Conversational AI for Electrical Measurement Education: An Open-Source LLM Framework

Branislav Fecko, Jozef Dziak, Tibor Vince, Ján Molnár
Technical University of Košice
Department of Theoretical and Industrial Electrical Engineering
Letná 1/9, 042 00 Košice, Slovakia
{branislav.fecko, jozef.dziak, tibor.vince, jan.molnar}@tuke.sk

Abstract. *Large language models are increasingly used in engineering education, yet their deployment in measurement laboratories remains limited by the lack of grounding in actual instrument data. This paper identifies three scenarios for LLM integration into electrical measurement education, defines seven criteria for their systematic comparison, and presents a functional open-source implementation of Scenario 2. The implemented framework connects two digital multimeters and a programmable power supply to a Custom GPT via a FastAPI backend and Cloudflare Tunnel, grounding conversational feedback in student-measured values.*

Keywords. electrical measurement, laboratory education, instrument integration, conversational AI

1 Introduction

Large language models (LLMs) have been adopted rapidly in higher education. A systematic review of 88 empirical studies confirmed their positive effects on academic performance and student engagement, while also highlighting the risks of overreliance and the need for longitudinal evaluation (Shi et al., 2026). In engineering education, these tools support dynamic content generation and personalized explanations.

Students naturally turn to AI tools because of their immediate availability and the ability to ask questions at their own pace (Shi et al., 2026). However, growing interest in the use of LLMs in laboratory courses suggests that this potential can also be extended to the practical component of instruction (Wang, 2026). As class sizes grow and individualized instructor feedback becomes less available, students naturally turn to LLMs as a source of assistance (Hasan et al., 2025). A key limitation, however, is that LLMs generate incorrect or misleading responses in more than one out of ten cases when the experimental context is not explicitly provided (Kregear et al., 2025). Unlike an instructor, an LLM cannot infer the course of an experiment from the surrounding situation and therefore requires the experimental context to be explicitly included in the prompt. One solution is to ground the LLM automatically in values measured directly by laboratory instruments, supplemented with relevant theoretical background and procedural guidance. This principle forms the core of the proposed system.

This paper addresses three questions: What are the main scenarios for integrating LLMs into electrical measurement education? What criteria enable their systematic comparison? And what architectural design can connect laboratory instruments with a conversational LLM to provide data-grounded feedback? The paper also presents a functional open-source implementation of one of these scenarios. Quantitative pedagogical validation with students is beyond the scope of the present study and is planned for the second phase of the project.

The remainder of the paper is organized as follows. Section 2 reviews the related literature and formulates the research questions. Sections 3 and 4 identify and compare the proposed integration scenarios. Section 5 describes the architectural design and implementation of the framework. Finally, Sections 6 and 7 discuss the results and conclude the paper.

2 Background and Related Work

The proposed system lies at the intersection of three domains: the application of generative AI in engineering education, laboratory automation, and efforts to integrate LLMs with physical laboratory instruments. The convergence of these domains reveals a research gap that this paper seeks to address.

2.1 Generative AI in Engineering Education

Generative AI is making its way into engineering and STEM education in several forms. The first is the direct support of conceptual learning. For example, in thermodynamics, ChatGPT-assisted instruction led to

measurable improvements in student understanding compared to traditional teaching approaches (El Fathi et al., 2025). The second is the enhancement of learning materials themselves. The concept of intelligent textbooks uses AI methods to enrich digital content with interactive modules, automated assessment, embedded chatbots, and adaptive student-content interaction (Sosnovsky et al., 2025).

A specific example of such an intelligent textbook in electrical engineering education is the GenAI-T system, built on GPT-4 and integrated into an introductory electrical engineering course. A four-year comparative study involving 736 students demonstrated improved performance in formative assessments (Fecko et al., 2026). Efforts to extend the use of LLMs beyond self-directed learning and into formal instruction have also reached laboratory education. In a quasi-experimental study, the integration of LLMs into an undergraduate laboratory course improved students' target competency, while also revealing several practical challenges associated with their implementation (Wang, 2026).

