Athens Journal of Technology & Engineering

Quarterly Academic Periodical, Volume 13, Issue 3
Published by the Athens Institute
URL: https://www.athensjournals.gr/ajte Email: journals@atiner.gr
e-ISSN: 2241-8237 DOI: 10.30958/ajte

September 2026

Athens Journal of Technology & Engineering

Quarterly Academic Periodical, Volume 13, Issue 3, September 2026

Published by the Athens Institute

URL: https://www.athensjournals.gr/ajte Email: journals@atiner.gr

Athens Journal of Technology & Engineering

Published by the Athens Institute

Athens Institute is an Athens-based World Association of Academics and Researchers based in Athens. Athens Institute is an independent and non-profit Association with a Mission to become a forum where Academics and Researchers from all over the world can meet in Athens, exchange ideas on their research and discuss future developments in their disciplines, as well as engage with professionals from other fields. Athens was chosen because of its long history of academic gatherings, which go back thousands of years to Plato's Academy and Aristotle's Lyceum. Both these historic places are within walking distance from Athens Institute's downtown offices. Since antiquity, Athens was an open city. In the words of Pericles, Athens"…is open to the world, we never expel a foreigner from learning or seeing". ("Pericles' Funeral Oration", in Thucydides, The History of the Peloponnesian War). It is Athens Institute's mission to revive the glory of Ancient Athens by inviting the World Academic Community to the city, to learn from each other in an environment of freedom and respect for other people's opinions and beliefs. After all, the free expression of one's opinion formed the basis for the development of democracy, and Athens was its cradle. As it turned out, the Golden Age of Athens was in fact, the Golden Age of the Western Civilization. Education and (Re)searching for the 'truth' are the pillars of any free (democratic) society. This is the reason why Education and Research are the two core words in Athens Institute's name.

The *Athens Journal of Technology & Engineering (AJTE)* is an Open Access quarterly double-blind peer reviewed journal and considers papers from all areas engineering (civil, electrical, mechanical, industrial, computer, transportation etc), technology, innovation, new methods of production and management, and industrial organization. Many of the papers published in this journal have been presented at the various conferences sponsored by the Engineering & Architecture Division of the Athens Institute. All papers are subject to Athens Institute's Publication Ethical Policy and Statement.

Athens Journal of Technology & Engineering
ISSN NUMBER: 2241-8237 - DOI: 10.30958/ajte
Volume 13, Issue 3, September 2026
Download the entire issue (PDF)

Athens Journal of Technology & Engineering
Editorial and Reviewers' Board

President's Message

All Athens Institute's publications including its e-journals are open access without any costs (submission, processing, publishing, open access paid by authors, open access paid by readers etc.) and is independent of presentations at any of the many small events (conferences, symposiums, forums, colloquiums, courses, roundtable discussions) organized by Athens Institute throughout the year and entail significant costs of participating. The intellectual property rights of the submitting papers remain with the author. Before you submit, please make sure your paper meets the basic academic standards, which includes proper English. Some articles will be selected from the numerous papers that have been presented at the various annual international academic conferences organized by the different divisions and units of the Athens Institute for Education and Research. The plethora of papers presented every year will enable the editorial board of each journal to select the best, and in so doing produce a top-quality academic journal. In addition to papers presented, Athens Institute will encourage the independent submission of papers to be evaluated for publication.

The current issue is the third of the thirteenth volume of the *Athens Journal of Technology & Engineering (AJTE),* published by the Engineering & Architecture Division of Athens Institute.

Gregory T. Papanikos
President
Athens Institute

Athens Institute for Education and Research

A World Association of Academics and Researchers

17th Annual International Conference on Civil Engineering
21-26 June 2027, Athens, Greece

The Civil Engineering Unit of ATINER is organizing its **17th Annual International Conference on Civil Engineering, 21-26 June 2027, Athens, Greece** sponsored by the Athens Journal of Technology & Engineering. The aim of the conference is to bring together academics and researchers of all areas of Civil Engineering other related areas. You may participate as stream leader, presenter of one paper, chair of a session or observer. Please submit a proposal using the form available (https://www.atiner.gr/2027/FORM-CIV.doc).

Academic Members Responsible for the Conference

- **Dr. Dimitrios Goulias**, Head, Civil Engineering Unit, ATINER and Associate Professor & Director of Undergraduate Studies Civil & Environmental Engineering Department, University of Maryland, USA.

Important Dates

- Abstract Submission: **24 November 2026**
- Acceptance of Abstract: 4 Weeks after Submission
- Submission of Paper: **24 May 2027**

Social and Educational Program

The Social Program Emphasizes the Educational Aspect of the Academic Meetings of Athens Institute.

- Greek Night Entertainment (This is the official dinner of the conference)
- Athens Sightseeing: Old and New-An Educational Urban Walk
- Social Dinner
- Mycenae Visit
- Exploration of the Aegean Islands
- Delphi Visit

Conference Fees

Conference fees vary from 400€ to 2000€
Details can be found at: https://www.atiner.gr/fees

Athens Institute for Education and Research
A World Association of Academics and Researchers

15th Annual International Conference on Industrial, Systems and Design Engineering, 21-26 June 2027, Athens, Greece

The Industrial Engineering Unit of ATINER will hold its **15th Annual International Conference on Industrial, Systems and Design Engineering, 21-26 June 2027, Athens, Greece** sponsored by the Athens Journal of Technology & Engineering.The aim of the conference is to bring together academics, researchers and professionals in areas of Industrial, Systems, Design Engineering and related subjects. You may participate as stream leader, presenter of one paper, chair of a session or observer. Please submit a proposal using the form available (https://www.atiner.gr/2026/FORM-IND.doc).

Important Dates

- Abstract Submission: **24 November 2026**
- Acceptance of Abstract: 4 Weeks after Submission
- Submission of Paper: **24 May 2027**

Academic Member Responsible for the Conference

- **Dr. Theodore Trafalis**, Director, Engineering & Architecture Division, ATINER, Professor of Industrial & Systems Engineering and Director, Optimization & Intelligent Systems Laboratory, The University of Oklahoma, USA.

Social and Educational Program

The Social Program Emphasizes the Educational Aspect of the Academic Meetings of Athens Institute.

- Greek Night Entertainment (This is the official dinner of the conference)
- Athens Sightseeing: Old and New-An Educational Urban Walk
- Social Dinner
- Mycenae Visit
- Exploration of the Aegean Islands
- Delphi Visit
- Ancient Corinth and Cape Sounion

More information can be found here: https://www.atiner.gr/social-program

Conference Fees

Conference fees vary from 400€ to 2000€

Details can be found at: https://www.atiner.gr/fees

*Athens Journal of Technology and Engineering - Volume 13, Issue 3,
September 2026 – Pages 151-170*

Educating Future Engineers to be Responsible Citizens

By Mareen Derda[*] & Marco Wedel[±]

In a time of rapid technological and societal transformation, universities are called upon to prepare future engineers not only for professional success, but also for their societal responsibilities. In addition to technical expertise, engineers need ethical awareness, sustainability competencies, and the ability to reflect critically on the consequences of their actions. Topics such as digitalization, artificial intelligence, and data literacy are becoming increasingly essential. Engineering education must therefore integrate socially relevant skills alongside scientific and methodological foundations. This article presents the ongoing transformation of engineering curricula at TU Berlin, beginning with the adoption of a mission statement on teaching in 2018. It outlines key developments over the past six years, including structural measures and teaching formats designed to foster ethics, sustainability, and interdisciplinary thinking. Practical examples demonstrate how these topics can be embedded into existing programs through interdisciplinary collaboration and curriculum development. The article also reflects on obstacles encountered and lessons learned along the way. By sharing these insights, it aims to contribute to the international dialogue on engineering education and invites readers to discuss strategies, methods, and structures that support the development of responsible and socially engaged engineers.

