# AI-Enhanced Science Education: From Data to Physical Laws

Ervin Szabolcs Murvai[1], Krisztián Hriczó[2], and Attila Károly Varga[3]

[1] Institute of Automation and Infocommunication, University of Miskolc, Miskolc, Hungary
ervin.szabolcs.murvai@student.uni-miskolc.hu
[2] Institute of Mathematics, University of Miskolc, Miskolc, Hungary
krisztian.hriczo@uni-miskolc.hu
[3] Institute of Automation and Infocommunication, University of Miskolc, Miskolc, Hungary
attila.varga@uni-miskolc.hu

**Abstract.** This research explores the integration of Artificial Intelligence (AI) and Internet of Things (Iot) technologies into modern science education to bridge the "abstraction gap" between digital data and physical reality. By leveraging data-driven algorithms, the project transforms everyday smart devices into professional-grade scientific instruments, moving beyond the "black box" usage of 21st-century technology. The methodology employs advanced models such as Long Short-Term Memory (LSTM) networks for thermodynamic forecasting and the SINDy algorithm for the sparse identification of underlying physical laws from noisy kinematic data.
Furthermore, the study demonstrates how AI-enhanced sensors can uncover imperceptible environmental phenomena, such as nocturnal background radiation fluctuations, thereby expanding the scope of human perception in the classroom. Utilizing a Docker-based containerization, the research ensures that complex computational environments are accessible and reproducible for both students and educators. Ultimately, this work provides a scalable framework for interdisciplinary learning, empowering students to shift from passive memorization to active scientific discovery by "rediscovering" the fundamental laws of nature through their own data.

**Keywords:** SINDy, IoT sensors, bridging Physics & IT, physical laws

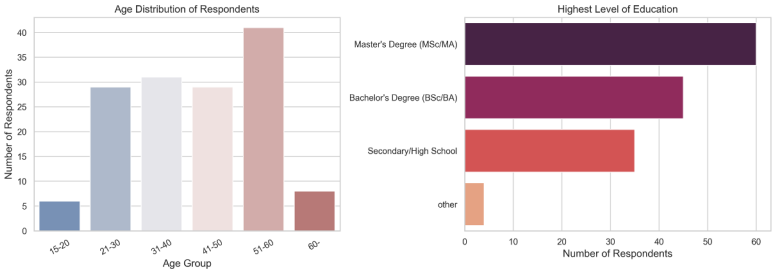
## 1 Introduction

A persistent challenge in science education is the increasing disconnect between theoretical knowledge and direct physical experience. While digital technologies are widely available in classrooms, they are often used as opaque tools rather than as instruments for understanding underlying physical principles.

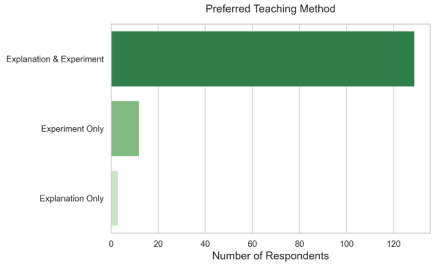
This work explores how AI can be used to bridge this gap by enabling students to interpret and model experimental data. The motivation stems from

© The Author(s) 2026
S. Árvai-Homolya et al. (eds.), *Proceedings of the 2nd Innovation in Engineering Education Conference (IEE 2026)*, Atlantis Highlights in Social Sciences, Education and Humanities 58,
https://doi.org/10.2991/978-94-6239-823-8_13

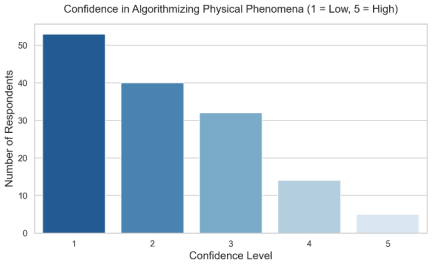
both professional experience in physics education and observations of student engagement, which indicate a strong demand for more hands-on, exploratory learning environments.