Across these approaches, AI operates exclusively at the level of text and pre-prepared content. It works with what students write or read, rather than with what they actually measure. As a result, these solutions lack feedback grounded in the real measurement values produced by the laboratory instruments.

2.2 Instrument Control and Automation in Measurement Laboratories

Software-based control of laboratory instruments is a well-established area in engineering practice. Commercial tools such as LabVIEW and MATLAB are powerful and widely used, but they are also expensive and difficult to transfer across different instruments and platforms (Letowski et al., 2020). These limitations have motivated a shift toward open-source solutions. For example, the remote laboratory project Laborem was originally built on National Instruments' LabVIEW platform but was later redesigned as a lightweight open-source system due to data communication issues (Letowski et al., 2020). Similarly, the high cost of commercial solutions motivated the development of a fully open-source remote laboratory based on Raspberry Pi and Python and support for real-time web-based control (Álvarez & Nomesqui, 2023).

Open-source approaches rely almost exclusively on Python and its ecosystem. For example, measurement automation implemented in Python has been shown to improve the reproducibility of experimental results by continuously recording all relevant information about the experiment (Koerner et al., 2019). Instrument communication is typically based on the Standard Commands for Programmable Instruments (SCPI) protocol and the PyVISA library, which provides a unified Python interface for instruments from different manufacturers (Koerner et al., 2019,Fébba et al., 2024). Libraries such as NumPy and SciPy then provide tools for the subsequent analysis of measurement data.

From an educational perspective, a meta-analysis found that virtual and automated laboratories contribute significantly to student learning outcomes, particularly by increasing motivation and engagement. However, they cannot fully replace hands-on work with physical instruments (Li & Liang, 2024). The combination of low cost, reproducibility, and vendor-independent instrument access therefore provides a mature technological foundation. Operating such systems still requires programming expertise, and the measured data remains largely devoid of pedagogical interpretation. Integrating this infrastructure with a conversational LLM would enable not only automated measurement execution but also immediate feedback on measured results within the context of the ongoing laboratory exercise.

2.3 AI-Driven Instrument Control and Conversational Laboratory Assistants

The integration of LLMs with experimentation is a rapidly evolving research area. The GPT-4–driven Coscientist system demonstrated the ability to autonomously design, plan, and execute complex experiments, by leveraging tools such as documentation retrieval, code execution, and laboratory automation (Boiko et al., 2023). Traditionally, developing control software for existing laboratory instruments has required substantial programming expertise, creating a barrier for researchers and educators without a computational background. LLMs have the potential to significantly lower this barrier (Xie et al., 2025).

This shift is increasingly reflected in laboratory measurement systems. Through interactions with ChatGPT-4, researchers developed a control interface for the Keithley 2400 source-measure unit, a widely used instrument for current–voltage (I–V) measurements, with only minimal manual code revisions. The resulting framework included both a graphical user interface and prompt templates for additional instruments (Fébba et al., 2024). The LABIIUM system extends this concept by integrating an LLM-based code generation assistant with Lab-Automation-Measurement Bridges, enabling zero-configuration instrument connectivity through standard tools such as Python and VSCode (Olowe & Chitnis, 2025). Remotely accessible platforms have also emerged as environments for prototyping automation and LLM workflows within self-driving laboratories (Kitchin, 2025).

These studies connect LLMs with laboratory instruments, yet their primary focus remains code generation and autonomous experimentation rather than pedagogically oriented feedback. The Adaptive Lab Mentor framework represents a notable exception by providing real-time personalized feedback through sensor fusion and Internet of Things (IoT)

technologies. Even in this case, however, the conversational language model is proposed only as a future extension of the framework (Hasan et al., 2025). Evidence from a physics laboratory setting also suggests that the reliability of feedback declines when it is not grounded in data describing the measurement process (Kregear et al., 2025).

