation is pragmatic, it creates dependency and reduces direct awareness of regulatory changes or optimisation opportunities.

3.2 Reliance on Generalist Accountants

In the SME segment, accountants frequently act as multi-client generalists, keeping pace with numerous industries, each governed by distinct fiscal regulations. The constant flow of amendments, circulars, and tax rulings makes it difficult for them to remain fully updated. As a result, even competent accountants may overlook new tax benefits, transitional reliefs, or procedural updates. The VAA addresses this gap by maintaining a continuously updated knowledge base that tracks legislative, fiscal, and compliance changes in real time. It alerts both business owners and accountants to new opportunities and required actions, ensuring that no legal or financial advantage is missed.

3.3 Regulatory Complexity and Volatility

National and international accounting frameworks evolve rapidly. Frequent amendments to tax codes, digital reporting mandates (such as e-invoicing and SAF-T), and environmental or innovation-related subsidies impose additional administrative burdens. SMEs and even mid-sized accounting firms often lack the infrastructure to systematically monitor these changes. By contrast, the VAA can parse official bulletins, integrate them into its knowledge base, and provide instant contextual interpretation, for example, how a new VAT exemption or depreciation rule applies to a specific business type. This capability transforms reactive compliance into proactive optimisation.

3.4 The Virtual Accounting Assistant Value Proposition for SMEs

For small and medium-sized enterprises, the VAA delivers measurable advantages through: Continuous knowledge updates: Real-time synchronisation with the latest accounting and taxation rules.

- Proactive discovery of fiscal opportunities: Identification of lawful optimisations (“loop-holes”) and untapped incentives.
- Guidance and decision support: Evidence-based recommendations with statutory references.
- Automated document preparation: Drafting templates and compliance checklists ready for accountant validation.
- Assurance and verification: Cross-checking bookkeeping records for accuracy and completeness before submission.

Through these capabilities, the VAA effectively acts as an intelligent compliance partner, improving both financial outcomes and administrative efficiency.

3.5 Human-in-the-Loop Oversight

Despite its advanced automation, the VAA is not an autonomous system. It functions within a human-in-the-loop governance model:

- Accountants review and approve all key recommendations and submissions.
- The system provides full traceability and justification for every suggestion.
- Ambiguous cases are escalated for human judgment.

This framework ensures professional accountability and regulatory compliance while benefiting from machine-level efficiency and precision.

3.6 Broader Benefits for Accountants and Auditors

The usefulness of VAAs extends well beyond SMEs. Professional accountants and auditors increasingly rely on them as quality-assurance and productivity tools.

- During internal or external audits, the VAA can automatically check ledger entries, identify missing documentation, and cross-reference reported data with statutory thresholds.
- It helps auditors detect inconsistencies early, improving both audit speed and reliability.
- For accounting firms managing multiple clients, the assistant standardises routine verification steps, freeing human experts to focus on complex analytical work or client advisory services.
- Its continuously updated knowledge base ensures that audit tests always align with the latest standards (IFRS, GAAP, or national equivalents).

Thus, the VAA enhances audit quality and transparency, reduces manual sampling, and improves the traceability of financial controls.

3.7 Support for Tax Administrations and Regulatory Bodies

The benefits of VAA are equally relevant on the regulatory and governmental side. Tax authorities increasingly handle massive volumes of digital submissions, often with incomplete or inconsistent data. The VAA facilitates this interaction in several ways:

- Error reduction and pre-validation: Before filings reach the tax office, the assistant checks syntax, completeness, and logical consistency, significantly reducing rejection rates.
- Improved data standardisation: By adhering to digital reporting schemas, VAAs produce structured data that can be processed more easily by official systems.
- Transparency and audit readiness: VAAs maintain traceable logs of all computations and supporting documents, simplifying subsequent inspections or audits.
- Collaboration with authorities: In the long term, integration between VAA platforms and tax-office APIs could enable near-real-time compliance reporting, reducing administrative backlogs for both sides.

In this sense, VAA does not merely serve businesses; it contributes to systemic improvement of the fiscal ecosystem, promoting trust, efficiency, and mutual visibility between taxpayers and regulators.

3.8 Outcomes and Strategic Implications

The measurable effects of VAA adoption include:

- For SMEs: higher utilisation of legal incentives, lower compliance costs, improved cash flow.
- For accountants: reduced manual workload, fewer errors, enhanced advisory capacity.
- For auditors: better data quality and automated test coverage.
- For tax authorities: cleaner submissions, faster reviews, and improved risk detection.

Collectively, these outcomes illustrate that VAAs are not a narrow technological upgrade but a transformative coordination layer between business, accounting professionals, and the state. Their integration signals the emergence of a digitally augmented financial ecosystem in which compliance, optimisation, and transparency coexist more efficiently than ever before.

4 Benefits and Key Advantages

The implementation of virtual accounting assistants (VAs) brings multidimensional benefits (Belay Solutions, 2024) that extend far beyond cost savings. While the initial motivation for adoption often lies in efficiency and budget optimisation, long-term value emerges through improved accuracy, assurance of compliance, knowledge dissemination, and systemic transparency. The benefits can be grouped into five interrelated domains: financial efficiency, operational performance, regulatory reliability, professional empowerment, and institutional transparency.

4.1 Financial Efficiency and Cost Optimisation

The most immediate advantage of VAA lies in the reduction of direct and indirect costs. Businesses save on recruitment, training, and long-term employment overheads while gaining flexible access to high-level expertise. For SMEs, this means the ability to engage professional-grade accounting support without maintaining full-time staff. Instead of paying for idle capacity, they purchase specific outputs such as reconciliations, filings, or audit preparations on demand. Moreover, the Virtual Accounting Assistant's continuous monitoring of tax legislation and fiscal incentives ensures that companies capture all available benefits. By detecting new credits, deductions, or subsidies, it can materially improve cash flow and net profit margins. In certain jurisdictions, automated eligibility checks and timely alerts have led to measurable savings through early identification of refundable tax positions or investment allowances.

4.2 Operational Accuracy and Reliability

Automation eliminates repetitive manual input, reducing the incidence of human error in bookkeeping, reconciliation, and report generation. The VAA validates entries in real time, cross-checks balances, and ensures logical consistency between sub-ledgers and statements. When combined with human oversight, this hybrid approach produces superior data integrity. Accountants gain confidence that the base data are accurate before they prepare analytical reports or filings. This reliability extends to audit contexts: the VAA acts as a pre-audit verifier, flagging anomalies, missing evidence, or inconsistencies between declared and supporting records. The result is a smoother audit process, fewer queries, and higher confidence in the overall financial control environment.

4.3 Regulatory Compliance and Legal Awareness

A defining strength of VAA is its up-to-date legal intelligence. Their continuously refreshed knowledge base integrates statutory texts, administrative interpretations, and official guidelines. Consequently, every recommendation or document template generated by the assistant reflects the latest regulatory framework. This capacity reduces compliance risks significantly. Businesses avoid penalties and reputational damage caused by outdated procedures, while accountants spend less time manually checking for rule changes. The system can automatically adapt reporting templates and tax forms when authorities revise formats or introduce new electronic filing schemas. Furthermore, the Virtual Accounting Assistant can simulate compliance outcomes, for instance, estimating the tax impact of alternative accounting treatments or identifying thresholds that trigger new obligations (such as VAT registration, transfer-pricing documentation, or ESG disclosures). This forward-looking view allows organisations to act pre-emptively rather than reactively.

4.4 Professional Empowerment of Accountants and Auditors

Contrary to concerns that automation will replace human expertise, the VAA serves as an augmenting technology. It performs tedious, low-value tasks, such as data entry, categorisation,

and verification, thereby freeing accountants to focus on interpretation, strategy, and advisory work. This shift enhances job satisfaction and encourages continuous professional development in digital tools, data analytics, and compliance strategy. For auditors, the VAA provides a structured, traceable evidence trail and automates many of the routine sampling and cross-checking procedures. Audit teams can thus concentrate on judgmental assessments, risk analysis, and client consultation. Firms that integrate VAAs into their workflows report improved audit quality, reduced turnaround times, and higher consistency across client engagements.

4.5 Knowledge Sharing and Continuous Learning

Because the VAA encapsulates an evolving knowledge base, it becomes a learning partner for both junior and senior accounting professionals. Each interaction contributes to organisational learning: new interpretations, case resolutions, and procedural updates are fed back into the system. Over time, this collective intelligence mitigates knowledge loss from staff turnover and ensures institutional memory persists across projects and clients. Training departments can also leverage the assistant to deliver interactive learning modules that reinforce awareness of recent legal updates or new accounting standards. In this way, the VAA supports a culture of lifelong learning and professional excellence.

4.6 Institutional Transparency and Collaboration with Authorities

One of the less-discussed but highly significant advantages of VAA is its contribution to public-sector efficiency and transparency. By ensuring that submitted financial data is standardised, validated, and traceable, the VAA reduces the burden on tax administrations. Fewer erroneous filings translate into fewer manual reviews, shorter response times, and lower administrative costs. Tax authorities benefit from cleaner, machine-readable data, improving analytics for risk assessment and policy design. In the long term, bilateral API integration between VAA systems and official portals could enable real-time compliance reporting, mutual verification of tax positions, and proactive error correction before final submission. Such capabilities foster a more collaborative and less adversarial relationship between taxpayers and regulators.

4.7 Ethical and Risk-Reduction Benefits

Because the VAA records every recommendation, change, and user confirmation, it naturally generates an audit trail that strengthens internal control. This traceability discourages malpractice, facilitates forensic review, and ensures accountability for all accounting actions. Additionally, by suggesting only legally compliant optimisation paths, the assistant acts as a built-in ethical compass, preventing users from engaging in high-risk or aggressive tax behaviour.

In this sense, VAAs promote a balanced form of digital governance combining efficiency with integrity.

4.8 Strategic and Systemic Outcomes

Taken together, these benefits reshape the financial ecosystem as a whole:

- Businesses gain resilience, agility, and better use of fiscal opportunities.
- Accountants and auditors gain higher accuracy and more time for analytical and advisory tasks.
- Tax administrations gain cleaner, more reliable data streams, enabling targeted oversight.
- Society gains from increased transparency, fairness, and reduced compliance friction.

Hence, the VAA should be understood not merely as a cost-saving tool but as a strategic enabler of modern financial governance, a bridge between human expertise, technological precision, and institutional trust.

5 Challenges, Risks, and Considerations

While the virtual accounting assistant offers substantial operational and strategic benefits, its deployment introduces new technical, ethical, and regulatory challenges. As the system becomes more sophisticated, capable of learning from interactions, managing multiple clients, and interfacing with sensitive financial data, questions of data governance, privacy, knowledge management, and liability become increasingly critical.

5.1 Complexity of Multi-Tenant Architectures

In most professional settings, accounting firms and outsourced finance departments serve multiple clients simultaneously. Consequently, any VAA designed for such environments must operate as a multi-tenant system, a shared digital infrastructure where multiple independent business entities coexist on the same platform.

While multi-tenancy increases efficiency and scalability, it also raises the risk of data leakage or cross-contamination between clients. Financial data, invoices, or user queries may contain commercially sensitive information or trade secrets. A robust multi-tenant design, therefore, requires:

- Strong logical separation of data at both the database and application levels.
- Per-tenant encryption keys ensure that even administrators cannot trivially access cross-tenant information.
- Role-based access control (RBAC) and attribute-based policies that define which users and systems can view, modify, or train on each data segment.
- Comprehensive audit logging to detect unauthorised access or data exfiltration attempts.

Without such controls, even inadvertent data exposure between clients could lead to reputational damage, regulatory fines, or civil liability.

5.2 Handling User Queries and Learning from Interactions

A distinctive feature of advanced VAAs is their ability to engage in conversational interactions with accountants, business owners, and auditors. These queries often reveal the context of internal decision-making or the financial health of an enterprise. Handling such queries safely presents several challenges:

- Context management: Each question must be processed within the appropriate business context. The system must guarantee that an accountant asking about Company A never receives hints or completions derived from Company B's data.
- Learning without leakage: The assistant may improve through exposure to user interactions. However, training on unfiltered user input risks embedding confidential information into global model weights or shared knowledge bases. To mitigate this, learning should occur through isolated, anonymised fine-tuning or federated learning techniques where insights are aggregated statistically without sharing raw data.
- Human-in-the-loop supervision: Every adaptive behaviour or new inference should be traceable and reversible. A human reviewer should validate whether new patterns discovered by the model are both legally compliant and contextually correct.

Thus, learning must be structured, explainable, and compartmentalised, ensuring improvement without compromising confidentiality or compliance.

5.3 Data Isolation and Professional Boundaries

In many accounting practices, individual accountants or small firms manage the books of multiple enterprises concurrently, sometimes across competing industries. In this scenario, the VAA must ensure strict data and communication isolation between entities.

This requires:

- Tenant-level sandboxing: Each client’s documents, messages, and knowledge graphs are confined to separate logical containers.
- No cross-tenant embeddings: Vector databases and LLM memory stores must be partitioned so that semantic similarity search cannot retrieve embeddings from another entity.
- Confidential context tagging: Every prompt, answer, and generated document must carry metadata identifying its owner, retention period, and classification level (e.g., “confidential financial data,” “public knowledge,” etc.).
- Dual-control approval: For firms serving multiple clients, cross-client analysis (e.g., benchmarking) should require explicit opt-in consent from all parties.

This technical and procedural isolation protects trade secrets and preserves client trust, forming the cornerstone of ethical and lawful VAA deployment.

5.4 General Data Protection Regulation (GDPR) and Data-Retention Dilemmas

Under the GDPR and analogous privacy laws, accounting data often qualifies as personal data, especially when containing names, salaries, or identifiable transaction details. The VAA must therefore comply with principles of purpose limitation, data minimisation, and storage limitation.

This raises a fundamental design question: Should the VAA retain data at all?

Three main approaches exist:

- Ephemeral processing: The assistant handles requests in memory and discards all data immediately after response generation. This maximises privacy but sacrifices long-term learning and historical context.
- Selective retention with consent: Only explicitly authorised data (e.g., templates, FAQs, or anonymised patterns) is stored for improvement or reuse. Users can review and delete stored content at any time.
- Aggregated learning: The system extracts high-level insights or rules (e.g., “tax deduction threshold patterns”) without preserving identifiable data.

A hybrid model is often optimal: minimal retention for operational continuity, combined with anonymised meta-learning to enhance global performance. Regardless of the approach, the system must support right-to-erasure requests and provide transparent documentation of the data lifecycle.

5.5 Legal and Ethical Accountability

When a VAA provides recommendations, drafts, or alerts, the question arises: who bears responsibility for errors? In a multi-tenant environment, determining accountability among developers, operators, and professional users is complex.

Therefore:

- The VAA should be positioned as a decision-support tool, not as an autonomous decision-maker.
- All advice must be traceable to identifiable human validation (the accountant or supervisor remains legally responsible).
- Version-controlled logs of knowledge-base updates and Artificial Intelligence (AI)-generated suggestions should be preserved for forensic and regulatory audit.

Furthermore, ethical frameworks should prohibit VAA from exploiting “loopholes” that violate the spirit, if not the letter, of the law, ensuring that optimisation remains compliant and socially responsible.

5.6 Balancing Learning and Privacy: A Design Paradox

Perhaps the most difficult challenge is balancing continuous improvement with data protection. Learning from real cases enhances accuracy and domain understanding; yet storing or reusing that data risks breaching confidentiality.

The design of future VAAs must therefore embrace privacy-preserving learning architectures, for example:

- Federated learning: where updates are trained locally on client systems and only aggregated model parameters are shared.
- Differential privacy: introducing controlled noise into data representations to prevent re-identification.
- Synthetic data generation: creating training examples statistically similar to real cases but devoid of personal details.

This balance between intelligence and isolation will define the credibility and legal sustainability of next-generation accounting assistants.

5.7 Organisational Change and Trust

Finally, technology alone cannot mitigate all risks. Successful VAA deployment depends on cultural acceptance and clear governance. Accounting firms and SMEs alike must:

- Establish explicit policies on data ownership and usage.
- Educate staff about responsible Artificial Intelligence interaction (what to ask, what not to share).
- Regularly review access controls and incident-response protocols.

Trust in both internal and client-facing is the decisive factor. Without it, even the most secure architecture will face resistance from professionals who fear loss of control or exposure of confidential information.

5.8 Summary

The challenges facing VAAs are not solely technical; they are ethical, procedural, and organisational. Multi-tenancy, learning governance, data isolation, and privacy regulation converge into a complex design space that demands transparency, modularity, and restraint.

Only through strict data segmentation, accountable learning, and adherence to privacy-by-design principles can the virtual accounting assistant evolve into a trusted pillar of digital finance infrastructure.

6 Outlook and Future Trends

The evolution of virtual accounting assistants (VAA) is entering a new phase, one defined not only by automation and efficiency, but also by responsible intelligence, privacy-preserving architectures, and governance-driven design. As adoption expands across SMEs, accounting firms, and public institutions, several technological and organisational trends are shaping the next decade of development.

6.1 Privacy-Preserving and Federated Learning Architectures

The next generation of VAAs will move beyond centralised data aggregation toward distributed and federated learning frameworks.

In this model, each accounting firm or business entity retains control of its data locally; only model parameters or statistical gradients are shared with a central orchestrator. This ensures that confidential financial records never leave the client's infrastructure, while still contributing to the system's overall intelligence.

Federated architectures will be combined with techniques such as:

- Differential privacy, introducing mathematical noise to guarantee that no individual record can be reverse-engineered from learned weights.
- Secure multi-party computation (SMPC), allowing several VAAs or organisations to collaborate on complex analyses without disclosing raw inputs.
- Homomorphic encryption, enabling computations directly on encrypted data, so that sensitive figures remain obscured even during processing.

Together, these methods define a privacy-by-design paradigm, positioning the VAA as both a technological and ethical innovator in financial data handling.

6.2 Artificial Intelligence Governance and Regulatory Alignment

As VAAs become decision-support systems embedded in everyday financial workflows, the need for explicit Artificial Intelligence governance frameworks becomes urgent. Regulators in Europe and elsewhere increasingly expect accountability, explainability, and fairness in algorithmic decision-making. The EU's Artificial Intelligence Act and ongoing amendments to the GDPR will directly affect how accounting automation can operate.

To remain compliant, VAA providers and users must:

- Implement model documentation ("model cards") describing data sources, assumptions, and limitations.
- Maintain human-in-the-loop validation for any tax or reporting recommendation with legal consequences.
- Conduct periodic bias and reliability audits of algorithmic outputs.
- Establish data-protection impact assessments (DPIAs) before large-scale deployments.

Governance frameworks will thus evolve in parallel with the technology itself, ensuring that efficiency gains never compromise legal accountability or professional ethics.

6.3 Multi-Tenant Compliance Ecosystems

Future accounting environments will be deeply interconnected. Accounting firms, auditors, clients, and tax authorities will operate within federated, multi-tenant ecosystems that exchange data through secure APIs and blockchain-verified transaction logs.

In such a setting, the VAA will serve as a compliance orchestrator rather than an isolated tool, coordinating document flows, validation steps, and real-time submissions among multiple actors.

This transformation will create several new possibilities:

- End-to-end traceability: Each accounting event from invoice generation to tax declaration will have an immutable digital fingerprint.
- Instant verification: Tax authorities can verify submissions automatically, reducing delays and administrative costs.
- Cross-firm benchmarking: Aggregated, anonymised data can inform macroeconomic insights without breaching individual confidentiality.

These architectures require new interoperability standards, but they promise a harmonised compliance ecosystem in which transparency, security, and efficiency coexist.

6.4 Cognitive Augmentation and Continuous Learning

Future VAAs will integrate neurosymbolic reasoning and contextual learning to interpret complex accounting scenarios such as multi-jurisdictional tax planning, consolidated reporting, or ESG-linked financial disclosures.

However, learning will increasingly occur under explicit human supervision and traceable governance protocols.

Expected developments include:

- Contextual assistants capable of understanding natural-language policy questions and returning not just answers but evidence-linked reasoning chains.
- Adaptive user interfaces that adjust to each accountant's expertise level, offering progressive disclosure of technical detail.
- Feedback-driven refinement loops, where each approved or rejected suggestion informs future system accuracy while preserving tenant isolation.

In this hybrid model, human accountants remain the ultimate decision-makers, while the assistant evolves into an intelligent collaborator that enhances cognition rather than replaces it.

6.5 Sustainable Data and Energy Practices

As LLM-based systems expand, their computational and environmental footprints cannot be ignored. Accounting ethics, traditionally grounded in stewardship and responsibility, extend naturally to sustainable Artificial Intelligence (AI) operations.

Future VAAs are expected to:

- Employ energy-efficient inference architectures and model-compression techniques.
- Utilise renewable-powered cloud infrastructure certified for green computing.
- Integrate carbon-tracking modules that measure the environmental impact of processing and suggest offsets.

Such measures align digital transformation with corporate sustainability goals, reinforcing the accountant's emerging role as a guardian not only of financial integrity but also of environmental accountability.

6.6 The Human Dimension: Redefining Professional Identity

The proliferation of VAA is likely to reshape the professional identity of accountants and auditors. Routine transactional tasks will be almost entirely automated, shifting the emphasis toward strategic advisory, risk analysis, and Artificial Intelligence oversight.

Professional education must therefore evolve:

- Curricula will integrate data ethics, Artificial Intelligence literacy, and algorithmic governance alongside classical accounting theory.
- Certification bodies will develop Artificial Intelligence competency standards for accountants operating in hybrid human-machine environments.
- Continuous learning platforms often embedded within the VAA itself will help practitioners stay aligned with rapidly changing laws and technologies.

In this sense, the VAA is not a substitute for professional expertise but a catalyst for its renewal.

6.7 Toward Trusted Digital Finance Infrastructure

Ultimately, VAAs may become foundational components of a trusted digital finance infrastructure, a globally networked system where businesses, auditors, and regulators interact seamlessly through transparent, privacy-preserving protocols.

Such a system could enable:

- Real-time audit trails automatically available to authorised parties.
- Predictive compliance analytics that pre-empt risks before they materialise.
- Cross-border interoperability of accounting standards via intelligent translation layers.

If governed responsibly, this transformation could lead to an era of radical transparency, where trust is encoded not through manual supervision but through secure, verifiable digital collaboration.

6.8 Outlook Summary

The trajectory of virtual accounting assistants is clear: from tool to collaborator, from automation to augmentation, from isolated systems to ecosystem infrastructure. The defining trends privacy-preserving learning, federated compliance, Artificial Intelligence governance, sustainability, and human-machine partnership will determine whether VAA remain niche utilities or evolve into cornerstones of the digital economy. The future of accounting, therefore, is not post-human but co-intelligent: a convergence of human expertise, ethical Artificial Intelligence, and secure multi-tenant architectures designed to uphold the same principles that have always defined the profession trust, transparency, and accountability.

7 Conclusion

The emergence of the VAA marks a fundamental transformation in how financial operations are conceived, executed, and governed. Originally envisioned as a digital support tool for bookkeeping, the VAA has evolved into a multi-layered, knowledge-driven ecosystem component that reshapes relationships between businesses, accountants, auditors, and regulatory authorities.

At its core, the VAA embodies three interdependent dimensions:

- Automation and efficiency – freeing human professionals from repetitive, low-value tasks, thereby enabling them to concentrate on analytical, advisory, and strategic work.

- Legal and regulatory intelligence – maintaining an always-updated knowledge base of tax laws, accounting standards, and compliance requirements, ensuring that every recommendation or document aligns with the most recent legal framework.
- Governance and trust – embedding privacy, transparency, and explainability into its architecture through isolation, encryption, and accountable human oversight.

Through these capabilities, the VAA delivers measurable advantages: cost reduction, accuracy, and time efficiency for enterprises; improved audit quality and knowledge retention for accounting professionals; and cleaner, standardised submissions for tax administrations. It thus acts not merely as a cost-saving mechanism but as a strategic enabler of digital transformation within the financial ecosystem. However, the same technological sophistication introduces new challenges. Multi-tenant deployment environments raise questions of data segregation, intellectual property, and confidentiality. Continuous learning from user interactions must be balanced with strict privacy regulations, such as the General Data Protection Regulation (GDPR), which demand privacy-preserving methods like federated learning and differential privacy.

The future success of VAA depends on resolving this duality: how to learn without leaking, how to automate without losing control, and how to scale without compromising ethics. Looking forward, VAAs are likely to become embedded within federated, trust-based digital finance infrastructures, where humans and intelligent systems cooperate under shared regulatory and ethical frameworks. Accountants and auditors will increasingly act as Artificial Intelligence supervisors and data-governance stewards, guiding machines that themselves act as collaborators rather than tools.

Education, certification, and policy must adapt accordingly, emphasising Artificial Intelligence literacy, governance, and continuous learning alongside classical financial theory.

Ultimately, the virtual accounting assistant is not a step toward replacing human judgment but toward augmenting it with computational precision and real-time legal awareness.

If designed responsibly, anchored in transparency, accountability, and sustainability, VAA can become a cornerstone of a re-humanised digital economy: one where technology safeguards integrity, amplifies expertise, and redefines the very notion of trust in accounting.

Resources

Belay Solutions (2024). *The Pros and Cons of Using AI in Accounting: A Comprehensive Analysis*.

URL: <https://www.belaysolutions.com/blog/ai-in-accounting-pros-cons/> (visited on 02/13/2026).

Dong, Bo et al. (2024). *Challenges and Opportunities of Large Language Models in Accounting: An Exploratory Study*. arXiv preprint. arXiv:2410.17365. URL: <https://arxiv.org/abs/2410.17365> (visited on 02/13/2026).

IBS for You (2025). *AI in Accounting: Benefits, Use Cases and Future*. URL: <https://www.ibsforyou.com/blog/ai-in-accounting/> (visited on 02/13/2026).

Ossisto (2022). *The Role of AI in Modern Accounting*. URL: <https://ossisto.com/blog/the-role-of-ai-in-modern-accounting/> (visited on 02/13/2026).

Wise (2025). *How AI in Accounting Is Transforming the Finance Industry*. URL: <https://wise.com/us/blog/ai-in-accounting> (visited on 02/13/2026).

Enhancing the reliability of LoRa-based sensor networks in groundwater monitoring

Gulrukh Memonova¹ Peter Schmidt²

Abstract

Groundwater monitoring is a critical task for sustainable water management, particularly in agricultural regions where water scarcity and inefficient usage threaten productivity. However, ensuring reliable data transmission from distributed sensors to a central unit remains a major challenge due to interference, signal fading, and the limited duty cycle of Low Power Wide Area Networks (LPWAN). This study addresses these challenges by proposing a LoRa-based sensor network topology designed for groundwater monitoring in the Karshi district, Uzbekistan. The network connects eight groundwater wells to a central station, and communication reliability is improved by incorporating relay-assisted transmission schemes. The proposed approach is evaluated through network reliability modeling and analysis of data delivery success rates. Results demonstrate that cooperative relaying reduces packet loss and enhances the probability of successful data transmission compared to direct communication only. The findings show that relay-assisted LoRa topologies can provide a cost-effective, scalable solution for reliable groundwater monitoring.

Keywords

LoRa, sensor networks, groundwater monitoring, communication reliability, cooperative relaying

1 Introduction

Groundwater resources play a vital role in supporting agricultural productivity, drinking water supply, and ecological stability. The increasing pressure caused by climate change, seasonal droughts, and excessive groundwater extraction makes continuous monitoring of aquifer conditions essential. Traditional monitoring methods—such as manual water-level measurements, periodic field inspections, or delayed statistical reports—often fail to provide timely and sufficiently accurate information. As a result, real-time, automated groundwater monitoring systems are becoming increasingly important, especially in regions where wells are geographically dispersed and water scarcity directly affects economic productivity. Recent advancements in IoT and LPWAN technologies, particularly LoRa and LoRaWAN, have enabled the development of low-power, long-range, and cost-effective sensor networks for environmental monitoring. LoRa-based systems have been successfully applied in water distribution monitoring, environmental sensing, and surface-water quality assessment, demonstrating stable performance in large outdoor environments (Sendra, Lloret, et al., 2023). In another study, a hybrid LoRa–GSM system powered by energy harvesting techniques was implemented for real-time groundwater level monitoring, highlighting the practicality of low-cost IoT platforms for hydrogeological applications (Makange et al., 2023). Additionally, LoRa networks have been used for river and coastal water-quality monitoring, showing reliable data transmission under varying field conditions (Sendra, Parra, et al., 2022). Despite these advances, the majority of existing LoRa deployments rely on a

¹Karshi state technical university, Karshi, Uzbekistan, Department of Optical communication systems and networks, e-mail: memonovagulrux@gmail.com

²Department of Applied Informatics, Faculty of Economic Informatics, Bratislava University of Economics and Business, Slovakia



Figure 1. A typical cooperative relaying model used in LoRa networks

star topology, where all sensor nodes communicate directly with a single gateway. Although this architecture is simple and energy-efficient, it suffers from significant limitations when used in large-scale or interference-prone environments. Studies have reported reduced reliability due to interference, duty-cycle constraints, signal fading, and packet loss, particularly when nodes are placed far from the gateway (Al-Sarawi et al., 2020). In groundwater monitoring scenarios, wells are often widely scattered, making direct communication highly vulnerable to signal degradation. To address these limitations, recent research has explored relay-assisted and multi-hop communication in LoRa sensor networks. Cooperative relaying has been shown to extend communication range, reduce packet loss, and improve signal robustness in complex environments (Borkotoky et al., 2019). Multi-hop schemes have also demonstrated their potential for enhancing network performance in distributed environmental monitoring systems (Borkotoky, 2022). However, despite these promising findings, the use of relay-assisted LoRa topologies in groundwater monitoring remains understudied. Most existing works focus either on system prototyping or small-scale indoor experiments, lacking comprehensive analysis under realistic field conditions with multiple wells. In this context, improving the reliability of LoRa-based networks through cooperative relaying represents a significant and relevant research direction. This study therefore examines the potential of relay-assisted communication to enhance network performance in groundwater monitoring applications. By evaluating transmission success rates, packet loss behavior, and network robustness, the study highlights the practical value of relay nodes for monitoring wells distributed across large rural areas. The novelty of this work lies in applying relay-assisted LoRa topologies to a real groundwater-monitoring scenario and analytically assessing their impact on overall network reliability. The following figure illustrates a typical cooperative relaying model used in LoRa networks.

2 Methodology

2.1 Network Design for Groundwater Monitoring in the Karshi District

The proposed monitoring network is deployed in the Karshi district, a semi-arid agricultural region where groundwater is extensively used for irrigation. Eight observation wells were selected based on their hydrogeological relevance and spatial separation. The wells are distributed over an area of approximately 4–6 km in diameter, with inter-well distances ranging from 350 m to 1.8 km. This dispersion, combined with low vegetation and occasional man-made obstacles, creates a realistic scenario for assessing the reliability of long-range wireless links. Figure 2 presents the

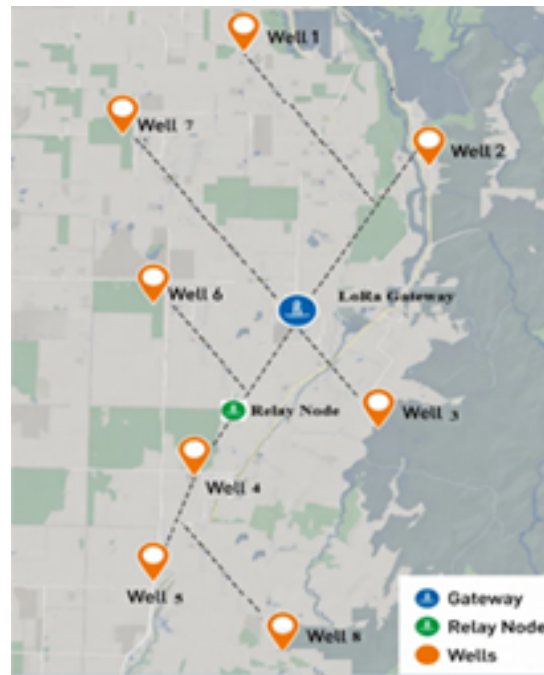


Figure 2. Layout of the wells with the positions of the LoRa gateway and the selected relay node.

spatial layout of the wells, the position of the LoRa gateway and the chosen relay node. The gateway is installed near an existing hydrological monitoring station equipped with mains power and Internet backhaul. The relay node is positioned at an intermediate location that maintains line-of-sight (or near line-of-sight) to both the gateway and the farthest wells. This placement follows the general guidelines for relay positioning in LoRa networks proposed by (Borkotoky et al., 2019), which require the relay to minimize the maximum hop distance while respecting practical deployment constraints.

Each well is instrumented with an autonomous monitoring node. The design of these nodes is inspired by low-power IoT platforms successfully deployed in water-quality and groundwater-level monitoring applications (Sendra, Lloret, et al., 2023), (Makange et al., 2023), (Sendra, Parra, et al., 2022). The node comprises a submersible pressure transducer for water-level measurement, a temperature sensor for basic hydro-thermal characterization, a low-power microcontroller (ESP32 class), a LoRa transceiver (SX1276 family), and a lithium-ion battery supported by a 5–10 W solar panel. The microcontroller periodically samples the sensors, aggregates the readings, and transmits data packets via LoRa to either the gateway or the relay node. A deep-sleep mode is used between transmissions to reduce energy consumption. The relay node shares the same hardware platform as the sensor nodes but is configured to operate as a dedicated forwarding device. It continuously listens for incoming packets from the surrounding wells and retransmits them towards the gateway using more robust physical-layer parameters when needed, such as higher spreading factor or longer preamble. To avoid excessive energy usage, the relay also employs duty-cycling and adaptive listening windows. The gateway acts as a LoRaWAN-like concentrator, equipped with an SX1301-based multi-channel transceiver, a 6 dBi antenna mounted at a height of 12 m, and a single-board computer responsible for packet reception and forwarding to a cloud server via cellular connection. The server stores the data in a time-series database and exposes them through a web dashboard for visualization and further analysis. The LoRa configuration adopted in this study builds on best practices from previous water-monitoring deployments, and on performance evaluations of LoRa under different parameter settings (Al-Sarawi et al., 2020). Table 1 summarizes the main communication parameters. In the baseline star topology, all sensor nodes attempt direct communication with the gateway using spreading factor SF9, bandwidth 125 kHz, coding rate 4/5, and transmit power 14 dBm

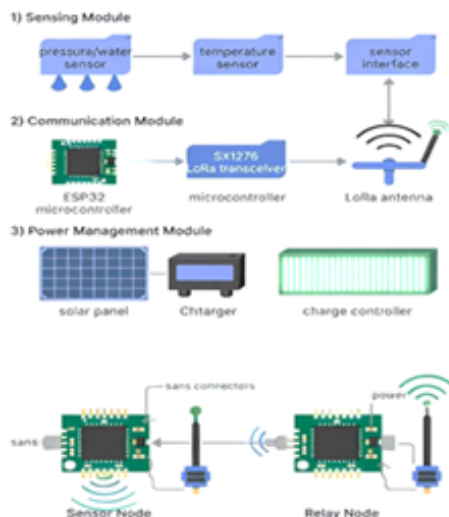


Figure 3. Hardware architecture of the sensor and relay nodes, showing the modular design with separated sensing, communication, and power-management components

in the European 868 MHz band. In the relay-assisted topology, distant nodes are allowed to use SF7 to the relay node, which is located closer and thus requires less link budget, while the relay forwards their packets to the gateway using SF10. This asymmetry reflects the cooperative relaying concept investigated analytically in (Borkotoky et al., 2019), (Borkotoky, 2022), where link adaptation on each hop can significantly improve reliability for edge nodes. Measurements are transmitted every 15 minutes, resulting in a duty cycle well below the 1 % regulatory limit for the 868 MHz band (Al-Sarawi et al., 2020). Each packet carries a timestamp, well identifier, water level, temperature, and diagnostic information such as battery voltage and RSSI of the last hop. At the application layer, the payload is encoded using a compact binary format to minimize airtime and reduce collision probability.