(a) Distribution of respondents by age group and educational attainment.




(b) Teaching preferences.




(c) Challenges in algorithmizing.

Fig. 1: Comprehensive analysis of the survey results: (a) demographic data, (b) teaching preferences, and (c) identified gaps in algorithmizing.

To ground the educational motivation of the proposed framework, a preliminary diagnostic survey was conducted among a highly qualified sample ($N = 144$), where the largest segment of respondents held a Master's degree ($n = 60$). The descriptive results strongly validate the necessity of implementing ready-to-use STEAM lesson plans.

Demand for Practicality: A striking 87.5% of participants ($n = 126$) stated that they learn best when concepts are presented through practical examples, and respondents highlighted that grasping physics concepts purely from textbooks without experiments is inherently difficult (Mean = 3.57 out of 5.0).

The Algorithmic Gap: Crucially, when asked about their confidence in algorithmizing a simple physical phenomenon or natural process, respondents reported a very low confidence level, with a mean score of only 2.15 out of 5.0 (Median = 2.0, SD = 1.13).

These preliminary findings expose a clear competency gap. There is an overwhelming demand for experiential learning, yet a significant lack of confidence

in translating physical observations into algorithmic structures. The interdisciplinary IoT and computer framework presented in this paper directly addresses this gap by providing structured, accessible for classrooms.

The central hypothesis of this research is that combining AI with accessible sensing technologies can transform everyday devices into platforms for scientific inquiry, thereby strengthening the connection between measurement, data and physical law.

## 2 Everyday Devices as Experimental Platforms

Modern smart devices continuously collect large amounts of sensor data, including motion, temperature, and location. Despite this, users typically interact only with processed outputs, which obscures the physical processes that generate the data.

In this work, we propose reinterpreting these devices as experimental instruments. Instead of treating them as "black boxes," students are encouraged to analyze raw measurements and build models that explain the observed phenomena and highlight the underlying physical principles.

As an illustrative example, a simple free-fall experiment can be conducted using a camera-based tracking system. By detecting the position of a falling object over time and fitting a kinematic model, it is possible to estimate gravitational acceleration directly from data. This approach emphasizes that physical laws are not merely theoretical constructs, but relationships that emerge from measurement. This experiment was tested using "vibe coding," and the results achieved were acceptable. However, it is true that the method can only be used effectively within proper constraints and with sufficient care.

## 3 Thermodynamic Modeling with Time-Series Data

In the first experimental module, environmental variables such as temperature, air pressure, humidity, and radiation are recorded using IoT sensors. The experimental setup is illustrated in Fig. 2. These measurements form time-series datasets suitable for predictive modeling.

We used the LSTM (Long Short-Term Memory) model in this experiment, which is a special type of recurrent neural network used for modeling time-series data. LSTM networks are applied to capture temporal dependencies and forecast system behavior. While the primary goal is prediction, the resulting models also provide insight into system dynamics when combined with domain knowledge. The measured results and the corresponding predicted values are shown in Fig. 3. The model's performance is characterized by Mean Absolute Error $MAE = 0.079$ and by Coefficient of Determination $R^2 = 0.91$ metrics.

In parallel, radiation levels are monitored using a Geiger–Müller counter. Measurements indicate a consistent increase in indoor radiation [**?**] levels during nighttime in poorly ventilated environments. This observation is consistent with the accumulation of radon and its decay products.