2.4 Research Gap

The literature review reveals that these components currently exist in isolation: pedagogically effective AI without access to measured data (2.1), instrument automation without a conversational layer (2.2), and LLM–instrument integration without a pedagogical focus (2.3). The Adaptive Lab Mentor framework comes closest to a comprehensive solution by combining real-time diagnostics with pedagogical feedback. Its conversational language model component, however, remains a proposed direction for future development. Based on the available literature, no existing system integrates a conversational LLM with physical measurement instruments to provide pedagogically oriented feedback specifically for electrical measurement education. These components can also be combined in multiple ways, differing primarily in the extent to which feedback is grounded in real measurement data and in the pedagogical transparency of the interaction. These observations give rise to three research questions:

- RQ1: What are the main scenarios for LLM integration into electrical measurement laboratory education and how do they differ in terms of data grounding and pedagogical transparency?
- RQ2: What criteria enable systematic comparison of LLM integration scenarios in the context of hands-on engineering education?
- RQ3: What architectural design connects laboratory instruments with a conversational LLM to deliver data-grounded student feedback?

3 LLM Integration Scenarios in Measurement Education

Three scenarios for integrating LLMs into electrical measurement laboratory education were identified and are described in terms of their degree of data grounding and pedagogical transparency. The scenarios are derived from two sources. The first is the literature review presented in Chapter 2, which identified three distinct layers: pedagogically oriented generative AI, instrument automation, LLM–instrument integration. The second is an informal analysis of student needs in the target course, conducted through a voluntary anonymous survey with open-ended questions about the challenges students encounter during laboratory measurements and how AI could support them. The purpose of this data collection was not to test hypotheses but to ground the scenario design in authentic user needs. A rigorous evaluation involving students is planned for the second phase of the project.

Three recurring categories of needs emerged from the survey responses. Students commonly rely on general-purpose conversational AI tools without providing structured measurement context, manually entering measured values into the chat or attaching photographs of circuit diagrams. At the same time, they repeatedly expressed a need for support during the measurement process itself, including verification of measured values, explanations of why a circuit behave as expected or unexpectedly, and assistance in identifying potential errors. Several students explicitly stated that they wanted the assistant to explain the procedure rather than simply provide the answer. This expectation highlights a fundamental pedagogical tension shared across all three scenarios. A higher degree of automation can improve measurement efficiency, yet it may also reduce the student's role to that of a passive observer. The goal of laboratory education is not merely to obtain a measured value but to develop an understanding of the measurement process. For this reason, the extent to which a system preserves students' active cognitive engagement is as important a dimension of comparison as its level of technical integration with laboratory instruments.

Based on the literature review and the identified student needs, seven criteria were defined for the systematic comparison of integration scenarios: context grounding, pedagogical transparency, degree of automation, deployment complexity, wiring error detection, longitudinal learning analytics, and instructor analytics. Three scenarios are analyzed in this paper: unstructured LLM use with manual data entry (3.1), an instrument-integrated LLM providing context-aware feedback (3.2), and an autonomous laboratory framework with interaction logging and adaptive analytics (3.3). Each scenario is illustrated by a block diagram showing the flow of information between the student, the laboratory instrument, and the language model.

3.1 Scenario 1 - Unstructured LLM Use

The first scenario represents the way LLMs are currently used by students in laboratory courses, as also reflected in the informal needs analysis. In this scenario, the laboratory instrument is not connected to the system. Measurements are performed independently, and the measured values are read directly from the instrument display and manually entered into a general-purpose conversational AI tool, often together with a photograph of the circuit diagram. The language model generates its response solely from the information provided by the student. The flow of information is illustrated in Figure 1.

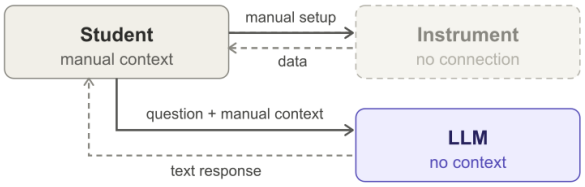

Figure 1. Unstructured LLM Use

3.2 Scenario 2 - Instrument-Integrated LLM via OpenAI Actions

The second scenario represents the main contribution of this paper. Students interact directly with a Custom GPT based on GPT-4. When a question is submitted, the Custom GPT invokes the appropriate backend endpoint through an OpenAI Action. The backend is implemented as a FastAPI application using the PyVISA library and serves as an intermediary between the language model and the laboratory instrument. The server runs locally and is exposed through a Cloudflare Tunnel under a public address, enabling communication with OpenAI servers without requiring a public IP address. Measurements are executed through SCPI commands, after which the instrument returns the measured values. These values are then passed back to the Custom GPT as part of the context of the ongoing conversation.