Keywords: *Sustainability, ethics, data literacy, responsible citizens, curriculum development*

Introduction

Not everyone is an engineer, but every engineer is a member of society – and as such, carries a responsibility that extends beyond the technical sphere. Engineers play a pivotal role in shaping the tools, systems, and infrastructures that influence how people live, work, and interact. Because their decisions can have far-reaching consequences – socially, economically, and ecologically – it is essential that they are not only technically proficient but also aware of the societal implications of their work.

Fostering this awareness is a key educational responsibility. Higher education institutions are therefore called upon to do more than just impart technical knowledge: they must also promote the development of ethical reasoning, critical reflection, and a sense of responsibility. Universities serve as spaces where students – as both future professionals and active members of society – learn to think systemically, weigh consequences, and act in ways that contribute to a sustainable and socially just future.

The Faculty V - Transportation and Mechanical Systems at the Technische Universität Berlin (TU Berlin) has reaffirmed its mission statement for teaching

[*]Vice Dean for Studies and Teaching as well as Equal Opportunities, Technische Universität Berlin/ Faculty V - Mechanical Engineering and Transport Systems, Germany.
[±]Technische Universität Berlin/Department Didactics of Work/Department Work, Technology and Participation, Berlin, Germany.

and learning at its latest faculty retreat in February 2025. This mission statement reads: *"We educate engineers who are aware of their social responsibility, act sustainably and reflectively, and develop motivated competencies that enable them to successfully enter the professional world."*

This requires, in addition to the development of professional, personal, and methodological competencies, also digital competencies, a reflective handling of AI, technology impact assessment, and consideration of the 17 Sustainable Development Goals of the United Nations (SDGs 2015). To achieve this, inter- and transdisciplinary, practice- and research-oriented teaching, encouragement of self-regulated learning, as a prerequisite for academic and professional success, and the continuous alignment and development of course content and curricula to meet the needs of the future labour market and society are necessary.

The argument presented here is situated within the value framework of liberal democracies shaped by European traditions, as outlined in the Charter of Fundamental Rights of the European Union (EU 2012). Accordingly, the meaning of "responsible and sustainable action" is not derived from ethical pluralism (Roberts et al. 2021) or individual viewpoints within democratic discourse, but from fundamental principles and core values that uphold citizens' rights and responsibilities under constitutionally anchored social contracts in the EU (Bundesverfassungsgericht 1986). Based on these foundations, there is growing consensus in higher education that both research and teaching should systematically address ethics, sustainability, digital transformation, diversity, and the principles of good scientific practice (EUA 2021).

The TU Berlin is also working to redesign its study programs and give due place to the promotion of socially relevant competencies in the curricula. This article aims to present the steps taken so far in curriculum development and invite discussion on the necessary steps and methods to develop a curriculum that educates students not only with technical competencies but also as responsible members of our society. The analysis presented in this paper is based on a practice-based expert perspective. The author was directly involved in the strategic development and implementation of the initiatives described in her role as Vice Dean for Studies and Teaching as well as Equal Opportunities at TU Berlin, and as a founding member of the Think Tank, which developed the presented framework. Consequently, the contribution draws primarily on participatory observation and experiential knowledge from ongoing curriculum development processes. Additionally, relevant policy documents, internal reports and higher education regulations were consulted throughout to contextualize and substantiate the described developments. Thus, the paper contributes an insider's perspective on institutional change processes against the backdrop of common European ethical and legal frameworks and policy initiatives, focusing on the articulation and structuring of practices that can subsequently be subjected to empirical evaluation.

In the course of this paper, we first clarify the crucial role of engineering education in and for society and explain the core competencies that responsible engineers should possess. Next, we discuss the special perspective of TU Berlin due to its founding mandate and the implementation steps taken by the university as a whole to integrate socially relevant competencies into teaching and curricula, as well as at the Faculty of Transportation and Mechanical Systems in particular. We also address concepts and implementations, such as possible formats for integrating socially relevant themes into study programs. Building on this, we present our future plans and also address previous obstacles and insights gained.

The Role of Engineers in and for Society

The role of engineers in and for society has undergone a fundamental transformation in the 21st century. Engineering sciences are now intricately linked with societal development. From infrastructure and mobility systems to digital technologies and energy solutions, the work of engineers fundamentally shapes the functioning and development of societies. Consequently, the expectations of future engineers extend far beyond the solution of technical problems. As highlighted in the Position Paper of the European Society for Engineering Education (SEFI 2025), engineers are increasingly seen as agents of change who must steer complex socio-technical systems and proactively contribute to addressing global challenges such as climate change, digitalization, and social justice.

This expanded societal mandate requires a redefinition of the engineering profession. This involves not only the ability to design and implement technical solutions but also the competence to critically reflect on their broader consequences. Engineers are expected to act responsibly, integrate sustainability and ethics into their practice, and engage constructively with the perspectives of various interest groups. Their role entails both freedom and responsibility – the freedom to shape the world through innovation and the responsibility to consider the long-term implications of their decisions.

In summary, the following expectations of society are placed on future engineers:

- Designing sustainable and responsible technologies: Engineers should actively contribute to addressing global challenges such as climate change, resource scarcity, and social justice. They are seen as designers of a sustainable and inclusive future (Pacher 2024).
- Reflecting on societal impacts: Society expects engineers to critically reflect on the social, ecological, and economic consequences of their work and to incorporate these into decision-making processes (Engineers Without Borders UK 2022).
- Innovation and adaptability: Engineers must be able to respond flexibly to new challenges and drive innovation, particularly in the context of digitalization, AI, and green transformation (CESAER 2024).
- Responsibility: Society demands that engineers recognize their special responsibility for the safety, health, and well-being of the general public and align their actions accordingly (Engineers Without Borders UK 2022).

The education of engineers thus stands at the intersection of technological innovation and societal responsibility. To meet the demands of society, future engineers must not only be excellent professionals but also responsible, reflective, and engaged designers of a sustainable future.