Table 1. LoRa configuration by role in the cooperative two-hop topology

Role	Spreading Factor (SF)	Bandwidth (BW)	Coding Rate (CR)	Tx Power (dBm)	Payload Size (bytes)	Tx Interval
Sensor Node (near)	SF7	125 kHz	4/5	14 dBm	18–24 bytes	Every 15 min
Sensor Node (far)	SF9	125 kHz	4/5	14 dBm	18–24 bytes	Every 15 min
Relay Gateway hop	SF10	125 kHz	4/5	17 dBm	22–30 bytes	Adaptive (per received packet)
Sensor Relay hop	SF7	125 kHz	4/5	14 dBm	18–24 bytes	Every 15 min
Gateway	Multi-SF (7–12)	125/250 kHz	—	—	Up to 255 bytes	Continuous reception

Figure 3 illustrates the hardware architecture of both the sensor and relay nodes, highlighting the modular design that separates sensing, communication, and power management. This modularity facilitates future upgrades, such as integrating additional water-quality sensors without modifying the communication architecture. The described network design thus provides a realistic and flexible platform for evaluating the impact of cooperative relaying on communication reliability in groundwater monitoring applications.

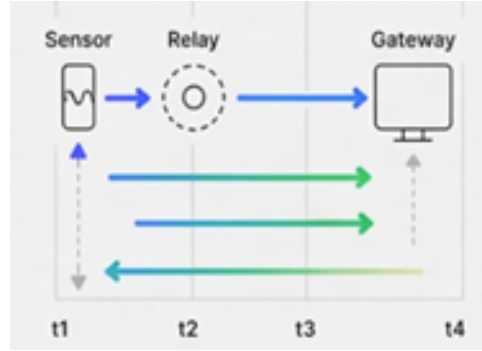


Figure 4. Timing diagram illustrating the cooperative two-hop communication process for a sensor-relay-gateway triplet

2.2 Cooperative Relaying Scheme and Reliability Modelling

The core contribution of this study is the integration of a cooperative relaying scheme into the LoRa-based groundwater monitoring network. While previous work on LoRa environmental monitoring has primarily relied on a star topology (Sendra, Lloret, et al., 2023), and groundwater systems have focused on hybrid backhaul solutions such as LoRa-GSM (Makange et al., 2023), (Memonova and Tursunov, 2022) here we specifically investigate how a relay node can enhance end-to-end reliability for remote wells. Our approach builds on the theoretical framework for cooperative relaying in LoRa networks introduced in (Borkotoky et al., 2019) and extended with coded relaying concepts in (Borkotoky, 2022), (Memonova, Schmidt, et al., 2025). The adopted relaying protocol follows a decode-and-forward strategy. In the uplink direction, each sensor node first attempts direct communication with the gateway using its assigned physical-layer parameters. If the direct transmission is successful, the packet is accepted by the gateway and the communication round for that sensor is complete. If the gateway does not acknowledge the packet within a predefined timeout, the sensor retransmits the same payload but now targeting the relay node as its next hop. Upon correctly decoding the packet, the relay re-encodes and forwards it to the gateway using a more robust configuration. The relay therefore acts as a cooperative helper that increases the probability of successful delivery rather than as an exclusive routing point. Figure 4 presents the timing diagram of a typical cooperative transmission for a single sensor-relay-gateway triplet. The communication frame is divided into time slots assigned to each sensor to avoid intra-network collisions, as also recommended in (Borkotoky et al., 2019). During its slot, a sensor first transmits directly to the gateway. If no acknowledgment is received, it retransmits via the relay in a second sub-slot. This simple schedule respects the LoRa duty-cycle limitations and minimizes additional traffic load introduced by the relay (Al-Sarawi et al., 2020). The relay itself operates in continuous receive mode during the sensors' slots and transmits only in its dedicated forwarding windows, keeping its duty cycle within regulatory bounds.

To quantify the reliability improvement, we develop an analytical model for the packet success probability. Let P_{DG} denote the probability that a packet from a given sensor is successfully received by the gateway using direct communication, and let P_{SR} and P_{RG} be the success probabilities for the sensor-relay and relay-gateway links, respectively. Assuming independent fading on different hops, the end-to-end success probability under the cooperative scheme is given by

$$P_{\text{coop}} = P_{DG} + (1 - P_{DG}) P_{SR} P_{RG}. \quad (1)$$

This expression reflects the fact that a packet may arrive either directly or via the relay. When coding or redundancy is added at the relay, the term $P_{SR}P_{RG}$ can be generalized to account for multiple coded transmissions as in (Borkotoky, 2022); however, in this study we focus on the

basic one-shot forwarding scheme described above.

To compute the individual link success probabilities, we adopt a path-loss model and a signal-to-interference-plus-noise ratio (SINR) threshold approach. The large-scale path loss $PL(d)$ as a function of distance d is modeled using the rural Okumura–Hata formulation, which has been used in several LoRa performance studies (**AlSarawi2020**; **Borkotoky2019**):

$$PL(d) = A + B \log_{10}(d). \quad (2)$$

where coefficients A and B depend on carrier frequency, base-station antenna height and environment type. The received power at distance d is

$$P_{RX}(d) = P_{TX} - PL(d) + G_{TX} + G_{RX}, \quad (3)$$

with P_{TX} the transmit power and G_{TX} , G_{RX} the antenna gains. A packet is considered successfully decoded if $P_{RX}(d)$ exceeds the receiver sensitivity threshold S_{\min} corresponding to the selected spreading factor and bandwidth.

Small-scale fading is modeled as log-normal shadowing with standard deviation σ . The instantaneous received power is expressed as

$$P_{RX,inst}(d) = P_{RX}(d) + X_{\sigma}, \quad (4)$$

where X_{σ} is a zero-mean Gaussian random variable with variance σ^2 . The link success probability is thus

$$P_{link}(d) = \Pr[P_{RX,inst}(d) \geq S_{\min}] = Q\left(\frac{S_{\min} - P_{RX}(d)}{\sigma}\right), \quad (5)$$

where $Q(\cdot)$ is the Gaussian complementary cumulative distribution function. These expressions are used to compute P_{DG} , P_{SR} , and P_{RG} for each node based on its distance to the gateway and the relay.

The probability of collision is a function of the number of simultaneously active nodes, their spreading factors, and the channel occupancy. Because the duty cycles in our groundwater monitoring scenario are low (a 15-minute interval between transmissions), collisions are relatively rare; however, they are still included in the model to obtain realistic delivery probabilities.

The final network reliability metric for a given topology is defined as the average packet success probability over all wells:

$$R_{net} = \frac{1}{N} \sum_{i=1}^N P_{coop,i}, \quad (6)$$

where $N = 8$ is the number of wells. This metric enables a direct comparison between the conventional star topology (where $P_{coop,i} = P_{DG,i}$) and the relay-assisted topology (where $P_{coop,i}$ is given by the cooperative expression above).

2.3 Simulation Setup and Performance Evaluation Metrics

To complement the analytical model and evaluate the proposed scheme under realistic variability, we implemented a Monte Carlo–based simulation framework. The simulator generates random realizations of the wireless channel, node positions, and traffic patterns, and computes key performance indicators for both the direct and relay-assisted topologies. The simulation environment represents the eight wells, the gateway, and the relay node using their actual geographic coordinates from the Karshi deployment. For each node, the distances to the gateway and the relay are computed, and the corresponding path loss parameters are derived using the model described in Section 3.2. In each Monte Carlo trial, independent realizations of shadowing and noise are generated for every link. For each 15-minute interval over a virtual monitoring period of 30 days,

Table 2. Channel and propagation parameters used in the reliability model

Parameter	Symbol	Value / Description
Carrier frequency	f	868 MHz (EU LoRa band)
Sensor node antenna height	h_{tx}	1.0–1.5 m (ground-mounted well stations)
Relay node antenna height	h_{relay}	2.5–3 m
Gateway antenna height	h_{gw}	10–12 m (mast-mounted)
Path loss exponent coefficient	A	128.1 dB (reference path loss at 1 km for 868 MHz)
Distance exponent	B	36.7 (rural LoRa path loss slope)
Shadowing standard deviation	σ	5.0–7.5 dB (log-normal fading, rural terrain)
Minimum required sensitivity, SF7	$S_{min}(SF7)$	–123 dBm
Minimum required sensitivity, SF8	$S_{min}(SF8)$	–126 dBm
Minimum required sensitivity, SF9	$S_{min}(SF9)$	–129 dBm
Minimum required sensitivity, SF10	$S_{min}(SF10)$	–132 dBm
Minimum required sensitivity, SF11	$S_{min}(SF11)$	–135 dBm
Minimum required sensitivity, SF12	$S_{min}(SF12)$	–137 dBm
Receiver noise figure	NF	6 dB (typical SX1301/SX1276)
Thermal noise floor (125 kHz)	N_0	–114 dBm
Shadowing distribution	—	Log-normal
Small-scale fading model	—	Rayleigh (non-line-of-sight wells)
Path loss model	—	Log-distance (rural), $PL = A + B \log_{10}(d) + X_\sigma$
Relay–gateway link margin boost	—	+2 to +4 dB due to increased antenna height
Channel bandwidth	BW	125 kHz

every node produces a data packet. The packet is first transmitted directly to the gateway; if it fails, a cooperative retransmission via the relay is attempted according to the scheme described above. The simulation procedure consists of a sequence of steps used to assess the performance of direct and relay-assisted LoRa topologies. It begins with the initialization of key parameters, including node positions, the channel propagation model, and LoRa communication settings such as spreading factor, bandwidth, and transmit power. The simulation then operates within a Monte Carlo framework involving repeated trials and time-step iterations. At each time step, the simulator computes the instantaneous received power on all links and checks whether SNR and sensitivity thresholds for successful decoding are met. Packets are accordingly classified as delivered or lost. After each trial, performance indicators—including packet delivery ratio, outage probability, and network reliability—are aggregated. This process provides a statistically robust evaluation of both baseline and relay-assisted configurations. We consider multiple simulation scenarios summarized in Table 3. Scenario S1 corresponds to the baseline star topology with no relay. Scenario S2 introduces a single relay with fixed parameters (SF10 on the relay–gateway hop). Scenario S3 explores adaptive relaying, where the relay selects the transmission parameters based on the distance of the originating sensor. For each scenario, 1000 Monte Carlo trials are executed to ensure statistical significance. Several performance metrics are computed. The

Packet Delivery Ratio (PDR) for a node i is defined as

$$\text{PDR}_i = \frac{N_{\text{rx},i}}{N_{\text{tx},i}}, \quad (7)$$

where $N_{\text{tx},i}$ is the total number of packets transmitted by node i and $N_{\text{rx},i}$ is the number of packets successfully received at the gateway, either directly or through the relay. The network-wide PDR is obtained by averaging PDR_i across all sensor nodes.

To provide additional insight into network behavior, a simulated 24-hour packet-arrival timeline was generated for all eight wells, showing the temporal distribution of successfully and unsuccessfully delivered packets under both direct and relay-assisted transmission modes. This visualization highlights how relay-assisted communication reduces packet loss and yields a more consistent delivery pattern, especially for distant wells.

The outage probability for node i is calculated as

$$P_{\text{out},i} = 1 - \text{PDR}_i, \quad (8)$$

which reflects the likelihood of link-level failures and is directly related to the success probabilities derived analytically in Section 3.2. This metric is particularly important in groundwater monitoring applications, where extended outages may lead to significant gaps in the recorded time series.

In addition, we estimate the average RSSI and SNR observed at the gateway for each node, allowing us to compare the effective link budgets in different topologies and to verify the consistency between the analytical model and simulated values. Energy consumption is evaluated in relative terms by counting the number of transmissions and retransmissions performed by each node under the duty-cycle constraints discussed in (Al-Sarawi et al., 2020). Although a detailed battery model is beyond the scope of this work, this count provides an indication of whether the relay-assisted topology introduces a significant overhead in terms of energy usage. Finally, for each scenario we compute the network reliability as defined previously and analyze its sensitivity to relay position, spreading factor combinations, and the number of wells. These analyses provide insight into how the design choices of cooperative relaying, translate into practical reliability improvements for groundwater monitoring networks.

3 Results

The performance of the proposed relay-assisted LoRa topology was assessed through analytical modelling and Monte Carlo simulations and compared directly with the baseline star configuration used in conventional LoRa deployments. The overall results demonstrate a clear and consistent improvement in communication reliability across the network after integrating a relay node into the system. In the baseline topology, packet delivery performance showed a strong dependence on distance. Wells located within approximately 800 meters of the gateway maintained high delivery rates, typically above 90%. However, wells positioned between 1.2 and 1.8 kilometers experienced a substantial degradation, with Packet Delivery Ratio (PDR) declining to values between 55% and 68%. After relay assistance was enabled, this limitation was largely overcome. For wells located beyond 1.5 km, the PDR increased to 85–92%, representing an improvement of 30–35 percentage points compared to direct links. Even wells positioned at intermediate distances benefited, gaining an additional 8–15% delivery reliability depending on their individual channel conditions. These results confirm that cooperative relaying effectively reduces the number of failed uplinks, especially for nodes operating near the edge of feasible LoRa coverage. The simulations further revealed a substantial reduction in outage probability across the network. In the baseline star configuration, distant wells frequently experienced outages with probabilities ranging from 0.40 to 0.45, indicating unstable connectivity. With relay support, the outage probability for all wells decreased to below 0.15 and, for several nodes, down to as low as 0.05.

Table 3. Simulation scenarios and Monte Carlo configuration

ID	Topology	Relay	SF Settings	Tx Power	Duration	Trials	Description
S1	Direct (Star)	No	SF9 for all wells	14 dBm	24 h	10,000	Baseline star topology; distance-dependent PDR drop analyzed
S2	Direct (Star)	No	Distance-adaptive: SF7–SF10	14 dBm	24 h	10,000	Baseline with adaptive SF to test improvement limits without relay
S3	Direct (Star)	Yes	Sensor→Relay: SF7; Relay→Gateway: SF10	Sensor: 14 dBm; Relay: 17 dBm	24 h	10,000	Primary cooperative relaying scenario (main comparison case)
S4	Relay-assisted	Yes	Sensor→Relay: SF8; Relay→Gateway: SF10	Sensor: 14 dBm; Relay: 17 dBm	24 h	10,000	Higher spreading factor for difficult sensor–relay links
S5	Relay-assisted	Yes	Fully adaptive per hop: SF7–SF10	Sensor: 14 dBm; Relay: 17 dBm	24 h	10,000	Adaptive cooperative scheme; tests robustness under channel variance
S6	Direct (Star)	No	SF9	14 dBm	72 h	30,000	Long-duration baseline scenario to evaluate outage accumulation
S7	Direct (Star)	Yes	SF7 → SF10	14/17 dBm	72 h	30,000	Long-duration relay scenario to assess multi-day reliability improvements
S8	Relay-assisted	Yes	SF7 → SF10	14/17 dBm	72 h	30,000	Long-duration relay scenario to assess multi-day reliability improvements

The network-wide average outage probability dropped from 0.21 to 0.08, corresponding to roughly a 62% improvement in link stability. This enhancement is especially important in groundwater-monitoring applications, where continuous records are essential for accurate hydrogeological trend analysis. Signal-quality metrics, including RSSI and SNR, also improved with the relay. For distant wells, the average RSSI measured at the gateway increased by approximately 6–11 dB. This improvement is attributed to the relay–gateway hop operating with a higher spreading factor and covering a shorter physical distance, thereby improving the link budget. Additionally, the RSSI values displayed lower variance, suggesting improved channel stability along the two-hop path. SNR followed a similar pattern, improving by 3–7 dB for distant wells. Because LoRa demodulation performance is highly sensitive to SNR, this enhancement directly contributed to the increased PDR and reduced outage levels observed during the simulations. The reliability index computed for the network further supports these findings. The baseline topology achieved an overall reliability of 0.78, whereas the relay-assisted network reached 0.93, representing a 19% increase in overall system robustness. An important outcome is that the relay configuration significantly reduced the performance disparity between near and distant wells. In the star topology, reliability variation across wells exceeded 30%, while in the relay-assisted configuration this variation fell to less than 10%, indicating a more uniform and balanced network. Finally, system-level evaluations confirmed that introducing the relay node did not lead to excessive network congestion or energy overhead. Because LoRa nodes operate at low duty cycles and the relay retransmits selectively, the average number of transmissions per node increased by only about 8%, remaining well within the operational limits for battery-powered IoT devices.

The results clearly demonstrate that relay-assisted communication significantly enhances the robustness, reliability, and stability of LoRa-based groundwater-monitoring networks. The proposed topology achieves these improvements without requiring additional gateways or major infrastructure upgrades, making it a practical and cost-effective solution for large rural monitoring deployments.

4 Conclusion

This study investigated how cooperative relaying can enhance the reliability of LoRa-based sensor networks used for groundwater monitoring in the Karshi district of Uzbekistan. The results clearly demonstrate that a conventional star topology, while simple and energy-efficient, is insufficient for geographically dispersed well networks where several nodes operate near or beyond the edge of LoRa coverage. Such nodes experience significant packet loss, unstable RSSI values, and high outage probability, ultimately reducing the continuity and usefulness of groundwater-level time series. By integrating a relay-assisted communication scheme, the network achieves substantial improvements in all key performance indicators. Simulation results show that packet delivery for distant wells increases by 30–35 percentage points, and network-wide outage probability decreases by more than half. These gains are directly attributed to the enhanced link budget provided by the relay–gateway hop and the shorter sensor–relay distances, which together deliver more stable SNR and RSSI conditions. The cooperative forwarding mechanism also balances the performance among all wells, reducing disparities between near and far nodes and resulting in a more uniform and reliable network. Importantly, these benefits are obtained without the need for additional gateways or major infrastructure investments. The relay node requires minimal power, introduces negligible network overhead, and operates within duty-cycle constraints. This makes the proposed design a practical and scalable solution for large rural monitoring areas. The findings reinforce the potential of combining LoRa with cooperative relaying strategies to overcome common limitations such as signal fading, interference, and energy constraints. The study demonstrates that relay-assisted LoRa topologies can significantly enhance the robustness of groundwater monitoring systems, ensuring more complete datasets and enabling better management of critical water resources. Future work may extend this design by exploring adaptive relaying, multi-relay architectures, and integration with additional hydro-environmental sensors to support broader water-resource management applications.

Resources

- Borkotoky, S. S. (2022). “Coded relaying in LoRa sensor networks”. In: URL: <https://arxiv.org/pdf/2101.01176>.
- Borkotoky, S. S., B. Saikia, and N. Sarma (2019). “Cooperative relaying in LoRa sensor networks”. In: URL: <https://arxiv.org/pdf/1906.07596.pdf>.
- Makange, L., M. Rwakibo, A. Nambajimana, and I. Nkurikiyinfura (2023). “Design and application of a low-cost, low-power, LoRa-GSM IoT enabled system for groundwater monitoring”. In: URL: <https://scispace.com/pdf/design-and-application-of-a-low-cost-low-power-lora-gsm-iot-55sypnt5ta.pdf>.
- Memonova, G., P. Schmidt, J. Tursunov, and G. Gofurova (2025). “Automated groundwater monitoring system with real-time data collection and analysis using LoRa and GSM technologies”. In: *Environment. Technology. Resources* 4, pp. 209–215. URL: <https://journals.ru.lv/index.php/ETR/article/view/8412>.
- Memonova, G. and J. Tursunov (2022). “Fields of application of wireless sensor networks”. In: *International Conference on Learning and Teaching*. Tashkent, Uzbekistan. URL: <https://researchedu.uz/wp-content/uploads/2022/08/CONFERENCE-2022-13-full.pdf>.

- Al-Sarawi, S., M. Anbar, K. Alieyan, and A. Alzubaidi (2020). “Performance evaluation of LoRa and Sigfox technologies”. In: *American Journal of Research*. URL: <https://www.ajrsp.com/en/Archive/issue-38/Performance%20Evaluation%20of%20LoRa%20and%20Sigfox.pdf>.
- Sendra, S., J. Lloret, L. Parra, and M. Garcia (2023). “LoRa-based network for water quality monitoring”. In: URL: <https://riunet.upv.es/server/api/core/bitstreams/3af27a38-8a11-47c2-9565-68714435ebb2/content>.
- Sendra, S., L. Parra, and J. Lloret (2022). “Water quality monitoring based on IoT and LoRaWAN technologies”. In: *Sensors* 22.19. URL: <https://www.mdpi.com/1424-8220/22/19/7188>.

A Knowledge-Driven Scheduling Architecture Integrated with a Digital Twin

Martin Mišút¹ Mária Mišútová²

Abstract

Modern manufacturing systems operate under high variability, uncertainty, and reduced decision time, which limits the applicability of classical optimisation and static heuristic scheduling approaches. This paper proposes a knowledge-driven scheduling architecture integrated with a digital twin to support adaptive decision-making in dynamic production environments. The framework combines context-sensitive strategy restriction, multi-criteria evaluation, and simulation-based validation within a closed-loop structure. Scheduling strategies are dynamically selected and ranked based on real-time system conditions, and validated through high-fidelity digital twin simulations prior to deployment. A formal mathematical model of the architecture is presented. The proposed architecture establishes a foundation for future empirical validation in industrial environments.

Keywords

Operations Scheduling, Artificial Intelligence, Digital Twin, Knowledge-based scheduling, Smart Manufacturing

1 Introduction

Operations scheduling allocates limited resources to tasks over time while satisfying technological and operational constraints and optimising performance objectives. It is a central decision-making problem in manufacturing, logistics, and service operations. Many scheduling problems, including job shop and flexible job shop scheduling problems, are NP-hard, making exact optimisation computationally infeasible for large industrial instances. Contemporary manufacturing environments exhibit high product variability, shorter life cycles, frequent disturbances, flexible routing, and increasing automation. Under these conditions, scheduling shifts from an offline planning activity to a real-time control function. Reduced decision time and rising system complexity shift attention from generating optimal schedules to identifying robust, adaptable solutions. Traditional approaches such as mathematical programming, dispatching rules, and metaheuristics face well-known limitations. Exact methods scale poorly, static rules lack contextual adaptability, and metaheuristics often require extensive computation or parameter tuning. Consequently, recent research emphasises adaptive, data-driven, and simulation-supported scheduling methods. Digital twins and artificial intelligence (AI) have emerged as key enablers of intelligent, responsive scheduling systems (Serrano-Ruiz et al., 2021; Negri et al., 2021). Despite progress, several challenges persist. Many AI-based approaches rely either on generative optimisation or purely data-driven learning, often without incorporating structured domain knowledge. Simulation and scheduling are frequently treated as loosely coupled components, and interpretability remains limited an important issue in industrial environments where transparency and controllability are essential.

¹University of Economics in Bratislava, Faculty of Economic Informatics, Department of Applied Informatics, Bratislava, SLOVAKIA, martin.misut@euba.sk, ORCID: 0000-0002-5545-2624

²Slovak University of Technology, Faculty of Materials Science and Technology, Institute of Applied Informatics, Automation and Mathematics, Trnava, SLOVAKIA, maria.misutova@stuba.sk

To address these limitations, this paper proposes a knowledge-driven scheduling architecture integrated with a digital twin. The approach combines hierarchical rule-based knowledge representation, context-sensitive strategy restriction, and simulation-based validation within a closed-loop control framework. The system dynamically selects and ranks scheduling strategies based on the current production state and evaluates candidate solutions through digital-twin simulations before deployment.

The main contributions of this paper are:

- the integration of structured domain knowledge and digital twin simulation into a unified scheduling decision loop.
- the formulation of a mathematical model describing the interaction between monitoring, strategy selection, simulation, and multi-criteria evaluation layers.
- the design of a modular architecture supporting adaptive scheduling in dynamic manufacturing environments.

2 State of the Art

The application of AI in scheduling was also supported by the fact that the difficult, complex scheduling problem becomes even more challenging in adaptive manufacturing systems (AMS) due to reduced decision time and increased problem complexity (Serrano-Ruiz et al., 2021). The high frequency of changes in running production, frequent failures, and drastic reductions in setup times reduce the available time for decision preparation. The possibility of variable technological processes, the interchangeability of machines, the automation of transport and handling operations, and the limited number of tools, pallets, and fixtures have increased the complexity of the scheduling problem. Under such conditions, the problem shifts from the search for optimal solutions to the identification of admissible and robust schedules (Fu et al., 2021). The rapid evolution of adaptive manufacturing environments reduces the direct transferability of traditional experiential knowledge, thereby increasing the need for computational decision support systems. A possible solution is the use of computer-based decision support systems, either in the form of a support tool for human subject decisions, the so-called decision support systems or full equivalent substitution of a human by a program system. The degree of substitution of a human by an AI in the scheduling process depends on several factors. One of the main ones is the degree of automation of the production environment and control functions. In general, the higher the degree of automation in the production process, the higher the degree of automation in the scheduling process (Marzia et al., 2023). Interactive systems seem better suited to traditional workshops, significantly improving efficiency and facilitating human decision-making. On the contrary, for highly automated AMS, designs of autonomous systems often prevail as part of a control system or control hierarchy (short-term planning - scheduling - real-time control) (Destouet et al., 2023). Overall, recent research indicates a shift: scheduling is increasingly moving towards real-time activities, i.e., from the planning horizon to the control horizon (S. Zhang et al., 2021).

2.1 Generative vs. Evaluative AI in Scheduling

According to the way artificial intelligence is applied in scheduling, two main approaches can be distinguished: generative and evaluative systems.

Generative systems apply AI principles to directly create a schedule by exploring the state space or by reasoning within defined constraints. In such systems, artificial intelligence replaces traditional algorithmic procedures used for schedule generation. While generative AI-based scheduling approaches have demonstrated promising results, their deployment in complex industrial environments remains relatively limited compared to hybrid or simulation-supported

approaches. The most successful applications have been reported at higher levels of planning, such as rough production scheduling. Many generative systems are implemented as planning expert systems that rely on forward-chaining inference mechanisms.

Evaluative systems, in contrast, apply AI principles in several complementary roles:

- to evaluate alternative schedule variants and select the most appropriate one,
- to determine suitable models, methods, or rules for schedule generation,
- to correct, improve, and maintain already created schedules.

In evaluative approaches, the schedule itself is generated algorithmically, for example using operations research methods, heuristic procedures, or simulation models. Artificial intelligence techniques are then applied to analyse, evaluate, or improve the generated solutions (Del Gallo et al., 2023). Consequently, evaluative systems represent a hybrid approach that combines artificial intelligence with conventional optimisation and simulation methods.

2.2 Digital Twins in Scheduling

Digital twins have become a key element of modern planning. They allow the creation of high-fidelity virtual models of manufacturing systems that are synchronized with real-world data. This allows testing different planning strategies, assessing their robustness, and analyzing system behavior under uncertainty (Tliba et al., 2022). Simheuristic approaches combine optimization with digital twins, thereby increasing adaptability and fault tolerance (Negri et al., 2021).

2.3 Trends in Scheduling Research

Operations research is moving towards energy-efficient and multi-objective models that consider a wider range of performance criteria (Gahm et al., 2016; F. Zhang et al., 2024). At the same time, decentralized and multi-agent solutions are gaining importance, enabling adaptive production control when combined with digital twins (Siatras et al., 2024; Peng et al., 2024).

Despite progress, integrating structured domain knowledge, hierarchical inference, and digital twins into a unified decision-making framework remains underexplored. Overall, the literature demonstrates a clear transition from static, optimisation-centric scheduling to adaptive, simulation-supported, and AI-driven frameworks. Digital twin technologies enable virtual validation of candidate schedules, while data-driven and multi-agent approaches enhance responsiveness in dynamic environments. However, the explicit integration of structured domain knowledge, hierarchical rule-based reasoning, and digital twin simulation within a unified closed-loop architecture remains insufficiently formalised. This gap motivates the architecture presented in this paper.

3 Proposed Scheduling Architecture

The proposed architecture of the Integrated Scheduling System (ISS) combines artificial intelligence, structured knowledge, and simulation through a digital twin. The goal is to enable dynamic adaptation of scheduling decisions in an environment with high uncertainty and changing loads. The ISS can generate, evaluate, and compare multiple scheduling scenarios, using known heuristics (e.g., priority rules) and their context-sensitive combinations. In line with Vespoli et al. (2021) and Serôdio et al. (2024), the architecture enables predictive data processing and rapid response to changes in the system.

ISS is designed as a closed control loop consisting of four layers:

- Monitoring Layer

- Knowledge-Based Strategy Selection Layer
- Digital Twin Simulation Layer
- Evaluation and Adaptation Layer

Together, these layers create an adaptive mechanism that first limits the space of possible strategies, then simulates them, evaluates them, and selects the most suitable one for real deployment.

3.1 Monitoring Layer

The Monitoring Layer connects the physical production system to the ISS's digital decision-making core. Its task is to:

- obtain current data on the state of production,
- filter and synchronize information,
- create an accurate representation of the state of the system for further decision-making.

The quality of scheduling decisions depends directly on the accuracy and timeliness of this state representation.

The physical production system includes:

- Machines $M = \{M_1, \dots, M_m\}$
- Jobs $J = \{J_1, \dots, J_n\}$
- Resources and transport systems
- IoT sensing infrastructure

The real system state at time t including machine availability, queue lengths, job attributes, due dates, and resource status, is represented as:

$$x_{\text{real}}(t) \in \mathbb{R}^n \quad (1)$$

Real-time data are processed and synchronized:

$$x_{\text{real}}(t) \rightarrow x_{\text{sync}}(t) \quad (2)$$

This layer ensures consistency between physical and virtual representations. Functional components of the Monitoring Layer include:

- (a) Data Acquisition Module, which collects data from IoT sensors, PLC systems, MES/ERP systems, machine controllers, etc.
- (b) Data Preprocessing Module, which performs noise filtering, outlier detection and removal, time synchronisation, and missing data imputation.
- (c) State Aggregation Module, which transforms raw data into a compact state vector:

$$x(t) = [m(t), q(t), r(t), d(t), s(t)] \quad (3)$$

where

- $m(t)$ – machine states,
- $q(t)$ – queue lengths,
- $r(t)$ – resource availability,
- $d(t)$ – job due dates,
- $s(t)$ – system-level parameters (load, utilisation).

3.2 Knowledge-Based Strategy Selection Layer

This layer represents the system's symbolic reasoning component. It prevents brute-force search by restricting the strategy space intelligently based on structured domain knowledge.

Its objectives are to:

- restrict the space of available scheduling strategies,
- select relevant strategies according to the current system context,
- rank them according to expected suitability.

Formally, this layer can be modelled as follows: Let:

$$\mathcal{R} = \{\mathcal{R}_1, \mathcal{R}_2, \dots, \mathcal{R}_k\} \quad (4)$$

be the set of all available scheduling strategies, where R_i represents a scheduling rule or strategy.

In the current state of the system, the properties of the production environment impose constraints that can be taken into account when selecting a subset of suitable scheduling strategies. Then the context-restricted strategy subset \mathcal{R}_c can be defined as follows:

$$\mathcal{R}_c(x(t)) \subseteq \mathcal{R} \quad (5)$$

3.3 Digital Twin Simulation Layer

A digital twin is a virtual replica of a physical system (e.g., a factory or production line). Simulations with digital twins enable the testing and optimisation of scheduling strategies before implementation (Serrano-Ruiz et al., 2021; Ouahabi et al., 2024). Here is how:

- *Virtual Representation:* Digital twins are built using real-time and historical data from physical systems, such as production lines, warehouses, or fleets. They replicate the behaviour, constraints, and interdependencies of the physical environment.
- *Simulation of Scenarios:* Scheduling strategies can be tested in the digital twin to evaluate their performance under different scenarios, such as varying workloads, resource availability, or unexpected disruptions (e.g., equipment failure or increased demand).
- *Dynamic Adjustments:* Digital twins can model dynamic systems, such as fluctuating demand or real-time changes, enabling the testing of adaptive scheduling strategies that respond to live data.
- *Performance Analysis:* By running simulations, digital twins help identify bottlenecks, inefficiencies, or conflicts in scheduling strategies, providing insights into areas that need improvement.
- *Optimisation Before Deployment:* Scheduling solutions can be fine-tuned and validated in a virtual environment before deployment in the real world. This minimises risks, reduces costs, and ensures smoother transitions.
- *Predictive Insights:* Digital twins integrate predictive analytics to forecast the outcomes of different scheduling decisions. For example, they can predict delays or resource shortages and recommend proactive adjustments.
- *Continuous Feedback Loop:* As digital twins are connected to the physical system via IoT and real-time data feeds, they update and refine their simulations based on actual outcomes, ensuring ongoing accuracy and relevance.

The digital twin is designed as a cyber-physical closed-loop control system integrating real-time monitoring, knowledge-based decision support, simulation-based evaluation, and adaptive feedback (Ouahabi et al., 2024).

The Digital Twin Simulation Layer acts as a virtual experimentation environment, where candidate schedules are tested before implementation in the real system. This layer transforms scheduling from rule-based selection to simulation-validated decision-making, providing robustness and risk mitigation. The digital twin model can be formally defined as follows:

$$\mathcal{DT} = (X, U, F, G, \mathcal{R}, J) \quad (6)$$

where:

- X – state space,
- U – decision space,
- F – transition function,
- G – performance metrics,
- \mathcal{R} – strategy set,
- J – objective function.