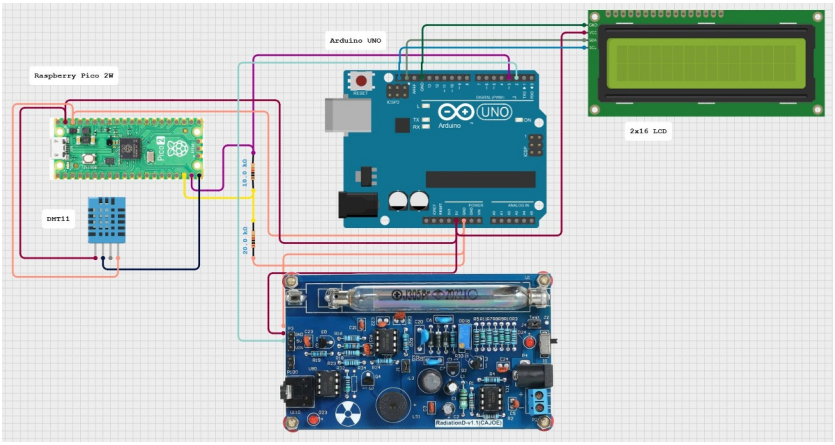



Fig. 2: The experimental setup

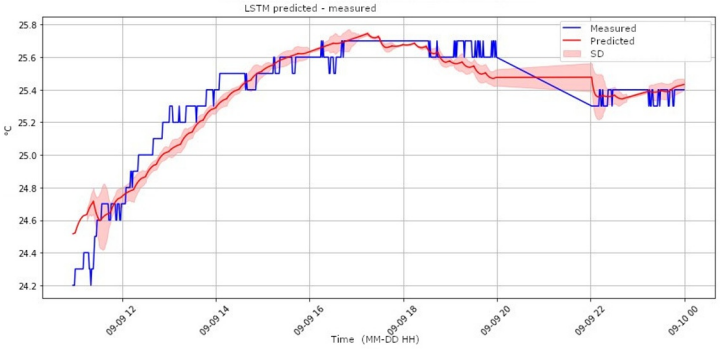



Fig. 3: Measured Data and LSTM Prediction

The data measured by the Geiger–Müller (GM) tubes are analyzed using an Allan deviation plot. In this representation, the segment with a slope of $-1/2$ corresponds to white noise, i.e., the background radiation, which can be considered steady in time. In contrast, the segment with a slope of $+1/2$ indicates random or unexpected fluctuations in the count rate. The plot clearly demonstrates that, as the number of measurements increases, the rising trend becomes observable.

It is important to note that these observations do not indicate any hazardous condition, as the maximum count rate remains around 50 CPM (Counts Per Minute), which is well below established safety limits.

We created a monitoring interface for the experiments using Node-RED. The dashboard displays both the current measured values and their variation over time. In the background, the data is stored in a SQLite database.

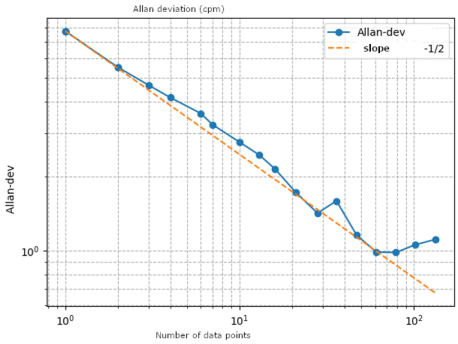



Fig. 4: The Allan deviation of the data acquired using the Geiger–Müller (GM) tube

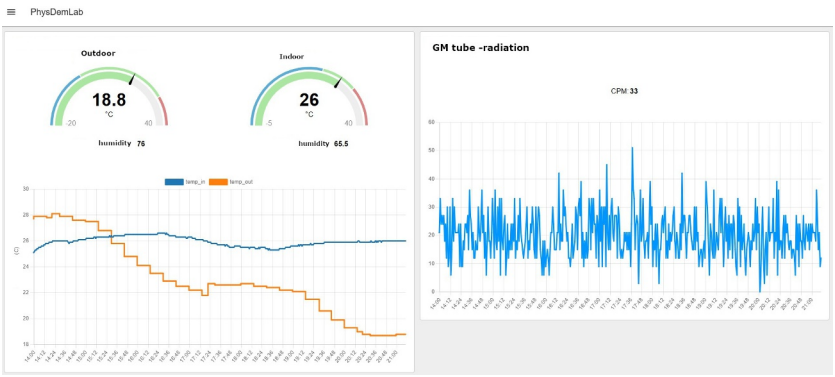



Fig. 5: The data were visualized using a Node-RED-based dashboard

From an educational perspective, this experiment demonstrates that important environmental processes may be imperceptible without instrumentation. By analyzing real data, students gain a deeper understanding of both measurement techniques and the physical phenomena involved.