Before the conversation begins, a system prompt is loaded into the Custom GPT from a configuration block. This block contains the theoretical background of the course, instructions for the laboratory exercise, and guidance directing the model to ask students guiding questions rather than provide direct answers. The Custom GPT therefore generates responses grounded in the values measured by the student and returns them as part of the ongoing interaction. The flow of information is illustrated in Figure 2.

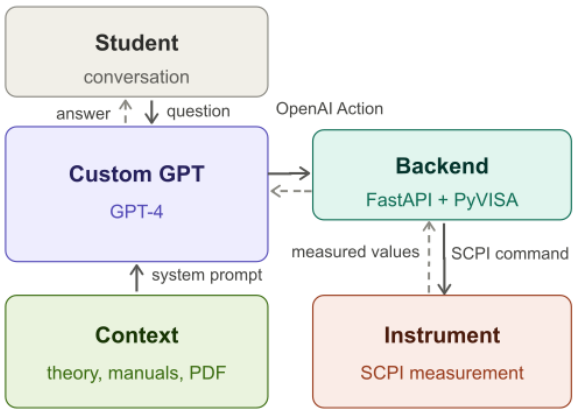

Figure 2. LLM via OpenAI Actions

3.3 Scenario 3 - Autonomous AI Laboratory Framework

The third scenario extends Scenario 2 with a longitudinal dimension and serves as a conceptual framework for the second phase of the research project. The entire system is encapsulated within a dedicated web application through which students submit questions and receive responses. As in Scenario 2, the backend layer provides programmatic communication with laboratory instruments through SCPI. Unlike Scenario 2, however, communication with the language model is handled directly through an API interface, enabling the deployment of a locally hosted model without dependence on external services.

A central component of this scenario is a database that serves two functions. The first is Interaction Logging, where every exchange between the student and the language model is recorded together with the measured values and their deviation from the expected results. The second is the storage of contextual resources and configuration data, including system prompts, theoretical materials, laboratory instructions, and model settings. Before each conversation, these resources are provided to the language model as a system prompt. Model responses are returned to the student through the backend layer. The recorded logs remain accessible to the backend and are used to generate adaptive feedback for students as well as aggregated reports for instructors. Empirical validation of this scenario is planned for the second phase of the project. The flow of information is illustrated in Figure 3.

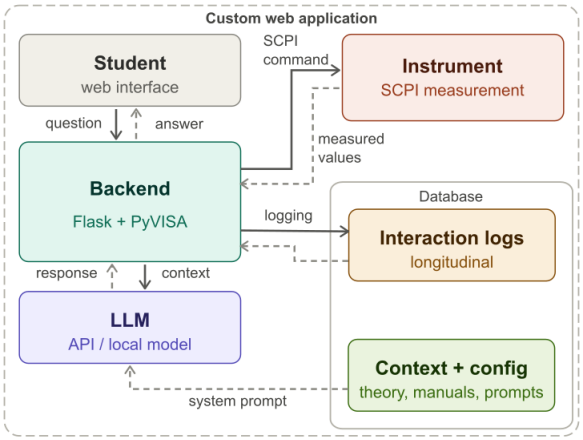

Figure 3. Autonomous AI Laboratory Framework

4 Comparative Analysis

Based on the literature review and the analysis of student needs, seven criteria were defined for comparing the proposed scenarios: Data Grounding, Pedagogical Transparency, Degree of Automation, Deployment Complexity, Wiring Error Detection, Longitudinal Learning Analytics, and Instructor Analytics. The first two criteria directly address RQ1 and are derived from the requirements identified in 3. The remaining five reflect the practical constraints and requirements of deployment in a university laboratory environment.

Scenario 1 requires no additional infrastructure and reflects current informal practice. The quality and relevance of feedback depend entirely on the context provided by the student, whether in the form of text, a circuit diagram image, or manually entered measurement values.