The evolving role of engineers as societal stewards aligns closely with central pillars of recent EU policy: e.g. the ecosystem of trust and excellence outlined in the European Commission's White Paper on Artificial Intelligence (European Commission 2020), the Ethics Guidelines for Trustworthy AI developed by the High-Level Expert Group on AI (High-Level Expert Group on Artificial Intelligence 2019); the subsequent legislative framework established within the EU Artificial Intelligence Act (European Union 2024); or the Union of Skills Communication (European Commission2025). Across these frameworks, the EU places citizens – not only in their roles as individuals but also as professionals and economic actors such as engineers and technologists – at the heart of Europe's ambition to shape a digital and sustainable future grounded in European values. Engineers are no longer seen merely as technical implementers but as critical agents of ethical, social, and environmental foresight. As articulated in the AI White Paper, the European approach to AI must be "grounded in fundamental rights such as human dignity and privacy protection" and should contribute to the Sustainable Development Goals, the ethical use of data, and democratic accountability (European Commission 2020). This vision resonates with the societal expectations placed on engineers to reflect on the broader impacts of their work and to act with responsibility toward public well-being and long-term inclusion.

Complementing this, the Union of Skills sets forth a transformative agenda for education and training, stressing the development of critical thinking, civic engagement, digital ethics, and interdisciplinary knowledge as foundational 21st-century skills (European Commission 2020). In particular, the forthcoming AI in Education initiative and the updated Digital Competence Framework promote AI literacy, responsible technology usage, and awareness of misinformation – competencies essential for engineers expected to navigate and co-create trustworthy AI systems. Moreover, the Union of Skills explicitly supports STEAM-based approaches, international cooperation in higher education, and the recognition of non-formal learning (e.g., ethical engagement, sustainability initiatives), all of which resonate with modern engineering roles that require collaboration across disciplines and societal sectors.

The Union of Skills Communication (European Commission 2020) outlines a series of concrete initiatives to be delivered over the coming years, forming the operational backbone of the EU's strategy for skills development. These include:

- an Action Plan on Basic Skills (by Q1 2025), providing a strategic framework to raise foundational competences in literacy, numeracy and digital skills;
- a Basic Skills Support Scheme (pilot) (by 2026), targeting adults with low skills, particularly in disadvantaged regions;
- a 2030 Roadmap on the Future of Digital Education and Skills (by Q4 2025), which will guide inclusive and effective digital learning strategies;
- an AI in Education Initiative (by 2026), including an EU AI literacy framework, teacher training, and ethical guidance on the use of AI in learning environments;
- a STEM Education Strategic Plan (by Q1 2025), designed to increase participation-especially among girls and underrepresented groups-in science, technology, engineering and mathematics;
- an EU Teachers and Trainers Agenda (by 2026), supporting the development of teaching professionals across all sectors;
- a European Competence Framework for Academic Staff (by 2026),

strengthening the pedagogical and interdisciplinary competences of university educators;
- measures for increasing accessibility of higher education (by 2027), aimed at reducing dropout and expanding access for disadvantaged learners (by 2027);
- a European Strategy for Vocational Education and Training (VET) (by 2026), modernizing VET systems and improving their quality, attractiveness and mobility options;
- and an Intergenerational Fairness Strategy (by Q1 2026), promoting social cohesion and addressing age-based skill divides.

Core Competencies for Responsible Engineers

In one of our previous contributions (Derda et al. 2023b) we stated that the postmodern, value-oriented engineer of the 21st century is a person who, in addition to technical knowledge, has also developed social skills during their studies; who is aware of the responsibility of their own work, their own decisions and their impacts, who is able to reflect on their decisions and their own actions from different perspectives, who can perceive and evaluate risks and risk groups and who can respond appropriately to constantly changing social and technical conditions.

In the same article, we also discussed what content should therefore be integrated into engineering courses and curricula. In the current article, we mainly refer to the latest SEFI position paper (SEFI 2025) and the latest findings on dealing with AI.

The European Society for Engineering Education (SEFI), headquartered in Brussels, is Europe's largest network dedicated to the advancement of engineering education. Its primary mission is to promote the development and continuous improvement of engineering education across Europe. SEFI brings together a diverse community of stakeholders, including universities, academics, educators, students, companies, and associated organisations from 47 countries.

In response to global uncertainty and complexity, the SEFI position paper (SEFI 2025) calls for a renewed focus on engineering competencies that goes beyond technical expertise. SEFI identifies three core dimensions of competence – the so-called 3Ts:

- *"Technical: Range from the foundation and application of science to the applied mathematical and computational skills necessary to develop technical solutions in the sense of innovation and current demands."*
- *"Transferable: The cognitive, social and emotional skills needed to thrive in various life contexts which people can potentially transfer from one social, cultural, or work setting to another. They include the ability to work creatively, effectively, and transversely with AI; planning, organizing, and executing projects successfully; as well as motivating and inspiring team members to achieve common goals, active listening, persuasive writing, and the ability to work collaboratively in diverse teams. Working on self-development and continuously improving oneself as a lifelong learner is also part of this category."*
- *"Transdisciplinary: Afford an outward facing, broader appreciation and active*

> *engagement with traditional and nontraditional sources of knowledge from both academic and non-academic domains. These encompass skills related to ethical decision-making, sustainability and social responsibility. These might include the full spectrum of engineering and science, humanities and social science, crafts and other practices from earlier civilisations as well as the business context of engineering solutions, beyond the project-level."*

SEFI further stresses that these competencies are context-specific and always emerge through the interplay of knowledge, skills, and attitudes. Engineering education must therefore offer challenge-based learning formats, support personalization of learning, and foster reflection, curiosity, and error literacy.

In addition to suitable learning formats that promote the development of the required competences, there is also a need, as already indicated in the section on transdisciplinary, for content that should enrich the engineering curriculum in addition to specialist content, such as ethics and sustainability as well as a critical approach to AI, which has recently been increasingly in demand. According to Thies Johannsen (Cooke & Wint 2024) teaching transdisciplinary skills further means 'that we leave the framework of disciplinary knowledge production and work on social challenges together with stakeholders from science, practice and the public.'

It becomes clear, according to the cited sources and the intentions reflected by the various stakeholders, that the role of engineers in the 21st century is expected to extend beyond technical expertise alone. Engineers are not only creators of innovative technologies but also responsible actors who shall integrate social, ethical, and ecological considerations into their work. The demands placed on engineering education reflect this expanded societal expectation: technical competence alone is no longer sufficient. Instead, critical reflection, sustainability awareness, and interdisciplinary thinking have become essential components of modern curricula. Recent European initiatives – embodied notably in the Union of Skills strategy – affirm and support this paradigm shift. They position engineers as key agents in shaping the digital and sustainable transformation of our societies. Engineering education thus plays a crucial role in ensuring that technological innovations are not only technically excellent but also ethically grounded and socially embedded – an indispensable foundation for addressing the complex challenges of our time.

These curricular goals resonate with empirical findings from engineering ethics education, which highlight that students benefit most when ethical issues are addressed repeatedly across different courses, using active and problem-based pedagogies rather than purely normative instruction (Lim 2021). Multi-level reviews also show that sustainable reform requires alignment between course-level learning activities, assessment practices and the broader institutional mission of engineering programmes (Martin et al. 2021). In parallel, the literature on responsible research and innovation underscores a shift of moral responsibility towards engineers and innovators themselves, calling for anticipatory, reflective and inclusive practices throughout the innovation process (Van de Poel & Robaey 2019).

Embedding AI & Data Literacy in the Curriculum

Because the topic of AI in teaching is highly controversial (which could i.a. be observed at our faculty's retreat), we would like to dedicate a short section to this topic.