## 4 Data-Driven Discovery of Physical Laws Using SINDy

We prepared a collision experiment in which the motion of objects is recorded and analyzed using computer vision techniques. The experimental setup consisted of a Raspberry Pi 4B module equipped with a 5-megapixel camera.

To move beyond prediction toward interpretability, the SINDy [**?**] (Sparse Identification of Nonlinear Dynamics) algorithm is employed to identify governing equations from experimental data. The SINDy algorithm is a data-driven method capable of identifying the governing equations of dynamical systems. The core idea is to use time-series measurements of state variables and their

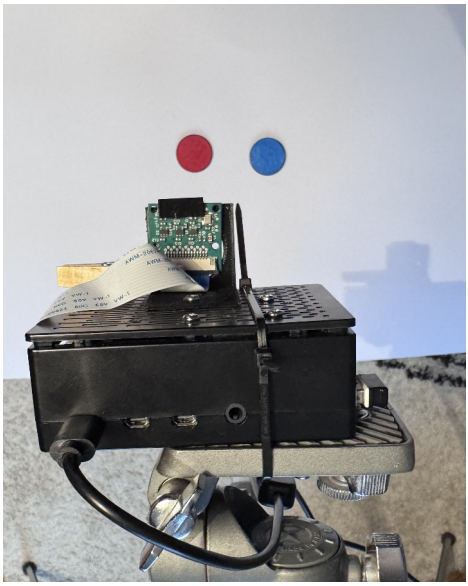


Fig. 6: A Raspberry Pi-based system equipped with a webcam was used for motion tracking experiments

derivatives, and apply sparse regression to determine which terms are actually present in the underlying physical model. The resulting trajectories are used as input for SINDy, which selects relevant terms from a predefined library of candidate functions. Through sparse regression, the algorithm eliminates irrelevant terms and constructs a minimal model that describes the observed dynamics. This allows students to see how familiar physical relationships — such as constant acceleration or conservation laws — can be recovered directly from data.

The objective is to identify the functional form of

$$\mathbf{f}(\mathbf{x}) = \frac{d\mathbf{x}}{dt} \tag{1}$$

from measured data.

The available data consist of time-series samples of the state variables, denoted by $\mathbf{X}$, and their time derivatives, $\dot{\mathbf{X}}$.

A library of candidate functions, $\Theta(\mathbf{X})$, is then constructed, containing possible terms such as $x$, $x^2$, $\sin(x)$, and others. The method proceeds by solving the regression problem

$$\dot{\mathbf{X}} = \Theta(\mathbf{X}) \cdot \Xi, \tag{2}$$

where $\Xi$ is a coefficient matrix determined using a sparse regression technique (e.g., LASSO Least Absolute Shrinkage and Selection Operator).

In the baseline setup, the motion of two discs was investigated. The discs were distinguished by color (red and blue) and were custom-designed, 3D-printed

objects with a diameter of 40 mm and a thickness of 2 mm (3 mm including the rim). The experiment was conducted on a flat surface, where one or both discs were given an initial velocity. Their motion was recorded, and the captured frames were processed using a Python-based analysis pipeline. The program detects the color and centroid of each disc, and these properties are tracked as a function of time. 9

In the present case, the state vector is defined as

$$\mathbf{x}(t) = \begin{bmatrix} x(t) \\ y(t) \end{bmatrix}. \tag{3}$$

The general form of the dynamical system identified by SINDy can be written as

$$\frac{d}{dt}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} a_1 x + a_2 y + a_3 \\ b_1 x + b_2 y + b_3 \end{bmatrix}. \tag{4}$$

After applying the algorithm, the learned systems of differential equations for the two discs are as follows.