Scenario 2 addresses Data Grounding through an SCPI/PyVISA backend and provides partial support for Wiring Error Detection, at the cost of moderate Deployment Complexity. Before the conversation begins, contextual information is loaded into the Custom GPT through a system prompt containing the theoretical background of the course and instructions for the laboratory exercise. The prompt also directs the model to ask guiding questions rather than provide direct answers, thereby preserving the student's active role in the measurement process. Dialogue history is retained within a single session, enabling context-aware follow-up questions but not longitudinal tracking of student progress.

Scenario 3 extends Scenario 2 by introducing long-term Interaction Logging and Instructor Analytics. This added functionality requires a more robust server infrastructure. The recorded interactions make it possible to identify recurring errors across an entire student cohort, creating opportunities for adaptive instruction.

Table 1. Comparison of LLM Integration Scenarios

Criterion	S1	S2 ★	S3
Data grounding	Partial	**Full**	Full
Pedagogical transparency	Low	**Moderate**	High
Degree of automation	None	**Full**	Full
Deployment complexity	None	**Moderate**	High
Connection error detect.	None	**Partial**	Full
Longitudinal analytics	None	**Partial**	Full
Instructor analytics	None	**None**	Full

5 Framework Design and Implementation - Scenario 2

The implemented framework realizes Scenario 2 described in 3.2 and directly addresses RQ3. Its primary pedagogical value lies not in measurement automation itself, but in the dialogue-driven configuration of measurements grounded in an understanding of the physical phenomena occurring within the investigated circuit. The framework is demonstrated through the measurement of the I–V characteristic of an unknown component. The task involves identifying a black box selected from a set of three components with distinct electrical characteristics: a resistor, a diode, and a capacitor.

5.1 System Architecture

The framework is organized into three layers. The physical layer consists of two Rigol DM3068 digital multimeters and a Hameg HMP4040 programmable power supply connected to a host computer via USB. The data layer is implemented as a FastAPI-based backend server responsible for instrument control, measurement configuration management, and the provision of measurement-specific endpoints for the AI layer. The server runs locally and is securely exposed through a Cloudflare Tunnel under a public address, enabling communication with OpenAI servers without requiring a public IP address or port forwarding. The AI layer is implemented as a GPT-4–based Custom GPT that communicates with the backend through OpenAI Actions. Through a sequence of guiding questions, the system encourages students to interpret and understand the measured results rather than merely receive them. The flow of information across all three layers is illustrated in Figure 4.

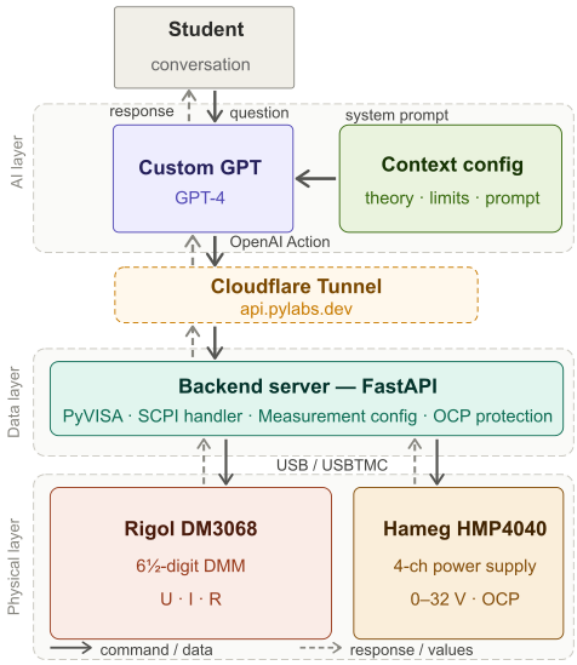

Figure 4. Three-Layer Framework Architecture

The framework is designed with extensibility in mind. Each laboratory instrument is represented by a dedicated driver, while each measurement task is implemented through a separate configuration file and FastAPI endpoint. As a result, new measurement tasks can be added without modifying the core architecture.

5.2 Instrument Communication Layer

The Hameg HMP4040 programmable power supply and two Rigol DM3068 digital multimeters are controlled via the PyVISA library using the USBTMC

protocol. One multimeter measures voltage, while the other measures current. Each instrument is identified at server startup using a standard SCPI query, which verifies availability and correct assignment before any measurement is performed.