Recent studies indicate a rapid increase in the use of AI tools such as ChatGPT in higher education. According to a longitudinal survey with over 5,000 students, the proportion of learners who frequently use AI tools for academic purposes rose from 31.8% in 2023 to 50.5% in 2025 (von Garrel & Mayer 2025). The most common applications include clarifying subject-specific concepts (73.2%), text-related tasks such as analysis and generation (56.8%), translation (53.8%), and support in research and literature review (50.8%). These findings highlight that AI is becoming a habitual part of students' learning routines. ChatGPT, alongside DeepL, remains the most widely used tool.

However, the use of artificial intelligence (AI) has a multifaceted impact on students' learning performance – with both positive and negative effects, depending on the application and usage behavior, as a compilation of several publications on the use of AI in higher education in Germany shows.

Positive Effects

- Adaptive learning systems: AI-supported learning platforms analyse individual strengths and weaknesses and adapt learning content and pace to the needs of students. This enables targeted learning and promotes motivation and learning success, especially in heterogeneous learning groups (Rommel n.d., de Witt 2020).
- Increased efficiency: AI can automate repetitive tasks such as literature research, summarizing, citing, or correcting assignments. This leaves more time for deepening specialist knowledge and self-directed learning ((PHWT 2023, de Witt 2020).
- Fast feedback and support: Virtual AI tutors offer 24/7 assistance, answering questions, and providing individual feedback, supporting continuous learning (Rommel n.d., de Witt 2020).
- Early detection of learning problems: AI-supported analyses can identify students at increased risk of performance decline or dropout early on, allowing targeted support measures to be initiated (de Witt et al. 2020.

Negative Effects and Risks

- Superficial learning: Studies show that unreflective or excessive use of AI tools can lead to a decrease in critical thinking, lower self-efficacy, and less in-depth engagement with learning content. Especially when students use AI to solve tasks automatically without engaging with the material, sustained learning performance declines (Campus Schulmanagement 2023).
- Reduced exam performance: Studies show that students who use AI unsupervised as a "solution aid" achieve better results in the short term, but

perform significantly worse in exams without AI support than control groups. A deeper understanding is lacking (Campus Schulmanagement 2023).

- Less self-efficacy and sustainable learning: The use of AI can lead to students being less immersed in the learning process ("flow") and developing less confidence in their own abilities, which negatively impacts motivation and long-term learning (Campus Schulmanagement 2023).
- Risk of dependency and plagiarism: A heavy dependence on AI tools can inhibit the development of problem-solving and analytical skills and increase the risk of plagiarism (Campus Schulmanagement 2023, Hochschule Harz, 2024).

An international study by Wang & Fan (Wang & Fan 2025) a meta-analysis of 51 studies and more than 10,000 participants, also comes to similar results. It shows that AI can improve learning, but only under certain conditions. What is crucial is systematic, planned use with didactic integration and the promotion of responsible, reflective use of AI in studies.

This increased integration as well as the results on the impact of the use of AI on learning underscores the need for higher education institutions to engage critically and constructively with AI-supported learning processes—both in terms of didactic integration and academic integrity.

An example of the didactic embedding of the use of AI in an engineering module was presented at our faculty teaching conference by Prof Kraft (2025) from the Department of Medical Engineering.

In a two-part module series on Medical Engineering, artificial intelligence tools such as ChatGPT were purposefully embedded into course design to foster critical reflection and responsible use. In the first course, Medical Engineering I, students worked in small groups to address a predefined question using ChatGPT–once with and once without a character limit. The generated responses were then evaluated in terms of scientific accuracy, depth, and relevance. Students compared the short and long versions, identified factual errors, redundancies, and missing content, and formulated suggestions for improving the associated lecture. They also reflected on the usefulness and limitations of ChatGPT for academic study and exam preparation.

In Medical Engineering II, students selected a topic from the course and prepared a group presentation that combined core lecture content with recent developments based on their own research. The use of AI tools was explicitly permitted but had to be documented transparently. The oral presentation was followed by a critical discussion to assess subject understanding and evaluate the quality of AI-assisted content creation.

This integrated approach illustrates how AI tools can be used not only to support learning but also to strengthen evaluative judgment, academic integrity, and digital literacy in engineering education.

The TU Berlin Approach: From Mission to Implementation

Democracy and market economy are not natural allies; rather, they often exist in a tense and complex relationship. Likewise, recent developments – particularly in the United States – demonstrate that the relationship between the state and science is far from inherently harmonious. Political agendas, economic interests, and ideological pressures can challenge or even undermine the independence of scientific inquiry. Against this backdrop, the integration of ethical reflection and social responsibility into the education of engineers gains particular importance. The Technische Universität Berlin has embraced this responsibility since its reestablishment after World War II and continues to see it as a core mission to educate future engineers not only with technical expertise but also with a grounded moral and societal awareness.

Due to its history and its founding mission, Technische Universität Berlin has a special perspective on the integration of ethics and social responsibility into the education of future engineers. When the university, formerly Technische Hochschule Charlottenburg, was re-established after the Second World War, the Allies decided that scientific education had to be inextricably linked with ethical reflection and social responsibility. This was to be emphasized by the aim of introducing the humanities, social sciences and education as an integral part of the curriculum. In his inaugural address on 9 April 1946, Major General Eric P. Nares emphasized that the institution should be "*a home of true education and not mere technology"* (Nares 1946) and underlined the importance of integrating ethical and social responsibility into technical education. He emphasised the university's new mission to serve humanity and to distance itself from its previous role of supporting the German war effort. All education, whether technical or humanistic, must be based on the ideal of a responsible attitude towards society (Hark 2023).

The Technische Universität Berlin remains committed to this founding mission and the associated responsibility to this day and will continue to do so in the future. Nevertheless, as outlined above, the demands on future engineers have changed in recent decades and curricula must therefore be constantly adapted to the needs of the world of work and society. We would like to summarize the main developments in this regard at the Technische Universität Berlin in recent years.

In 2018, after a year-long participatory development process involving all status groups, a mission statement for teaching was adopted, which is binding for all members of the university. The mission statement is understood as a constitution for teaching and formulates educational goals and content, some of which are relevant to society. Teaching should be modern, innovative, digital, practice-orientated and international. Students should be enabled to face technological change and its social impact with professional qualifications and a sense of responsibility and to consider the ethical consequences of their actions (Technische Universität Berlin n.d.).

In line with this mission statement for teaching, the General Study and Examination Regulations (AllgStuPO) were revised, with the new version entering into force in August 2021 and expanded in September 2023. This revision establishes a clear mandate for integrating these principles into the curricula of all study programs (Technische Universität Berlin 2024):

(3) In the degree programmes, the rules of good scientific practice are taught at the earliest possible stage in accordance with the applicable principles for safeguarding good scientific practice at the TU Berlin and trained on an ongoing basis.
(4)[1]Students learn to place their own and general knowledge and actions in an overarching historical, social and cultural context and to consider the ethical consequences of their actions in order to be able to contribute to sustainable development in the sense of the United Nations Sustainable Development Goals. [2]To this end, it must be ensured that all students have completed at least 12 CP of relevant course content by the time they complete the degree programme.