*First disc*

$$(x_0)' = -1.124\, x_0 + 1.890\, x_1 \tag{5}$$
$$(x_1)' = -0.102\, x_0 + 0.154\, x_1 \tag{6}$$

*Second disc*

$$(x_0)' = 0.810 - 48.522\, x_0 + 96.773\, x_1 - 0.184\, x_0^2 + 0.667\, x_0 x_1 - 0.618\, x_1^2 \tag{7}$$
$$(x_1)' = 21.800 \tag{8}$$

From the recorded trajectories, key parameters such as the point of collision and post-collision directions can be determined. While physical modeling often assumes point-like objects, this approximation is not valid in the present case due to the finite size of the discs. Instead, the Euclidean distance between the centers of the two discs is evaluated, and the collision event is defined by the point at which this distance reaches its minimum.

Using these equations, trajectory simulations were performed and compared to the measured motion. The results show that the motion of the first disc is captured accurately by the learned model. In contrast, the trajectory of the second disc is less well reproduced. This discrepancy is likely due to limited data availability and the relatively small magnitude of the observed displacements, which can reduce the robustness of the identification process.

This approach provides a transparent alternative to black-box models, emphasizing that physical laws can be discovered rather than simply applied.

Since the primary objective is to obtain indirect validation of physical laws, an approach was required in which the correctness of the physical state or equations of motion can be inferred from experimentally measured data [**?**]. This can be achieved through two complementary approaches: machine learning methods and statistical techniques.

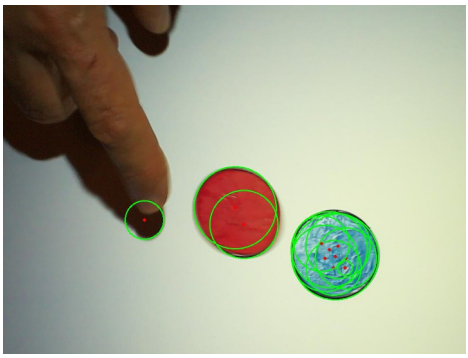


Fig. 7: The precise detection of the circular objects and their centroids is crucial for analysis

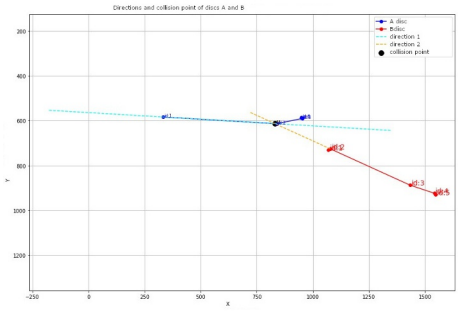



Fig. 8: The collision event between the two discs was determined from the recorded motion data.

## 5 Acoustics and Signal Analysis

While the previous sections focused primarily on hardware-based physics and environmental sensing, this experiment expands the framework into the domain of acoustics and computer vision, fully embracing the ethos of STEAM (Science, Technology, Engineering, Arts, and Mathematics) education. Modern pedagogy increasingly emphasizes the integration of arts into STEM fields to foster creativity and cross-disciplinary thinking. By combining the physical and acoustic principles of sound with real-time hand gesture recognition, this module demonstrates how abstract signal processing can be mapped to musical expression. Furthermore, from an instructional standpoint, the integration of user-friendly frameworks like MediaPipe provides a critical educational advantage: it allows informatics and physics students to build functional, AI-driven applications within the constraints of standard classroom hours, without requiring the complex training or heavy computational resources of deep custom neural networks.