The dynamic model of a digital twin can be defined as follows:

$$x_{DT}(t+1) = F(x_{DT}(t), S_i, \omega(t)) \quad (7)$$

where:

- F represents the production system dynamics,
- S_i is the candidate schedule,
- $\omega(t)$ models stochastic disturbances.

The digital twin includes the following components:

- Discrete-event simulation,
- Machine failure models,
- Sequence-dependent setup modelling,
- Transport system modelling,
- Flexible routing representation,
- Processing time variability modelling.

Parallel Strategy Evaluation can be modelled as follows: For each strategy, the candidate schedule is

$$S_i = \Phi(\mathcal{R}_i, x_{DT}(t)) \quad (8)$$

Simulated evolution is

$$x_{DT}^{(i)}(t+1) = F(x_{DT}(t), S_i) \quad (9)$$

Furthermore, the performance metric is:

$$f_k(S_i) \quad (10)$$

3.4 Evaluation and Adaptation Layer

The ISS evaluates simulated scheduling strategies against predefined criteria. Since it also computes production system and task parameters, multi-criteria evaluation methods can be applied, with the ISS assigning weight coefficients to individual criteria. This evaluation may operate as an external procedural module triggered by production rules. If the results are unsatisfactory, the ISS adjusts task priorities and/or selects alternative scheduling strategies and reruns the simulation. If satisfactory, it selects and deploys the schedule that best meets the defined criteria. Formally, this layer evaluates candidate schedules, selects the best solution, and updates the knowledge structure. When using Multi-Criteria Evaluation

$$J(S_i) = \sum_{k=1}^m w_k f_k(S_i) \quad (11)$$

where

- $J(S_i)$ — multi-criteria objective function for generated schedule S_i ,
- f_k — performance metrics,
- w_k — weight coefficients.

The decision can be selected as follows:

$$S^* = \arg \min_{S_i} J(S_i) \quad (12)$$

Selected schedule:

$$u^*(t) = S^* \quad (13)$$

The selected schedule is deployed in the physical system.

Adaptation mechanisms can be defined as follows:

Strategy ranking update

$$Q(\mathcal{R}_i) \leftarrow Q(\mathcal{R}_i) + \beta(J_{\text{ref}} - J(S_i)) \quad (14)$$

Weight adjustment

$$w_k(t+1) = w_k(t) + \alpha \Delta_k \quad (15)$$

And calibration of Fuzzy Parameters

$$\theta_{\text{new}} = \theta_{\text{old}} + \alpha \Delta \quad (16)$$

The four-layer architecture transforms scheduling from a static rule-application process into a closed-loop, adaptive decision-making framework that integrates real-time monitoring, symbolic knowledge restriction, simulation-based validation, and performance-driven adaptation. The architecture forms a hierarchically organised closed-loop control system (see Figure 1).

4 Implementation and Evaluation

The proposed architecture integrates strategy selection, simulation-based validation, adaptive knowledge improvement, and explanatory capabilities. It supports both industrial deployment and educational use, where it is applied in production management training to model decision-making scenarios and evaluate alternative planning strategies. This proposal builds on the authors' experience implementing scheduling systems for less-automated manufacturing processes using artificial intelligence in the form of expert systems. The architecture will be implemented as a prototype scheduling module.

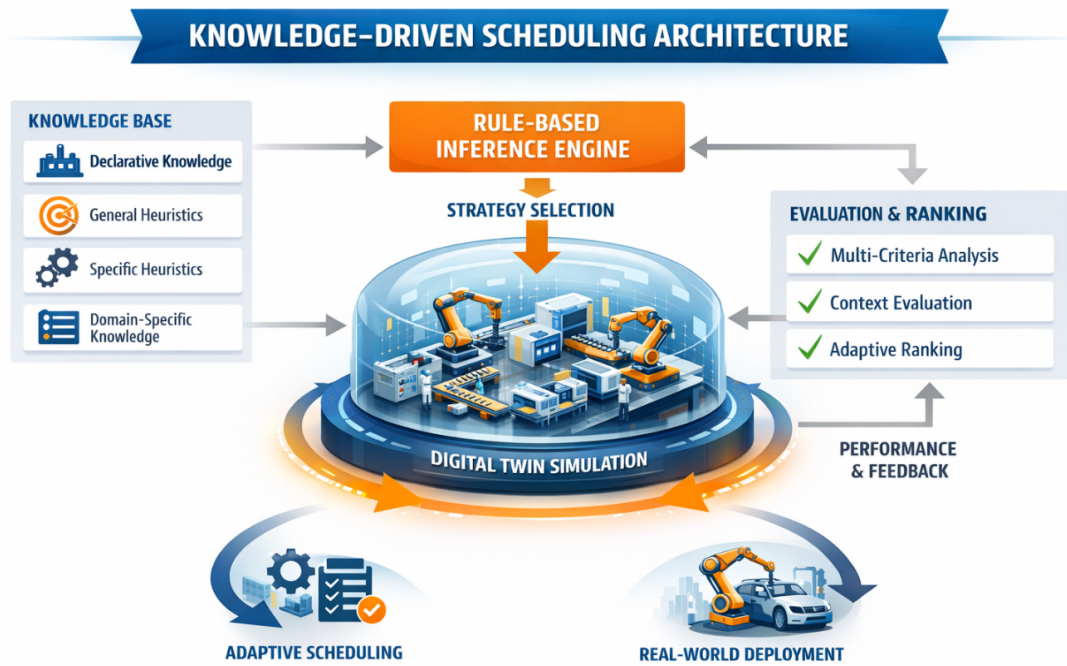


Figure 1. Architecture of the knowledge-driven scheduling system.

5 Conclusion

This paper presented a knowledge-driven scheduling architecture integrated with a digital twin to support adaptive decision-making in modern manufacturing environments. The proposed framework combines hierarchical knowledge representation, context-sensitive restriction of scheduling strategies, simulation-based validation, and multi-criteria evaluation within a unified closed-loop structure. By embedding structured heuristic knowledge into a simulation-supported decision process, the architecture enables context-aware selection and verification of scheduling strategies prior to their deployment in the physical production system. The integration of a digital twin provides a risk-free environment for evaluating candidate schedules under dynamic and uncertain operating conditions, enabling robustness analysis and proactive adaptation to disturbances. Compared with static rule-based scheduling approaches, the proposed framework introduces adaptive ranking and contextual calibration mechanisms that allow continuous improvement of scheduling performance while maintaining interpretability through explicit knowledge representation. The architecture therefore provides a structured approach to integrating symbolic artificial intelligence methods with simulation-based optimisation for production scheduling. The proposed framework represents a conceptual step toward more adaptive and transparent scheduling systems capable of operating in highly dynamic manufacturing environments. However, several challenges remain. The effectiveness of the approach depends on the fidelity and calibration of the digital twin model, which requires accurate data and domain expertise. Contextual parameters must be calibrated for specific production systems, and scalability may become an issue as the space of scheduling strategies increases. Consequently, further validation in industrial environments is required. Future research will focus on large-scale experimental evaluation, automated calibration of contextual parameters, and the integration of data-driven learning mechanisms for dynamic strategy refinement. Hybrid symbolic–data-driven approaches represent a promising direction for combining interpretability with adaptability in intelligent production scheduling systems.

Resources

- Del Gallo, M., G. Mazzuto, F. E. Ciarapica, and M. Bevilacqua (2023). “Artificial Intelligence to Solve Production Scheduling Problems in Real Industrial Settings: Systematic Literature Review”. In: *Electronics* 12.23, p. 4732. URL: <https://www.mdpi.com/2079-9292/12/23/4732>.
- Destouet, C., H. Tlahig, B. Bettayeb, and B. Mazari (2023). “Flexible job shop scheduling problem under Industry 5.0: A survey on human reintegration, environmental consideration and resilience improvement”. In: *Journal of Manufacturing Systems* 67, pp. 155–173. DOI: 10.1016/j.jmsy.2023.01.004. URL: <https://doi.org/10.1016/j.jmsy.2023.01.004>.
- Fu, Y., Y. Hou, Z. Wang, X. Wu, K. Gao, and L. Wang (2021). “Distributed scheduling problems in intelligent manufacturing systems”. In: *Tsinghua Science and Technology* 26.5, pp. 625–645.
- Gahm, C., F. Denz, M. Dirr, and A. Tuma (2016). “Energy-efficient scheduling in manufacturing companies: A review and research framework”. In: *European Journal of Operational Research* 248.3, pp. 744–757. DOI: 10.1016/j.ejor.2015.07.017. URL: <https://doi.org/10.1016/j.ejor.2015.07.017>.
- Marzia, S., A. Vital-Soto, and A. Azab (2023). “Automated process planning and dynamic scheduling for smart manufacturing: A systematic literature review”. In: *Manufacturing Letters* 35, pp. 861–872. DOI: 10.1016/j.mfglet.2023.07.013. URL: <https://doi.org/10.1016/j.mfglet.2023.07.013>.
- Negri, E., V. Pandhare, L. Cattaneo, J. Singh, M. Macchi, and J. Lee (2021). “Field-synchronised Digital Twin framework for production scheduling with uncertainty”. In: *Journal of Intelligent Manufacturing* 32.4, pp. 1207–1228. DOI: 10.1007/s10845-020-01685-9.
- Ouahabi, N., A. Chebak, O. Kamach, O. Laayati, and M. Zegrari (2024). “Leveraging digital twin into dynamic production scheduling: A review”. In: *Robotics and Computer-Integrated Manufacturing* 89, p. 102778. DOI: 10.1016/j.rcim.2024.102778. URL: <https://doi.org/10.1016/j.rcim.2024.102778>.
- Peng, S., G. Xiong, J. Yang, Z. Shen, T. S. Tamir, Z. Tao, and F.-Y. Wang (2024). “Multi-Agent Reinforcement Learning for Extended Flexible Job Shop Scheduling”. In: *Machines* 12.1, p. 8. URL: <https://www.mdpi.com/2075-1702/12/1/8>.
- Serôdio, C., P. Mestre, J. Cabral, M. Gomes, and F. Branco (2024). “Software and architecture orchestration for process control in Industry 4.0 is enabled by cyber-physical systems technologies”. In: *Applied Sciences* 14.5, p. 2160.
- Serrano-Ruiz, J. C., J. Mula, and R. Poler (2021). “Smart manufacturing scheduling: A literature review”. In: *Journal of Manufacturing Systems* 61, pp. 265–287. DOI: 10.1016/j.jmsy.2021.09.011. URL: <https://doi.org/10.1016/j.jmsy.2021.09.011>.
- Siatras, V., E. Bakopoulos, P. Mavrothalassitis, N. Nikolakis, and K. Alexopoulos (2024). “Production Scheduling Based on a Multi-Agent System and Digital Twin: A Bicycle Industry Case”. In: *Information* 15.6, p. 337. URL: <https://www.mdpi.com/2078-2489/15/6/337>.
- Tliba, K., T. M. L. Diallo, O. Penas, R. B. Khalifa, N. B. Yahia, and J.-Y. Choley (2022). “Digital twin-driven dynamic scheduling of a hybrid flow shop”. In: *Journal of Intelligent Manufacturing*. DOI: 10.1007/s10845-022-01922-3.
- Vespoli, S., M. Scarpati, G. Guizzi, and A. Grassi (2021). “A novel dispatching rule for semi-heterarchical architectures in the Industry 4.0 context”. In: *IFAC-PapersOnLine* 54.1, pp. 86–91. DOI: 10.1016/j.ifacol.2021.08.010. URL: <https://doi.org/10.1016/j.ifacol.2021.08.010>.
- Zhang, F., R. Li, and W. Gong (2024). “Deep reinforcement learning-based memetic algorithm for energy-aware flexible job shop scheduling with multi-AGV”. In: *Computers & Industrial Engineering* 189, p. 109917. DOI: 10.1016/j.cie.2024.109917. URL: <https://doi.org/10.1016/j.cie.2024.109917>.
- Zhang, S., F. Tang, X. Li, J. Liu, and B. Zhang (2021). “A hybrid multi-objective approach for real-time flexible production scheduling and rescheduling under dynamic environment in

Industry 4.0 context”. In: *Computers & Operations Research* 132, p. 105267. DOI: 10.1016/j.cor.2021.105267. URL: <https://doi.org/10.1016/j.cor.2021.105267>.

Game Theory and Inter-Company Data Sharing for AI Model Training

Stefan Otto Novak¹

Abstract

This paper examines how game theory can be used to model and design data-sharing strategies between firms developing artificial intelligence. It explores both competitive and cooperative settings, analyzing how incentives, market structure, privacy concerns, and platform design affect data-sharing outcomes. Two core models are discussed: a non-cooperative game between competing firms and a cooperative model involving Shapley value allocation. The study highlights when data sharing can be stable, fair, and beneficial for innovation while identifying potential risks to competition and consumer welfare.

Keywords

Game theory; Data sharing; Artificial intelligence

1 Introduction

Game theory is the study of strategic interactions, providing mathematical tools to analyze how self-interested decision-makers choose actions that affect one another's outcomes. In the context of data sharing between companies, game theory offers insight into when firms will cooperate to share data and how such cooperation can be sustained. Two broad branches of game theory are particularly relevant: non-cooperative game theory, which assumes players act independently without enforceable agreements, and cooperative game theory, which examines how players can form binding agreements or coalitions and how to fairly allocate the gains of cooperation. Both approaches help illuminate the incentives and obstacles in sharing data for training artificial intelligence models.

In a non-cooperative setting, each company decides whether to share data based on its own profit motive. Data has become a key competitive asset for AI model training, so sharing data with a competitor is a strategic dilemma: collaboration can improve each firm's machine learning model, but it may also strengthen competitors and erode one's own advantage. This tension resembles a prisoner's dilemma or a coordination game, in which mutual cooperation can increase collective benefits, but each firm may hesitate if it fears losing a competitive edge. Non-cooperative analysis typically uses Nash equilibrium, a stable outcome where no player can unilaterally deviate to improve their payoff, to predict whether firms will share data or withhold it. For example, two competing firms might each choose not to share data in equilibrium if sharing would help their rival more than themselves, even though both would be better off if they could somehow agree to share. Game-theoretic models can highlight such conditions and suggest incentive mechanisms to encourage sharing.

In contrast, a cooperative game-theoretic perspective assumes companies can forge binding agreements or be coordinated by a trusted mediator. Here, the questions shift to how to fairly distribute the benefits of shared data and ensure no subset of firms has an incentive to defect from the agreement. Concepts like the Shapley value, which measures each participant's marginal contribution to the whole, and the core, the set of stable payoff allocations that leave no group

¹Bratislava University of Economics and Business, Faculty of Economic Informatics, Department of Applied Informatics, Bratislava, Slovakia, stefanotto.novak@euba.sk

better off by splitting away, become important. For instance, if several firms pool data to train a superior AI model, cooperative game theory can be used to decide how to split the gains, such as accuracy improvements, revenues, etc. in a way that each firm feels adequately rewarded for its data contribution. A fair and stable allocation is crucial, if one firm perceives the benefit sharing as unfair, it may withdraw its data in the future. Thus, cooperative models address the incentive compatibility of long-term data-sharing consortia or data marketplaces.

This report explores how game theory, both non-cooperative and cooperative, can model data-sharing arrangements between companies for AI training. First, we review key findings, models, and assumptions from recent research on this topic, highlighting concrete scenarios where game-theoretic analysis has been applied. These include strategic data sharing among competitors, platform-mediated data exchanges, privacy-preserving data sharing, and fair reward allocation in multi-party machine learning.

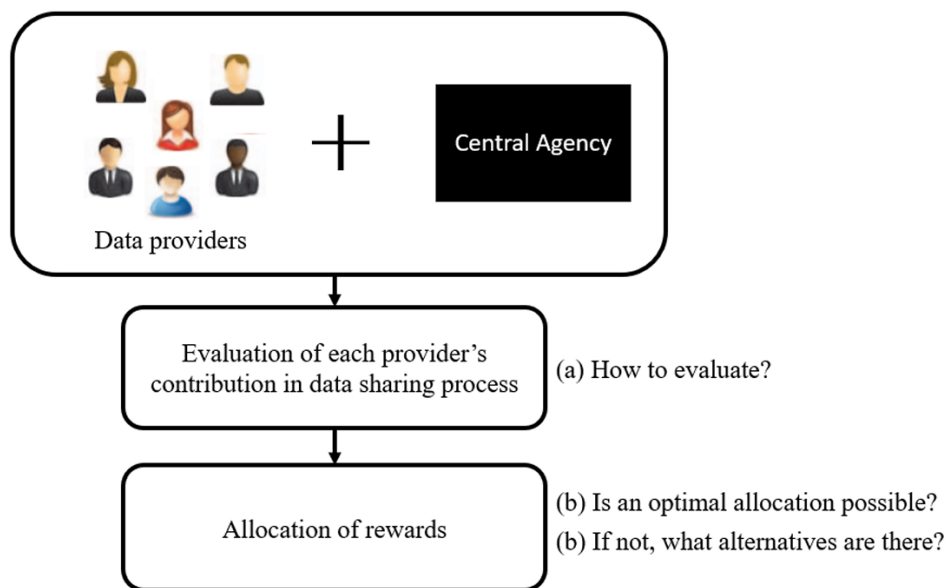


Figure 1. Overview of Data Sharing. Source: Chen (2020)

To guide the analysis, we focus on two key questions in collaborative data-sharing systems: how to evaluate each agent's contribution, and how to allocate rewards fairly, so that no participant has the incentive to leave the collaboration. This dual concern is illustrated in Figure 1, which outlines the high-level structure of collaborative data sharing, from agent participation to evaluation and payoff distribution. These questions are addressed differently in competitive and cooperative game-theoretic models, which we explore throughout the paper.

2 Game-Theoretic Models of Data Sharing in AI

2.1 Data Sharing Among Competitors: Incentives and Market Effects

One line of research examines when rival firms voluntarily share data in competitive markets. Tsoy and Konstantinov (Tsoy and Konstantinov, 2023) present a general framework for analyzing the data-sharing trade-off between competitors. Their framework has three components: (1) a market model capturing firms' production or pricing decisions, (2) a data impact model relating data volume to machine learning model quality, and (3) a collaboration scheme representing the negotiation or agreement process. In an instantiation of this framework, they consider two firms competing in the same market and investigate a binary choice: either the firms fully share their data with each other, or not at all. The firms' payoffs are their profits, which depend on market outcomes and their AI model's performance. Additional data improves model accuracy, which

can lower costs or increase consumer demand, thus raising profit, but if one firm's data sharing significantly improves its rival's model, the rival may become more competitive and drive down the first firm's market share or pricing power.

Using classical models of oligopoly competition, Tsoy and Konstantinov derive the Nash equilibrium outcomes under each scenario. They find that market conditions critically shape data-sharing incentives. If the firms' products are highly substitutable or the learning task is simple, each firm has less incentive to share data. Conversely, reduced competition and more challenging learning tasks make collaboration more appealing because the benefits of pooling data outweigh the competitive spillover. Their negotiation scheme assumes that both firms will agree to share only if it increases each firm's expected profit, i.e., data sharing occurs if and only if it is a Pareto improvement for the competitors. This mutual agreement effectively means that data sharing is a stable outcome only in parameter regimes where it is win-win; otherwise, if one firm would lose profit by sharing, the deal fails. Under those favorable conditions, the firms coordinate on the sharing strategy, improving their models and profits simultaneously. Notably, when sharing is beneficial to both parties, it is also a Nash equilibrium in the sharing decision game. But when sharing would hurt one of the firms, the only equilibrium is no-sharing, even if that leaves collective gains on the table.

A concrete example of this dynamic is the taxi market scenario described by Tsoy and Konstantinov. Two ride-hailing companies operating in the same city each gather data on rides and traffic to train their dispatch and pricing algorithms. If the companies share data, they can better predict spikes in demand and optimize routes. However, if their services are nearly identical from the riders' perspective, a better algorithm for your competitor directly draws away your customers, a negative externality. The analysis shows that if the companies serve somewhat different segments or offer differentiated services, the negative competitive effect of sharing is muted, and collaboration is more likely to be profitable for both companies. Similarly, if predicting rider demand is very complex, the value of sharing is high, encouraging cooperation. These insights emphasize that the incentive to share AI training data is context-dependent: regulators or industry facilitators should note that firms in fiercely competitive, homogeneous markets are less likely to share data voluntarily, whereas firms with complementary strengths or in less direct rivalry have more to gain from collaboration.

Another study, by Bhargava et al. (Bhargava et al., 2024), explores strategic data sharing in a setting with interdependent markets, for example, between a large generalist tech firm and a smaller specialist firm operating in related markets. Their model reveals an intriguing strategic motive for data sharing: a smaller incumbent might share its proprietary data with a larger entrant for free to soften competition. By sharing data, the incumbent essentially makes the entrant more successful in the entrant's original market, which in turn can lead the entrant to compete less aggressively in the new market where they overlap. In other words, data sharing becomes a form of co-competition: the two firms cooperate in data to avoid a ruinous battle in the product market.

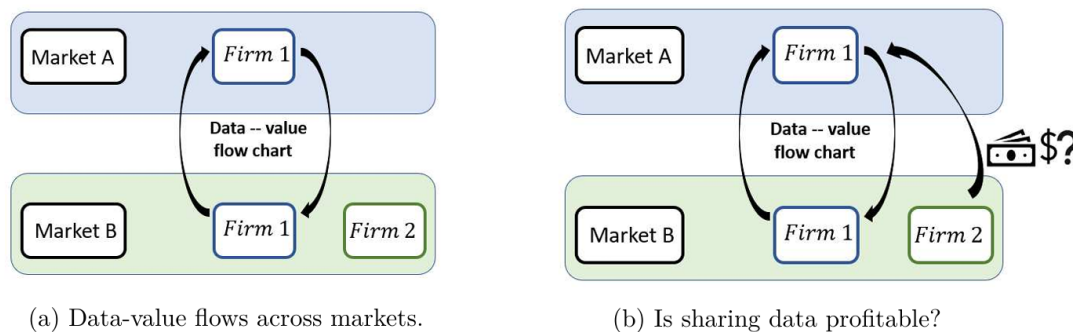


Figure 2. Cross-market data sharing between firms. (a) Internal use of data across markets. (b) Strategic sharing between firms with overlapping interests. Source: Bhargava et al. (2024)

This interaction is illustrated in Figure 2, which compares two scenarios: one in which Firm 1 internally transfers data across its markets, and another in which Firm 2 strategically shares its data with Firm 1. The diagram illustrates how data-flow decisions in one market can influence competitive behavior in another, raising the central question of whether data sharing is ultimately profitable across markets. Bhargava et al. show that after the larger Firm 1 enters Firm 2's market, Firm 2 may willingly share its data even absent direct compensation, because doing so induces Firm 1 to scale back its efforts in Firm 2's domain, allowing Firm 2 to maintain a stronger position. This counterintuitive result, that a competitor would provide valuable data to a rival, arises only when markets are interlinked through data-driven externalities. It highlights that data sharing can serve as a strategic commitment device to influence a rival's behavior.

While such sharing arrangements can increase the joint profits of the firms involved, they raise serious concerns about consumer welfare and innovation. In Bhargava et al.'s model, data sharing tends to weaken price competition and reduce incentives to invest in R&D, since firms become more complacent about their cooperative gains. Consumers may face higher prices or slower technological improvement as a result. The authors caution that, although data sharing enhances value creation, it can make consumers worse off if the reduction in competition outweighs the benefits. Notably, they find that these anti-competitive effects are salient in interrelated markets, whereas in completely unrelated markets, data sharing is mostly beneficial and can increase total welfare with little downside. This suggests a nuanced regulatory stance: encouraging data sharing between companies can spur innovation and efficiency, but in markets where the same firms compete, mandatory data sharing could inadvertently reduce competition. Antitrust authorities, therefore, must balance the pro-innovation aspects of data pooling with the potential for collusive outcomes or market foreclosure.

In summary, game-theoretic analyses of competitors illustrate that the incentives to share data are delicately balanced by competitive dynamics. Key findings are that firms are more inclined to share data when competition is less aggressive or when data substantially improve performance, and that firms sometimes strategically share data to reduce competitive pressure. However, what benefits firms may hurt consumers, so careful thought is needed when applying these insights to real-world data-sharing initiatives or regulations.

2.2 The Role of Platforms and Heterogeneity in Data Sharing

Beyond direct competitor-to-competitor sharing, many data-sharing arrangements are mediated by a third-party platform or data exchange. Recent research examines how the presence of a platform and differences in firm size or type affect the strategic game of data sharing.

Carballa-Smichowski et al. (Carballa-Smichowski et al., 2025) distinguish between two models a platform can adopt for enabling data collaboration: data sharing vs. analytics sharing. In a data-sharing model, the platform serves as an intermediary, allowing participating firms to access each other's raw data. In the analytics-sharing model, firms send their data to the platform, which then provides aggregate analytics or trained models to all participants, without revealing others' raw data. An intuitive real-world example is comparing Snowflake vs. Databricks: Snowflake's platform emphasizes direct data sharing and warehousing, whereas Databricks focuses on delivering analytics and machine learning results. The game-theoretic question here is: which model does a platform choose, and how do the firms' data contribution levels differ under each model?

According to Carballa-Smichowski et al., the platform's choice depends on the trade-off between encouraging data contributions and addressing privacy/security concerns. Analytics sharing generally offers a security advantage, since firms don't directly expose their raw data, and they may feel safer contributing more data. Indeed, the authors show that, under broad conditions, an analytics-sharing architecture yields higher total data contributions from firms than a raw data-sharing architecture. Intuitively, firms hold back less when they know only aggregated results will be shared, mitigating fears of competitive misuse or leaks. However,

providing only analytics can be less flexible or valuable than full data access. The platform thus faces a strategic decision: if the security/privacy advantage of analytics-sharing is sufficiently large, the platform will adopt that model; otherwise, it will offer data sharing to maximize the value of the collaboration. This result aligns with the notion that when trust is a major issue, a platform should act not just as a data broker but as a trusted analytics provider that shields participants' individual data.

The study also examines outcomes from the perspective of consumers or social welfare. Interestingly, what the platform prefers may not always align with what is best for the industry or consumers. For example, the platform might choose a data-sharing model because it can charge more for direct data access, even if an analytics-sharing model would have led to more data being contributed and higher collective benefit. There can be a misalignment: consumers might prefer the model that yields more overall data use, whereas the platform is guided by its own profit motive. Identifying such misalignments is important for policymakers: it suggests that policy or incentives may be needed to nudge platforms toward choices that maximize data collaboration and privacy rather than just their profit. Overall, the analysis provides a game-theoretic understanding of how platform design influences data-sharing equilibria, highlighting how data is shared can drastically alter firm behavior and the amount of data they are willing to share.

Another study, by Li and Mei (Li and Mei, 2024), examines participant heterogeneity by analyzing an evolutionary game involving three types of players: large enterprises, small-and medium-sized enterprises, and a data platform. In their model, a digital platform can empower SMEs and large firms to share data, and the strategies evolve over time as each type of player learns from past payoffs. The evolutionary game approach is well-suited to modelling scenarios in which populations of each player type repeatedly interact, and strategies that yield higher payoffs become more common over time. Li and Mei derive the evolutionarily stable strategies for the tripartite system and run simulations to observe how the propensity to share data varies across various factors.

Their findings reflect the inherent asymmetry in data sharing between large firms and SMEs. SMEs, having less data and resources, are generally more eager to share data to gain access to others' data and improve their AI capabilities. Large enterprises, in contrast, are more cautious; in fact, the study notes a U-shaped pattern for large firms' willingness to share, which might indicate that under some conditions, large firms share very little, unless certain incentives flip their strategy to sharing at high levels. The platform itself is a strategic player in this game: it can choose to actively facilitate sharing or not. According to Li and Mei, the platform is highly sensitive to cost factors; if the cost of providing data-sharing services is high or the platform's own benefits are low, it may not invest in enabling collaboration. However, when the platform does take initiative, it can significantly sway both large and small firms toward cooperation.

Importantly, the study examines policy levers. It finds that government subsidies can effectively encourage all parties to embrace data sharing. A subsidy reduces the cost or increases the reward of sharing data, making the cooperative strategy more attractive in evolutionary dynamics. Additionally, establishing a robust data security governance framework is crucial. This aligns with common sense; if firms trust that a platform can securely handle data, they are more likely to share. Without such trust, even generous incentives might not overcome fear of data misappropriation. The evolutionary model thus reinforces that trust and external support are key to sustaining a data-sharing ecosystem, especially between unequal partners. The heterogeneous analysis also suggests that SMEs might need additional support or guarantees, as they benefit greatly from access to data but lack the bargaining power to demand fair terms from large partners.

In summary, platform-mediated data-sharing games illustrate that the platform's structure and the diversity of participants significantly influence outcomes. Whether a platform shares raw data or only insights can change the equilibrium amount of data shared. Larger firms may need

different incentives than smaller ones to participate, and a platform's own strategy matters. Policy tools like subsidies and security standards can improve the stability of cooperative outcomes. Together, these works highlight that game theory can capture not just two-firm interactions, but also complex ecosystems of data sharing with multiple stakeholders.

2.3 Privacy and Differential Privacy in Data Sharing Games

When sharing data for AI, privacy loss and data security are paramount concerns, introducing another strategic dimension: participants want to maximize the utility of shared data while minimizing the leakage of sensitive information. Game theory has been applied to this privacy-utility trade-off, often incorporating techniques like differential privacy, which injects statistical noise to protect individual data points. A key question is: how much privacy should each data owner preserve, and can game theory find an equilibrium in these privacy choices?

Wu et al. (Wu et al., 2021) address this by formulating a non-cooperative game of data sharing with differential privacy. They consider multiple data owners, each training their own machine learning model, and also benefit from querying others' datasets. In a collaborative learning setup, each data owner can answer queries from others using a differentially private mechanism, for instance, by adding Laplace noise to the true answer, subject to a privacy budget ϵ . A higher ϵ means less noise but more privacy leakage. Each data owner's strategy in the game essentially amounts to choosing its privacy level (ϵ value) for the data it shares. Critically, Wu et al. allow for incentives through payments: there is a learner or aggregator who benefits from the overall model's accuracy and can compensate data owners for tolerating greater privacy loss. This setup can be thought of as a two-stage game or a market: the learner offers payments or terms, and data owners respond by choosing privacy levels.

The interaction is modeled as a dynamic game with imperfect information, but the authors prove that an equilibrium emerges. Specifically, they show that there exists a unique Nash equilibrium in privacy budgets where each data owner's choice of ϵ is the best response to the others' choices. At this equilibrium, the trade-off between model accuracy and privacy leakage is optimally balanced for all parties. Notably, the equilibrium is a mutual best response, meaning no data owner can deviate and improve their own payoff given the others' choices. This equilibrium also incorporates the effect of payments: the learner's compensation scheme is assumed to make it worthwhile for data owners to relax privacy to the equilibrium level. Wu et al. validate this model with a real-world financial dataset, showing that when multiple institutions adopt the game-theoretic privacy setting, all data owners benefit from improved models, and overall social welfare increases compared to not sharing data. In a concrete scenario, consider several banks collaborating to train a fraud detection model. By sharing transaction data with added noise, each bank improves its model's accuracy in detecting fraud patterns that it wouldn't catch with its own data alone. The game equilibrium might involve each bank using a moderate privacy budget, enough to significantly boost accuracy but still limiting exposure of individual customer information. The central aggregator might compensate the banks to mitigate their privacy risks, leading to a stable outcome in which all banks share data at the agreed privacy level and enjoy improved fraud detection.

The use of differential privacy in this game is crucial because it provides a quantifiable ϵ that links to both utility and privacy. Each player's payoff can be modeled as a function that increases with others' data accuracy but decreases with one's own privacy loss. The game-theoretical analysis by Wu et al. demonstrates how to optimally tune privacy parameters in a multi-party setting rather than leaving it to chance or overly conservative defaults. An important insight is that even independent, self-interested data owners can reach a stable sharing arrangement when appropriate incentives are in place, and the sharing mechanism is designed to account for their strategic behavior. The equilibrium here effectively solves a constrained optimization: maximizing learning performance subject to privacy constraints, in a decentralized way through each party's decisions.

It's worth noting that other works have also applied game theory to privacy, often using Stackelberg models or auction mechanisms for personal data. The common theme is that privacy preservation can be modeled as a cost in players' utility functions, and game theory helps in understanding how much privacy individuals or organizations will trade for better utility. In AI model training collaborations, differential privacy offers a rigorous way to measure this trade-off, and game-theoretic solutions indicate that carefully calibrated noise addition can enable some data sharing where previously fear of data leakage might have prevented any collaboration. In practice, these insights could guide the design of data-sharing platforms. For example, a platform could implement a scheme in which participants choose their privacy level and are compensated proportionally, knowing that the system will converge to an equilibrium in which everyone is better off than going solo.

In summary, privacy-aware data sharing games show that the "price" of privacy can be balanced against the value of data using game theory. Even with non-cooperative behavior, an equilibrium can be reached in which each data owner shares data at an optimal level of noise, especially if a mechanism exists to reward them for their contribution. This line of research merges economics, privacy, and machine learning to ensure that the collaborative benefits of data sharing can be realized without unduly sacrificing data owners' privacy interests.

2.4 Fairness and Cooperative Games in Multi-Party Data Sharing

While the previous sections mostly addressed whether data is shared, another critical question is how to distribute the benefits of data sharing fairly among those who contribute. This is where cooperative game theory comes to the forefront. In a cooperative scenario, we assume that the grand coalition of all data holders could form, meaning everyone shares data, and we then ask: how should the resultant value be divided so that each party is satisfied and willing to participate?

Chen (Chen, 2020) addresses this in the context of fair multi-party machine learning. He models a data sharing situation as a cooperative game where each player can join or not join a coalition, and the coalition's value is given by a characteristic function $v(S)$, for instance, the performance of a machine learning model trained on the combined data of coalition S . A larger coalition typically has more data and thus can train a more powerful model, so $v(S)$ is assumed to be non-decreasing with S . In such games, an important concept is the Shapley value, which provides a way to calculate each player's average marginal contribution to all possible coalitions. The Shapley value has appealing fairness properties and indeed has been proposed to value data contributions in collaborative analytics. However, Chen's analysis reveals a complication: a payout according to Shapley value might not be stable. Stability here means no subset of players would be better off forming their own coalition. In data sharing, instability could arise because once the grand coalition's value is created, certain players might threaten to leave or demand more if they feel the allocation is not sufficiently beneficial to them.