Although these regulations apply to all degree programs at TU Berlin across all faculties, the development of socially relevant competencies such as ethical reflection, sustainability, and social responsibility cannot be assumed to occur automatically within any disciplinary context. Depending on disciplinary traditions, curricular structures, and teaching cultures, these competencies may be addressed to varying extents or remain implicit. The present contribution therefore deliberately focuses on engineering education, as the transformation processes described here were initiated, piloted, and institutionally anchored at the Faculty of Transport and Mechanical Engineering.

In response to the mission statement for teaching and the new version of the AllgStuPO, which established centrally binding requirements for all faculties to revise their degree programmes, the Faculty of Transport and Mechanical Engineering held a teaching retreat in 2022. The aim was to inform faculty members about these new requirements. Building on this, a Think Tank Technology Reflection was founded at the same event, tasked with driving the curriculum development process forward through monthly meetings. This think tank involved participation from all status groups and, fortunately, a high level of student involvement.

The Think Tank Technology Reflection's first step was to analyse the existing modules at the faculty and in the module catalogues that cover topics such as ethics, sustainability, digitalization and technology assessment. This was done in order to establish the current state of affairs as a basis for further curriculum development.

Concepts and Formats from TU Berlin

In a second step, options were identified and developed as to how these topics could continue to be integrated into the curricula. The most easily realisable option is the establishment of a compulsory elective area from which modules on the above-mentioned topics can be selected for a certain number of credit points. Interestingly, this option was already implemented in the Physical Engineering degree programme at the insistence of the students before the AllgStuPO was amended. Here there is a compulsory elective area on the topics of *'ecological and social competences'*, in this respect this degree programme was ahead of its time. Nevertheless, this solution for integrating the topics is not the last choice, as it harbours the risk that students may learn about socially relevant topics, but that these remain unreflected and detached from the subject-specific content. It must therefore be ensured that integration and reflection also take place in the engineering modules. To this end, the Think Tank Technology Reflection developed a concept and various formats that are summarised in the ENG+

Framework, which has already been presented at various conferences (Ammon et al. 2023, Derda et al. 2023a). The ENG+ Framework is designed as a transferable model that can be adapted to other disciplines, provided appropriate subject-specific connections are developed systematically. It offers a systematic approach to anchoring ethics, reflection on science and technology and sustainability in the curriculum and strengthening core values such as diversity and good scientific practice and is based on the fundamental strategies of *advancing* and *complementing*, which should go hand in hand. It is important to emphasize that the ENG+ Framework does not primarily rely on the introduction of additional, isolated courses. Instead, its central objective is the discipline-integrated embedding of ethical, societal, and sustainability-related reflection into existing engineering modules (*advancing*), complemented by selected interdisciplinary formats where appropriate (*complementing*).

The *Advancing* strategy focuses on further developing existing engineering courses to highlight the relevance of ethics and reflection in engineering and their connection to sustainability. Central to this approach is providing targeted support and training for engineering educators. It comprises three core measures:

- *Emphasize* refers to the enhancement of courses that already incorporate ethical, scientific, or technological reflection by making these aspects more visible and explicitly linking them to questions of sustainable development.
- *Empower* describes the process of equipping educators – particularly those teaching courses that have not yet addressed such dimensions – with the knowledge and tools needed to integrate ethical and reflective content into their disciplinary frameworks.
- *Embed* involves the inclusion of external expertise from the humanities, ethics, or social sciences. In this model, short teaching units (2-6 hours) are incorporated into engineering courses to address ethical and reflective questions directly within the course topic.

The *Complementary* strategy, *Complementing*, focuses on the development of new courses dedicated specifically to ethics, science reflection, and technology reflection. These courses are tailored to the thematic contexts of the respective engineering disciplines and are typically co-taught in interdisciplinary formats. Three types of courses are observed:

- *Enable* courses provide foundational knowledge in ethics and introduce basic reflection practices, typically by contextualizing engineering topics within broader societal and ethical frameworks.
- *Enrich* courses offer more in-depth engagement with ethical or reflective content and may lead to formal recognition through certificates, such as the Berlin Ethics Certificate or the Sustainability Certificate (Ammon et al. 2022).
- *Encounter* formats create interdisciplinary learning environments where instructors and students from engineering fields and the humanities or social sciences work together. These settings allow participants to experience different disciplinary perspectives and underscore the value of cooperation across academic cultures.

While the ENG+ framework offers a coherent and modular approach to integrating ethics, sustainability and technology reflection into engineering curricula, it also exhibits characteristic strengths and weaknesses. A key strength lies in its hybrid design which combines the advancement of existing disciplinary modules with complementary, stand-alone learning opportunities. This aligns with empirical findings that ethics and societal issues are most effective when taught both in dedicated courses and embedded across the curriculum, rather than in isolated "add-on" formats (Martin et al. 2021, Martin et al. 2020).

At the same time, the approach is vulnerable to several bottlenecks. First, successful implementation relies heavily on motivated individual instructors who are willing and able to redesign their courses, a challenge widely documented in engineering ethics education where faculty often report lack of time, pedagogical training and institutional incentives. Second, the framework presupposes sufficient institutional resources for coordination, staff development and inter-faculty collaboration, which are not always available. Third, the complementary courses and certificates risk attracting primarily already interested students, thereby reproducing a "preaching to the converted" effect instead of reaching the full student body (Martin 2024, Telsang & Kulkarni 2014).

When contrasted with similar initiatives at other universities, ENG+ shares important features but also follows a distinct trajectory. Case-based and project-based ethics integration, as reported for instance in North American and European engineering programs, similarly emphasises contextualised ethical reasoning within technical design courses. However, the TU Berlin model places comparatively stronger emphasis on institutional anchoring through faculty-level governance structures and on linking ethics and sustainability to broader university-wide missions and certificates, thereby combining bottom-up experimentation with top-down curriculum regulation. This dual anchoring increases robustness but also makes the process slower and more exposed to shifts in institutional priorities (Ammon et al. 2023, Mott & Peuker 2024).

As the development of various modules at the faculty in recent times has affected almost all of these measures recently, providing a list of examples would be excessive at this point. Instead we will rather refer to the overarching development of the degree programmes using the Mechanical Engineering degree programme as an example. This section aims to demonstrate how socially relevant competencies are integrated into the programme structure. A detailed documentation of individual syllabuses, lesson plans or examination formats is beyond the scope of this contribution, which focuses specifically on curriculum architecture and institutional transformation processes.

The following concept was recently presented for the revision of the Bachelor of Mechanical Engineering programme, which ensures the integration of the topics of sustainability, ethics and social responsibility to the extent of 12 CP:

For this purpose, the existing module '*Introduction to Mechanical Engineering*', which is strongly practice-orientated, will be expanded by 3 CP to cover the basics of the topics mentioned, which can be built upon in the further course of the degree programme. This module is specifically intended for the start of the degree programme in order to use the students as multipliers for this content on the one hand and to avoid students dropping out at an early stage through the integration and relevance of socially relevant topics on the other.

Furthermore, a 6 CP compulsory elective area *sets* (sustainability, ethics, transfer, social issues) will be created, in which competences in the areas of sustainability, ethics, transfer and social issues will be taught.