A third experimental module focuses on acoustics, where students analyze sound signals using microphones and numerical tools. By computing frequency

spectra and applying perceptual scales, such as the Mel [**?**] scale, students can compare the harmonic structures of different instruments. In this case, we used a harmonium (pump organ) and a guitar. This reveals how physical properties of sound sources influence the observed signal.

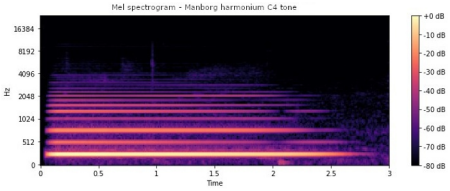



Fig. 9: The Mel spectrum of the harmonium signal

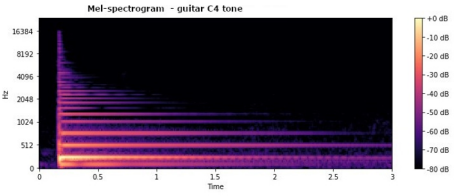



Fig. 10: The Mel spectrum of the guitar signal

A wide range of interesting physical properties relevant to acoustics can be demonstrated to students, such as Fourier transforms, envelope curves, noise characteristics, and others. In this paper, we do not discuss these in detail, as the techniques are readily accessible. We aim instead to draw attention to the rich variety of possibilities they offer.

Beyond its relevance to physics education, this module introduces concepts widely used in engineering applications, including signal processing and predictive maintenance.

In the field of music, we have developed an application that may be particularly engaging for students, and something similar could also be implemented within computer science education. At the same time, it provides an excellent opportunity to become familiar with computer vision and machine learning. The interactive graphical application serves as an intuitive tool for exploring musical sounds and supporting both the writing and reading of musical notation. In addition, it provides a visual and kinesthetic approach to learning solmization.

The application designed to support solmization learning is based on the MediaPipe framework [**?**]. MediaPipe is an open-source framework developed by Google for real-time multimedia processing tasks, including computer vision,

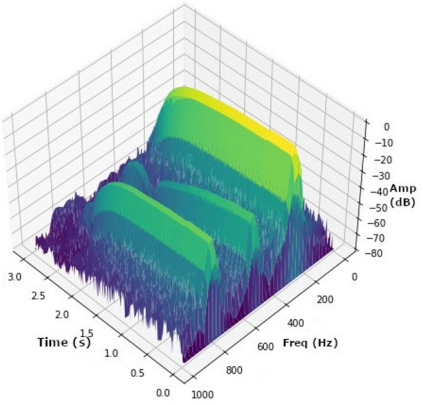



Fig. 11: Complex representation of the harmonium sound

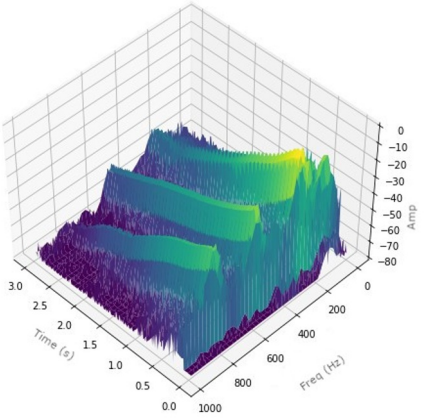



Fig. 12: Complex representation of the guitar sound

audio processing, and pose tracking. It provides pre-trained models and can be efficiently used in programming languages such as Python and C++.

In this work, the hand-tracking model is utilized. This model identifies 21 key points on the hand and tracks their relative positions in real time. Using this representation, it becomes possible to establish a mapping between solmization hand signs and the fundamental notes of the C major scale, as illustrated in Fig. 13.

The hand detection model assigns 21 landmark points, as shown in the Fig. 14. These points are defined based on anatomical considerations rather than arbitrary selection. Specifically, they correspond to the structural and functional units of the hand that enable its motion, providing a physically meaningful representation of hand gestures.
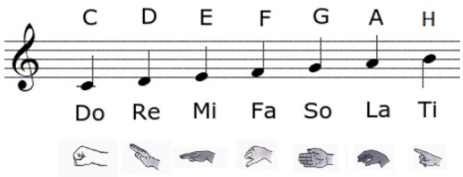



Fig. 13: Musical notes and solfège hand signs.