Before each measurement, the backend configures the overcurrent protection of the HMP4040 power supply according to the limits defined for the active measurement task. This hardware-level protection prevents damage to the component under test regardless of the course of the dialogue. The sweep sequence used for measuring the I–V characteristic is implemented as a FastAPI endpoint (Listing 1).

```
@app.post("/measure_va_characteristic")
async def measure_va_characteristic(
        params: MeasurementParams):
    psu.write(f":CURR {params.i_limit}")
    psu.write(":OUTP ON")
    dmm_v.write(":FUNC:VOLT:DC")
    dmm_i.write(":FUNC:CURR:DC")
    results = []
    for v in params.voltage_steps:
        psu.write(f":VOLT {v}")
        await asyncio.sleep(params.step_interval)
        U = float(dmm_v.query(":READ?"))
        I = float(dmm_i.query(":READ?"))
        results.append({"U": round(U, 4), "I":
round(I, 6)})
    psu.write(":OUTP OFF")
    return {"points": results,
            "config": params.model_dump()}
```

Listing 1. I–V characteristic sweep, FastAPI endpoint

5.3 Measurement Configuration and Guided Execution

Each measurement task is defined through a configuration file prepared by the instructor. The file contains a measurement identifier, instrument safety limits, a system prompt for the Custom GPT, and reference values used for anomaly detection.

The measurement workflow consists of three phases. In the first phase, students identify the task through a natural-language request. The Custom GPT loads the corresponding configuration and guides them in defining the measurement parameters in the context of the physical properties of the component under investigation. This phase represents the pedagogical core of the system: students actively make methodological decisions, developing an understanding of the underlying physical phenomena before the measurement begins. In the second phase, the backend verifies the instrument configuration through SCPI queries and prompts the student to make corrections if inconsistencies are detected. In the third phase, the backend executes the measurement and returns the results as a structured JSON object.

The following example illustrates the first phase of the workflow for measuring the I–V characteristic of an unknown component:

***Student**: "I want to identify an unknown component by measuring its voltage-current characteristics."*
***Custom GPT**: "What current limit will you choose, and why?"*
***Student**: "I'll set it to 10mA as a safe limit."*
***Custom GPT**: "What voltage range and step size do you propose?"*
***Student**: "0 to 2V, 0.1V steps—for an initial exploration of the characteristics."*
***Custom GPT**: "Measurement complete. The measured data points are available in Table 2."*
***Student**: "Exponential curve — looks like a diode. I'll run a new sweep from 0 to 1 V, step 0.02 V, limit 15 mA to verify the threshold voltage."*

The JSON payload returned by the backend and processed by the Custom GPT is shown in Listing 2.

```
{
  "measurement": "VA characteristic",
  "parameters": {
    "U_max": 2.0,
    "step": 0.1,
    "i_limit": 0.010
  },
  "points": [{"U": 0.0, "I": 0.000},
             "...",
             {"U": 2.0, "I": 0.008}],
  "analysis": {
    "threshold_voltage": 0.62,
    "type_hint": "exponential"
  }
}
```

Listing 2. JSON payload returned to Custom GPT

5.4 Open-Source Availability

The Scenario 2 implementation will be made publicly available under the PolyForm Non-Commercial License 1.0.0 upon acceptance. Educational and research use is unrestricted; commercial use requires a separate license agreement. The repository link is omitted to preserve double-blind review. Scenario 3 will be released in a subsequent research phase.

6 Discussion

The three research questions form a progression from the conceptual to the technical level. A common theme runs through all three: to what extent should the system take over tasks traditionally performed by students and where should space remain for their own decision-making. The findings suggest that this boundary is not technical but pedagogical. The system should automate data collection, not thinking.

6.1 LLM Integration Scenarios

The three scenarios differ in the extent to which feedback is automatically grounded in measurement data, in how effectively they preserve students' active cognitive engagement, and in the reliability of the model's output. Reliability depends directly on whether the context is provided by the instructor or assembled independently by the student (Kregear et al., 2025). A higher degree of instrument integration does not necessarily translate into greater pedagogical value. A system that performs measurements automatically and simply returns the result may replace the instructor without contributing to the development of student

competence. The added value of Scenario 2 therefore lies not in automated data collection, but in requiring students to actively determine measurement parameters before the experiment and justify their choices in the context of the physical properties of the component under investigation.