The Projects module area, which is included in the Bachelor's degree with 6 CP, is now also intended to impart competences in the area of *sets*. In addition to the basic requirements, such as 1. practical complex planning engineering activities in mechanical engineering, e.g. product development or production planning, 2. teamwork (3 to a maximum of 6 team members), 3. development of a project plan, 4. project management, 5. development and documentation of solutions, 6. colloquium/ presentation; the modules attended here must add the following additional requirements:

- System environment analysis to identify sustainability-relevant requirements
- Development of appropriate sustainable solutions
- Holistic, critical thinking
- Application of methods for evaluating sustainability
- Self-reflection

In order to ensure that students also learn to deal with AI in a reflective manner during their studies, the compulsory elective area of information technology has been divided into two areas, *programming* and *AI and data competence*. Modules are to be selected from both areas. The concept is currently being further developed by the degree programme working group, with the involvement of students.

It should be emphasized that the introduction of these 12 CP does not increase the overall credit volume of the degree program. Instead, these competencies are structurally redistributed within existing program requirements. The intention is not to add further workload, but to enhance relevance, motivation, and coherence of learning experiences through meaningful integration.

It is also worth mentioning that the *Gesellschaft von Freunden der TU Berlin e. V.* (Technische Universität Berlin n. d.) awards an annual prize for exemplary teaching. This prize recognises and promotes innovations and examples of good teaching practice at TU Berlin, as well as making them more visible both internally and externally. The award, which is endowed with 4,000 euros, is presented annually with a different focus. In 2024, the focus was on the integration of sustainability aspects into basic courses of the Bachelor's degree programmes. Among other things, the award recognised the lecture on drive technology in the field of mechanical engineering, which was continuously developed by the lecturer into the lecture '*Sustainable Drive Technology*' with the support of the students. Here, what was called for in the AllgStuPO above was achieved: integrating the cross-cutting topic of sustainability in such a way that the subject-related learning objectives are retained, but nevertheless systematic references to climate protection, energy saving and waste avoidance are established. The lively, open-ended discussions on the topics in the course led to an above-average evaluation of the lecture by the students.

This year, the prize will be awarded on the topic of AI in teaching. The aim is to honour courses that focus on a critical, reflective approach to AI. It will be interesting to see.

So far, promising concepts and formats have been developed, and many courses have been revised to include the teaching of transferable and transdisciplinary skills alongside technical skills. More will follow because, as mentioned earlier, our goal is to train engineers who are aware of their social responsibility and can act sustainably and reflectively, developing the skills they need to start their careers successfully. Some of our degree programmes are currently being revised to adapt to the evolving skill sets required of future engineers. Once these have been updated, the next curricula will be revised, as this ongoing goal requires courses and curricula to be constantly further developed and aligned with the current needs of the world of work and society.

Limitations and Transferability

The ENG+ framework is intentionally designed as a modular and competence-oriented model, which in principle allows for adaptation to different institutional contexts. Nevertheless, its transferability is subject to several limitations. At TU Berlin, ongoing budget constraints and hiring freezes have considerably slowed down implementation, as curriculum redesign, cross-faculty cooperation and new teaching formats all require time, staffing and organisational support. This illustrates a broader pattern in engineering education reforms, where ambitious visions for ethics and sustainability frequently encounter structural barriers such as limited resources, rigid regulations and competing priorities in faculties (Martin et al. 2021, Ammon et al. 2023, Telsang & Kulkarni 2014).

Moreover, the TU Berlin model is embedded in a specific regulatory and cultural environment, including a university-wide study regulation that mandates ethics and reflection components and a strong tradition of collaboration between engineering, humanities and social sciences. Institutions operating under different governance structures, accreditation regimes or disciplinary cultures may not be able to replicate this configuration one-to-one. Instead, they might selectively adopt elements of ENG+ - for example, the distinction between advancing and complementing strategies or the six integration measure – while developing their own organisational arrangements. Future research could therefore focus on comparative case studies that examine how similar frameworks evolve in contexts with different resource levels, incentive structures and national higher-education policies (Ammon et al. 2023).

Lessons Learned and Challenges

In the following section, we want to describe the main obstacles in the development of the curricula, as well as our measures to deal with them and give same recommendations based on our experiences with the development of curricula for socially responsible acting engineers.

These are the main obstacles we have encountered in recent years:

1. Not all technical subjects are suitable for integrating ethical or social reflection in a meaningful way (e.g. basic maths).
2. Conservative subject-specific understandings make it difficult to open up the curriculum to new topics; teachers who see the teaching of technology as the sole educational goal often underestimate the importance of reflection and responsible behaviour.
3. Teachers need time, support and further training in order to integrate ethical content in a high-quality manner.
4. A lack of incentive systems makes it difficult to motivate further development and existing performance measurement models hinder freedom for innovative teaching.

And here are our ways to deal with them:

Re 1: There is no need for integration in every module. Forced integration does not have the desired learning effect on students. On the other hand, the overall concept of the degree programme should ensure that knowledge of socially relevant topics and the necessary skills described are acquired during the course of the degree programme.

Re 2: Constant dripping wears away the stone. In recent years, we have never tired of describing the relevance of the content and skills described for the engineering degree programme, holding discussions and presenting best practices. The faculty's annual teaching retreat with expert input, workshops on transfer and discussion rounds, as well as the monthly Think Tank Technology Reflection, which took place at the beginning, served this purpose.

Re 3: Yes, it takes time. Fortunately, many of our faculty's lecturers are highly motivated and constantly developing their courses. To back this up, the university's own central organisation, Scientific Continuing Education and Cooperation (ZEWK) (Technische Universität Berlin n.d.), and the Berlin Centre for Further Education (BZHL) (Technische Universität Berlin n.d.), with which we liaise frequently, provide appropriate additional instruction for instructors.

Re 4: In the meantime, obstacles to the implementation of interdisciplinary teaching formats, such as crediting to the teaching load, have been removed. Consideration is currently being given to further incentivisation, for example through the faculty's own recording of student numbers for the distribution of resources. What is certain is that the prize for exemplary teaching has already created a certain incentive through the visibility achieved. The introduction of a similar award within the faculty could therefore also provide a clear incentive.

We would like to summarise a few recommendations based on our experience over the last few years of developing the degree programme towards the competence profile of a responsible engineer.

- A central mission statement for the entire university and the amendment to the General University Regulations (AllgStuPO) set a clear commitment to the relevance of socially relevant topics in all degree programmes and created binding rules for all university members. This creates a certain pressure to change and develop the degree programmes.

- At the beginning of a change process, it seems important to gain an overview of what already exists and to visualise best practices from which others can learn.
- On the other hand, it seems important to involve as many participants as possible and to enable low-threshold discussion spaces in order to gain acceptance from teachers and students and to develop new concepts together. In particular, this means identifying the needs of students and taking them seriously.
- Our annual faculty teaching retreats as well as the cross-status group and cross-faculty think tank described above served both purposes.
- Exchanges with other universities, such as at international conferences on engineering education or in smaller groups, or advice from experts, were also extremely useful for us to gain ideas for curriculum development. In addition to attending conferences on engineering education, we at the Faculty of Transport and Mechanical Systems took advantage of a year-long consultation process organised by the Hochschulforum Digitalisierung (n.d.) to further develop our degree programmes, as well as a low-threshold digital exchange with University College London as part of our Think Tank Technology Reflection, for example.
- And last but not least: do good and talk about it. The constant communication of development processes, including small steps, within and outside the faculty creates transparency and motivation for everyone by emphasising what has already been achieved.