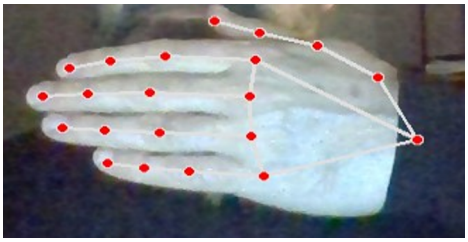


Fig. 14: Hand detection was performed using the MediaPipe framework. The system identifies 21 key landmarks on each hand and tracks their positions in real time.

## 6 Implementation and Reproducibility

To ensure accessibility and reproducibility, all computational components are deployed using containerization technologies. This eliminates the need for complex software installation and allows experiments to be replicated across different environments.

The use of standardized workflows also supports scalability, making it possible to integrate these modules into diverse educational settings.

## 7 Future Work

Future developments will focus on extending the framework to more advanced physical phenomena. Planned experiments include the detection of cosmic-ray muons and the implementation of AI-based control systems for magnetic levitation setups.

These extensions aim to further demonstrate how complex physical concepts can be explored using accessible, data-driven methods.

## 8 Conclusion

The interdisciplinary framework proposed in this study is designed as a pedagogical blueprint that fosters collaboration between physics and informatics

teachers. By combining hardware-based physics experiments with computer vision tools, such as the MediaPipe-driven solmization module, the curriculum follows a deliberate progression. It moves from traditional environmental data collection to advanced, AI-supported human-computer interaction, demonstrating the versatile educational applications of modern technology within a unified STEAM (Science, Technology, Engineering, Arts, and Mathematics) framework.

It is important to note that the primary scope of this work is foundational and instructional; it delivers fully developed, reproducible lesson materials rather than empirical student data. Consequently, a comprehensive quantitative evaluation of learning outcomes was not within the scope of this initial phase.

While the conceptual framework is strongly supported by the preliminary teacher survey, the lack of empirical classroom validation represents a limitation of the current study. Future research will focus on the field-testing and implementation of these modules in secondary and higher education settings. We plan to conduct a rigorous quantitative evaluation, utilizing pre- and post-tests to measure the framework's actual impact on student motivation, conceptual retention, and interdisciplinary problem-solving skills.

## References

1. Varga, K..: Felszín alatti vizekből származó radon gáz a természetes és épített környezetben. Debreceni Egyetem Tudásközpont repozitórium (2023) https://dea.lib.unideb.hu/server/api/core/bitstreams/d0ee6dec-d9ef-45e4-be1b-8b0ec076d4a5/content
2. Brunton, S.L., Proctor, J.L., Kutz, J.N.: Discovering governing equations from data by sparse identification of nonlinear dynamical systems. Proceedings of the National Academy of Sciences 113(15), 3932–3937 (2016). https://doi.org/10.1073/pnas.1517384113
3. Budó, Á.: Kísérleti Fizika I. – Mechanika, Hangtan, Hőtan. Tankönyvkiadó Vállalat, Budapest (1979). Egyetemi tankönyv ISBN:9631752623
4. Fletcher, N.H., Rossing, T.D.: The physics of musical instruments (1998) https://doi.org/10.1007/978-1-4612-2980-3
5. Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., Zhang, F., Chang, C.-L., Yong, M., Lee, J., Chang, W.-T., Hua, W., Georg, M., Grundmann, M.: MediaPipe: A Framework for Perceiving and Process- ing Reality. Let¨oltve:2025.10.30. https://mixedreality.cs.cornell.edu/s/NewTitleMay1MediaPipeCVPRCV4ARVRWorkshop2019.pdf

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