Greatest benefits emerge in measurement tasks where time-consuming automated data collection directly illustrates fundamental principles of electrical engineering. Examples include the nonlinear I–V characteristic of an incandescent lamp, which demonstrates the relationship between resistance and filament temperature, and transient phenomena during capacitor charging. In such cases, the conversational assistant transforms raw measurement data into a pedagogical dialogue that would be impractical to sustain through manual data collection alone. The level of system autonomy is not fixed. It can range from step-by-step verification of instrument settings through SCPI commands before measurement execution to fully automated data collection followed by result interpretation. This flexibility enables multiple implementation approaches tailored to the objectives of a particular course and student cohort.

6.2 Criteria for Systematic Comparison

Systematic comparison of scenarios in hands-on laboratory education requires criteria that capture not only the technical level of integration but also the pedagogical quality of the interaction. Criteria derived exclusively from the literature on LLM-based tutors are insufficient in this context because they overlook the physical measurement layer. It is precisely this layer that introduces two criteria absent from existing frameworks: Data Grounding in real measurement data and Wiring Error Detection. Deviations between measured and expected values make it possible to identify issues such as instrument misconfiguration or an open circuit without requiring visual input. Deployment Complexity also distinguishes Scenario 2 from Scenario 3. Scenario 2 depends on OpenAI's Custom GPT platform, including its availability and service conditions. Scenario 3 removes this dependency through direct API-based integration, at the cost of increased infrastructure requirements.

6.3 Architectural Design of the Proposed Framework

Integrating laboratory instruments with a conversational LLM requires an architecture that separates the physical measurement layer from the AI interaction layer while supporting extensibility without modifications to the system core. The proposed three-layer architecture, built around measurement-specific endpoints, satisfies these requirements and enhances operational safety by preventing students from interacting directly with low-level SCPI commands. The demonstrated use case of identifying an unknown component serves as a deliberate proof of concept that validates all layers of the system while minimizing hardware dependencies. The same architecture can be scaled to measurement tasks involving multiple instruments, where automation offers substantially greater benefits. Empirical validation of Scenario 3 is planned as part of the next phase of the research project.

7 Conclusion and Future Work

Three scenarios for integrating LLMs into electrical measurement education were identified, seven comparison criteria were defined, and an open-source implementation of Scenario 2 was presented. The framework connects two Rigol DM3068 digital multimeters and a Hameg HMP4040 programmable power supply to a GPT-4–based Custom GPT through a FastAPI backend, enabling conversational feedback grounded in student-generated measurement data.

The primary contribution of the proposed framework is not measurement automation itself, but the dialogue-driven configuration of measurement parameters prior to experiment execution. Students' active cognitive engagement is preserved by encouraging them to justify methodological decisions in the context of the physical properties of the component under investigation. The demonstrated use case of identifying an unknown component through measurement of its I-V characteristic validates all layers of the system and illustrates its potential for extension to more complex diagnostic tasks, where automation offers substantially greater benefits than manual data collection. Wiring Error Detection was also included among the comparison criteria. This criterion is systematically absent from existing frameworks for LLM-based tutoring systems and emerged directly from practical experience gained during implementation.

The current implementation has two main limitations. Dependence on OpenAI’s Custom GPT platform introduces operational risks related to service availability and platform policies. In addition, quantitative pedagogical validation involving students enrolled in the target course has not yet been conducted.

Future research will focus on three areas: empirical validation with students, implementation of Scenario 3 with support for a locally hosted language model, and expansion of the measurement portfolio to include tasks requiring coordination among multiple laboratory instruments.

Acknowledgments

Supported by the Cultural and Educational Grant Agency of the Ministry of Education, Science,

Research and Sport of the Slovak Republic (KEGA)
under the project No. 038TUKE-4/2025.