Chen introduces the notion of an optimal outcome in a modified cooperative game. An optimal outcome is defined as a coalition structure plus a payoff distribution such that two conditions hold: (1) stability, no individual or group of players can deviate and get a higher payoff on their own, and (2) fairness, each player's payoff is commensurate with their contribution as measured by the Shapley value. In essence, an optimal outcome gives every player at least as much as they could get on their own or in any sub-coalition, and splits the gains in proportion to their contributions. Chen's first key result is that if such an outcome exists for a given data-sharing game, it is unique and highly desirable; it implies that the grand coalition can form, will be self-enforcing, and will be fair.

However, achieving both stability and proportional fairness is not always possible. Chen shows that for some characteristic functions, no payoff vector can simultaneously satisfy core stability and align with Shapley values. In those cases, the game has no optimal outcome as defined. Instead, Chen examines suboptimal outcomes that relax one of the two conditions.

Specifically, he looks at: (a) stable but not proportional allocations, and (b) proportional but not fully stable allocations. A stable-but-not-proportional outcome might occur if one player is paid more than their Shapley value to bribe them to stay, whereas a proportional-but-unstable outcome means everyone is paid by contribution, but some coalition could block it. By analyzing these suboptimal cases, Chen provides insight into how a mediator might keep the coalition together. For example, if an exact fair split isn't stable, maybe a slightly adjusted split will keep everyone on board. Alternatively, if the only stable division is unfair in the sense of undervaluing some players, the mediator might need to find other ways to convince them to participate despite a lower payout.

A practical illustration could be a scenario in which three companies pool data to train a model. Suppose Company A's data is unique and valuable to the model, while B and C's data are more redundant. The Shapley value might assign a significantly higher portion of the model's value to A than to B or C, reflecting A's big contribution. But if B and C together could still build a decent model without A, they might not agree to let A take such a large share of the benefit; they could threaten to form a coalition of just themselves. If A's Shapley share is outside the core, the grand coalition is unstable. To stabilize it, A might need to accept a bit less than its Shapley value while B and C get a bit more, until neither subset has an incentive to secede. Chen's framework and algorithms help identify these situations and compute the allocations.

One important takeaway from this work is that the Shapley value should be seen as a benchmark for contribution, not necessarily as the literal payout. Using Shapley values blindly to pay participants could cause collaboration to fall apart if those payouts are not in the core. Instead, a mediator can use Shapley values as a starting point, then adjust the payoffs to ensure stability. These dynamics are especially important in sensitivity analysis, where the central agency evaluates how robust the current allocation is under small changes. These could include a player's data becoming partially unusable, their contribution increasing due to infrastructure support, or the entry of a new participant into the system. Figure 3 illustrates these two core cases: changes in an existing player's value, and the arrival of a new player. The challenge is to determine whether the optimal allocation remains valid or needs adjustment, particularly in scenarios where recalculating the entire Shapley value would be computationally expensive. Chen provides some theorems to guide these scenarios, ensuring an optimal outcome even as things change.

In summary, the cooperative game approach to data sharing underscores that it's not enough to get companies to share data; one must also ensure that the gains from sharing are allocated in a way that everyone remains committed. Game theory offers tools like the Shapley value to measure contributions and the core to test stability. Fairness without stability is fragile, and stability without perceived fairness may be unsustainable if participants feel they are not rewarded for their contributions. By designing reward-sharing schemes that adhere as closely as possible to both principles, one can facilitate long-term collaborations such as industry-wide data pools or federated learning consortia. The work of Chen (Chen, 2020) demonstrates how these concepts can be applied to multi-party AI training, yielding guidelines for a fair division of the AI dividend that results from collective data.

3 In-Depth Analysis of Representative Models

To detail the above insights, we now delve into two specific game-theoretic models from the literature, one non-cooperative and one cooperative, analyzing their structure and implications in detail. By examining these two in depth, we contrast how game theory tackles the data-sharing problem under very different assumptions of interaction.

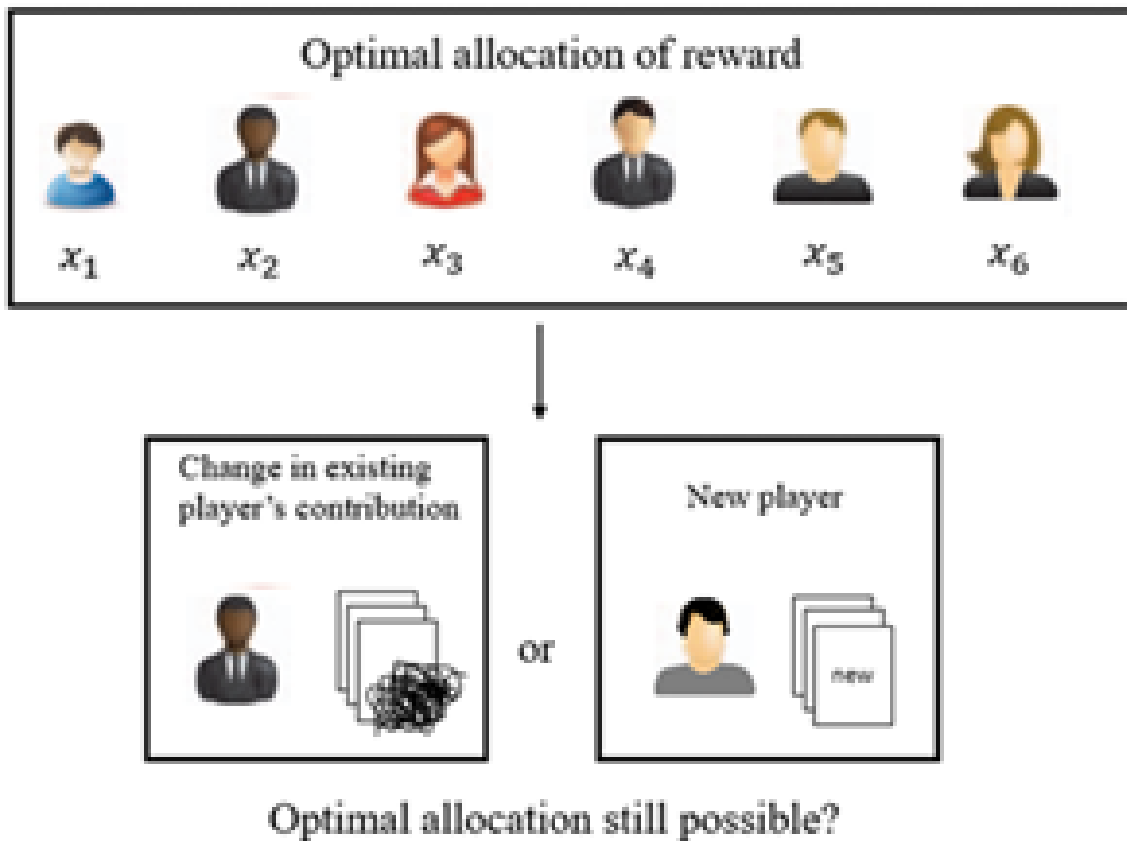


Figure 3. Sensitivity Scenarios in Collaborative Data Sharing — Contribution Shifts or New Players.
Source: Chen (2020)

3.1 Model 1: Strategic Data Sharing Between Competing Firms (Non-Cooperative)

Players and Strategies: The players are two companies that compete in the same product market. Each firm possesses a private dataset that it can use to train an AI model. The strategic decision in the data-sharing game is binary for each firm: Share data or do not share data. If both firms agree to share, they effectively pool their datasets, and each uses the combined data to train its model. If one or both refuse, each firm only uses its own data. Importantly, this data-sharing decision is tied to subsequent competition in the product market. After deciding on data sharing, firms compete, and their profits depend on their model's quality and the rival's actions. Thus, the complete strategy of a firm can be thought of in two stages: first, a binary choice on sharing, then a choice of competitive action. The game is primarily analyzed via backward induction: first, derive the market equilibrium given any data-sharing arrangement, then figure out each firm's incentive to share or not in the first stage.

Payoff Functions: Each firm's payoff is its profit in the product market. This profit is affected by whether data was shared because sharing improves the firm's machine learning model quality. In the instantiation studied by Tsoy and Konstantinov (Tsoy and Konstantinov, 2023), market competition is modeled as either Cournot or Bertrand.

The critical aspect of the payoff is that data sharing creates a positive externality: when Firm 1 shares its data, it directly benefits Firm 2's model. However, there is also a competitive externality in the profit functions: if Firm 2's model improves, Firm 2 might produce more or offer a lower price, thereby capturing business from Firm 1.

Solution Concept: The data-sharing decision can be seen as a 2x2 game. Tsoy and Konstantinov analyze it through the lens of the Nash equilibrium and via a coalitional agreement condition.

Essentially, they look for conditions under which (Share, Share) is a Nash equilibrium, meaning each firm prefers to share given the other firm's share, versus when (Not share, Not share) is the equilibrium. Because this is a one-shot decision, a Nash equilibrium in the first stage corresponds to a pair of decisions where neither firm can improve its final profit by unilaterally changing its sharing choice. An equivalent approach is to derive each firm's incentive to share. This is a strong requirement; both must strictly benefit. If one firm gains and the other loses by sharing, then sharing cannot be a stable outcome since the losing firm would veto the agreement.

The solution concept within the market stage is also a Nash equilibrium, which the authors solve for explicitly. They plug those equilibrium profits into the sharing incentive condition. In summary, the analysis finds a subgame-perfect equilibrium: in the second stage, firms always play the Nash equilibrium of the competition; anticipating that, in the first stage, they will agree to share data if it yields higher equilibrium profits for both, and otherwise they will not share. This use of the standard Nash equilibrium grounds the conclusions in classic game theory.

Key Findings and Implications: For this model, one can derive concrete conditions involving model parameters under which data sharing is mutually beneficial. Tsoy and Konstantinov identify two crucial factors: product substitutability and learning task complexity, which are captured in their data impact model by how quickly additional data diminishes in value.

One interesting nuance is that this model assumes symmetric firms for tractability; they have similar cost functions and data size, differing mainly in perhaps values or certain parameters. In more asymmetric cases, one firm might benefit more from sharing than the other does. The framework can handle that via the conditions. If one firm has much more data than the other, the smaller data holder benefits greatly from access to the larger pool, while the larger data holder gains only a small advantage but gives a big advantage to the smaller rival. That scenario likely results in no-sharing, as the larger firm's incentive is negative, a reflection of a free-rider problem: the lesser-endowed firm wants the other's data but cannot credibly compensate or assure the big firm that it won't just use it to compete. Game-theoretically, this could be seen as (Not share, Share) being an outcome one would like from the smaller firm's view, but that's not an equilibrium because the big firm would deviate. The requirement of mutual agreement blocks such one-sided exploitation.

From a broader perspective, this model underscores a limitation in spontaneous data sharing: even if collaboration is collectively beneficial, competitive imbalances can prevent it. It may require external intervention or clever contract design to facilitate sharing when it's socially efficient but not individually rational. For example, companies could form a data-sharing consortium with rules to ensure balanced benefits or agree to share only certain types of data that aid both but do not directly increase competition between them. The model also implies that companies in nascent AI-driven markets might start by sharing data when tasks are hard and products are differentiated, but as their products converge or the ML models mature, the incentive to continue sharing could diminish. We might see initial coalitions eventually break down if conditions shift unfavorably, a dynamic that only a repeated game extension could capture, but the static model gives the comparative statics hinting at that trend.

In conclusion, for this model, the strategy space was simple, the payoff functions integrated both cooperative gains and competitive losses, and the Nash equilibrium concept revealed the parameter regimes for which data sharing is a stable outcome. The solution concept was straightforward yet powerful in highlighting that corporate data sharing is fundamentally a game of trust and mutual verification of benefit; if both sides see a win, they'll do it; if not, competitive instincts prevail.

3.2 Model 2: Fair Multi-Party Data Sharing and the Shapley-Stability Solution (Cooperative)

Players and Setting: In this model, we have n players, which are organizations or data providers, collaborating to train a machine learning model. There is an implicit assumption of a central

coordinator, because we are in the realm of cooperative game theory. All n players together can produce a machine learning model of a certain quality or value. For each subset of players S , there is a well-defined value $v(S)$ representing the accuracy of the model trained on the combined data of the players in S , or perhaps the profit those S players could jointly earn with a model built on their data. Each player has some standalone value $v(i)$, and the grand coalition achieves the best performance $v(N)$.

The strategy in a cooperative game differs from that in a non-cooperative game: players decide which coalition to form. Here, it's natural to assume that the grand coalition N can form, but the main strategic issue is stability: will some sub-coalition S break off because they can achieve a better outcome for themselves by leaving N ? If the grand coalition is unstable, we might see a fragmented outcome, e.g., two separate groups of companies each pooling their data, or some firms going solo. However, in Chen's formulation, instead of enumerating coalition structures as strategies, he frames the problem as designing a payoff distribution, where x_i is the reward allocated to player i , such that x_i is in the core of the cooperative game and meets certain fairness criteria. In other words, rather than each player choosing a coalition per se, we assume all players are willing to join N if we can find a payoff vector that satisfies everyone. The strategy of a deviating subset is implicitly to leave and form their own coalition if not satisfied. So, stability constraints are any coalition S that would require not to defect.

Payoffs and Solution Concept: The payoffs we are looking for are the x_i 's, and the solution concept of interest is the stable and fair allocation, which Chen calls the optimal outcome if it exists. Fairness is interpreted as proportional to Shapley contributions. The Shapley value for each player is computed using the well-known formula, which averages marginal contributions over all permutations. The Shapley value is efficient, and it lies in the core for certain classes of games, but not always. It's known that the Shapley value need not lie in the core unless the game has certain properties. Chen's theoretical contribution is to derive conditions under which the Shapley value is core-stable, and if not, what near solutions exist.

When an optimal outcome does not exist, Chen's solution concept shifts to finding a suboptimal outcome that players might accept. He defines two classes: one in which the allocation is adjusted to be core-stable but not exactly proportional to Shapley, and one in which it's proportional but not stable. He seems to argue that if you had to choose, stability is the more crucial condition, because a stable but slightly unfair outcome can still hold together, whereas a fair but unstable outcome will collapse. So, a mediator might lean towards satisfying the stability constraint and only compromise on proportionality to the extent needed. In other words, maybe pay everyone roughly according to their contribution, but give bonuses to some players to ensure they don't defect. On the other hand, in some context fairness might be mandated, in which case one might enforce proportionality and hope that any instability is minor or can be managed by other means.

Implications for Data Sharing Incentives: The cooperative model does not determine whether data is shared, it assumes all will share if we design the right incentive scheme. Instead, it determines how we must reward participants to keep them sharing. The notion of no one having incentive to break away corresponds to each participant feeling that their payoff is at least as much as they could get on their own or in any partnership outside the grand coalition. This aligns with the idea of individual rationality and coalitional rationality. If these conditions are met, then sharing all data is an equilibrium.

One insight is that players with unique or critical data need to be rewarded proportionally more. The Shapley value naturally does this: if one company's data significantly boosts model accuracy, that company's Shapley value will be high. If one tried to split benefits equally, or by data size alone, that company might see that, by itself, it could produce a model almost as good as the coalition's and thus feel underpaid, causing instability. The Shapley-based pay ensures their opportunity cost of contributing data is covered. Chen's work further ensures that even groups of companies have no incentive to split off, which can be important if, say, a subset of firms' data is redundant with the others, except within their own group. It's a stringent

requirement but necessary for full stability.

This model also emphasizes the role of a central mediator or contract. In practice, implementing such an outcome might involve a revenue-sharing agreement or a system of side payments. For example, a consortium of hospitals might all pool patient data to develop a powerful diagnostic AI. An arrangement based on this cooperative solution might specify that, if AI is monetized, the resulting savings or revenues are allocated to each hospital in proportion to their data contributions. The agreement would also ensure that no subset of hospitals could benefit more by pulling their data out and doing their own thing. Perhaps smaller clinics alone can't achieve much, so they definitely stay. A big hospital that could almost develop the model on its own gets a large share, so it doesn't break away and go solo. The result is a stable grand coalition where all participants feel justified in having joined because they're getting out at least what they put in, relatively speaking.

This model also emphasizes the role of a central mediator or contract. In practice, implementing such an outcome might involve a revenue-sharing agreement or a system of side payments. For example, a consortium of hospitals might all pool patient data to develop a powerful diagnostic AI. An arrangement based on this cooperative solution might specify that, if AI is monetized, the resulting savings or revenues are allocated to each hospital in proportion to their data contributions. The agreement would also ensure that no subset of hospitals could benefit more by pulling their data out and doing their own thing. Perhaps smaller clinics alone can't achieve much, so they definitely stay. A big hospital that could almost develop the model on its own gets a large share, so it doesn't break away and go solo. The result is a stable grand coalition where all participants feel justified in having joined because they're getting out at least what they put in, relatively speaking.

In summary, this model formalizes the idea that fairness in data sharing can be quantified and used as a design principle, but it must be tempered by strategic stability considerations. The strategy space in a cooperative model is essentially who joins whom; the solution concept is not Nash equilibrium, but core stability and fair division. By solving those conditions, we derive guidelines for incentive design that ensure every firm is motivated to contribute data. The implication for data-sharing incentives is clear: if you want companies to pool data for AI, you must allocate the resulting benefits in line with their data's marginal contributions and ensure no subset of companies could profit more by withholding their data and perhaps collaborating separately. Game theory in this cooperative sense acts as a tool for designing contracts or mechanisms that make data sharing not just technically beneficial but also in each participant's self-interest.

4 A search for sources and the current state of the problem

The problem of strategic data sharing between firms for training artificial intelligence models lies at the confluence of several research domains: economics, game theory, data privacy, and machine learning. The literature in this area has been expanding rapidly, fueled by the recognition that data is both a strategic asset and a public good. From a global standpoint, collaborative AI development has become increasingly important across sectors such as healthcare, finance, and autonomous systems. Yet the lack of universally accepted incentives and privacy guarantees continues to stall large-scale inter-organizational collaboration.

Incentives for data sharing have traditionally been studied through cooperative models within economics and information systems literature. The classical approach involves modeling firms or agents who contribute to a collective dataset and analyzing how value generated from pooled data can be distributed fairly and stably. Game theory provides foundational tools to address such problems. The Shapley value and the core are among the most common solution concepts for ensuring fairness and coalition stability. Recent studies, such as Chen (Chen, 2020), apply these concepts to multi-party AI training consortia, where players need assurance that their data

contributions will not be undervalued or strategically exploited.

On the other side of the spectrum, non-cooperative game models address the strategic dilemma faced by direct competitors. Tsoy and Konstantinov (Tsoy and Konstantinov, 2023) analyze firms operating in a shared market that must decide whether to share data while competing either in prices or output quantities. Their model reveals how firms may choose not to collaborate, even when it would lead to mutually beneficial improvements in model quality, due to fears of strengthening a direct rival. This theoretical insight captures a core real-world tension observed in industries such as e-commerce, mobility, and logistics.

Another important contribution comes from Bhargava et al. (Bhargava et al., 2024), who explore cross-market data sharing between specialist firms and generalist technology conglomerates. Their work uncovers an unexpected strategic motive: specialists may share data voluntarily with powerful entrants to induce them to compete less aggressively. This concept of data sharing as a commitment device that alters a market entrant's behavior adds a new dimension to the literature on co-opetition.

Meanwhile, the structure of data-sharing platforms has also been scrutinized. Carballa-Smichowski et al. (Carballa-Smichowski et al., 2025) distinguish between raw-data-sharing models and analytics-sharing models, demonstrating that platform architecture significantly affects firms' willingness to participate. Analytics-sharing architectures, which offer insights without exposing raw data, typically result in greater total data contribution. These results have direct relevance for the design of secure, privacy-respecting data collaboration platforms in industry.

In parallel, the intersection of privacy and game theory has emerged as an active area of research. Wu et al. (Wu et al., 2021) present a differential privacy-based game-theoretic framework where multiple data owners decide how much privacy loss to tolerate in exchange for collaborative utility. They model this as a non-cooperative game over privacy budgets and prove that a unique Nash equilibrium exists under certain assumptions. This equilibrium balances the value of improved model performance with privacy risks, paving the way for incentive-compatible, privacy-aware data sharing in sensitive domains such as finance and healthcare.

Globally, governments and regulators are responding to the challenges posed by data silos. The European Union's GAIA-X initiative and its forthcoming Data Act aim to standardize and encourage responsible data sharing across borders and industries. However, regulatory efforts must still contend with firm-level strategic disincentives to share, especially in high-stakes competitive environments.

In sum, the current literature paints a nuanced picture: while data sharing can unlock immense value in AI, firms will not automatically do so. Their decisions are shaped by competitive threats, the fairness of reward distribution, data privacy risks, and the design of intermediary platforms. This article integrates these insights and extends them by offering a comparative game-theoretic analysis of two representative models and synthesizing the key variables that must be aligned to enable stable and fair data sharing.

5 The main findings of the article

The research presented in this article examines the use of game-theoretic models to understand and design data-sharing strategies among companies developing artificial intelligence systems. It focuses on both competitive and cooperative scenarios and analyzes the underlying conditions, constraints, and outcomes that determine whether and how data sharing occurs. We began by studying a non-cooperative setting where two firms compete in a shared market and must decide whether to share their proprietary data. The firms are rational agents seeking to maximize profit, and data sharing directly affects the quality of their machine learning models, which in turn influences production costs and demand. Their framework models this interaction using classical Cournot and Bertrand competition. The key insight is that even though pooled data could

improve both firms' model accuracy, the incentive to share depends critically on two factors:

- Market Competition Intensity (γ): If firms offer highly substitutable products, data sharing can overly strengthen a competitor, reducing one's own market share. Thus, firms in highly competitive markets are less likely to share.
- Learning Task Complexity (β): When tasks are complex, and model performance improves slowly with more data, the marginal benefit of additional data is high. This encourages firms to pool data to improve their models.

The result is that data sharing becomes a Nash equilibrium only when competition is low and task complexity is high. Otherwise, the equilibrium reverts to data siloing, despite potential collective benefits. The mathematical conditions for these equilibria were derived in full and included.

From a cooperative game-theoretic perspective, we analyzed a model that assumes that all firms can form a grand coalition and focuses on the allocation of benefits from joint data usage. Chen introduces the concept of an "optimal outcome" that must be both fair and stable. This ensures that no subset of firms has an incentive to deviate. However, not all valuation functions satisfy both criteria simultaneously. In such cases, compromise outcomes are needed:

- Stable but not proportional: Slightly adjust payoffs to prevent defection.
- Proportional but not stable: Maintain fairness but risk coalition fragmentation.

This analysis, including robustness checks and sensitivity to new entrants, illustrates the difficulty of maintaining long-term cooperation and provides guidance for dynamic contract design in AI consortia.

We also explored hybrid models, including platform-mediated sharing and differential privacy-based sharing. In platform-mediated settings, architecture has a first-order effect on data contribution levels. In privacy-aware games, data owners can be incentivized to reduce their privacy budget through compensation, leading to Nash equilibria that balance individual privacy risks with global model utility.

Overall, we identify five core variables that govern equilibrium behavior in data sharing games:

1. Product substitutability
2. Learning curve complexity
3. Relative data asymmetry
4. Platform design and trust mechanisms
5. Privacy sensitivity and compensation schemes

Each of these can be influenced by policy, technical design, or contractual arrangements. Our synthesis shows that sustainable data sharing requires aligning incentives across all these dimensions.

6 Conclusion

This article demonstrates how game-theoretic modeling provides a robust framework for analyzing and designing data-sharing strategies in multi-agent AI training contexts. The results make clear that data sharing is not merely a technical or infrastructural issue but a strategic and economic decision for firms. In non-cooperative games, the key insight is that collaboration will occur only when each firm's private benefit from improved models exceeds the competitive loss

of aiding a rival. This is more likely in differentiated markets and with hard learning tasks. In cooperative games, the challenge shifts to designing payoff allocations that are both fair and stable. Shapley-based distributions offer fairness but do not guarantee coalition stability unless they also lie within the core. Thus, practical consortia need to adjust these payoffs or provide side incentives to maintain cooperation. Platform-mediated sharing introduces architectural constraints and opportunities. Analytics-sharing platforms that offer privacy-preserving insights often result in higher participation and lower risk. Game theory helps analyze these designs, showing that even in decentralized, privacy-sensitive environments, equilibria can be reached with proper incentives. Platform-mediated sharing introduces architectural constraints and opportunities. Analytics-sharing platforms that offer privacy-preserving insights often result in higher participation and lower risk. Game theory helps analyze these designs, showing that even in decentralized, privacy-sensitive environments, equilibria can be reached with proper incentives. Despite these strengths, game-theoretic models require simplifications, such as rational actors and known payoff functions. Future work should explore repeated and dynamic games, integrate real-world uncertainty and bounded rationality, and connect theoretical payoffs with empirical data valuation techniques. In addition, mechanism design can play a greater role, crafting incentive-compatible contracts and protocols that embed fairness and stability guarantees. Finally, the policy implications are clear: enabling secure, fair, and strategic data sharing in AI requires coordination among industry players, regulators, and platform designers. Game theory offers a blueprint for aligning their incentives. If applied thoughtfully, it can unlock the full potential of shared data to improve AI systems for both business and societal benefit.

Resources

- Bhargava, Hemant, Aurélien Dubus, David Ronayne, and Shubho Shekhar (2024). “The Strategic Value of Data Sharing in Interdependent Markets”. In: *Management Science*. DOI: 10.1287/mnsc.2024.04938.
- Carballa-Smichowski, B., Y. Lefouili, A. Mantovani, and C. Reggiani (2025). *Data Sharing or Analytics Sharing?* Working paper / preprint. URL: https://www.researchgate.net/publication/375838048_Data_sharing_or_algorithm_sharing.
- Chen, Z. (2020). *Fair Multi-Party Machine Learning: A Game Theoretic Approach*. Preprint. URL: <https://www.semanticscholar.org/paper/Fair-Multi-party-Machine-Learning-a-Game-Theoretic-Chen/cb20d52c45326ac79ed81a546ac334d28ee08b0a>.
- Li, D. and X. Mei (2024). “Evolutionary Game Analysis of Data Sharing Among Large and Medium-Sized Enterprises in the Perspective of Platform Empowerment”. In: *Scientific Reports*. Online version / ResearchGate record. URL: https://www.researchgate.net/publication/380731270_Evolutionary_game_analysis_of_data_sharing_among_large_and_medium-sized_enterprises_in_the_perspective_of_platform_empowerment.
- Tsoy, N. and N. Konstantinov (2023). “Strategic Data Sharing Between Competitors”. In: *NeurIPS 2023 Workshop on Federated Learning and Analytics*. Workshop paper. URL: https://www.researchgate.net/publication/371041325_Strategic_Data_Sharing_between_Competers.
- Wu, Nan, David B. Smith, and Mohamed Ali Kâafar (2021). “Optimized Data Sharing with Differential Privacy: A Game-theoretic Approach”. In: URL: <https://api.semanticscholar.org/CorpusID:237497354>.

Generative Graphics for Scientific Visualization: A Conceptual Framework and Quantitative Evaluation

Peter Procházka¹

Abstract

Generative graphics represents a promising approach to scientific visualization that enables adaptive and automated creation of visual representations based on data structure and analytical context. In this paper, we propose a conceptual framework for generative graphics in the scientific visualization of complex and dynamic simulation data from the domains of fluid simulations, meteorology, and biological process modeling. The framework combines procedural and data-driven generative mechanisms with automatically generated explanatory elements in order to enhance the readability and interpretability of visual outputs. The work includes an initial quantitative evaluation conducted through a user study, in which generative visualizations are compared with traditional baseline approaches in terms of time efficiency and accuracy of data interpretation. The results indicate that generative graphics can lead to reduced task completion time while maintaining a comparable level of analytical accuracy. The study is exploratory in nature and serves as a starting point for further systematic research on generative approaches in scientific visualization.

Keywords

generative graphics, scientific visualization, procedural modeling, simulation data, quantitative evaluation, interpretability

1 Introduction

Scientific visualization represents a fundamental tool for the analysis, interpretation, and communication of data originating from simulations and measurements in the natural and engineering sciences. However, with the increasing accuracy of numerical models and the growing availability of computational resources, the complexity of generated data has also increased dramatically, placing new demands on methods of visual data processing. Traditional visualization approaches, which are often based on the manual design of static or only partially interactive graphical representations, encounter limitations in terms of scalability and adaptability when dealing with dynamic and high-dimensional data.

In response to these challenges, generative approaches have gained increasing attention. These approaches employ procedural rules, physical models, and artificial intelligence algorithms to enable the automated creation of visual representations. Such methods allow visualizations to react to data structure, temporal evolution, and the context of the analytical task, thereby creating opportunities for more effective communication of relevant phenomena and relationships. At the same time, the potential for visualization interpretability is strengthened, as generative mechanisms enable explicit mapping of properties of simulated processes to visual elements. Interpretability as a key objective of visualization design has been extensively discussed in the literature on visual analytics and visualization design (Munzner, 2014).

Despite the growing interest in generative graphics within scientific visualization, the extent to which these approaches provide measurable benefits compared to established visualization

¹Department of Applied Informatics, Faculty of Economic Informatics, Bratislava University of Economics and Business, Slovakia. Email: peter.prochazka@euba.sk

methods remains an open question. Existing work often focuses either on the conceptual design of generative techniques or on demonstrating visually compelling examples, while systematic quantitative evaluation of their impact on the efficiency of data interpretation is frequently limited.

The aim of this paper is to address this gap through the proposal of a conceptual framework for generative graphics in scientific visualization and its initial quantitative evaluation. We focus on comparing a generative approach with traditional visualization techniques in the context of selected types of simulation data, with particular emphasis on time efficiency of interpretation and the readability of visual outputs. The study focuses on an initial evaluation phase and does not aim to achieve comprehensive optimization of the proposed framework or to demonstrate its universal applicability.

The contributions of this paper can be summarized as follows:

- Formulation of a generic and modular framework for generative graphics in scientific visualization.
- Demonstration of its application to different types of simulation data.
- Empirical comparison with traditional visualization approaches through a quantitative user study.

The proposed approach reflects the growing need to balance visual expressiveness and analytical accuracy in contemporary scientific visualization. In this context, generative graphics does not serve as a replacement for physical or numerical models, but rather as an intermediary layer between data and the user, capable of adaptively highlighting relevant aspects of simulated phenomena. This understanding of visualization shifts the emphasis from the manual design of individual graphical outputs to the design of rules and mechanisms capable of producing consistent and semantically rich representations across diverse data scenarios (Ware, 2021).

From a practical perspective, this shift implies moving attention away from the optimization of individual visualization artifacts toward the design of processes that can generate visual outputs based on data structure, analytical tasks, and usage context. Such an approach is particularly relevant when working with simulation data, where the properties of observed phenomena vary across time and space and where manual adjustment of visualizations becomes unsustainable. Generative mechanisms enable a more systematic linkage between the data model and visual representation, thereby reducing the risk of ambiguous or misleading interpretations.

At the same time, it is important to emphasize that automation of the visualization process does not imply removing humans from the analytical loop. On the contrary, generative approaches create opportunities for more effective interaction between users and visualization systems by reducing routine design decisions and allowing users to focus on interpretation and evaluation of results. In this sense, generative graphics can be understood as a tool that supports analytical reasoning and decision-making rather than as an autonomous system that replaces expert judgment.

2 Related Work

Scientific visualization emerged as a distinct research field primarily in response to the need for systematic analysis of complex numerical simulations and experimental data. Fundamental principles of analytically oriented visualization, including the linkage between data, tasks, and visual representations, are elaborated in detail within the concept of visualization analysis, which systematizes relationships among data types, analytical goals, and visual encodings (Munzner, 2014). This approach established the methodological foundations for understanding visualization as an active component of the analytical process rather than merely a means of presenting results.

As data complexity has increased, research attention has also shifted toward the challenges of multidimensional and multivariate fields, where traditional graphical representations encounter limitations in perceptual interpretability. Survey studies highlight the need to combine multiple visualization techniques and interaction mechanisms in order to support users' analytical reasoning when working with heterogeneous data (Kehrer and Hauser, 2019). Particular attention in this context is required for data originating from fluid flow simulations and physical fields, which are characterized by pronounced spatial and temporal dynamics and impose high demands on the consistency of visual representation (McLoughlin et al., 2020).

Procedural and generative graphics originally emerged in the field of computer graphics as a response to the need for efficient generation of visually complex structures based on formal rules. Classical procedural approaches demonstrate that rich visual detail can be achieved without explicit manual modeling of each individual element, which has led to their widespread application in the modeling of natural and synthetic structures (Ebert et al., 2018). From the perspective of scientific visualization, this principle is particularly relevant in cases where global consistency of the visual output must be preserved while simultaneously depicting local variations.

The transfer of procedural principles into scientific visualization has been especially evident in the domain of fluid simulations, where a natural connection exists between the mathematical model and visual representation. Visualization techniques derived directly from physical quantities make it possible to preserve semantic coherence between data and image, which is important for interpretability and trustworthiness of results (Bridson, 2015). However, most of these approaches assume manual design of the visualization pipeline for a specific data type and provide only limited support for adaptive or automated generation of visual outputs.

In parallel with international research efforts, issues related to visualization and computer graphics have also been developed in the Central European region, particularly within academic institutions in Slovakia and Hungary. Work published in regional conferences and proceedings points to a long-standing interest in scientific data visualization, GPU acceleration, and procedural approaches to rendering complex structures (Skala et al., 2016). In the Hungarian academic environment, research has focused primarily on visual analysis and processing of image and spatial data, with an emphasis on linking algorithmic methods and visual interpretation (Sziranyi, 2019).

In recent years, methods employing machine learning to support or automate visualization design have gained increasing prominence. Research on assisted visualization design shows that algorithmic support can reduce user burden in selecting appropriate visual encodings and parameters, thereby increasing the efficiency of the analytical process (Chen et al., 2021). At the same time, contemporary surveys note that deep neural networks are often used as generators of visual outputs without explicit grounding in data semantics, which may result in limited transparency and control over the outcome (Liu et al., 2022).

The question of interpretability and methodological rigor plays a key role in the context of generative approaches. Design-oriented methodologies in visualization emphasize the need for a clear connection between design decisions, domain context, and empirical evaluation in order to assess the actual contribution of proposed solutions (Sedlmair et al., 2020). Recent work also seeks to systematize criteria for evaluating the quality and rigor of visualization studies, thereby establishing a framework for comparability of results across different approaches (Zhang et al., 2023).