We hope that these recommendations will inspire you to take the first steps and make it easier for you to further develop the study programmes at your university.

Conclusion

The challenges of the 21st century require engineers who are not only technically competent but also socially responsible, ethically reflective, and capable of acting within complex and dynamic systems. Engineering education must therefore evolve beyond a narrow focus on disciplinary knowledge. As this article has argued, fostering competencies in areas such as ethics, sustainability, digital literacy, diversity, and scientific integrity is not an optional add-on, but a core component of educating future engineers.

The ENG+ framework is intentionally designed as a modular and competence-oriented model, which in principle allows for adaptation to different institutional contexts. Nevertheless, its transferability is subject to several limitations. At TU Berlin, ongoing budget constraints and hiring freezes have considerably slowed down implementation, as curriculum redesign, cross-faculty cooperation and new teaching formats all require time, staffing and organisational support. This illustrates a broader pattern in engineering education reforms, where ambitious visions for ethics and sustainability frequently encounter structural barriers such as limited resources, rigid regulations and competing priorities in faculties.

Moreover, the TU Berlin model is embedded in a specific regulatory and cultural environment, including a university-wide study regulation that mandates ethics and reflection components and a strong tradition of collaboration between engineering, humanities and social sciences. Institutions operating under different governance structures, accreditation regimes or disciplinary cultures may not be able to replicate this configuration one-to-one. Instead, they might selectively adopt elements of ENG+ – for example, the distinction between advancing and complementing strategies or the six integration measures – while developing their own organisational arrangements. Future research could therefore focus on comparative case studies that examine how similar frameworks evolve in contexts with different resource levels, incentive structures and national higher-education policies.

Against this backdrop, the approaches discussed in this paper – ranging from curriculum development and interdisciplinary collaboration to structural reforms – illustrate that meaningful change is possible when academic institutions commit to cultural as well as pedagogical transformation. Even though resistance and structural barriers persist, the examples presented demonstrate that new forms of teaching and learning can be implemented in ways that are both context-sensitive and scalable. At the present stage, systematic evaluation data on learning outcomes or long-term impact are not yet available, as many of the curricular reform steps are still in early phases of implementation. Nevertheless, strong qualitative indicators for relevance and acceptance can already be observed, particularly through sustained student engagement in curriculum committees and in the Think Tank Technology Reflection. Students have repeatedly emphasized the importance of ethics, sustainability, and digital responsibility for their future professional roles, thereby actively reinforcing the reform agenda.

In light of these developments, it becomes evident that universities can no longer rely solely on static curricula. Instead, they must create learning environments that are flexible, inclusive, and globally oriented – supported by a shared commitment to continuous renewal and responsiveness to societal change. This requires shared responsibility, structural support, and institutional spaces for experimentation, dialogue, and reflection.

We invite all colleagues and institutions engaged in engineering education to share ideas, experiences, and practices–because shaping the engineers of tomorrow is a collective task that can only be successfully addressed together.

References

Ammon S, Derda M, Hildebrandt T, Kühne S, Klein N (2023) *ENG+ | Ethics for Sustainable Development in Engineering Programmes. A Praxis Report of TU Berlin's Think Tank Technology Reflection*. In: Proceedings of the SEFI Annual Conference. Engineering Education for Sustainability, 11.–14. September 2023, TU Dublin, UK.

Ammon S, Kljagin A, Rettschlag J, Vortel M (2022) *The Berlin Ethics Certificate: Conceptualizing Interdisciplinarity as a Core Building Block of Ethics in Engineering Education.* In Towards a New Future in Engineering Education, New Scenarios That European Alliances of Tech Universities Open u*p*, 913–924. Universitat Politècnica de Catalunya. https://doi.org/10.5821/conference-9788412322262.1422

Campus Schulmanagement (2023 ChatGPT als Gedankenbremse: Wie KI das Lernen erschweren kann. Available at: https://www.campus-schulmanagement.de/magazin/chatgpt-als-gedankenbremse-wie-ki-das-lernen-erschweren-kann

Cooke N, Wint N (2024) *Transdisciplinary* (Season 5, Episode 4) [Audio podcast episode]. In European Engineering Educators. SEFI. https://www.sefi.be/2024/12/16/podcast-seas on-5-episode-4-transdisciplinary/

CESAER (2024) *White paper Engineer of the future*. Zenodo. https://doi.org/10.5281/zenodo.10972834

Derda M, Ammon S, Hildebrandt T, Klein N (2023a) *Transformation der Studiengänge für sozial verantwortliche und nachhaltig handelnde Ingenieure und Ingenieurinnen*. Presentation at the TURN Conference, 13.–15. September 2023, Köln.

Derda M, Wedel M, Albrecht M (2023b) *Curriculum Development for Socially Responsible and Sustainably Acting Engineers.* In: Proceedings of the Conference 2023 46th ICT and Electronics Convention (MIPRO), Opatija, Croatia, 1862–1867. https://doi.org/10.23919/MIPRO57284.2023.10159936

de Witt C, Rampelt F, Pinkwart N, Bernd M, Brandt S, Burchardt A, Dufentester C, Etsiwah B, Gloerfeld C, Kravčík M, Mah D, Pinkwart N, Rampelt F, Renz A, Schwaetzer E, Wrede S (2020) *Künstliche Intelligenz in der Hochschulbildung: Whitepaper*. KI-Campus. https://ki-campus.org/sites/default/files/2020-10/Whitepaper_KI_in _der_Hochschulbildung.pdf

Engineers Without Borders UK (2022) *Global responsibility of engineering - The extent to which global responsibility is embedded into engineering practice: an exploratory study. London: Engineers Without Borders UK*. Available at: https://www.ewb-uk.org/global-responsibility-of-engineering-an-exploratory-study/

European Commission (2020) *White Paper on Artificial Intelligence: A European approach to excellence and trust* (COM(2020) 65 final). Publications Office of the European Union. https://commission.europa.eu/system/files/2020-02/commission-white-paper-artificial-intelligence-feb2020_en.pdf

European Commission (2025) *Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions: The Union of Skills* (COM(2025) 90 final). Publications Office of the European Union. https://www.cedefop.europa.eu/files/communication_-_union_of_skills.pdf

European University Association (2021) *Universities without walls: A vision for 2030*. https://www.eua.eu/publications/positions/universities-without-walls-eua-s-vision-for-europe-s-universities-in-2030.html

European Union (2012) *Charter of Fundamental Rights of the European Union.* https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:12012P/TXT

European Union (2024) Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending certain Union legislative acts (Artificial Intelligence Act). *Official Journal of the European Union*, L, 202, 1–151. https://eur-lex.europa.eu/legal-content/EN/TXT/?Uri=CELEX:32024R1689

High-Level Expert Group on Artificial Intelligence (2019) *Ethics guidelines for trustworthy AI. European Commission*. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

Hark S (2023) TU-Geisteswissenschaften schließen? Die Senatorin stellt den Gründungsauftrag der Universität infrage. *Der Tagesspiegel*. https://www.tagesspiegel.de/wissen/tu-geisteswissenschaften-schliessen-die-senatorin-stellt-den-grundungsauftrag-der-universitat-infrage-13284490.html

Hochschulforum Digitalisierung (n. d.) Peer-to-Peer-Fachbereichsberatung. https://hochschulforumdigitalisierung.de/peer-to-peer-fachbereichsberatung/

Lim JH (2021) "In our own little world": Invisibility of the social and ethical dimension of engineering in higher education. *Studies in Engineering Education*, 2(1), 1–18.