References

Álvarez Ariza, J., & Nomesqui Galvis, C. (2023). RaspyControl Lab: A fully open-source and real-time remote laboratory for education in automatic control systems using Raspberry Pi and Python. *HardwareX*, *13*. https://doi.org/10.1016/j.ohx.2023.e00396

Boiko, D. A., MacKnight, R., Kline, B., & Gomes, G. (2023). Autonomous chemical research with large language models. *Nature*, *624*(7992), 570–578. https://doi.org/10.1038/s41586-023-06792-0

El Fathi, T., Saad, A., Larhzil, H., Lamri, D., & Al Ibrahmi, E. M. (2025). Integrating generative AI into STEM education: enhancing conceptual understanding, addressing misconceptions, and assessing student acceptance. *Disciplinary and Interdisciplinary Science Education Research*, *7*(1). https://doi.org/10.1186/s43031-025-00125-z

Fébba, D., Egbo, K., Callahan, W. A., & Zakutayev, A. (2024). From text to test: AI-generated control software for materials science instruments. *Digital Discovery*, *4*(1), 35–45. https://doi.org/10.1039/d4dd00143e

Fecko, B., Dziak, J., Vince, T., & Molnár, J. (2026). Generative AI interactive textbook in electrotechnics: A four-year comparative study on student performance and inclusion. *Computers and Education: Artificial Intelligence*, *10*. https://doi.org/10.1016/j.caeai.2026.100611

Hasan, M. S., Ahmed, A., Rasool, N., Jabe, M. M. N., Zeng, X., & Ali Pirzado, F. (2025). The Adaptive Lab Mentor (ALM): An AI-Driven IoT Framework for Real-Time Personalized Guidance in Hands-On Engineering Education. *Sensors*, *25*(24). https://doi.org/10.3390/s25247688

Kitchin, J. R. (2025). The evolving role of programming and LLMs in the development of self-driving laboratories. *APL Machine Learning*, *3*(2). https://doi.org/10.1063/5.0266757

Koerner, L. J., Caswell, T. A., Allan, D. B., & Campbell, S. I. (2019). A Python Instrument Control and Data Acquisition Suite for Reproducible Research. In *IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT*.

Kregear, T., Babayeva, M., & Widenhorn, R. (2025). Analysis of Student Interactions with a Large Language Model in an Introductory Physics Lab Setting. *International Journal of Artificial Intelligence in Education*, *35*(5), 2993–3016. https://doi.org/10.1007/s40593-025-00489-3

Letowski, B., Lavayssière, C., Larroque, B., Schröder, M., & Luthon, F. (2020). A fully open source remote laboratory for practical learning. *Electronics (Switzerland)*, *9*(11), 1–16. https://doi.org/10.3390/electronics9111832

Li, J., & Liang, W. (2024). Effectiveness of virtual laboratory in engineering education: A meta-analysis. *PLoS ONE*, *19*(12). https://doi.org/10.1371/journal.pone.0316269

Olowe, E. A., & Chitnis, D. (2025). *LABIIUM: AI-Enhanced Zero-configuration Measurement Automation System*. https://doi.org/10.1109/I2MTC62753.2025.11079200

Shi, Y., Yu, K., Dong, Y., & Chen, F. (2026). Large language models in education: a systematic review of empirical applications, benefits, and challenges. In *Computers and Education: Artificial Intelligence* (Vol. 10). Elsevier B.V. https://doi.org/10.1016/j.caeai.2025.100529

Sosnovsky, S., Brusilovsky, P., & Lan, A. (2025). Intelligent Textbooks. In *International Journal of Artificial Intelligence in Education* (Vol. 35, Number 3, pp. 967–986). Springer. https://doi.org/10.1007/s40593-024-00451-9

Wang, Y. (2026). Integrating large language models into medical undergraduate laboratory course to enhance bioethical competence: a quasi-experimental study. *Frontiers in Medicine*, *12*. https://doi.org/10.3389/fmed.2025.1745975

Xie, Y., He, K., & Castellanos-Gomez, A. (2025). Toward Full Autonomous Laboratory Instrumentation Control with Large Language Models. *Small Structures*, *6*(8). https://doi.org/10.1002/sstr.202500173