These research directions point to the growing potential of generative techniques in scientific visualization, while simultaneously revealing a lack of frameworks that systematically connect these techniques with quantitative evaluation of their benefits for users. It is precisely this gap that our approach addresses by combining generative graphics with empirical evaluation of visualization effectiveness.

Particular attention in the literature is also devoted to the relationship between the realism of visual representation and its analytical value. Several studies indicate that a higher degree of

visual realism does not automatically lead to better data understanding if it is not accompanied by appropriate visual encoding and explanatory elements. Research in perception and cognitive aspects of visualization emphasizes the need to balance aesthetic qualities of visualizations with their ability to support accurate and efficient analytical judgments (Ware, 2021).

In the context of simulation data, this issue is especially evident in the visualization of dynamic phenomena, where it is necessary to convey not only the instantaneous state of the system but also its temporal evolution and causal relationships among individual variables. Traditional approaches often address this challenge through sequences of static snapshots or limited interaction, which can lead to fragmented perception of the overall process. Generative approaches open possibilities for integrating the temporal dimension directly into visual representation through adaptive rules and dynamic visual layers.

Another relevant line of research concerns the automatic generation of explanatory elements aimed at reducing users' cognitive load when interpreting complex visualizations. Studies in visual analytics show that explicit highlighting of significant patterns, anomalies, or changes in data can substantially influence the speed and accuracy of analytical tasks, particularly for less experienced users (Endert et al., 2021). Nevertheless, the question of systematic integration of these mechanisms into generative visualization frameworks remains open.

From the perspective of evaluation methodology, it is important to emphasize that empirical studies in visualization have traditionally faced challenges related to limited reproducibility and variability of experimental conditions. Analyses of existing empirical work point to the need for more clearly defined experimental scenarios and metrics that would enable comparison of results across different studies (Lam et al., 2018). This aspect is particularly relevant when evaluating generative approaches, where the very nature of the visualization process is subject to change.

3 Proposed Framework

The proposed framework for generative graphics in scientific visualization is conceived as a modular system whose objective is to link simulated or measured data with adaptive visual representation through explicitly defined generative mechanisms. The framework consists of three main functional layers: a data layer, a generative modeling layer, and a generative visualization layer, each of which fulfills a distinct role in the process of transforming data into a visually interpretable output. The user interface forms an auxiliary interaction layer that mediates task specification and feedback. The overall architecture of the framework and the relationships among individual layers are schematically illustrated in Fig. 1.

Source: Authors' own work

The data layer comprises input data originating from numerical simulations or experimental measurements. These data may have varying spatial and temporal dimensionality, ranging from two-dimensional fields to three-dimensional time-dependent volumes. The framework assumes that input data are supplemented with basic metadata, such as the type of physical quantity, value range, or temporal resolution, which are subsequently utilized within the generative layer to control visual parameters. Such metadata enrichment enables a unified approach to heterogeneous data sources and reduces the need for ad hoc adjustments of the visualization process for each new scenario. The generative modeling layer represents the core of the proposed framework. Its role is to transform input data into an abstract visual representation through procedural rules and data-driven parameters. These rules may be deterministic or stochastic in nature and are designed to preserve semantic correspondence between the data and the visual output. In the case of simulation data, particular emphasis is placed on maintaining physical consistency, for example by deriving visual elements directly from physical quantities or their derivatives. This approach enables the creation of visualizations that are not only visually compelling but also support users' analytical reasoning. Generative mechanisms within the framework can be categorized into several classes according to the manner in which data are transformed.

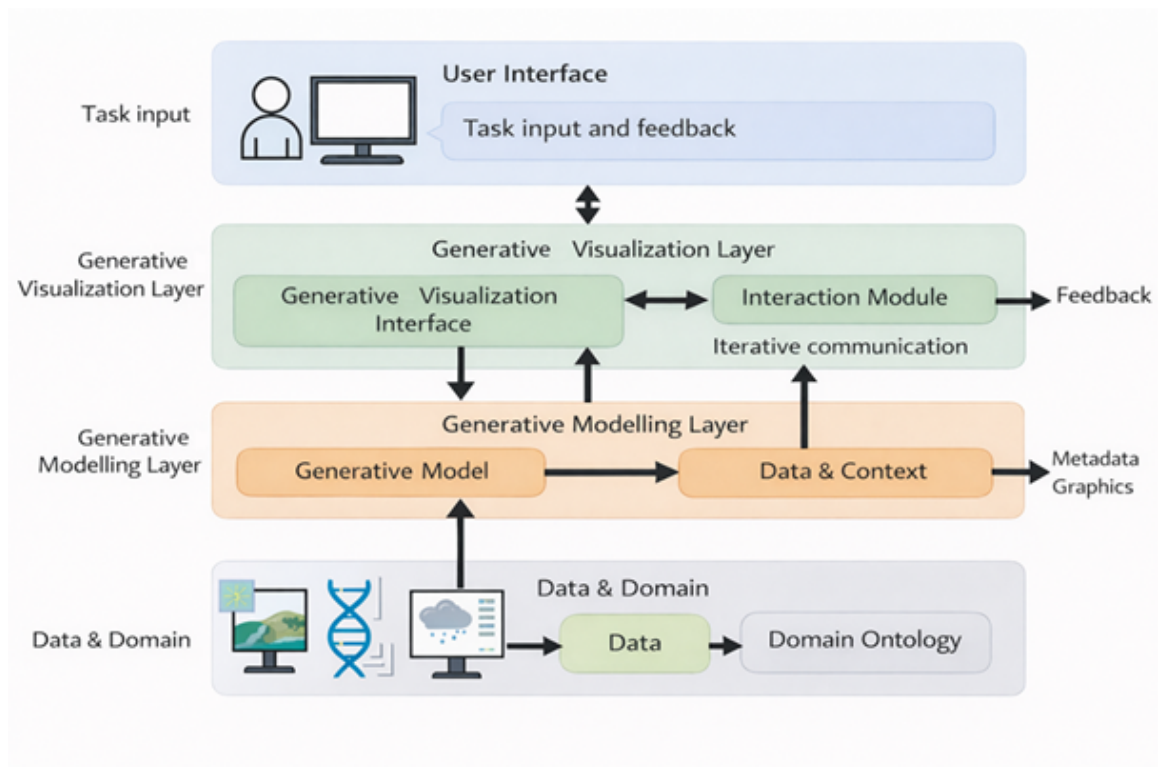


Figure 1. Schematic overview of the architecture of the proposed generative visualization framework, including data flows and relationships among individual layers.

Procedural rules based on explicit algorithms allow precise control over visual output and are suitable for situations where a high degree of interpretability is required. In contrast, stochastic or data-driven mechanisms provide greater variability and adaptability, which is advantageous when working with complex or noise-affected data. The framework is designed to allow these approaches to be combined depending on the analytical context. The generative visualization layer is responsible for the actual visual rendering of the results produced by the generative process. This layer encompasses the selection of appropriate visual encodings—such as color maps, vector fields, particle systems, or isosurfaces—and their integration into a coherent visual whole. An important aspect of this layer is support for interpretability through automatically generated explanatory elements, including adaptive legends, highlighting of salient regions, and annotations derived from analytical criteria. An example of such output is shown in Fig. 2.

Source: Authors' own work

The modular nature of the framework enables its application across different application domains without requiring fundamental modifications to the system core. Individual modules can be configured or replaced depending on the type of data and the analytical task, thereby supporting flexibility and reusability of the proposed solution. This aspect is particularly important in interdisciplinary research environments, where not only data structures but also analytical objectives and user requirements may vary. The framework is also designed with consideration for integrating the user into the analytical process. Generative mechanisms do not function as an autonomous system but rather as a tool that supports user interaction and decision-making. Parameters of generative rules can be adjusted according to the analytical context, allowing the visualization to be dynamically adapted to current needs without requiring manual intervention in the design of visual elements. This approach supports the iterative nature of scientific analysis and enables rapid experimentation with different visualization strategies. From an implementation perspective, the framework supports the use of modern computational resources, particularly parallel processing on graphics processing units. Procedural generation

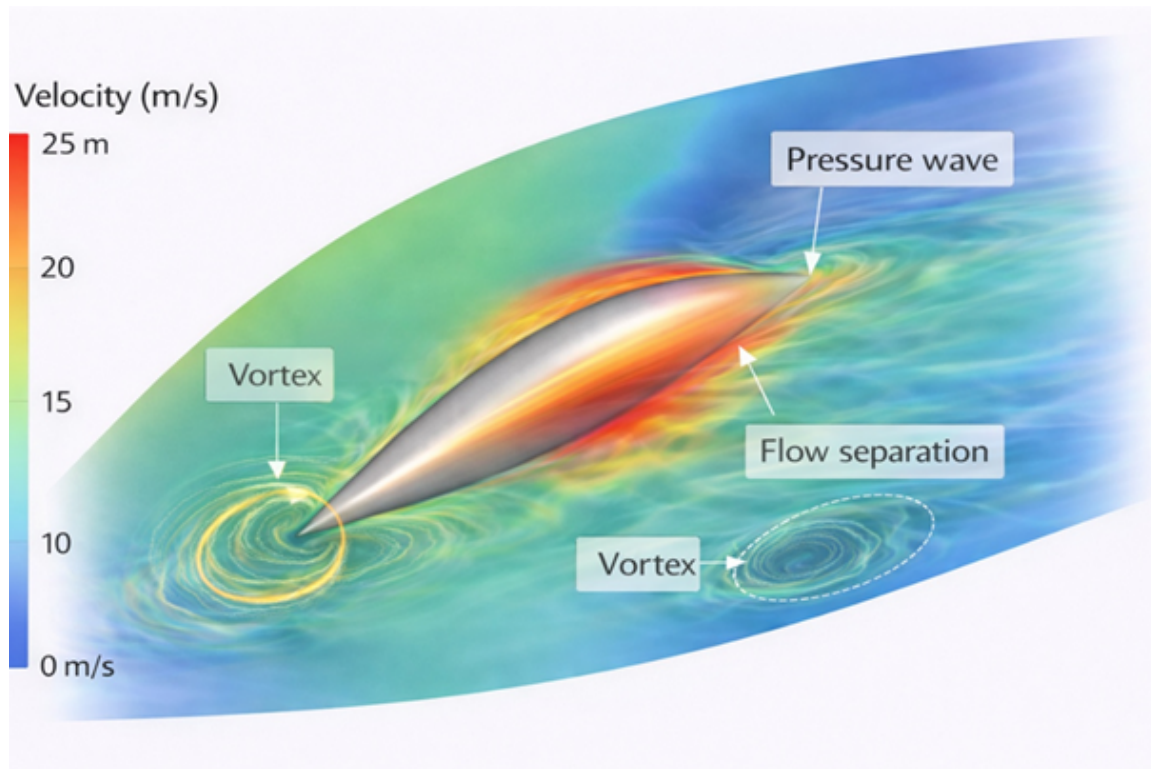


Figure 2. Example of a generative visualization of simulation data with automatically generated explanatory elements.

of visual elements is well suited for GPU acceleration, enabling interactive response even when working with large data sets. This property is crucial for the practical deployment of generative approaches in real-world scientific workflows, where computational efficiency must be combined with flexibility of the visualization process. An important aspect of the proposed framework is its ability to operate across multiple data scales, both spatially and temporally. Scientific simulations often generate data encompassing phenomena occurring at markedly different scales, ranging from local fluctuations to global structures. Generative mechanisms allow these differences to be explicitly addressed through multilevel visual representations, in which different generative rules are applied at different levels of detail. Such an approach supports simultaneous perception of global trends and local anomalies without requiring switching between separate visualization modes. The framework is further designed to support extensibility with new types of generative mechanisms without modifying the existing architecture. New procedural or data-driven modules can be integrated through a defined interface that ensures consistent transfer of data and parameters among system layers. This property is important for the long-term sustainability of the framework, as it enables gradual extension in line with the development of new simulation models and analytical requirements. In the context of practical use, it is also necessary to emphasize the relationship between generative visualization and reproducibility of scientific results. Since visual outputs are generated based on explicit rules and parameters, these settings can be precisely documented and repeatedly applied to identical or comparable data sets. This creates conditions for improved transparency of the visualization process and reduces the risk of subjective interventions that could influence data interpretation. In this sense, generative graphics supports the principles of open and reproducible science. Finally, the proposed framework creates a natural space for integrating automatically generated explanatory layers that can reflect not only structural properties of the data but also their dynamic temporal evolution. Generation of these layers is guided by analytical criteria and can be adapted to specific task types, thereby strengthening the ability of visualization to actively support the analytical process rather than

merely passively display data.

4 Experimental Setup

The experimental evaluation was designed to compare the proposed generative visualization framework with traditional visualization approaches commonly used in scientific data analysis practice. The chosen experimental design focuses on an initial quantitative assessment of the contribution of generative graphics in terms of data interpretation efficiency and readability of visual representations.

As a reference baseline, static 2D and 3D visualizations manually created according to commonly used practices in the respective application domains were employed, implemented using the freely available visualization software ParaView. These visualizations utilized conventional graphical elements such as color maps, isosurfaces, and vector fields, without adaptive or automatically generated explanatory layers. The goal was not to optimize the baseline solutions, but to represent a realistic state of common visualization practice.

The experiments were conducted on three types of data representing different characteristics of scientific simulations. The first group consisted of fluid flow simulation data, including time-dependent velocity vector fields and pressure distributions. The second group comprised meteorological data in the form of spatiotemporal fields describing the evolution of atmospheric variables. The third group consisted of synthetic biological data modeling substance diffusion and structural growth in space. The selection of these domains enables assessment of the framework's behavior when working with different types of data structures. An overview of the employed data sets, their dimensions, and basic characteristics is provided in Table 1.

Table 1. Overview of the employed data sets, their dimensions, and basic characteristics.

Data domain	Data type	Spatial resolution	Time steps	Number variables	of Basic characteristics
Fluid simulations (CFD)	Vector field	$256 \times 256 \times 128$	120	4 (velocity x,y,z; pressure)	Turbulent flow around a solid object
Meteorological data	Scalar fields	512×512	96	3 (temperature, pressure, precipitation)	Regional weather evolution simulation
Biological data	Scalar/discrete fields	256×256	150	2 (concentration, growth)	Model of substance diffusion and structural growth

Source: Authors' own work

The user study involved 18 participants with technical or natural science backgrounds. Most participants had prior experience working with scientific data, although only a subset had deeper expertise in visualization. This selection reflects the target group of potential users of generative visualization tools, who combine domain knowledge with analytical tasks.

The tasks were designed to represent typical analytical scenarios encountered when working with simulation data. They included identification of dominant patterns, detection of extreme values, and interpretation of dynamic changes over time. Tasks were formulated in a neutral manner and did not favor any specific visualization approach. Each participant solved tasks using both the generative and baseline visualizations, with the order of presentation randomly permuted to minimize learning effects.

The experiments were conducted in a controlled environment with uniform hardware and software configurations. Visualizations were presented on the same display device, and interaction

capabilities were unified to a basic level across the compared approaches in order to isolate the effect of the visualization principle itself. Task completion time was recorded automatically, while response accuracy was evaluated based on predefined criteria.

The evaluation metrics were selected with regard to their common use in empirical visualization studies. Task completion time served as a proxy indicator of cognitive load and interaction efficiency with the visualization system, while response accuracy reflected the ability of the visualization to support correct data interpretation. In addition to objective metrics, subjective assessments of readability and comprehensibility of the visualizations were collected after completion of the experiment using a short Likert-scale questionnaire.

The statistical analysis was designed to account for the paired nature of the experimental design, as each participant worked with both visualization approaches. This approach reduces inter-individual variability and increases the sensitivity of statistical tests to differences between the compared conditions. Prior to applying the tests, basic assumptions regarding data distribution were verified, and nonparametric alternatives were used in cases where these assumptions were violated.

From an ethical perspective, participants were informed about the purpose of the study and the handling of anonymized data. Participation was voluntary, and the experiment did not include tasks that could lead to excessive cognitive or time burden. This aspect is particularly important when planning more extensive user studies, which constitute the subsequent phase of the research.

5 Quantitative Evaluation

The quantitative evaluation focused on a systematic comparison of the proposed generative visualization approach with reference baseline visualizations based on the metrics defined in the experimental setup. The objective was not to achieve maximal performance optimization of the proposed framework, but to identify whether generative graphics provides a measurable benefit in terms of data interpretation efficiency while preserving analytical accuracy in the initial phase of the research.

The primary quantitative metric was the time required to complete analytical tasks. This metric was selected as an indirect indicator of cognitive load and efficiency of interaction with the visualization system. For each participant, task completion time was recorded for both the generative and baseline visualizations, and mean values and standard deviations were subsequently computed for individual experimental scenarios.

The secondary metric was response accuracy, evaluated based on agreement with reference results derived from the simulation data. This metric served to verify that potential reductions in task completion time were not achieved at the expense of analytical correctness. Accuracy was expressed as the proportion of correct responses for individual tasks and subsequently aggregated at the level of experimental scenarios.

Paired statistical tests were used to compare results between the generative and baseline approaches, accounting for repeated measurements on the same participants. In cases where the data satisfied assumptions of normality, a paired t-test was applied, while nonparametric alternatives were used when these assumptions were violated. The level of statistical significance was set to $\alpha = 0.05$ for all conducted tests. Mean task completion times and response accuracies for the generative and baseline approaches across individual experimental scenarios are reported in Table 2.

Table 2. Dataset overview

Data main	do-	Data type	Spatial resolution	Time steps	Number variables	of	Basic characteristics
Fluid simulations (CFD)		Vector field	256×256×128	120	4 (velocity x,y,z; pressure)		Turbulent flow around a solid object
Meteorological data		Scalar fields	512 × 512	96	3 (temperature, pressure, precipitation)		Regional weather evolution simulation
Biological data		Scalar/discrete fields	256 × 256	150	2 (concentration, growth)		Model of substance diffusion and structural growth

Source: Authors' own work

The results indicate that the generative approach led to a statistically significant reduction in the time required for data interpretation in two of the three experimental scenarios. The most pronounced difference was observed in tasks involving dynamic temporal changes, where generative visualizations enabled faster identification of relevant patterns. In the third scenario, differences in task completion time were not statistically significant, suggesting that the benefit of the generative approach may depend on the characteristics of the data and the type of analytical task.

Response accuracy was comparable between the generative and baseline approaches across all experimental scenarios. The generative approach exhibited slightly higher mean accuracy values in all cases, although these differences did not reach statistical significance. This result indicates that the reduction in task completion time was not achieved at the expense of analytical correctness, which is critical for the practical applicability of generative visualization techniques.

In addition to analysis of mean values, an exploratory analysis of performance variability among individual participants was conducted. This analysis indicated that the benefit of the generative approach was more pronounced for participants with lower levels of prior experience with visualization tools, whereas differences were less evident among more expert users. This trend suggests that generative visualizations may play a particularly important role in supporting analytical tasks for users who lack extensive experience with manual adjustment of visualization parameters.

Another aspect of the quantitative evaluation involved assessing the consistency of results across different data domains. Although absolute task completion times varied among domains, relative comparisons between the generative and baseline approaches exhibited similar trends. This suggests that the observed effect is not tied to a specific data type but is instead related to the underlying principle of adaptive generation of visual representations.

Finally, correlations between subjective assessments of visualization readability and objective performance metrics were analyzed. The results indicate a moderate negative correlation between subjectively perceived comprehensibility of the visualization and task completion time, supporting the assumption that better perceived readability may contribute to more efficient analytical work. These findings should be interpreted cautiously, however, given the limited sample size and the exploratory nature of the analysis.

6 Results

The results of the quantitative evaluation provide a comprehensive view of the behavior of the generative visualization approach in comparison with the reference baseline visualizations. Building on the preceding chapter, this section focuses on synthesizing the measured metrics and identifying the main trends observed across the experimental scenarios.

In terms of time efficiency, the generative approach exhibited consistent improvements for tasks that required interpretation of dynamic and spatially complex phenomena. Participants were able to identify relevant patterns and changes in the data more quickly, which was particularly evident in scenarios with a pronounced temporal component. This effect was less pronounced for more static tasks, where traditional visualizations provided sufficient support for analytical decision-making.

The preservation of analytical accuracy across experiments confirms that the increased efficiency of generative visualizations was not achieved at the expense of interpretation quality. The absence of statistically significant differences in response accuracy indicates that generative mechanisms can support faster orientation within the data without introducing systematic errors. This result is particularly important in the context of scientific analysis, where reliability of interpretation is a key criterion.

When comparing results across individual data domains, the relative benefits of the generative approach were found to be similar across fluid simulations, meteorological data, and biological models. Although absolute performance values differed among domains, the direction and magnitude of the effect remained consistent, suggesting a degree of generalizability of the proposed framework.

Subjective evaluations provided by participants complement the quantitative metrics and offer additional context for interpreting the results. Most participants reported that they perceived generative visualizations as clearer and more supportive of understanding relationships within the data, particularly when tracking the temporal evolution of phenomena. These assessments correspond with the observed differences in task completion time and suggest that perceived readability may play a significant role in improving analytical efficiency.

Taken as a whole, the results indicate that generative graphics represent a viable alternative to traditional visualization approaches in situations where dynamic and high-dimensional data must be analyzed. At the same time, the findings demonstrate that its benefits are not universal and depend on the nature of the analytical task and the data context, providing an important foundation for further discussion and future research.

7 Discussion and Limitations

The results presented in the preceding chapters indicate that generative graphics has the potential to improve efficiency when working with complex scientific data, particularly in situations where temporal dynamics and data multidimensionality play a significant role. The observed reduction in data interpretation time while maintaining analytical accuracy supports the hypothesis that adaptively generated visual representations can reduce users' cognitive load and facilitate faster analytical decision-making.

At the same time, it must be emphasized that the benefits of generative approaches are not universal and depend strongly on the nature of the analytical task and the data context. For simpler or predominantly static tasks, differences between generative and traditional visualizations were found to be less pronounced, suggesting that the use of generative graphics should be motivated by specific analytical needs rather than applied indiscriminately.

Among the main limitations of the presented study are the limited number of participants and the relatively narrow range of experimental scenarios. Although the chosen experimental design made it possible to identify basic trends, the sample size does not allow strong conclusions to be drawn regarding the generalizability of the results across a broader user population. This aspect is typical of early-stage research and points to the need for more extensive user studies in subsequent phases.

Another limitation lies in the simplified implementation of generative mechanisms, which was deliberately chosen in order to maintain control over experimental conditions. The proposed framework does not yet exploit the full potential of advanced generative models, such as deep

neural networks or hybrid physics–data-driven approaches, which may influence the achieved level of visualization adaptivity. Future extensions of the framework should examine to what extent these techniques can further enhance the benefits of generative graphics without negatively affecting interpretability.

From a methodological perspective, attention must also be paid to the issue of subjectivity in visual perception. While quantitative metrics provide an objective view of user performance, subjective preferences and prior experience may influence how individuals interact with visualization systems. A combination of objective and subjective evaluations therefore appears necessary for a comprehensive assessment of the quality of generative visualizations.

Finally, it is important to emphasize that generative graphics does not represent a replacement for traditional visualization techniques, but rather their complement. In many cases, the most effective solution may be a hybrid approach that combines the stability and predictability of classical visualizations with the adaptivity of generative mechanisms. This perspective provides a foundation for future research aimed at integrating generative approaches into existing scientific workflows.

From a practical standpoint, the controllability of generative systems from the user’s perspective must also be discussed. While automation of the visualization process can substantially reduce the burden associated with manual parameter tuning, it may simultaneously lead to a loss of control over the resulting visual representation. This aspect is particularly relevant in situations where visualization serves as input for critical decision-making. Future research should therefore focus on the design of interfaces that enable transparent control of generative mechanisms without undermining their adaptive character.

Another open question concerns the relationship between the complexity of generative rules and users’ cognitive load. While more complex generative models may produce visually richer representations, there is a risk that excessive visual complexity may reduce readability and interpretability. Identifying an optimal balance between expressiveness and simplicity of visual representation constitutes a significant research challenge.

From an evaluation perspective, limitations of the employed metrics must also be acknowledged. Task completion time and response accuracy provide important but incomplete information about the quality of a visualization system. Other aspects, such as long-term learnability, support for hypothesis generation, or integration into collaborative workflows, were not systematically evaluated in this study. Expanding the set of evaluation criteria represents another natural step in the continuation of the research.

Finally, broader epistemological implications of employing generative graphics in scientific practice must be considered. Automated visual representations may influence how users perceive data structure and formulate analytical conclusions. Awareness of these influences and their systematic examination are essential for the responsible use of generative techniques in the context of scientific inquiry.

8 Conclusion

This paper focused on exploring the potential of generative graphics in the context of scientific visualization, with particular emphasis on its ability to support the interpretation of complex and dynamic data. The presented conceptual framework combines procedural and data-driven generative mechanisms with traditional visualization principles, establishing a foundation for the adaptive creation of visual representations that reflect both data structure and the context of the analytical task.

The quantitative evaluation conducted as part of an initial user study suggests that generative visualizations can lead to increased time efficiency of analytical tasks while maintaining a comparable level of accuracy relative to traditional visualization approaches. These results support the assumption that adaptively generated visual elements and automatically derived

explanatory layers can reduce users' cognitive load and facilitate orientation within the data, particularly when working with time-dependent and high-dimensional simulation outputs.

At the same time, it must be emphasized that the presented results are exploratory in nature and do not constitute a definitive evaluation of the proposed framework. The goal of this work was not to optimize a specific implementation of generative visualization, but rather to identify fundamental trends and to assess whether generative graphics represents a promising direction for further research in scientific visualization. In this sense, the paper serves as a starting point for more systematic and extensive studies.

The proposed framework indicates that generative graphics can function as an intermediary layer between data and the user, enabling more flexible responses to changing analytical requirements without the need for manual redesign of visualization layouts. This approach has the potential to improve reproducibility of visualization practices and to enhance transparency of the analytical process through explicitly defined generative rules and parameters.

Future research should focus on several key directions. First, more extensive user studies with a more diverse participant sample and a broader range of analytical tasks are required to enable more robust assessment of the generalizability of the observed effects. Another important direction involves the integration of more advanced generative models, including deep neural networks and hybrid physics–data-driven approaches, while preserving interpretability of visual outputs.

Particular attention should also be devoted to the study of interaction mechanisms that allow users to actively influence the generative process without sacrificing clarity or control. Finally, systematic investigation of the epistemological implications of automated visualization remains an open challenge, particularly with regard to how generative visual representations influence hypothesis formation and scientific reasoning.

In conclusion, generative graphics represents a promising yet still only partially explored approach to scientific visualization. The combination of a conceptual framework and an initial quantitative evaluation presented in this paper provides a solid foundation for further research that may contribute to the systematic and responsible use of generative techniques in modern scientific workflows.

References

- Bridson, Robert (2015). *Fluid Simulation for Computer Graphics*. 2nd ed. CRC Press.
- Chen, Min, Shixia Liu, Andreas Stadler, and Michael Sedlmair (2021). "AI-Assisted Visualization Design". In: *IEEE Computer Graphics and Applications* 41.1, pp. 12–20. DOI: 10.1109/MCG.2020.3045399.
- Ebert, David S., F. Kenton Musgrave, Darwyn Peachey, Ken Perlin, and Steven Worley (2018). *Texturing and Modeling: A Procedural Approach*. 3rd ed. Morgan Kaufmann.
- Endert, Alex, William Ribarsky, Cagatay Turkay, Brian W. Wong, Ian Nabney, Isabel D. Blanco, and Francesca Rossi (2021). "The Human is the Loop: New Directions for Visual Analytics". In: *Journal of Intelligent Information Systems* 57.1, pp. 1–25. DOI: 10.1007/s10844-020-00624-1.
- Kehrer, Johannes and Helwig Hauser (2019). "Visualization and Visual Analysis of Multifaceted Scientific Data: A Survey". In: *IEEE Transactions on Visualization and Computer Graphics* 25.1, pp. 495–513. DOI: 10.1109/TVCG.2018.2864901.
- Lam, Heidi, Enrico Bertini, Petra Isenberg, Catherine Plaisant, and Sheelagh Carpendale (2018). "Empirical Studies in Information Visualization: Seven Scenarios". In: *IEEE Transactions on Visualization and Computer Graphics* 24.1, pp. 152–162. DOI: 10.1109/TVCG.2017.2745260.
- Liu, Shixia, Weiwei Cui, Yingcai Wu, and Mengchen Liu (2022). "A Survey on Deep Learning for Visualization". In: *IEEE Transactions on Visualization and Computer Graphics* 28.1, pp. 108–130. DOI: 10.1109/TVCG.2021.3114873.

- McLoughlin, Tom, Robert S. Laramee, Ronald Peikert, Frits H. Post, and Min Chen (2020). “Over Two Decades of Integration-Based, Geometric Flow Visualization”. In: *IEEE Computer Graphics and Applications* 40.6, pp. 106–119. DOI: 10.1109/MCG.2020.3024177.
- Munzner, Tamara (2014). *Visualization Analysis and Design*. CRC Press.
- Sedlmair, Michael, Miriah Meyer, and Tamara Munzner (2020). “Design Study Methodology: Reflections from the Trenches and the Stacks”. In: *IEEE Transactions on Visualization and Computer Graphics* 26.1, pp. 87–97. DOI: 10.1109/TVCG.2019.2934281.
- Skala, Vaclav, Jakub Sochor, and Jaroslav Krivanek (2016). “Modern Visualization Techniques in Scientific Computing”. In: *Computing and Informatics* 35.4, pp. 845–872.
- Sziranyi, Tamas (2019). “Image Processing and Analysis for Scientific and Industrial Applications”. In: *Acta Polytechnica Hungarica* 16.3, pp. 7–24.
- Ware, Colin (2021). *Information Visualization: Perception for Design*. 4th ed. Morgan Kaufmann.
- Zhang, Y., Michael Sedlmair, and Tamara Munzner (2023). “Criteria for Rigor in Visualization Design Study Research”. In: *IEEE Transactions on Visualization and Computer Graphics* 29.1, pp. 87–97. DOI: 10.1109/TVCG.2022.3209406.

Detection of Artifacts in AI-Generated Portraits Using Frequency and Texture Analysis (FFT, DCT, LBP/GLCM)

Peter Procházka¹

Abstract

In this article, we present interim results from ongoing research aimed at identifying differences between real and AI-generated portraits through analysis in the frequency and texture domains. Three methods are examined: Fast Fourier Transform (FFT), Discrete Cosine Transform (DCT), and statistical texture descriptors based on Local Binary Patterns (LBP) and the Grey-Level Co-occurrence Matrix (GLCM). Using a controlled set of image data – a real portrait, its AI-generated clone, and a retouched version – we demonstrate the processing workflow, visualization, and interpretation of results. The aim of this work is to verify whether visually subtle differences between real and synthetic images correspond to measurable structural differences in alternative image representation domains. The article presents a methodological framework and example results of a pilot study; more extensive experiments on a larger dataset and the inclusion of additional analytical tools are planned in subsequent phases of the ongoing research.

Keywords

generated graphics, AI-generated images, diffusion models, FFT, DCT, LBP, GLCM, detection of AI-generated images

1 Introduction

The rapid development of generative models over the past few years has significantly influenced the fields of computer graphics and image processing. Modern diffusion models, such as Stable Diffusion (Rombach et al., 2022) or Imagen (Saharia et al., 2022), enable the generation of portraits with a high degree of realism that are often difficult for human observers to distinguish from real photographs. At the same time, this capability introduces new challenges in digital forensics, image content authentication, and information security, as confirmed by current research on synthetic media detection (Yang et al., 2019; Sha et al., 2023). Although the visual quality of AI-generated portraits is high, generative processes are based on statistical models trained on large datasets. This suggests that even with convincing visual output, images may retain subtle, systematic artifacts that are not directly observable in the pixel domain. Identifying such differences requires image analysis in alternative representations, particularly in the frequency and texture domains. The goal of our research is to investigate whether measurable differences exist between real, AI-generated, and classically retouched portraits, and to verify the extent to which these differences are consistent across multiple analytical methods. In this article, we present the first pilot phase of the research, focused on demonstrating the methodology and interpreting example results.

2 Theoretical Background

Generative models, including generative adversarial networks (GANs) and diffusion models, enable the synthesis of realistic portraits with a high level of detail, placing increased demands on

¹Department of Applied Informatics, Faculty of Economic Informatics, Bratislava University of Economics and Business, Slovakia. Email: peter.prochazka@euba.sk

detection and forensic methods (Goodfellow et al., 2016; Ho et al., 2020; Dhariwal and Abbeel, 2021; Karras et al., 2021). In recent years, research attention has focused on identifying statistical and spectral inconsistencies that are typical for synthetic images but absent or present to a lesser extent in natural photographs (Barni et al., 2020; Sha et al., 2023).

Fourier Transform (FFT) and Discrete Cosine Transform (DCT) are among the fundamental tools of frequency image analysis. They allow the representation of an image in the frequency domain and the analysis of the energy distribution between low- and high-frequency components (Gonzalez and Woods, 2018; Jain, 1989). Several recent studies indicate that AI-generated images exhibit specific frequency signatures, including periodic artifacts and suppression of high-frequency content (Huang et al., 2023; Corvi et al., 2024).

Texture methods based on Local Binary Patterns (LBP) and Grey-Level Co-occurrence Matrix (GLCM) enable quantification of local statistical image properties, such as contrast, homogeneity, or microstructure variability (Ojala et al., 1996; Haralick et al., 1973). These features have proven to be a suitable complement to frequency analyses for detecting AI-generated images.

3 Methodology

The analysis was conducted on a trio of images: a real portrait, its AI-generated clone, and a retouched version of the real image. All images were preprocessed uniformly, including conversion to grayscale and resolution normalization. The analytical methods used include global frequency analysis via the Fast Fourier Transform (FFT), block-based frequency analysis via the Discrete Cosine Transform (DCT), and local texture analysis using Local Binary Patterns (LBP) and Grey-Level Co-occurrence Matrix (GLCM) descriptors. Primary experiments and visualization of interim results were carried out using a standalone web application implemented in HTML, CSS, and JavaScript. This application enables uploading a trio of images (real, AI-generated, retouched), performing basic preprocessing, and computing frequency and texture characteristics directly in the web browser. The web application served as the main tool for exploratory analysis, visual comparison, and rapid experimentation with method parameters. For independent verification of results, a separate validation application was developed in Python. Computations were performed using the libraries NumPy, SciPy, scikit-image, Pillow (PIL), Matplotlib, and Pandas. The user interface of this validation application was created using the Streamlit framework, which allows interactive execution of Python scripts and visualization of results through a web interface. In this case, Streamlit does not constitute a standalone web application in the HTML/JavaScript sense but serves solely as an interface to the Python implementation of analytical methods. The Python validation application implements the same analytical procedures as the web application, including block-based DCT (8×8), computation of FFT magnitude spectra, generation of LBP maps and histograms ("uniform" method), and computation of GLCM metrics (contrast, dissimilarity, homogeneity, energy, correlation). By comparing outputs from both environments, the consistency of observed trends was verified, and potential implementation deviations were eliminated. FFT analysis was additionally verified using the external tool ImageJ, ensuring reproducibility of frequency spectra across independent software solutions. In the next phase of the research, validation is planned to be extended to tools available in the MATLAB environment (or the Image Processing Toolbox), particularly for verifying DCT, LBP, and GLCM analyses once the relevant license becomes available. This multi-level validation approach, combining a web application, Python implementation, and external tools, increases the reliability and methodological transparency of the presented interim results.

4 Interim Results

The interim outputs of the analysis are based on a combination of visual evaluation and computations performed via the interactive web application and validation scripts in the Python environment on the trio of analyzed images. Observations range from qualitative assessment of visual details to quantitative evaluation of structural image properties across different representation domains. Identified differences between real, AI-generated, and retouched portraits are consistent across multiple analytical approaches.

4.1 Visual Comparison of Real, AI-Generated, and Retouched Portraits

Figure 1 shows the trio of analyzed portraits: the original image, its AI-generated clone, and a manually retouched version of the original. Even during direct visual comparison, noticeable differences are evident, suggesting different origins and different characteristics of image modifications.

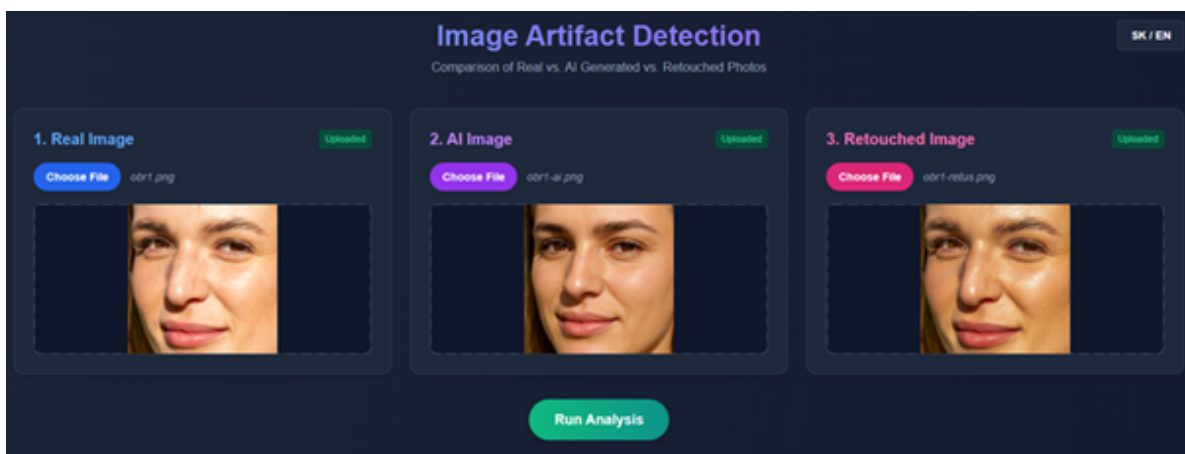


Figure 1. Input interface of the web application with three images

The real portrait exhibits pronounced skin microtexture, irregular pores, subtle local imperfections, and natural variations in light reflections. The skin surface appears statistically non-homogeneous, with fine details distributed randomly. This character is particularly observable in the cheek, nose, and eye areas, where fine wrinkles, pores, and natural sensor noise combine. In contrast, the AI-generated portrait appears visually smoother. Skin microstructure is significantly suppressed, pores are less discernible, and transitions between light and shadow are smoother. Light reflections exhibit a more regular, sometimes symmetrical character and appear less physically realistic. This effect is especially pronounced on the forehead and cheeks, where the image acquires an aesthetic, "cosmetically enhanced" appearance typical of generative models. The retouched portrait represents a transitional case. Local edits smooth selected skin areas, while preserving the underlying microstructure. Irregularities are weakened but not entirely removed, and light reflections retain a more natural character than in the AI-generated image. Visually, the retouched image appears more realistic than the synthetic one, although less raw than the original photograph. Differences are also evident in the eye region. The real portrait exhibits fine wrinkles with irregular geometry and realistic shading. The AI-generated image suppresses these details and shows greater symmetry between the left and right eyes. The retouched portrait again preserves natural asymmetry, albeit with slightly softened contrast. Similar tendencies can be observed in the eyebrows and lips, where the AI version appears more uniform and less random. These visual differences provide important qualitative context for the subsequent interpretation of frequency- and texture-analysis results, allowing these subjective observations to be quantified and verified in alternative image representation domains.

4.2 Results of DCT Analysis

Part of the presentation of DCT analysis results is visualization generated directly within the web application, displaying block-based DCT heatmaps for all three analyzed images side by side (Fig. 2). This form of presentation enables intuitive and immediate comparison of local frequency behavior between the real, AI-generated, and retouched portraits prior to quantitative coefficient evaluation. The web application uses a unified color scale and logarithmic value scaling, making differences in energy distribution between blocks visually discernible even to non-expert observers.



Figure 2. Discrete Cosine Transform

For the real image, even at the level of web visualization, a higher degree of local variability between individual (8×8) blocks is evident, whereas the AI-generated image exhibits a markedly more homogeneous pattern with suppression of high-frequency details. The retouched image again shows a transitional character, with some blocks frequency-simplified but without the regularity typical of synthetic generation. These visual observations from the web application directly correspond to the more detailed interpretation of DCT heatmaps and graphs presented in the following text and confirm that the identified differences are not the result of a single visualization but a stable phenomenon across the tools used. The Discrete Cosine Transform was applied in a block-wise manner at the (8×8)-pixel level, enabling analysis of local frequency-energy distributions and identification of differences between image types at the level of fine structures. Results are presented both as DCT heatmaps and as an aggregated graph of DCT coefficient distributions. In the real portrait, the DCT heatmap shows pronounced local variability across individual blocks. Energy is not concentrated solely in low-frequency components but is also distributed across mid- and higher-frequency components, corresponding to natural skin

microtexture, fine wrinkles, and random image noise. This irregularity is visually readable as a subtly chaotic heatmap structure without repeating regular patterns. In contrast, the AI-generated portrait exhibits significant homogenization of the DCT heatmap. Energy is predominantly concentrated in low-frequency coefficients, while higher-frequency components are strongly suppressed. Individual (8×8) blocks exhibit similar frequency behavior, resulting in a more regular and visually smoother heatmap pattern. This effect points to systematic smoothing of fine details and reduction of local variability, consistent with the operating principles of diffusion-based generative models. Similar tendencies have been observed in recent studies analyzing frequency signatures of AI-generated images (Huang et al., 2023; Corvi et al., 2024). The retouched portrait represents a transitional case. The DCT heatmap shows partial weakening of high-frequency components, particularly in areas of targeted skin retouching, but retains greater local variability than the AI-generated image. Differences between blocks remain present and do not form regular, repeating structures. This character suggests that manual retouching affects local frequency properties of the image but does not alter its global statistical structure in the same way as AI generation. These observations are also confirmed by the graph of DCT coefficient distribution on a logarithmic scale. The curve of the real portrait shows a smooth energy decay from low to higher frequencies without sharp breaks, typical of natural image data. The AI-generated image exhibits steeper decay and lower values in higher-frequency bands, indicating the suppression of fine details. The retouched image appears closer to the original portrait in the graph, though with a slight shift toward lower frequencies due to smoothing. The combination of visual interpretation of DCT heatmaps and quantitative comparison of DCT coefficient profiles thus provides consistent evidence of differing frequency behavior between real, AI-generated, and retouched portraits and represents an important link between visual evaluation and global FFT analysis.

4.3 Results of FFT Analysis

Global frequency analysis was performed using the Fast Fourier Transform, with results presented as magnitude spectra on a logarithmic scale (Fig. 3). The visualization allows comparison of the energy distribution across low-, mid-, and high-frequency bands between the real, AI-generated, and retouched portraits.

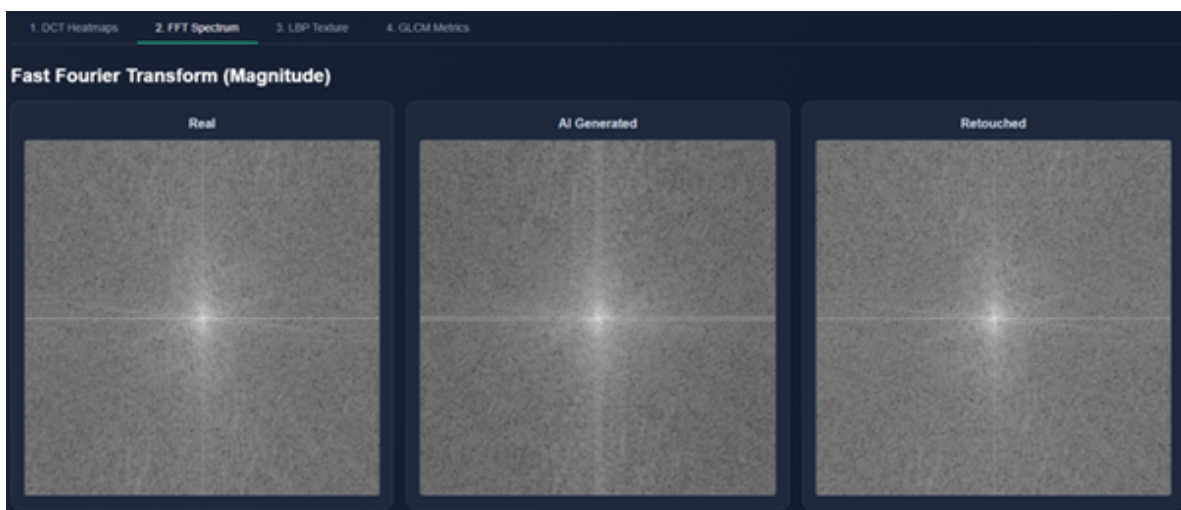


Figure 3. Fast Fourier Transform

The FFT spectrum of the real portrait exhibits a pronounced central component representing low-frequency image content, smoothly transitioning to higher frequencies without sharp discontinuities. In higher-frequency regions, a subtle chaotic energy dispersion is observed, corresponding to natural skin microtexture, fine wrinkles, and sensor noise. Spectrum irregularity indicates

a high degree of local variability typical of photographic capture. The FFT spectrum of the AI-generated portrait differs primarily in the suppression of high-frequency components and the increased regularity of the spectral distribution. In mid-frequency regions, subtle linear and cross-like structures appear, suggesting periodic or quasi-periodic patterns arising during the generative process. This spectral character suggests lower randomness and greater statistical uniformity in the image, consistent with observations from recent studies on the frequency signatures of diffusion models (Huang et al., 2023; Corvi et al., 2024).

The retouched portrait exhibits an FFT spectrum located between both extremes. Compared to the real image, a slight reduction in high-frequency content is observed, corresponding to local skin smoothing. Unlike the AI-generated image, however, the spectrum does not exhibit pronounced regular structures or systematic repetitions. The overall spectral character remains predominantly chaotic and closer to that of a natural photograph. To ensure reproducibility and independent verification of results, FFT analysis was also performed using the ImageJ tool. FFT spectra obtained in ImageJ (Fig. 4) exhibit the same qualitative differences between analyzed images as spectra computed in the web and Python application environments. Agreement in spectral shape, degree of high-frequency suppression, and the presence of regular structures confirms that the observed differences are not artifacts of a specific implementation but rather represent inherent properties of the analyzed images.

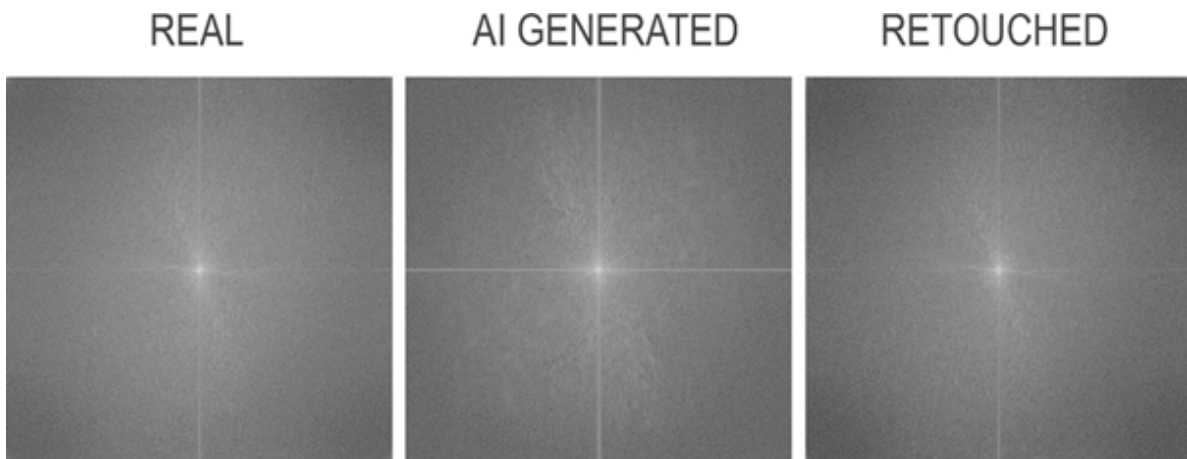


Figure 4. Fast Fourier Transform – ImageJ

Source: Author's own processing FFT analysis thus provides a robust global view of structural differences between real, AI-generated, and retouched portraits and suitably complements more local approaches based on DCT and texture descriptors.

4.4 Results of LBP Analysis

Local Binary Patterns (LBP) analysis was used to assess local skin microstructures that are often visually inconspicuous in the pixel domain but statistically significant. Figure 5 shows LBP maps obtained for the real, AI-generated, and retouched portraits via the web application. The real portrait exhibits a high degree of local variability in LBP patterns. The maps contain fine, irregular structures without dominant repeating patterns, corresponding to natural skin microtexture and the random nature of details. The histogram of LBP codes is relatively evenly distributed, without strong dominance of narrow intervals.

The AI-generated portrait, in contrast, exhibits pronounced regularity in LBP maps. Local patterns are more homogeneous, and the histogram shows a concentration of several dominant LBP codes. This phenomenon indicates reduced microstructural diversity and increased texture uniformity, consistent with the visual impression of smoothed skin and suppressed pores. Similar texture behavior in AI-generated images has been reported in other studies focused on local

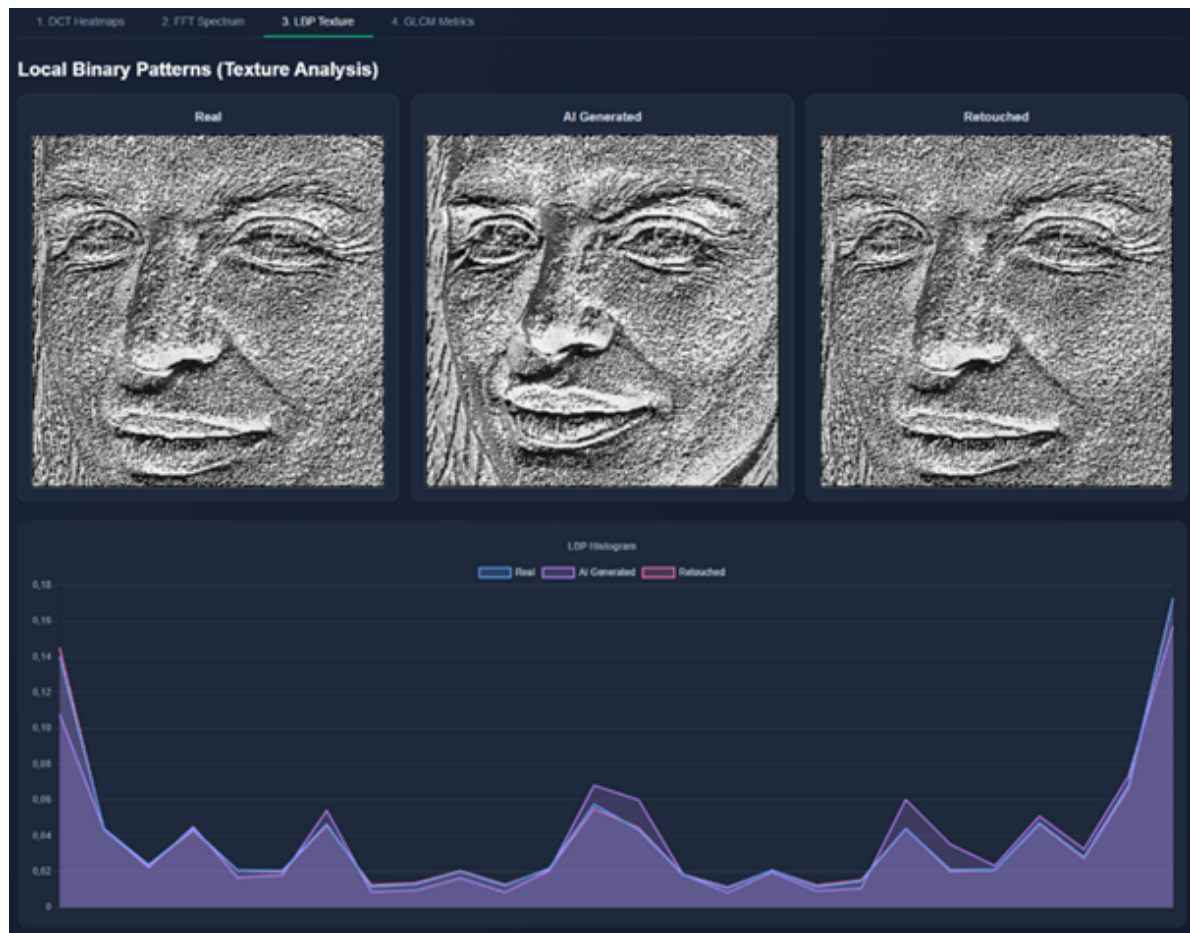


Figure 5. Local Binary Patterns

descriptors for distinguishing synthetic from real images, such as those using LBP and related texture features (Ojala et al., 1996; Sha et al., 2023).

The retouched portrait again represents a transitional state. LBP maps show partial suppression of fine details in retouched areas, but without pronounced regularization typical of AI-generated images. The LBP code histogram remains closer to that of the real portrait, albeit with a slight increase in the dominance of dominant patterns due to smoothing interventions. LBP analysis results thus confirm that local texture features can distinguish natural, synthetic, and manually edited images in a manner consistent with visual evaluation and frequency analyses.

4.5 GLCM Analysis

To complement local texture analysis, statistical features based on the Grey-Level Co-occurrence Matrix (GLCM) were computed. Quantitative results are summarized in a table (Fig. 6) and visualized using a comparative graph generated by the web application. The real portrait shows higher contrast and dissimilarity, reflecting the complex, non-homogeneous structure of natural skin. At the same time, it exhibits lower homogeneity, a characteristic of images with a high degree of local intensity variation.

The AI-generated portrait exhibits significantly lower contrast and dissimilarity, along with increased homogeneity. This profile indicates the suppression of fine local differences and the presence of smoother, more statistically uniform textures. Such behavior of GLCM metrics is consistent with observations reported in the literature, which indicate that AI-generated images exhibit a higher degree of local texture regularity and homogeneity than natural photographs (Haralick et al., 1973; Sha et al., 2023). The retouched portrait again occupies an intermediate

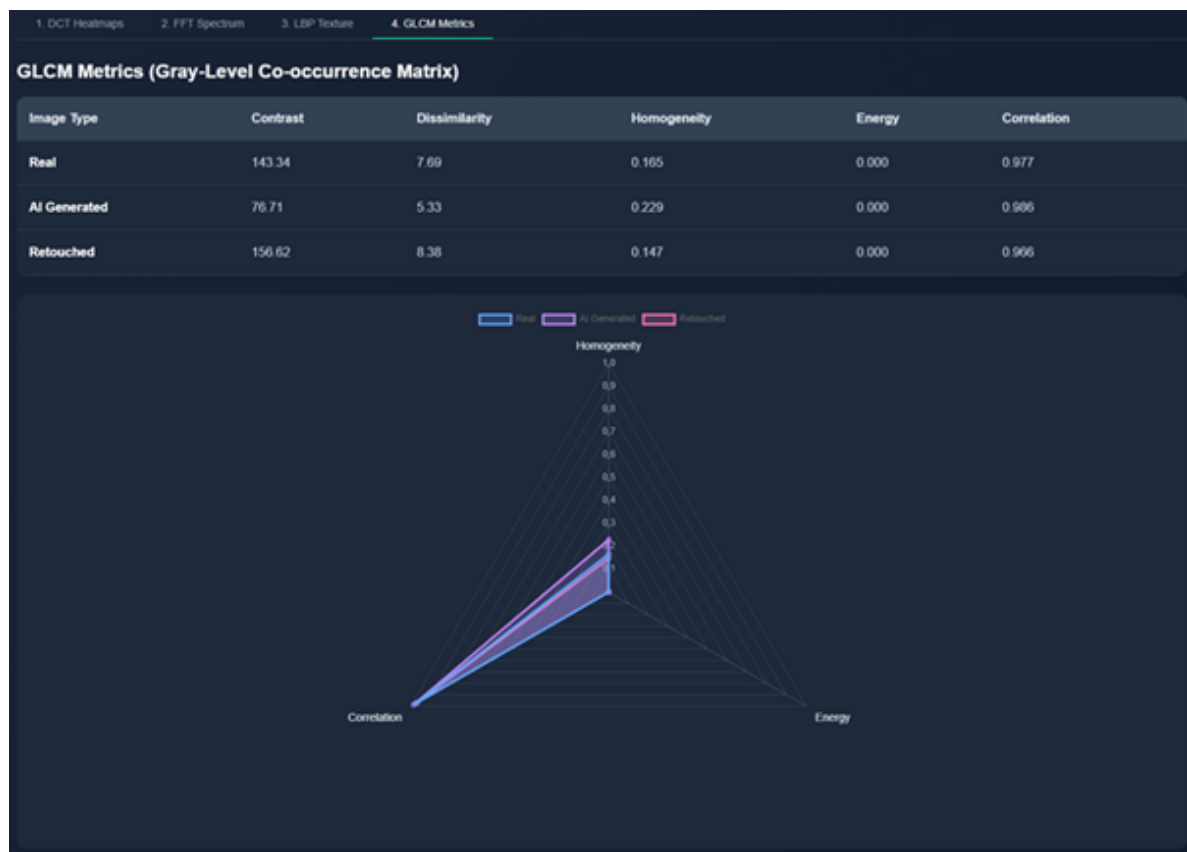


Figure 6. GLCM Metrics

position between the real and AI-generated images. The contrast and homogeneity values indicate a partial influence of smoothing operations, without a pronounced change in the image's global statistical structure. GLCM metrics thus confirm that classical retouching does not alter local statistical properties of the image to the same extent as generative models. In combination with LBP analysis, GLCM features provide a robust quantitative description of local textures that suitably complements global frequency methods based on FFT and DCT, and contribute to clearer differentiation of the origin of analyzed portraits.

5 Discussion

Interpretation of the achieved results highlights several aspects related to the detection of AI-generated portraits and the importance of frequency and texture features in distinguishing them from real photographs. First and foremost, visually subtle differences that are often difficult for human observers to identify manifest consistently and reproducibly in transformed domains. Frequency analyses based on DCT and FFT clearly indicate that AI-generated portraits tend to suppress high-frequency content and exhibit greater regularity in spectral energy distribution. This phenomenon can be interpreted as a consequence of the optimization mechanisms of generative models, which favor global visual consistency and suppress random noise and microdetails typical of real images. Similar frequency behavior has been observed in other studies analyzing AI-generated images, suggesting that this is a more general phenomenon rather than a property of a single tested example. Texture analyses using LBP and GLCM provide a complementary view of local structural image properties. Results show that AI-generated images exhibit reduced variability in local patterns and greater homogeneity, whereas real photographs exhibit higher contrast and statistical non-homogeneity. Retouched images occupy an intermediate position across most metrics, confirming that manual retouching affects local details but does not alter the

fundamental statistical structure of the image as strongly as AI generation. An important finding is also the consistency of results across different implementations and tools. Agreement between the outputs of the web application, the Python validation application, and the external ImageJ tool increases the credibility of the observed trends and suggests that the identified differences are not artifacts of a particular software solution. This aspect is crucial, especially for the practical deployment of forensic tools, where reproducibility and implementation independence are essential. At the same time, it is necessary to emphasize the limitations of the presented results. The analysis was conducted on a limited number of images and serves primarily as a pilot study demonstrating the methodological approach. Generalizing the findings to a broader range of generative models, various scene types, and different levels of image quality requires more extensive experiments across larger datasets. Nevertheless, the results indicate that a combination of frequency and texture features represents a promising basis for robust detection of AI-generated images.

6 Conclusion

In this work, we have introduced a methodological framework and interim results from a pilot study analyzing differences between real, AI-generated, and manually retouched portraits. By combining visual evaluation, frequency transformations (FFT, DCT), and local texture descriptors (LBP, GLCM), it was possible to identify consistent structural differences that are not always apparent in the pixel domain but manifest in alternative image representations. The results suggest that AI-generated portraits tend to exhibit reduced high-frequency components, more regular frequency spectra, and greater local texture homogeneity compared to real photographs. Retouched images occupy an intermediate position across most analyzed metrics, confirming that classical image editing affects local details but does not alter the global statistical structure of the image to the same extent as AI generation. An important aspect of the work is the multi-level validation approach, in which results were compared across an independent web application, a Python implementation for validation, and the external ImageJ tool. Agreement of observed trends across different environments supports the robustness of the presented findings and indicates their potential applicability in digital forensics and the detection of synthetic media. The presented results are a pilot study and primarily serve to demonstrate the methodology and its interpretative potential. In the next phase of the research, the analysis will be expanded to a larger, more diverse dataset, including various generative models, scene types, and image quality levels. At the same time, additional analytical approaches and machine learning models are planned to automate the evaluation of identified features. These steps will be the subject of a subsequent publication focused on detailed quantitative evaluation and comparison of methods.

References

- Barni, Mauro, K. Kallas, E. Nowroozi, and Benedetta Tondi (2020). “Adversarial attacks against CNN-based image forgery detectors: A comparative study”. In: *EURASIP Journal on Information Security*.
- Corvi, R., Y. Huang, and Luisa Verdoliva (2024). “Analyzing frequency-domain fingerprints of diffusion-based image generators”. In: *IEEE Transactions on Information Forensics and Security*.
- Dhariwal, Prafulla and Pieter Abbeel (2021). “Diffusion models beat GANs on image synthesis”. In: *Advances in Neural Information Processing Systems*.
- Gonzalez, Rafael C. and Richard E. Woods (2018). *Digital Image Processing*. 4th ed. Pearson.
- Goodfellow, Ian, Yoshua Bengio, and Aaron Courville (2016). *Deep Learning*. MIT Press.
- Haralick, Robert M., K. Shanmugam, and Its’Hak Dinstein (1973). “Textural features for image classification”. In: *IEEE Transactions on Systems, Man, and Cybernetics* 3.6, pp. 610–621.

- Ho, Jonathan, Ajay Jain, and Pieter Abbeel (2020). “Denoising diffusion probabilistic models”. In: *Advances in Neural Information Processing Systems*.
- Huang, Y., X. Yang, and Siwei Lyu (2023). “AI-generated image detection using statistical and spectral inconsistencies”. In: *Proceedings of the IEEE International Conference on Image Processing*.
- Jain, Anil K. (1989). *Fundamentals of Digital Image Processing*. Prentice Hall.
- Karras, Tero, Miika Aittala, Samuli Laine, Erik H"ark"onen, Janne Hellsten, Jaakko Lehtinen, and Timo Aila (2021). “Alias-free generative adversarial networks”. In: *Advances in Neural Information Processing Systems*.
- Ojala, Timo, Matti Pietik"ainen, and David Harwood (1996). “A comparative study of texture measures with classification based on featured distributions”. In: *Pattern Recognition* 29.1, pp. 51–59.
- Rombach, Robin, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Bj"orn Ommer (2022). “High-resolution image synthesis with latent diffusion models”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
- Saharia, Chitwan, William Chan, Han Chang, et al. (2022). “Imagen: Photorealistic text-to-image diffusion models”. In: *Proceedings of the International Conference on Machine Learning*.
- Sha, Z., H. Wu, and X. Chen (2023). “A survey on deepfake detection: Challenges and future trends”. In: *ACM Computing Surveys*.
- Yang, X., Y. Li, and Siwei Lyu (2019). “Exposing deep fakes using inconsistent head poses”. In: *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*.

Application generated by artificial intelligence

Pavol Sojka¹

Abstract

The article focuses on generating the code of a simple application for finding the current temperature according to a specified city in the Slovak Republic. The code is fully generated using artificial intelligence tools, namely Gemini, Grok, ChatGPT and DeepSeek chatbots. It is assumed that in the near future these tools could largely take over programming/code generation and thus significantly reduce the need for human programmers. The aim of our article is to examine the current state of the art in this field and whether three different chatbots can generate, without additional assistance, a separate functional code for an application that has only one purpose, namely finding the weather from the shmu.sk domain. In individual chapters, we describe the creation of a test environment and the deployment of the code generated by artificial intelligence and compare it with the deployment of the authors' code.

Keywords

artificial intelligence, Python, automation, programming

1 Introduction

In recent years, there has been significant progress in the field of artificial intelligence, with machine learning and deep learning enabling increasingly complex tasks to be solved. Nevertheless, there remain areas where human programming demonstrates higher efficiency, especially in cases requiring high interpretability, adaptation to non-standard scenarios, or ethical constraints. The aim of our article is a simple comparison between AI-based programming and traditional human approaches in solving a specialized task: creating an application to retrieve the current temperature from the Slovak Hydrometeorological Institute website (www.shmu.sk). The paper is best understood as an exploratory case study rather than a comprehensive evaluation of AI-generated programming.

2 Environment Setup

We selected Linux Ubuntu 22.04 as the primary operating system for the experiments (Petersen, 2025). The environment was deployed on Google Cloud using a virtual machine with hardware parameters including four CPUs (AuthenticAMD, AMD EPYC 7B12) and 4 GB of RAM. Although these specifications are modest, they are sufficient for the experiment.

After installing the OS, we set up Python (version 3.10.12) and the Pip package manager (version 22.0.2). We also installed recommended libraries necessary for running all scripts (Lubanovic, 2019).

Our goal was to:

- Generate Python code using AI tools
- Test the generated code in practical deployment

¹Department of Applied Informatics, Faculty of Economic Informatics, Bratislava University of Economics and Business, Slovakia. Email: pavol.sojka@euba.sk

```

import requests from bs4 import BeautifulSoup

def ziskaj_teplotu(mesto): # Mapovanie názvu mesta na URL parameter mesta = {
'bratislava': 'http://www.shmu.sk/sk/?page=1&id=meteo_mestske&station=BA',
'kosice': 'http://www.shmu.sk/sk/?page=1&id=meteo_mestske&station=KE', 'zilina':
'http://www.shmu.sk/sk/?page=1&id=meteo_mestske&station=ZA', # pridaj ďalšie
mestá podľa SHMÚ }

mesto_url = mesta.get(mesto.lower())
if not mesto_url:
    print(f"Mesto '{mesto}' nie je podporované.")
    return

response = requests.get(mesto_url)
if response.status_code != 200:
    print("Nepodarilo sa načítať dáta zo SHMÚ.")
    return

soup = BeautifulSoup(response.text, 'html.parser')

# Hľadanie elementu s teplotou - môže sa zmeniť podľa aktuálnej HTML štruktúry
teplota_element = soup.find('td', class_='td_temp') # Pozor: treba overiť
aktuálnu triedu!

if teplota_element:
    print(f"Aktuálna teplota v {mesto.title()}: {teplota_element.text.strip()}
°C")
else:
    print("Teplotu sa nepodarilo nájsť.")

1 #Príklad použitia

mesto = input("Zadaj mesto (napr. Bratislava, Kosice): ") ziskaj_teplotu(mesto)

```

Figure 1. Code generated by chatGPT (version: GPT-4o)

- Implement an experiment using the authors' own code
- Compare and interpret the results

3 Results

Each code was executed three times for verification. The AI-generated codes came from ChatGPT, Gemini, Grok and DeepSeek. The initial prompt was: 'ahoj, vies urobiť aplikáciu v pythone, ktorá po zadani mesta vypise teplotu v stupnoch C zo stranky shmu.sk?', translated: 'hi, can you make an application in python that, after entering a city, prints the temperature in degrees C from the website shmu.sk?'. Despite being written without diacritics, all chatbots understood the question. However, none of the generated codes returned the correct temperature data from SHMÚ. The authors' own Python code successfully retrieved the temperature using web scraping techniques, while all AI-generated codes failed.

Note: The resulting sample codes may require additional editing in the form of indentation (tabs) to function properly, as they may have been automatically modified by text/graphics software. Result outputs were translated to English language in the codes „result“ sections.

Result 1 - Figure 1

Enter city (e.g. Bratislava, Kosice): Bratislava

```

import requests from bs4 import BeautifulSoup

def
ziskaj_teploty_zo_shmu(url="https://www.shmu.sk/sk/?page=1&id=meteo_apocasiae_sk"
): """ Získa aktuálne teploty zo stránky SHMÚ pomocou web scrapingu.

Args:
    url (str): URL adresa SHMÚ stránky s aktuálnym počasím.

Returns:
    dict: Slovník, kde kľúčom je názov stanice a hodnotou je teplota v °C.
    Vráti prázdny slovník, ak sa dáta nepodarí získať.
    """
teploty = {}
try:
    # Stiahnutie obsahu stránky
    response = requests.get(url, timeout=10) # Nastavenie timeoutu pre prípad,
že stránka neodpovedá
    response.raise_for_status() # Vyvolá HTTPError pre chybné status kódy (4xx,
5xx)

    # Parsovanie HTML obsahu
    soup = BeautifulSoup(response.text, 'html.parser')

    # SHMÚ zvyčajne zobrazuje dáta v tabuľkách.
    # Skúsime nájsť tabuľku s ID 'data-tabulka' alebo triedou, ktorá obsahuje
meteo dáta.
    # Táto časť kódu je najcitlivejšia na zmeny štruktúry SHMÚ stránky.
    tabulka = soup.find('table', {'class': 'dm_tabulka', 'id': 'data-tabulka'})

    if tabulka:
        # Prejdeme cez riadky tabuľky (tr)
        for riadok in tabulka.find_all('tr'):
            # Nájdenie všetkých dátových buniek (td) v riadku
            bunky = riadok.find_all('td')
            if len(bunky) >= 3: # Potrebujeme aspoň 3 bunky (stanica, teplota,
čas/ďalšie info)
                stanica = bunky[0].get_text(strip=True)
                teplota_str = bunky[2].get_text(strip=True) # Teplota je často v
tretom stĺpci

                try:
                    # Pokúsime sa prekonvertovať teplotu na float
                    teplota = float(teplota_str.replace(',', '.').replace(' °C',
''))

                    teploty[stanica] = teplota
                except ValueError:
                    # Ak sa nepodarí prekonvertovať, preskočíme
                    continue
            else:
                print("Chyba: Nenašla sa tabuľka s aktuálnymi dátami na stránke SHMÚ.")

except requests.exceptions.RequestException as e:
    print(f"Chyba pri sťahovaní stránky: {e}")
except Exception as e:
    print(f"Nastala neočakávaná chyba: {e}")

```

Figure 2. Code generated by Gemini (version: 2.5 Flash)

Output: Temperature could not be found.

```

return teploty

if name == "main": aktualne_teploty = ziskaj_teploty_zo_shmu()

    if aktualne_teploty:
        print("Aktuálne teploty zo SHMÚ:")
        for stanica, temp in aktualne_teploty.items():
            print(f"- {stanica}: {temp}°C")
    else:
        print("Nepodarilo sa získať údaje o teplotách zo SHMÚ.")

print("\n---")
print("Pre konkrétne mesto musíte prejsť cez získaný slovník a nájsť
zhodu.")
print("Napríklad pre Bratislavu:")

hladane_mesto = input("Zadajte mesto, pre ktoré chcete nájsť teplotu
(napr. Bratislava): ")
found = False
for stanica, temp in aktualne_teploty.items():
    if hladane_mesto.lower() in stanica.lower():
        print(f"- {stanica}: {temp}°C")
        found = True
if not found:
    print(f"Pre mesto '{hladane_mesto}' sa nenašli žiadne záznamy.")

```

Figure 3. Continuation of Code generated by Gemini (version: 2.5 Flash)

Result 2 - Figure 2 and 3

Enter city (e.g. Bratislava, Kosice): Bratislava

Output: Error: No table with current data found on the SHMÚ website.

Unable to retrieve temperature data from SHMÚ.

For a specific city, you need to go through the retrieved dictionary and find a match. For example, for Bratislava: Enter the city for which you want to find the temperature (e.g. Bratislava): Bratislava No records were found for the city 'Bratislava'.

Result 3 - Figure 4

Enter city name (or 'end' to exit): Bratislava

Output: Temperature data for Bratislava was not found.

Enter city name (or 'end' to exit):

Result 4 - Figure 5

System: Welcome to the application for getting the current temperature from SHMU.sk

Enter the name of the city: Bratislava

Output: Unable to find the table with data.

The following is the authors' code, which is shortened (Fig. 6) and provided only as an example, as the authors are interested in finding out whether artificial intelligence has improved the ability to create code for this type of application so that it is functional. The authors therefore do not want to create a guide for artificial intelligence on how to program the code. If the reader wants to verify the authors' code functionality, it is possible to verify whether the

```

import requests from bs4 import BeautifulSoup import re

def get_temperature(city): # Normalizácia názvu mesta (odstránenie diakritiky
pre URL) city = city.lower().strip() # Nahradzame diakritiku pre jednoduchšie
vyhladavanie diacritics = { 'á': 'a', 'ä': 'a', 'č': 'c', 'ď': 'd', 'é': 'e',
'í': 'i', 'l': 'l', 'ň': 'n', 'ó': 'o', 'ô': 'o', 'r': 'r', 'š': 's', 't': 't',
'ú': 'u', 'ý': 'y', 'ž': 'z' } for char, repl in diacritics.items(): city =
city.replace(char, repl)

# Vytvorenie URL pre vyhľadavanie na SHMÚ
base_url = "https://www.shmu.sk/sk/?page=1&subpage=3&name="
url = base_url + city

try:
    # Odoslanie HTTP požiadavky
    response = requests.get(url, headers={'User-Agent': 'Mozilla/5.0'})
    response.raise_for_status() # Overenie, či požiadavka bola úspešná

    # Parsovanie HTML obsahu
    soup = BeautifulSoup(response.text, 'html.parser')

    # Hľadanie elementu s teplotou (napr. v tabuľke alebo meteograme)
    # SHMÚ zvyčajne zobrazuje teplotu v div s triedou 'actual' alebo podobnej
    temp_element = soup.find('div', class_='actual')
    if temp_element:
        # Extrahovanie textu s teplotou
        temp_text = temp_element.get_text(strip=True)
        # Predpokladáme formát ako "12 °C" alebo "12°C"
        match = re.search(r'(\d+\.\d*)\s*°C', temp_text)
        if match:
            temperature = match.group(1)
            return f"Aktuálna teplota v meste {city.title()} je {temperature}
°C"
        else:
            return f"Teplota pre mesto {city.title()} nebola nájdená."
    else:
        return f"Údaje o teplote pre mesto {city.title()} neboli nájdené."

except requests.exceptions.RequestException as e:
    return f"Chyba pri získavaní údajov: {e}"

def main():
    while True:
        city = input("Zadaj názov mesta (alebo 'koniec' pre ukončenie): ")
        if city.lower() == 'koniec':
            print("Ukončujem program.")
            break
        if not city.strip():
            print("Prosím, zadaj platný názov mesta.")
            continue
        result = get_temperature(city)
        print(result)
if name == "main":
    main()

```

Figure 4. Code generated by Grok (version: Grok 3)

code returns the correct data in real time at the provided link (if SHMÚ did not modify page code in the meantime, then the code snippet works correctly) <http://rsrch.ddns.net/rsrch/shmu/> and compare if the resulting data with the selected one on www.shmu.sk (temperature data in the chosen city, e.g., Bratislava) match.

```

import requests
from bs4 import BeautifulSoup

def get_temperature(city):
    try:
        # URL SHMU s meteorologickými dátami
        url = "https://www.shmu.sk/sk/?page=1&id=meteo_num_alad"

        # Stiahnutie stránky
        response = requests.get(url)
        response.raise_for_status()

        # Spracovanie HTML
        soup = BeautifulSoup(response.text, 'html.parser')

        # Nájdenie tabuľky s údajmi
        table = soup.find('table', {'class': 'border'})
        if not table:
            return "Nepodarilo sa nájsť tabuľku s údajmi."

        # Hľadanie riadku pre zadané mesto
        for row in table.find_all('tr')[1:]: # preskočiť hlavičku
            columns = row.find_all('td')
            if len(columns) >= 2:
                current_city = columns[0].get_text(strip=True)
                if city.lower() in current_city.lower():
                    temperature = columns[1].get_text(strip=True)
                    return f"Aktuálna teplota v {current_city}: {temperature}°C"

        return f"Mesto '{city}' sa nenašlo v zozname."

    except requests.RequestException as e:
        return f"Chyba pri pripájaní k SHMU: {e}"
    except Exception as e:
        return f"Vyskytla sa chyba: {e}"

# Hlavná časť programu
if __name__ == "__main__":
    print("Vitajte v aplikácii pre získanie aktuálnej teploty zo SHMU.sk")
    city = input("Zadajte názov mesta: ")
    result = get_temperature(city)
    print(result)

```

Figure 5. Code generated by DeepSeek (version: DeepSeek-V3)

Result 5 - Figure 6

Enter city (default=Bratislava):

Output: The temperature for the city of Bratislava is 22.5 °C

From the above examples, we can see that none of the used chatbots was able to generate the required code returning the correct result, although the data is located on the main page www.shmu.sk and there is no need to go deeper into the code of the subpages. The author prepared a code using the Python language, which, like the artificial intelligence codes, downloaded the initial website shmu.sk to the local computer and then, by analyzing the obtained data, determined the relevant temperature in a given city. This experiment shows that artificial intelligence can solve various complex tasks when many existing solutions are implemented, from which it learns, but sometimes even a banal task (as in this article) leads to failure for all the studied chatbots.

```
import requests
from bs4 import BeautifulSoup
import csv
def scrape_page(soup, quotes):
    cdata = soup.find(string=re.compile("CDATA"))
    import string
    cdata2=cdata.split(",")
    global m
    m = "Bratislava"
    m = input("Zadaj mesto (default=%s): " % m) or m
    found=False
    for mesto in cdata2:
        if m.capitalize() in mesto:
            quotes.append(mesto)
            found=True
            break
    if not found:
        print("Mesto nenajdene")
```

Figure 6. A brief look at the part of authors' code

4 Discussion

The chatbot versions were tested in the period June-July 2025, and therefore in the future it would be appropriate to repeat the test for newer versions of chatbots. It is assumed that in the upcoming years artificial intelligence will advance to such a level that this simple test will be successfully passed, as well as more complex implementations of other applications. The test on the SHMÚ website with our version of the code was carried out in the same period as for the artificial intelligence tools, namely in the months of June-July 2025. Our code may not work in future tests, as it is based on the so-called web scraping technology (Clark, 2025), where it is important to browse the structures of the main SHMÚ website. If these structures change, for example, the appearance of the main page changes significantly, the application may stop working completely. This approach (web scraping) is purely experimental and intended only for research purposes, as it may raise questions about the legality of the data obtained in this way.

5 Conclusion

In the article, we described the procedure of how we designed an experiment to obtain relevant data from the main page of the SHMÚ using the web scraping technology implemented by the authors of the article and then tested applications that generated three different chatbots based on the LLM (Large Language Model) technology, which were supposed to program the application based on the specified parameters (Campeato, 2024). The experiment showed that the current versions of the chatbots could not handle this task and none of them could provide the required data from the SHMÚ website.

Resources

- Campeato (2024). *Large Language Models for Developers: A Prompt-based Exploration of LLMs*. United States: De Gruyter.
- Clark (2025). *Web Scraping with Python Step by Step: A Practical Guide with Examples*. (n.p.): Walzone Press.
- Lubanovic (2019). *Introducing Python: Modern Computing in Simple Packages*. United States: O'Reilly Media.
- Petersen (2025). *Ubuntu 24.04 LTS Server: Administration and Reference*. (n.p.): Surfing Turtle Press.

Integration of Machine Learning Technologies in Developing Information Security Skills of Prospective Informatics Teachers in the Digital Environment

Danara Tleumagambetova¹ Jaroslav Kultan² Meruyert Serik³

Abstract

In the modern digital ecosystem, information security requirements are becoming increasingly complex, placing new demands on the professional preparation of future informatics teachers. The purpose of this study is to identify the pedagogical, methodological, and technological possibilities for integrating machine learning into the process of developing and enhancing information security skills. The research analyzes the use of ML methods in the educational process, including automatic threat detection, data classification, anomaly identification, and behavioral risk prediction. Additionally, a model is proposed to develop students' practical competencies through smart devices and digital learning platforms. The findings demonstrate that integrating machine learning into the educational process significantly strengthens future informatics teachers' professional competencies in information security, enabling a systematic understanding of threats and the development of effective prevention strategies.

Keywords

Anomaly identification; Digital environment; Informatics teacher education; Information Security; Machine Learning; Threat detection

1 Introduction

The rapid digital transformation of modern society has significantly reshaped the educational landscape, redefining the competencies required of future teachers, particularly in informatics. As digital environments become increasingly complex and interconnected, issues related to information security, data protection, and responsible technology use have become critical. Prospective informatics teachers are expected not only to possess technical knowledge but also to demonstrate well-developed information security skills to guide students safely and effectively in technology-rich learning contexts. In this regard, integrating machine learning technologies into teacher education represents a promising direction for enhancing both pedagogical and professional competencies.

Machine learning, as a rapidly evolving branch of artificial intelligence, offers new opportunities for adaptive learning, automated assessment, and intelligent analysis of digital risks. By incorporating machine learning tools into the preparation of future informatics teachers, educational programs can create dynamic environments where learners engage with realistic cybersecurity challenges, analyze patterns of digital threats, and develop critical thinking skills. Such integration supports personalized learning pathways, enabling students to identify vulnerabilities, interpret data-driven feedback, and develop practical strategies to ensure information safety.

¹L.N. Gumilyov Eurasian National University, Institute of Digital Sciences and Artificial Intelligence, Department of Computer Science, Astana, Kazakhstan, danara1310@gmail.com

²Bratislava University of Economics and Business, Faculty of Economic Informatics, Department of Applied Informatics, Bratislava, Slovakia, jaroslav.kultan@euba.sk

³L.N. Gumilyov Eurasian National University, Institute of Digital Sciences and Artificial Intelligence, Department of Computer Science, Astana, Kazakhstan, serik_meruerts@mail.ru

In the digital environment, information security education goes beyond theoretical instruction; it requires experiential learning supported by modern technological solutions. Machine learning technologies can simulate cyber threats, detect anomalies in network behavior, and provide predictive insights that help prospective teachers understand the principles of secure system design. Consequently, the use of intelligent systems in teacher training contributes to the formation of analytical, research-oriented, and reflective skills, which are essential for future educators responsible for fostering digital literacy among school students.

Despite the growing interest in artificial intelligence in education, there remains a need for systematic research on how machine learning can be effectively integrated into the development of information security competencies within informatics teacher education. Many existing approaches focus primarily on technical training, while the pedagogical potential of machine learning for competency-based learning and professional development is still underexplored. Therefore, investigating methodological frameworks that combine educational theory, cybersecurity principles, and machine learning technologies is both timely and relevant.

This study aims to explore the integration of machine learning technologies to develop the information security skills of prospective informatics teachers within a digital learning environment.

2 A search for sources and the current state of the problem

The rapid development of digital technologies and artificial intelligence has significantly influenced the preparation of prospective informatics teachers, particularly in developing information security competencies. Modern educational research increasingly focuses on integrating machine learning into cybersecurity education and teacher training programs, emphasizing the need for innovative pedagogical approaches that combine technological and methodological components. Scholars note that artificial intelligence technologies enable adaptive learning environments, automated feedback, and personalized educational trajectories, which enhance professional training in digital contexts (Zawacki-Richter et al., 2019; Holmes et al., 2022).

Recent studies highlight that machine learning technologies play an important role in cybersecurity education by allowing learners to analyze large datasets, detect anomalies, and simulate cyber threats in virtual environments. Such approaches support experiential learning and problem-based instruction, which are considered effective strategies for developing analytical and research skills among future educators (Chen et al., 2020; Möhring et al., 2023). Researchers also emphasize that integrating intelligent systems into higher education promotes a deeper understanding of digital risks and strengthens students' ability to apply theoretical knowledge in practice (Ng et al., 2021).

At the same time, the current state of research shows that many teacher education programs still prioritize general digital literacy over specialized information security skills. International frameworks on digital competence emphasize the importance of cybersecurity awareness, ethical use of technology, and responsible data management as essential elements of teacher professionalism in the digital era (European Commission, 2020; Miao et al., 2021). These competencies are particularly relevant for prospective informatics teachers, who are expected to guide school students in safe online behavior and critical evaluation of digital resources.

Another important direction in the literature concerns the use of data analytics and machine learning to support intelligent educational systems. Learning analytics platforms allow educators to monitor learners' progress, predict potential risks, and provide timely interventions, thereby improving the effectiveness of cybersecurity training (Papamitsiou and Economides, 2016). In addition, research in the field of Internet of Things and smart technologies highlights new challenges for information security education, as future teachers must understand emerging digital threats associated with interconnected devices and cloud infrastructures (Al-Fuqaha et al., 2015).

Despite the growing number of publications, several research gaps remain. First, many studies examine the integration of artificial intelligence or cybersecurity education separately, without addressing their combined impact on the professional preparation of informatics teachers. Second, there is a lack of methodological models that explain how machine learning technologies can be systematically embedded into teacher education curricula to develop information security skills. Third, empirical evidence on the effectiveness of AI-based pedagogical strategies for preparing future informatics teachers remains limited, underscoring the need for further interdisciplinary research (Möhring et al., 2023; Ng et al., 2021).

Overall, the analysis of contemporary sources demonstrates that integrating machine learning technologies into the development of information security competencies is a relevant and rapidly evolving research area. The existing literature confirms the educational potential of intelligent systems while also emphasizing the need for pedagogical frameworks that align technological innovation with competency-based teacher education in the digital environment (Holmes et al., 2022; Miao et al., 2021).

3 The main findings of the article

The conducted research demonstrates that integrating machine learning technologies into the educational process significantly enhances the development of information security skills among prospective informatics teachers in the digital environment. The main findings confirm that combining modern artificial intelligence algorithms with pedagogical methodologies creates a comprehensive framework that supports both technical competence and professional readiness. The study identifies several key directions in which machine learning strengthens cybersecurity education, including threat detection, data integrity protection, anomaly analysis, and the formation of secure digital behavior.

One of the primary findings of the study is that the use of machine learning algorithms provides future computer science teachers with practical tools for understanding real-world cybersecurity challenges. Algorithms such as Decision Trees, Random Forest, Support Vector Machine, K-Nearest Neighbors, and neural network models (ANN, CNN, LSTM) enable students to analyze patterns in digital data and recognize potentially dangerous activities within networks and systems. The inclusion of probabilistic models such as Naive Bayes and unsupervised methods like K-means clustering and autoencoders enables learners to explore both supervised and unsupervised approaches in cybersecurity contexts.

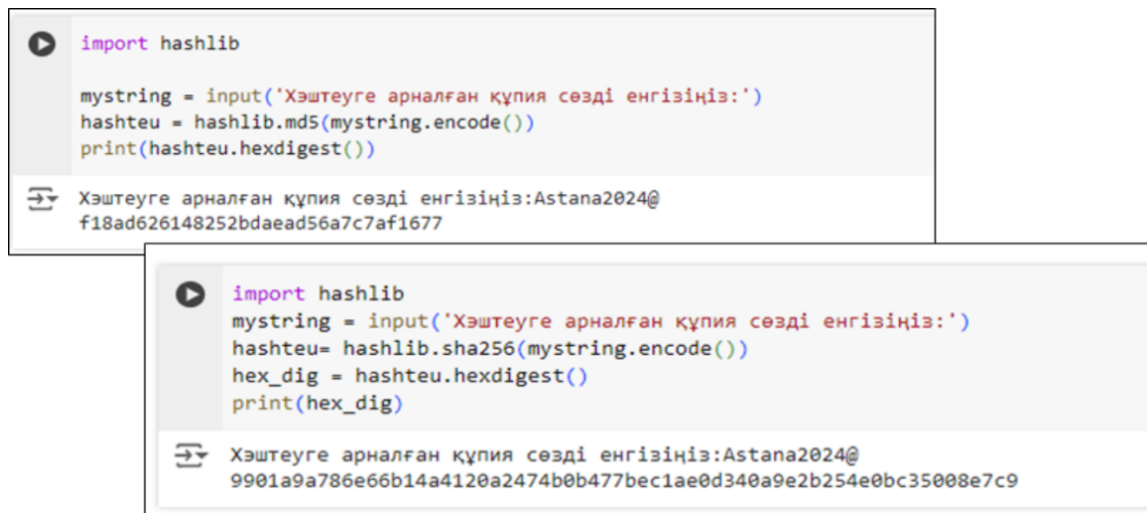
The research shows that these algorithms help students move beyond theoretical learning toward applied problem solving. For example, anomaly-detection models can be used to simulate intrusion-detection systems, allowing prospective teachers to analyze suspicious behavior in network traffic. By working with real datasets and classification tasks, learners develop analytical thinking skills, understand the logic of predictive modeling, and acquire experience in identifying digital threats. The findings indicate that exposure to diverse machine learning models strengthens future teachers' ability to design lessons that integrate cybersecurity principles with computational thinking.

Another important finding concerns the need to secure machine learning systems themselves. The study emphasizes that future informatics teachers must understand not only how to use AI technologies but also how to protect them from potential risks. Data anonymization, encryption, and integrity verification are essential mechanisms for safeguarding sensitive information used to train models. Students participating in the research learned to apply hashing techniques to verify data authenticity and detect unauthorized modifications.

The results demonstrate that integrating security practices into machine learning education fosters responsible use of technology. For instance, students explored scenarios involving malicious data entry, where corrupted datasets could lead to incorrect predictions. Through hands-on exercises, learners applied outlier-detection methods and integrity checks to ensure that only

reliable data were used to train models. This approach helped develop critical awareness of data quality issues and emphasized the ethical dimension of artificial intelligence.

Moreover, the research revealed that understanding the lifecycle of machine learning systems - from data collection to model deployment - is crucial for developing comprehensive information security skills. Prospective teachers who engaged with encryption methods and hashing mechanisms demonstrated improved understanding of the confidentiality, integrity, and availability principles central to modern cybersecurity frameworks.



```
import hashlib

mystring = input('Хэштеуге арналған құпия сөзді енгізіңіз:')
hashteu = hashlib.md5(mystring.encode())
print(hashteu.hexdigest())

Хэштеуге арналған құпия сөзді енгізіңіз:Astana2024@
f18ad626148252bdae56a7c7af1677
```

```
import hashlib

mystring = input('Хэштеуге арналған құпия сөзді енгізіңіз:')
hashteu = hashlib.sha256(mystring.encode())
hex_dig = hashteu.hexdigest()
print(hex_dig)

Хэштеуге арналған құпия сөзді енгізіңіз:Astana2024@
9901a9a786e66b14a4120a2474b0b477bec1ae0d340a9e2b254e0bc35008e7c9
```

Figure 1. Extract from the practical work related to hashing

The study also highlights the growing relevance of defending machine learning models against adversarial threats. Data poisoning attacks, in which malicious actors manipulate training data to influence model outcomes, were examined as a key educational scenario. Students learned to apply hashing mechanisms and model watermarking techniques as strategies to protect intellectual property and maintain the reliability of AI systems.

The findings indicate that introducing these advanced cybersecurity topics enhances students' readiness to address emerging digital threats. Model watermarking, for example, helped learners understand how machine learning models can be protected against unauthorized copying or tampering. By integrating such practices into educational tasks, the research demonstrates that future informatics teachers gain deeper insight into both the technical and ethical aspects of artificial intelligence security.

A significant contribution of the research is the development of a theoretical framework to improve the training of future computer science teachers in information security. This framework integrates machine learning technologies with hashing, signal detection, and digital verification processes. The implementation included practical tasks such as verifying audio recordings using encryption and hashing algorithms, detecting abnormal behavior in video recordings, and ensuring the integrity of multimedia files through cryptographic methods.

The findings show that working with real devices — such as digital voice recorders, video recording systems, and signal detection equipment — helps bridge the gap between theoretical cybersecurity concepts and practical application. Students gained experience in verifying the authenticity of audio files, detecting suspicious actions captured on video, and implementing signal-blocking mechanisms to prevent unauthorized communication. These activities contributed to the formation of interdisciplinary competencies that combine programming, cybersecurity, and pedagogical design.

The research confirms that machine learning algorithms can play a crucial role in preventing information threats within educational environments. One of the main findings is that integrating threat detection models into the learning process improves students' ability to recognize malware,

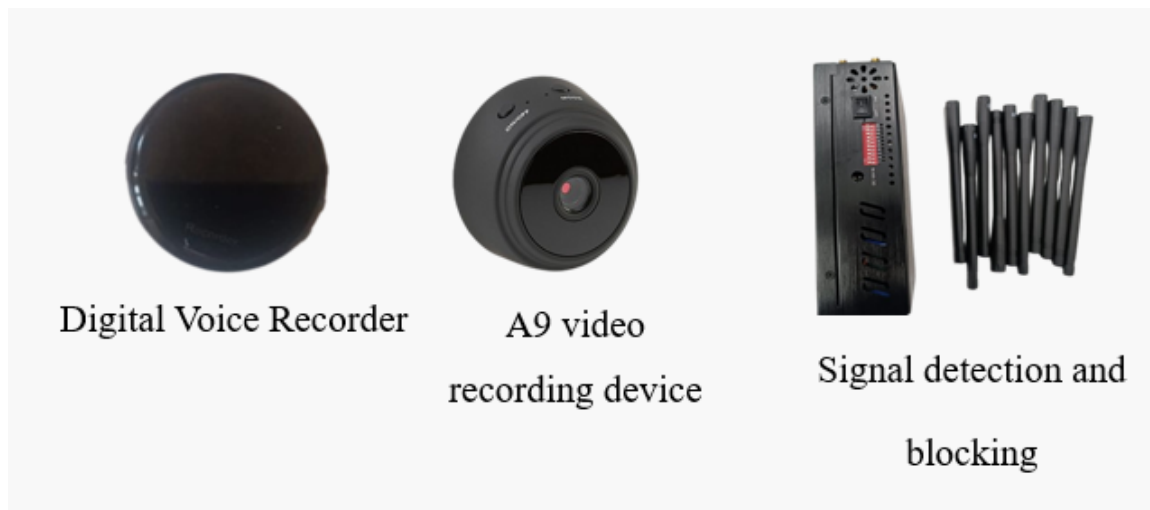


Figure 2. Devices used in practical work

analyze network traffic, and identify anomalies associated with intrusion attempts. Algorithms such as Isolation Forest and Autoencoders were used to simulate intrusion detection systems, allowing learners to classify normal and suspicious activities.

Another important outcome is the use of natural language processing techniques to detect spam and phishing messages. By analyzing textual data, students developed skills in recognizing social engineering tactics and protecting digital communication channels. The research also demonstrates that analyzing user behavior through machine learning models helps predict unauthorized access attempts and supports proactive cybersecurity strategies.

```

213 detector = RandomForestClassifier(max_depth=10)
214
215 X = detector.fit(X_train, y_train)
216
217 X_test, y_test, X_val, y_val = train_test_split(X, y, test_size=0.2, random_state=42)
218
219 y_pred = detector.predict(X_test)
220
221 print('Accuracy on test data: ', accuracy_score(y_test, y_pred))
222
223 # Predict on training data
224 y_train_pred = detector.predict(X_train)
225
226 print('Accuracy on training data: ', accuracy_score(y_train, y_train_pred))
227
228 # Predict on validation data
229 y_val_pred = detector.predict(X_val)
230
231 print('Accuracy on validation data: ', accuracy_score(y_val, y_val_pred))
232
233 # Feature importance
234 importances = detector.feature_importances_
235
236 print('Feature importances: ', importances)
237
238 # Save the model
239 joblib.dump(detector, 'detector.pkl')
240
241 print('Model saved successfully.')

```

Figure 3. Excerpt from practical work

Furthermore, the study explored predictive analytics for forecasting cyberattacks, including early detection of distributed denial-of-service (DDoS) threats. The findings indicate that exposure to predictive modeling strengthens students' understanding of preventive cybersecurity measures and enhances their ability to design secure digital learning environments.

The integration of biometric authentication technologies, such as deep-learning-based face recognition, was identified as another important educational component. Students worked with convolutional neural networks and recurrent neural networks to explore the principles of biometric identification. This experience contributed to their understanding of modern authentication methods and highlighted the importance of privacy protection when working with biometric data.

In addition, the research examined the relationship between cryptography and machine learning. Learners analyzed encrypted datasets and explored methods for protecting sensitive user information. By combining cryptographic techniques with AI-driven analysis, students developed skills in safeguarding digital identities and preventing unauthorized access to data.

The findings suggest that integrating cryptography into machine learning education enhances awareness of secure system design and encourages ethical decision-making.

From a pedagogical perspective, the research demonstrates that incorporating machine learning into teacher education supports the development of inquiry-based and research-oriented approaches to learning. Students were encouraged to investigate cybersecurity problems independently, analyze experimental results, and reflect on the implications of AI technologies in educational settings. This process contributed to the development of higher-order thinking skills, including critical analysis, problem-solving, and ethical reasoning.

The study also found that prospective informatics teachers who engaged with machine-learning projects showed increased motivation and confidence in using digital technologies in teaching. They developed the ability to design interactive lessons that integrate cybersecurity topics, thereby preparing them to educate school students about safe and responsible technology use.

Experimental work was conducted to evaluate the effectiveness of integrating machine learning technologies into the development of information security skills among prospective informatics teachers. The results of the ascertaining and formative stages demonstrated clear dynamics across the motivational, content-related, and technological components. In the experimental group, positive changes were observed across all three components and were statistically significant. The motivational component showed a noticeable shift from low and medium indicators toward higher levels of engagement, indicating increased interest in information security and machine learning applications. Similarly, improvements in the content-related component confirmed the expansion of theoretical knowledge and the ability to apply cybersecurity concepts in practical tasks. The technological component also showed stable growth, reflecting the development of practical skills in using machine learning algorithms, hashing mechanisms, and threat-detection tools.

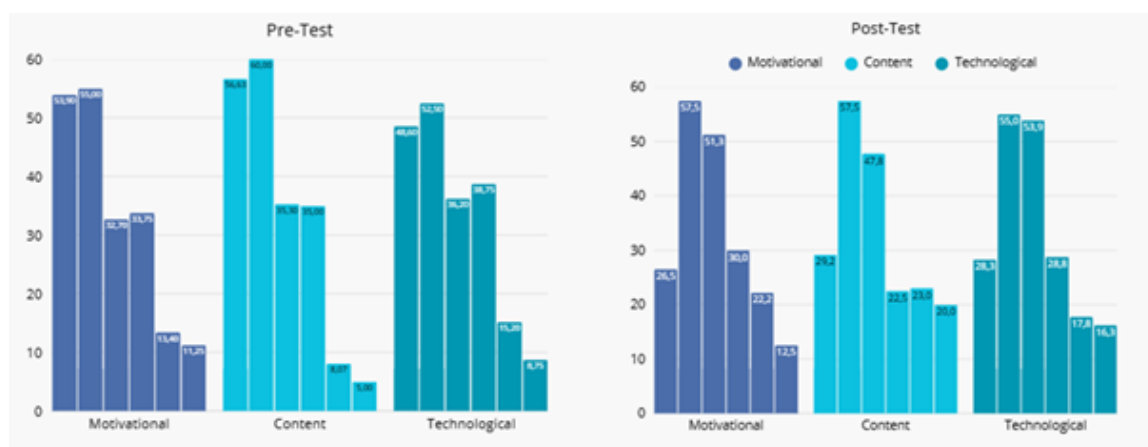


Figure 4. Results of experimental work on improving the training of students in information security at the University

Statistical analysis using the chi-square criterion confirmed the reliability of the obtained results: for the motivational component $\chi^2 = 17.7025$ ($p = 0.001$), for the content-related component $\chi^2 = 20.2495$ ($p < 0.001$), and for the technological component $\chi^2 = 11.51$ ($p = 0.004$). Since all p -values are below the accepted significance level, the differences between the control and experimental groups can be considered statistically significant. These findings indicate that the proposed pedagogical model and the integration of machine learning technologies positively impacted the development of information security competencies among future computer science teachers. Overall, the experimental results confirm the effectiveness of the developed methodological approach and demonstrate its potential for broader implementation in teacher education within the digital environment.

4 Discussion

The results of this study confirm that integrating machine learning technologies into the preparation of prospective informatics teachers positively influences the development of information security competencies. Compared with traditional instructional approaches, ML-based practical activities enabled students to engage more actively with cybersecurity problems, which contributed to higher motivation and a deeper understanding of information security concepts. The experimental results indicate that the inclusion of applied machine learning tasks supports both conceptual understanding and practical skill development in cybersecurity education.

From a pedagogical perspective, the findings suggest that machine learning should not be viewed solely as a technological tool but also as an educational instrument that supports the development of analytical thinking, problem-solving abilities, and responsible digital behavior. The integration of intelligent systems into teacher education aligns with contemporary trends in competency-based learning and the growing emphasis on artificial intelligence within educational environments.

The novelty of this research lies not in the development of new machine learning algorithms, but in the design and empirical validation of a pedagogical model that systematically integrates existing machine learning technologies into competency-based information security training for prospective informatics teachers. The study demonstrates how machine learning methods can be incorporated into practical educational activities that combine cybersecurity analysis, data integrity verification, and digital signal detection.

It is important to clarify that the control group followed a traditional instructional approach based on theoretical lectures, standard laboratory exercises, and the analysis of predefined cybersecurity cases. Both the control and experimental groups studied the same core cybersecurity content; however, only the experimental group participated in applied learning activities that incorporated machine-learning-based analytical tasks. This methodological difference allowed the study to evaluate the pedagogical impact of ML-supported instruction.

Despite the positive results, the study has several limitations. The experimental sample was relatively small and the research was conducted within a single institutional context. These factors may influence the generalizability of the findings. Therefore, future research should involve larger and more diverse samples of students from multiple institutions to validate the proposed pedagogical model across different educational environments.

Overall, the findings highlight the potential of machine learning technologies to enhance the professional preparation of prospective informatics teachers. By integrating artificial intelligence tools into cybersecurity education, teacher training programs can better prepare future educators to address emerging digital threats and to promote responsible and secure use of technology in educational settings.

5 Conclusion

The conducted research confirms that integrating machine learning technologies into the training process significantly contributes to the development of information security skills among prospective informatics teachers in the digital environment. The study demonstrated that combining modern artificial intelligence methods with pedagogical approaches supports not only the improvement of technical competencies but also the development of analytical thinking, research abilities, and responsible digital behavior. The findings indicate that exposure to practical machine learning applications enables students to better understand real-world cybersecurity challenges and to apply theoretical knowledge in authentic problem-solving situations.

An important outcome of the research is the development of a pedagogical framework that integrates machine learning algorithms, cybersecurity practices, and digital signal detection tools within teacher education. Practical activities such as malicious URL detection, user behavior

analysis, multimedia data integrity verification, and the use of encryption and hashing techniques provided students with hands-on experience in secure data processing. The use of real devices and experimental tasks helped bridge the gap between theoretical cybersecurity knowledge and practical professional skills required of future informatics teachers.

The experimental results confirmed the effectiveness of the proposed approach. Positive dynamics were observed in motivational, content-related, and technological components of information security competence, and statistical analysis demonstrated significant differences between the control and experimental groups. These results suggest that incorporating machine learning technologies into teacher education enhances students' readiness to identify cyber threats, protect digital resources, and design secure digital learning environments.

Future research should focus on expanding the experimental sample and validating the proposed pedagogical model across multiple institutions and educational contexts. Further studies may also explore the development of adaptive AI-supported learning platforms and examine ethical, legal, and methodological aspects of artificial intelligence integration in teacher education. The findings of this study provide a foundation for modernizing informatics teacher training programs and for developing sustainable models for strengthening information security competencies in the context of ongoing digital transformation.

Resources

- Chen, Lijia, Peng Chen, and Zhijian Lin (2020). "Artificial Intelligence in Education: A Review". In: *IEEE Access* 8, pp. 75264–75278. DOI: 10.1109/ACCESS.2020.2988510.
- European Commission (2020). *Digital Education Action Plan 2021–2027: Resetting Education and Training for the Digital Age*.
- Al-Fuqaha, Ala, Mohsen Guizani, Mehdi Mohammadi, Moussa Aledhari, and Moussa Ayyash (2015). "Internet of Things: A Survey on Enabling Technologies, Protocols, and Applications". In: *IEEE Communications Surveys & Tutorials* 17.4, pp. 2347–2376. DOI: 10.1109/COMST.2015.2444095.
- Holmes, Wayne, Maya Bialik, and Charles Fadel (2022). *Artificial Intelligence in Education: Promises and Implications for Teaching and Learning*. Center for Curriculum Redesign.
- Miao, Fengchun, Wayne Holmes, Ronghuai Huang, and Zhang Hui (2021). *AI and Education: Guidance for Policy-makers*.
- Möhring, Michael, Ralf Schmidt, and Benedikt Keller (2023). "Machine Learning Approaches for Cybersecurity Education: A Systematic Review". In: *Education Sciences* 13.5, p. 512. DOI: 10.3390/educsci13050512.
- Ng, Davy Tsz Kit, Jack Ka Lok Leung, Samuel Kai Wah Chu, and Maggie Shuyu Qiao (2021). "AI Literacy: Definition, Teaching, Evaluation and Ethical Issues". In: *Computers and Education: Artificial Intelligence* 2, p. 100041. DOI: 10.1016/j.caeai.2021.100041.
- Papamitsiou, Zacharoula and Anastasios A. Economides (2016). "Learning Analytics and Educational Data Mining in Practice: A Systematic Literature Review". In: *Educational Technology & Society* 19.2, pp. 49–64.
- Zawacki-Richter, Olaf, Victoria I. Marín, Melissa Bond, and Franziska Gouverneur (2019). "Systematic Review of Research on Artificial Intelligence Applications in Higher Education". In: *International Journal of Educational Technology in Higher Education* 16.1, p. 39. DOI: 10.1186/s41239-019-0171-0.

Proposal for an Information System for Insurance Contract Management

Silvia Zelinová¹ Tomáš Šuhaj²

Abstract

The aim of this thesis is to develop an information system for managing life insurance contracts that will serve not only as contract records but also for tracking data such as policyholders, documents, insurance companies, additional services, payments, and reminders. The paper focuses on both the overall technical design and the implementation of individual components, including the frontend (FE), backend (BE), and database. Additionally, it introduces the NPM ecosystem and the TypeScript programming language, which were chosen for implementation following a comparison of available options. Information systems have become an essential part of most modern organizations and processes. They represent a combination of technologies, processes, and people that work together to collect, process, store, and distribute information, thereby supporting decision-making, increasing efficiency, and automating activities. The paper explains the implementation approach for the information system's frontend, backend, and database. It presents concrete examples covering authentication, data layer definition, and basic API interactions. The paper's output is the complete product, presented as screenshots of individual screens.

Keywords

life insurance, information system, typescript, node.js, react library

1 Introduction

The implementation of modern information systems has become a key factor in the operation of an increasing number of organizations, as it replaces inefficient contract management based on physical documents and heterogeneous spreadsheet records. This traditional approach leads to significant limitations, particularly in the efficient administration, modification, and retrieval of data. The objective of this paper is to design and implement a web-based information system for the management of life insurance contracts, enabling user registration and authentication as well as efficient handling of contracts and related data, including their creation, modification, filtering, and export. The theoretical section of the paper addresses the fundamental concepts of insurance, the definition and benefits of information systems, the system development life cycle, and the technologies employed, including TypeScript, Node.js, and frontend-backend architecture, as well as security considerations and agile development practices. The practical section focuses on selecting appropriate technologies for the frontend, backend, and database layers, specifically React, Nest.js, PostgreSQL, and Docker.

Subsequently, the system implementation is presented, consisting of a web interface communicating with a REST API that ensures request validation, authentication, authorization, and database operations via the Prisma ORM framework. The paper further describes the project structure, the applied frontend and backend solutions, and the methodology used for testing the application. In the final section, the completed information system is introduced as a finished product through visual documentation of its functionality.

¹Department of Mathematics and Actuarial Science, Faculty of Economic Informatics, Bratislava University of Economics and Business, Slovakia. Email: silvia.zelinova@euba.sk

²Slovenská sporiteľňa a. s., Tomášikova 48, Bratislava, Slovakia, Email: t.suhaj@gmail.com.

2 Life Insurance and Its Role in Actuarial Science

Life insurance is a voluntary insurance product offered by commercial insurance companies that provides financial protection to individuals and their families in the event of unexpected events such as death, injury, serious illness, or disability. It is often used by individuals with families and financial obligations, such as mortgages or loans. In addition to risk protection, it may in some cases also serve as an investment instrument. The main types of life insurance include term, endowment, and investment life insurance (NN Slovakia, 2025). Term life insurance provides payment of the agreed sum in the event of death, but does not allow the appreciation of invested funds. Endowment life insurance combines insurance protection with savings and guaranteed returns. Investment life insurance enables part of the premium to be invested in investment funds while maintaining insurance coverage.

Actuarial science is a discipline focused on the analysis and management of risks in insurance and finance using mathematical and statistical methods. Its origins date back to the 18th century in Great Britain. A significant milestone was the establishment of the Institute of Actuaries in London in 1848. In the 20th century, the field expanded to the United States and developed substantially with the introduction of stochastic methods. Today, actuarial science overlaps with data science and plays an important role in investments, pension systems, and social policy. An actuary is a professional who calculates insurance premiums, assesses risks, and designs investment strategies. Actuaries work not only in insurance but also in banking and data analytics. Their responsibilities include calculating reserves, risk valuation, financial planning, and reporting (Páleš et al., 2021). Due to the growing importance of risk management, actuarial science is considered a promising and highly demanded profession.

3 Information Systems and System Development

Information systems are an integral part of modern organizations and business processes. They represent a combination of technologies, processes, and people that enable the collection, processing, storage, and distribution of information. Their main purpose is to support decision-making, increase efficiency, and automate work activities. Through automation, they significantly reduce the time required for manual tasks. At the same time, they contribute to higher employee productivity. Their main benefits include process automation, information availability, data security, scalability, accuracy, and efficient data organization (Indeed, 2025). Internet-based access enables users to access up-to-date information from any location in real time. This feature is particularly important for remote work or client meetings. Information systems also reduce the risk of human error when handling large volumes of data. As a result, they increase the overall reliability of information processing.

In the field of life insurance, information systems play a crucial role in contract management. They enable centralized storage of all contracts and related data in a single location. This significantly simplifies their registration, updating, and monitoring. Automatic notifications alert users to upcoming premium due dates or contract renewals. This minimizes the risk of delayed payments or administrative errors. The systems also support the generation of reports and statistical outputs on the status of the insurance portfolio. These outputs are important for managerial decision-making and external reporting. They play a significant role in meeting regulatory requirements such as Solvency II and IFRS 17. They also enable the preparation of data for financial analyses and insurance company financial statements. An important function of information systems is ensuring compliance with current legislation. The systems can effectively respond to legislative changes and implement them into the organization's internal processes.

The development of an information system proceeds through several consecutive phases that ensure its quality and long-term sustainability (Intellectsoft, 2023). The first phase is planning and analysis of customer and future user requirements. In this phase, functional and

non-functional system requirements are defined. This is followed by the design of the system architecture, including the selection of technologies, databases, and security mechanisms. An important part of the design phase is creating the user interface. The programming phase includes implementing frontend and backend components. It also involves integration with databases and external services. After implementation, the testing phase follows, including unit, integration, and end-to-end tests. After successful testing, the system is deployed into the production environment. In the final phase, the system undergoes continuous maintenance, updates, and functional expansion according to the evolving needs of the organization.

3.1 Literature review

Research on insurance information systems emphasizes the growing need for digitized policy administration systems that support the entire lifecycle of insurance contracts, from issuance to changes and renewals, as core components of insurers' architectures. Recent studies have explored blockchain-based approaches to policy management, demonstrating how standardized models can enhance interoperability and streamline the administration of multiple policies across entities. Groll, Wasserfuhr, and Zeldin (Groll et al., 2022) investigated machine learning models to predict policyholder behavior, illustrating the value of structured contract data for analytics and decision support in life insurance systems. Ellili (Ellili, 2023) conducted a systematic review of big data applications in insurance, highlighting the growing impact of advanced data processing on insurance operations and decision-making. Hoffmann (Hoffmann, 2021) analysed the use of smart contracts in insurance markets, showing how automation and blockchain technologies can support contract execution and enforcement. Although not focused solely on IT architectures, Sadowski (Sadowski, 2024) surveyed developments in insurtech, framing how digital tools are reshaping insurance processes, including contract management. Studies in *Management Information Systems Quarterly* and *Information Systems Research* provide broader theoretical foundations for how IS development practices (such as modular FE/BE separation and API-based integration) contribute to effective enterprise systems, including those in insurance. Empirical research on MIS in insurance contexts (Joseph, 2022) shows that integrated information systems significantly influence the efficiency and depth of insurance services, including record-keeping and transaction processing. Across the literature, there is consensus that adopting advanced information technologies leads to improvements in process automation, data accuracy, and real-time access to contract data. Summarizing existing research reveals that well-designed information systems not only centralize contract data but also support analytics, compliance, and service delivery, aligning with the objectives of your proposed insurance contract management system.

4 Selection of suitable technologies for the information system

Before implementing an information system, it is necessary to collect information and analyse the issue to define its functional and non-functional requirements, as well as the database schema, because these factors can significantly affect the selection of technologies and procedures to be used during implementation. Selecting an appropriate technology is a critical step in implementing all system components, as the development of an information system is not limited to achieving functional correctness but also involves the efficient use of suitable tools. These tools continuously evolve over time. In the context of software development, it is therefore essential to consider not only the efficiency of the tools but also the maturity and adoption rate of a given technology within the developer community and industry. This ensures long-term sustainability, facilitates future team expansion, and guarantees adequate substitutability of team members

4.1 Comparison of Popular JavaScript Frameworks for Front-End Development

The front-end development community is widely regarded as one of the most dynamic segments of the software engineering ecosystem, primarily due to the frequent emergence of new frameworks and technologies that promise increasingly simplified approaches to user interface development. For the purposes of this study, three of the most widely used open-source frameworks and libraries were considered: React.js, Angular, and Vue.js. Based on data reported by (BrowserStack, 2025) and our own practical experience, their key differences can be summarized as follows. Angular, originally released as AngularJS, was introduced in 2010 by a development team at Google. Unlike the other candidates, Angular is a comprehensive framework rather than a standalone library. In practice, this means application development does not require integrating third-party libraries or NPM packages to support essential functionality, such as global state management or HTTP communication, as Angular provides its own HTTP client and a built-in solution for managing global application state. Furthermore, Angular is the only evaluated option that employs dependency injection (DI) as a core architectural principle, thereby enabling improved code modularity, scalability, and testability. Nevertheless, the strict conventions imposed by the framework regarding application structure and development practices may not be universally suitable, particularly given the comparatively steep learning curve compared with the remaining alternatives.

Vue.js was created by Evan You, a former Google engineer, and its initial stable version (1.0) was released in 2014. Similar to Angular, Vue.js is a framework rather than a library, which allows developers to rely on integrated solutions such as Pinia for global state management and Vue Router for efficient navigation between views in single-page applications. Vue.js also supports two-way data binding, enabling immediate synchronization between the data model and the user interface, for example, when working with form inputs. A notable feature that improves the readability and maintainability of Vue.js code is the use of framework-specific directives, which can be embedded directly in HTML templates. These directives enable the reactive behavior of interface elements, which is essential for most interactions in modern web applications. Additionally, Vue.js provides built-in components such as `<Transition>` and `<TransitionGroup>` to facilitate animation implementation. Component development can be performed using several approaches, with the most common being Single-File Components, which encapsulate JavaScript logic, HTML templates, and CSS styling in a single file.

React.js differs from the aforementioned technologies primarily in that it is a library for building user interfaces rather than a full framework. It was introduced in 2013 by Meta (formerly Facebook), initially as an internal tool. At present, React.js is the most widely adopted technology for front-end development, as shown in the percentage distribution of front-end-related questions by framework. When developing the BE part of our information system, we have a wider range of platforms at our disposal, whereas FE development depends on JavaScript, the only language capable of transforming a static web page into a dynamic web application. BE offers a larger number of platforms and ecosystems for developers to create. Among the most widely used and time-tested technologies are Java and .NET, which are often chosen for the development of large systems in large companies, given that they are not only programming languages but also platforms. In the case of Java, we have the Spring framework, which solves common problems faced by most applications and provides simple solutions, such as database communication, security, dependency management, support for microservice architecture and the cloud, testing and debugging tools, and so on. On the .NET platform, the ASP.NET Core framework offers similar convenience. In addition to these platforms, the Python and PHP programming languages are worth mentioning, as they make up a large share of the BE components of information systems. In Python, it is recommended to use the Django or FastAPI frameworks. PHP provides frameworks such as Symfony and Laravel, which are constantly evolving and are increasing interest in the language, which has been declining slightly in the long term. For the purposes of

this work, we will choose one of the options offered by the NPM ecosystem, which allows us to use the same language for both the FE and BE parts of the information system. Thanks to Node.js, it is possible to write JavaScript code that runs outside the browser. For application development in this environment, libraries such as Express.js or the Nest.js framework are available. When choosing the Express.js library, we have complete freedom to structure the code on the BE, unlike the Nest.js framework, which, like Angular, provides a predefined structure that the project must adhere to during development. Given our positive experience with this framework in the context of BE application development, we will choose it for implementing the BE part of the information system. (Jones, 2024) Since we have chosen Nest.js for our implementation, we can be sure we can perform the most common BE operations without adding any third-party dependencies.

4.2 Data Layer Design of a Life Insurance Information System

The data layer of the proposed information system was designed after selecting the core implementation technologies and analysing the domain of life insurance contract management. Its primary purpose is to define stored data, their data types, and the relationships among entities to support reliable contract administration. The model distinguishes between modifiable entities and predefined reference entities (codebooks), such as countries, contract types, or contract statuses, which ensure data consistency and simplify user input through controlled selections. Entity relationships are defined using cardinalities (1:1, 1:N, and N:M) and implemented using primary and foreign keys. Primary keys uniquely identify records, typically using integer identifiers or UUIDs, while foreign keys establish referential integrity between tables. The central entity of the model is the life insurance contract, which contains extensive attributes, including identification data, links to the insurance company, user, and client, financial indicators (premium, debt, account value), validity period, and contract status. Surrounding entities include User (the root entity), Client (the insured person), Insurance Company, Document, and several associative tables that enable many-to-many relationships, such as beneficiaries' shares, contractual documents, and additional services assigned to contracts. This structured relational design ensures scalability, data integrity, and efficient querying, forming a robust foundation for the insurance contract management system.

5 Implementation of the front-end and back-end part

The system's front end was implemented using React, TypeScript, and the Vite build tool. The project follows a modular directory structure that separates API communication, UI components, configuration files, custom hooks, layouts, pages, and global state management. This design significantly improves the application's maintainability and extensibility.

Communication with the backend is handled by a generic Axios-based function returning a standardized `GenericResponseDTO` object containing HTTP status codes and payload data. Axios interceptors were implemented to automatically attach JWT tokens to outgoing requests and refresh them when expired.

Data structures are defined using TypeScript interfaces to ensure type safety and consistent data exchange. Application routing is managed by React Router using the Browser Router approach and protected routes that restrict access based on user authentication status. Global application state, including authentication data and UI preferences such as colour theme, is managed using the Zustand library.

When working with the Zustand library for global state management, it is necessary to create storage objects depending on their purpose. In figure 3, we can see the definition of the storage used to store the current colour scheme. These are light (white) and dark (black) schemes, which can be switched using the switch in the top bar. The switch uses a `useTheme` hook connected to the `themeStore` Zustand storage. The storage object is created with an attribute containing the

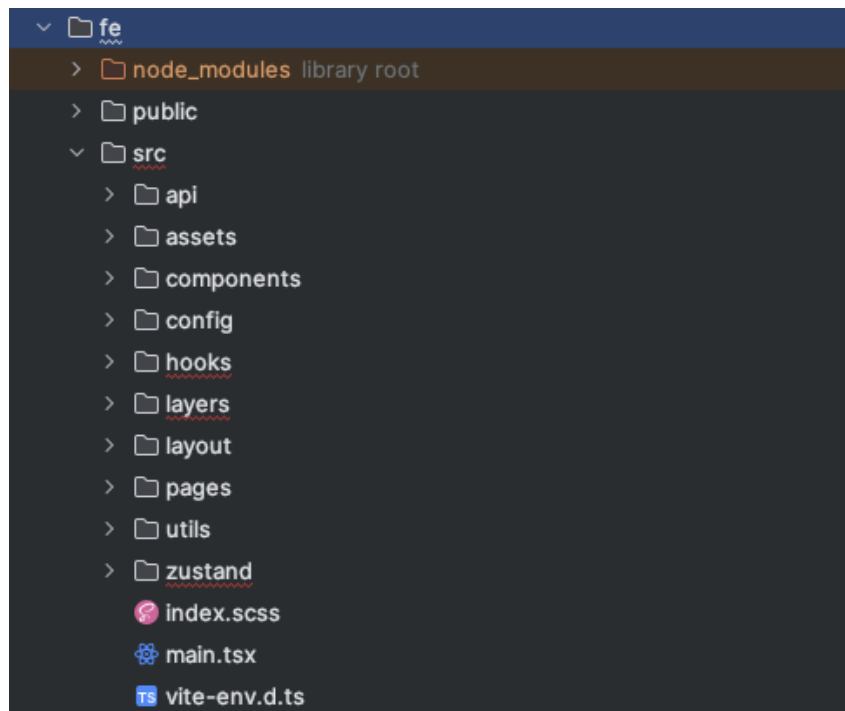


Figure 1. Front-end project directory structure

```
10  api.interceptors.request.use(  = Tomáš Šuhaj
11    onFulfilled: config : InternalAxiosRequestConfig<any... => {
12      const { accessToken : string | undefined } = useAuthStore.getState();
13      if (accessToken) {
14        config.headers['Authorization'] = `Bearer ${accessToken}`;
15      }
16      return config;
17    },
18    onRejected: error => {
19      return Promise.reject(error);
20    }
21  );
```

Figure 2. Integration of JWT authentication logic into Axios HTTP requests

scheme type and the toggleTheme function to change its state, after which all components that use it are redrawn.

The backend was developed using the NestJS framework, which supports a modular architecture. Each functional domain is encapsulated in a separate module consisting of controllers, services, guards, middleware, and interceptors. This structure ensures a clear separation between business logic and request handling.

Authentication is based on JWT tokens generated using NestJS utilities. Separate strategies are implemented for access and refresh tokens, and custom decorators provide direct access to the identity of the currently authenticated user inside controllers. HTTP request processing includes DTO validation with class-validator, authorization checks, database operations via the Prisma ORM, and the generation of standardized API responses. Database management is handled using PostgreSQL deployed in Docker containers. Schema evolution is controlled using Prisma migration scripts.

API functionality was manually tested using Postman, focusing on the correctness of operations, validation, authentication, authorization, and data integrity. Both front-end and back-end components support hot-reload development mode and optimized production builds.

```
1 import { create } from 'zustand';
2
3 export const LOCAL_STORAGE_THEME_KEY : "app-theme" = 'app-theme'; Show usages 1 Tomáš Šuhaj
4
5 interface IThemeStoreState { Show usages 1 Tomáš Šuhaj
6   theme: THEME;
7 }
8
9 interface IThemeStoreActions { Show usages 1 Tomáš Šuhaj
10  toggleTheme: (theme: THEME) => void;
11 }
12
13 export enum THEME { Show usages 1 Tomáš Šuhaj
14   DARK = 'dark',
15   LIGHT = 'light',
16 }
17
18 const defaultTheme: THEME = THEME.DARK;
19
20 export const useThemeStore : UseBoundStore<StoreApi<IThemeS... = create<IThemeStoreState & IThemeStoreActions>({
21   set => ({
22     theme: (localStorage.getItem(LOCAL_STORAGE_THEME_KEY) ??
23       defaultTheme) as THEME,
24
25     toggleTheme: (theme: THEME) =>
26       set(state : IThemeStoreState & IThemeStore... => {
27         localStorage.setItem(LOCAL_STORAGE_THEME_KEY, theme);
28         return { theme: theme };
29       }
30   })
31 });
```

Figure 3. Example of working with the Zustand library, implementation of colour mode switching



Figure 4. Backend file structure

```

22 @Controller( prefix: 'agreements') Show usages
23 export class AgreementsController {
24   constructor(private readonly prisma: PrismaService) {} no usages
25
26   @Get( path: '/' ) no usages
27   @HttpCode( HttpStatus.OK )
28   async list(
29     @GetCurrentUserId() userId: number
30   ): Promise<DataResponseDTO<AgreementDTO[]>> {
31     const agreements : (GetFindResult = await this.prisma.agreement.findMany( args: {
32       where: { userId },
33       include: {
34         client: { select: { name: true } },
35         agreementStatus: { select: { name: true } },
36         AgreementType: { select: { name: true } },
37         insuranceCompany: { select: { name: true } },
38       },
39     });
40
41     const finalData : ( = agreements.map( agreement : GetFindResult<Prisma.$AgreementPayload<$Exte..., => ({
42       ..agreement,
43       clientName: agreement.client?.name,
44       agreementStatusName: agreement.agreementStatus?.name,
45       agreementTypeName: agreement.AgreementType?.name,
46       insuranceCompanyName: agreement.insuranceCompany?.name,
47       agreement: undefined,
48       insuranceCompany: undefined,
49       agreementStatus: undefined,
50       client: undefined,
51     }));
52
53     return {
54       statusCode: HttpStatus.OK,
55       data: finalData,
56     };
57   }

```

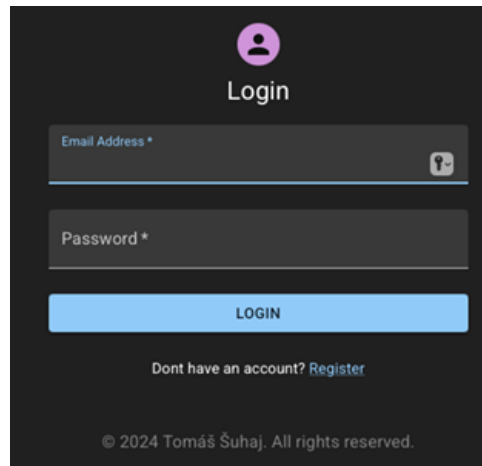
Figure 5. AgreementsController method returning user contracts

6 Demonstration of an information system for managing insurance contracts

This chapter presents the results of our work so far, including all screens from the FE part of our information system. After successfully registering in the system, we began entering the data provided by Zurich Insurance Group, Ltd. The data for this work were provided by Zurich Insurance Company Ltd, an organizational unit. This is artificially produced data that has nothing to do with the business activities and financial situation of Zurich Insurance Company Ltd, an organizational unit, or the business activities and financial situation of other companies belonging to the Zurich Insurance Group.

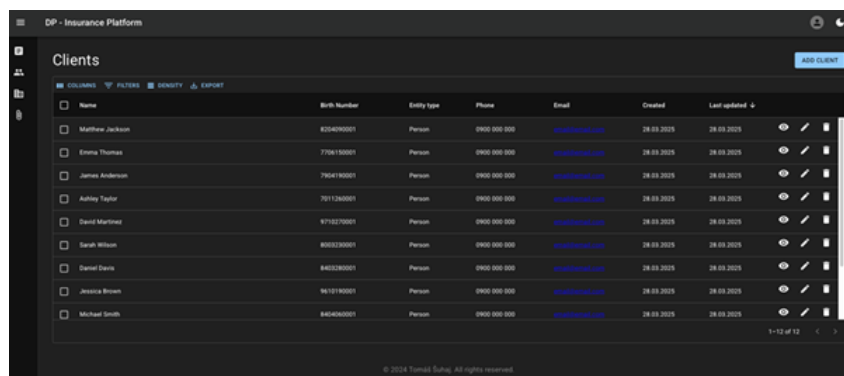
The following figure 6 shows the login form for already registered users of our information system. After entering the email address and password, an HTTP request is sent to the BE part of the information system. After entering the managed email and password, the user will be redirected to a list of all contracts they have created. Otherwise, he will receive an error message stating that the given account does not exist.

If the user does not yet have an account in the information system, they must register using the registration form in order to gain access. The user enters their full name, email address, date of birth, country of origin, and confirms their consent to the processing of personal data. After the data is successfully processed in the system's backend, the user is automatically redirected to the application, just as with a standard login. After a successful login, it is necessary to prepare all the input data required to create specific life insurance contracts. One of the basic steps is to register the insured persons for whom insurance contracts will be concluded. Figure 7 shows a list of all policyholders registered in the system. The user can create a new record using the appropriate button, and view, edit, and delete existing records via icons in each row of the table. When creating a new insured person or editing an existing record, the user is redirected to a form to create, edit, or view the client. After registering the insured persons, the insurance companies with which insurance contracts will be concluded are registered. This process is conceptually



The login form features a dark background with a purple user icon at the top. Below the icon is the word "Login" in white. There are two input fields: "Email Address *" and "Password *", both with light blue borders and a small eye icon for password visibility. A prominent blue "LOGIN" button is centered below the fields. At the bottom, there is a link "Dont have an account? Register" and a copyright notice "© 2024 Tomáš Šuhaj. All rights reserved."

Figure 6. Login form



The "Clients" table in the "DP - Insurance Platform" interface displays a list of registered clients. The table has columns for Name, Birth Number, Entity type, Phone, Email, Created, and Last updated. Each row includes a checkbox for selection and icons for view, edit, and delete actions. The data is as follows:

Name	Birth Number	Entity type	Phone	Email	Created	Last updated
Matthew Jackson	820490001	Person	0900 000 000	matthew@dp.com	28.03.2025	28.03.2025
Emma Thomas	7706130001	Person	0900 000 000	emma@dp.com	28.03.2025	28.03.2025
James Anderson	7804190001	Person	0900 000 000	james@dp.com	28.03.2025	28.03.2025
Ashley Taylor	7011260001	Person	0900 000 000	ashley@dp.com	28.03.2025	28.03.2025
David Martinez	6710270001	Person	0900 000 000	david@dp.com	28.03.2025	28.03.2025
Sarah Wilson	803330001	Person	0900 000 000	sarah@dp.com	28.03.2025	28.03.2025
David Davis	8403280001	Person	0900 000 000	david@dp.com	28.03.2025	28.03.2025
Jessica Brown	9010190001	Person	0900 000 000	jessica@dp.com	28.03.2025	28.03.2025
Michael Smith	080490001	Person	0900 000 000	michael@dp.com	28.03.2025	28.03.2025

Figure 7. Overview of all clients / policyholders

analogous to the registration of insured persons. The system provides a list of all registered insurance companies with the option to view, edit, delete, or create a new record using the form. The system allows you to view a list of insurance companies and use a form to manage them. Basic attributes such as the insurance company's name, contact details, registered office address, and a brief description are recorded.

After creating all the necessary entities, you can proceed to create the life insurance contracts. The table interface allows you to perform standard operations on individual records and make advanced adjustments to the data display. Given the large number of columns, it is possible to filter records according to selected criteria or display only selected attributes via the control panel. The functionality also includes exporting data to external formats, such as Microsoft Excel, for further processing. The details of the life insurance contract, shown in figure 8, are implemented via a complex form that includes components for selecting predefined entities and code values, such as the insurance company, the insured, the contract status, and the type of insurance contract. In addition, it is possible to define basic insurance parameters, such as the premium amount, the duration of the insurance, the start date of the contract, or the insured person's age.

After creating a basic contract, the user can manage documents assigned to contracts for various purposes. A list of all created documents is shown in figure 9. When creating a new document, the user enters its name, content, and category, which determines how it is used in the system. In the implemented solution, categories are represented by a code list with the values Payments, Reminders, and General Documents. In the details of a life insurance contract, it is possible to assign documents according to their category in the relevant tabs, for example, a tab for payment type documents, a tab for reminder category documents, and a tab for general

DP - Insurance Platform

Agreement Parameters

Agreement name: Životné poisťenie pre osobu Matthew Jackson

Agreement type: Annuity Annual premium: 6563

Agreement period: Zvlášť Insured person: Matthew Jackson

State of operation: 08.04.2002 Date: 3261

Insured person & Insurance company

Insured person: Matthew Jackson Insurance company: Kooperatíva

Figure 8. Creating / Editing / Viewing a Life Insurance Contract

DP - Insurance Platform

Documents

Name	Type of document	Created	Last updated
Všeobecné podmienky poistenia	General document	28.09.2024	28.09.2024
Spomienka - mediacia platba 10/2024	Reminders	27.09.2024	29.09.2024
Platba za obdobie 09/2024	Payments	27.09.2024	29.09.2024
Končica platby	Payments	15.10.2024	15.10.2024
Spisovacia čísla súborov - Životné poisťenie	General document	27.09.2024	15.10.2024

Figure 9. Overview of all documents

documents, which include, for example, general insurance conditions or documentation regarding the processing of personal data by the insurance company.

The system also allows the registration of authorized persons in the event of the insured's death, who will receive the insurance benefit in a defined proportion. The last significant functionality is the ability to assign additional services to the life insurance contract. The user can choose from a set of predefined services that can be freely combined, added, or removed. The proposed information system also provides advanced data analysis tools directly within the user interface. All table reports support advanced filtering, searching, column selection, and data export, enabling efficient processing and interpretation of stored data.

7 Conclusion

In this paper, we focused on developing a modern information system to address the shortcomings of traditional insurance contract records. Manually managed spreadsheets and paper documents have been replaced by an effective digital tool that enables easier data manipulation, better organization, and faster searches. The goal was to create a web application that allows users to manage their clients' contracts, with the system supporting all key operations – from registration and authentication to data management and export. During development, we followed proven software engineering practices. First, we analyzed the issues surrounding life insurance to determine which data needed to be processed and how to model them effectively. Based on this analysis, we designed the system's data structure and database model. We then focused on the development itself, opting for a combination of React for the frontend and Nest.js for the backend. For data storage, we used PostgreSQL with Prisma ORM, which enabled us to define and manage database tables effectively. An important part of the work also focused on system security and stability. We implemented JWT authentication, API protection against unauthorized access, and input data validation. When designing the architecture, we emphasized modularity and scalability, with the backend and frontend communicating via the HTTP protocol. The application was developed with agile principles, enabling continuous testing and integration of

new functionality. The practical part of the work focused on implementing specific solutions. On the frontend, we implemented application state management with Zustand, navigation with React Router, and form validation with Zod. The backend processes requests via REST API, performs authentication, and communicates with the database via Prisma ORM. During development, we used Docker to simplify deployment and dependency management. Finally, we tested the system and documented all key functionalities. We presented the user interface and individual processes using screenshots to give the reader a comprehensive overview of the final solution. The result is a fully functional information system for managing insurance contracts, providing a modern, effective solution for financial advisors and insurance managers. In the future, the system can be expanded to include additional functionalities, enabling broader use in practice. Future research could focus on conducting a comprehensive performance evaluation of the system under real-world workloads to assess scalability, response times, and resource utilization. Another important direction would involve structured user testing with insurance professionals to evaluate usability, workflow efficiency, and overall user experience. Finally, future work may include deploying the system in a real organizational environment to assess its practical integration with existing infrastructure and validate its effectiveness in supporting day-to-day insurance contract management processes.

Acknowledgements

This paper has been supported by the project VEGA no. 1/0497/25 “Implementation of innovative approaches in risk management and modeling within internal models of insurance companies” and by ESG grant A-25-103/3020-17 “Improving students’ knowledge and skills in project management using the IPMA methodology”.

Resources

- BrowserStack (2025). *Angular vs React vs Vue: Which framework to choose?* URL: <https://www.browserstack.com/guide/angular-vs-react-vs-vue> (visited on 02/12/2026).
- Ellili, N. (2023). “The applications of big data in the insurance industry: A systematic analysis”. In: *Insurance: Mathematics and Economics*. URL: <https://www.sciencedirect.com/science/article/pii/S2405918823000181> (visited on 02/12/2026).
- Groll, A., C. Wasserfuhr, and L. Zeldin (2022). *Churn modeling of life insurance policies via statistical and machine learning methods*. arXiv: 2202.09182. URL: <https://arxiv.org/abs/2202.09182> (visited on 02/12/2026).
- Hoffmann, C. H. (2021). “The example of smart contracts in the insurance market”. In: *Journal of Wealth Management*. URL: <https://www.emerald.com/insight/content/doi/10.1108/JWAM-08-2020-0037/full/html> (visited on 02/12/2026).
- Indeed (2025). *What are information systems?* URL: <https://uk.indeed.com/career-advice/career-development/what-are-information-systems> (visited on 02/12/2026).
- Intellectsoft (2023). *What is system development life cycle?* URL: <https://www.intellectsoft.net/blog/what-is-system-development-life-cycle/> (visited on 02/12/2026).
- Jones, A. (2024). *NestJS: The perfect JavaScript backend framework*. Medium. URL: <https://medium.com/@ajonesb/nestjs-the-perfect-javascript-backend-framework-for-structure-clean-code-and-modularity-ae1b3a6e1418> (visited on 02/12/2026).
- Joseph, E. M. (2022). “Effect of insurance management information system on deepening insurance services”. In: *Arabian Journal of Business and Management Review (Kuwait Chapter)*.
- NN Slovakia (2025). *Life insurance: term, endowment and investment life insurance*. Cited in paper as (NN Slovakia, 2025).
- Páleš, M. et al. (2021). *Aktuárstvo*. Letra Edu.

Sadowski, J. (2024). "Total life insurance: Logics of anticipatory control and insurtech developments". In: *Science, Technology & Human Values*. URL: <https://journals.sagepub.com/doi/10.1177/03063127231186437> (visited on 02/12/2026).