Martin DA (2024) *Teaching methods in engineering ethics education*. In S. Chance, T. Børsen, D. A. Martin, R. Tormey, T. T. Lennerfors, & G. Bombaerts (Eds.), The Routledge International Handbook of Engineering Ethics Education (pp. 339–344). Routledge, Taylor & Francis Group. https://doi.org/10.4324/9781003464259-23

Martin DA, Conlon E, Bowe B (2020) Exploring the curricular content of engineering ethics education in Ireland. In *2020 IFEES World Engineering Education Forum – Global Engineering Deans Council (WEEF–GEDC)* (pp. 1–5). IEEE. https://doi.org/10.1109/WEEF-GEDC49885.2020.9293664

Martin DA, Conlon E, Bowe B (2021) A Multi-level Review of Engineering Ethics Education: Towards a Socio-technical Orientation of Engineering Education for Ethics. *Sci Eng Ethics*, 27(5):60. https://doi.org/10.1007/s11948-021-00333-6

Mott J, Peuker S (2024) *Integrating ethics into engineering education: A case-based learning approach*. In 2024 ASEE Annual Conference & Exposition, Portland, Oregon (pp. 1–xx). American Society for Engineering Education. https://doi.org/10.18260/1-2--47655

Nares EP (1946) *Opening speech at the inauguration of the Technical University of Berlin*. British Military Government.

PHWT (2023) *Künstliche Intelligenz im Studium nutzen – aber wie?* https://www.phwt.de/ki-im-studium-nutzen/14922/

Pacher C, Woschank M, Zunk BM, Gruber E (2024) Engineering education 5.0: a systematic literature review on competence-based education in the industrial engineering and management discipline. *Production & Manufacturing Research*, 12(1). https://doi.org/ 10.1080/21693277.2024.2337224

Rommel L (n.d.) *Wie beeinflusst Künstliche Intelligenz unser Studium?* Universität Jena – eLearning. https://www.elearning.uni-jena.de/1695/einfluss-von-ki-auf-das-studium

Sustainable Development Goals of the United Nations (2015) https://sdgs.un.org/goals

SEFI Position Paper (2025) *Engineering Competencies and Skills for an Uncertain Future.* https://www.sefi.be/wp-content/uploads/2025/03/SEFI-Skills-position-paper-2025-1.pdf

Telsang M, Kulkarni SS (2014) Institutional transformation in action of a private engineering college in Western Maharashtra: A case study. *Journal of Engineering Education Transformations*, 27(4), 41–47.

van de Poel I, Robaey Z (2019) Responsible innovation in business: A critical reflection on conceptions and approaches. *Journal of Responsible Innovation*, 6(2), 1–24.

von Garrel J, Mayer J (2025) Künstliche Intelligenz im Studium – Eine quantitative Längsschnittstudie zur Nutzung KI-basierter Tools durch Studierende. https://doi.org/10.48444/h_docs-pub-533

Wang J, Fan W (2025) The effect of ChatGPT on students' learning performance, learning perception, and higher-order thinking: insights from a meta-analysis. *Humanities and Social Sciences Communications*, 12, 621. https://doi.org/10.1057/s41599-025-04787-y

Wühle M (2022) *Ökobilanz und CO_2-Fußabdruck: Entscheidende Instrumente für nachhaltiges Wirtschaften* [Life cycle assessment and carbon footprint: crucial tools for sustainable business practices]. Available at: https://www.nachhaltigkeit-management.de/okobilanz-und-co2-fussabdruck

VDI Center for Resource Efficiency (2019) Ökobilanz - DIN EN ISO 14040/44. Leitfaden Ressourceneffizienz [Life cycle assessment - DIN EN ISO 14040/44. Resource efficiency guidelines]. Available at: https://www.ressource-deutschland.de/leitfaden-re/methoden/oekobilanz-din-en-iso-14040/44

World Resources Institute (2025) Greenhouse Gas Protocol. Available at: https://www.wri.org/initiatives/greenhouse-gas-protocol

Athens Journal of Technology and Engineering - Volume 13, Issue 3, September 2026 – Pages 171-194

Teaching Financial Literacy with a STEAM Approach

By Michihiro Sakai[*]*, Masatoshi Okita*[±]*, Katsutoshi Kawashima*[•] *& Mitsuharu Aizawa*[○]

In Japan, children have traditionally had limited exposure to financial literacy education. However, recent societal changes—such as low interest rates, wider internet access, and the lowering of the legal age of adulthood—have increased the need for financial education, which was introduced to high schools in 2022. In line with Society 5.0, which emphasizes problem-solving and interdisciplinary thinking, Kurume KOSEN has incorporated STEAM education into its "Liberal Arts Special Lecture" since 2019. This course promotes collaboration among students from various departments and fosters deep learning. As part of this initiative, we developed STEAM-based financial education materials on simple and compound interest, integrating economics and mathematics. These materials were used in an open course for elementary and junior high school students and their parents, with KOSEN students serving as teaching assistants. The course was divided into two parts: the first explored future financial needs and money growth using arithmetic and geometric sequences; the second explained the differences between simple and compound interest through simulations, also addressing the risks of debt. The program was well received, offering valuable social engagement opportunities. Although still in the early stages, the initiative shows promise and calls for further refinement of teaching methods.

Keywords: *financial education, simple interest and compound interest, isometric sequence and geometric sequence, economics, mathematics*

Introduction

According to Van Rooij, Lusardi, and Alessie (2011, 2012), financial literacy encompasses not only understanding interest rates but also inflation, time value of money, and money illusion. Their standard questionnaire has been widely used internationally, allowing for cross-country comparisons. Incorporating these elements into our STEAM-based financial education could enhance its comprehensiveness and relevance.

STEAM education, introduced by Yakman (2008), is a learning approach that utilizes science, technology, engineering, the arts, and mathematics as entry points to promote inquiry, dialogue, and critical thinking in solving real-world problems. While a

[*]Professor, Department of Liberal Arts (Natural Science and Mathematics), National Institute of Technology, Kurume College, Japan.
[±]Associate Professor, Department of Liberal Arts (Natural Science and Mathematics), National Institute of Technology, Kurume College, Japan.
[•]Associate Professor, Institute of Education, Center of Advanced Education, Osaka Sangyo University, Japan.
[○]Professor, Faculty of Economics, Ryutsu Keizai University, Japan.

wide range of teaching materials related to STEAM education exists, those incorporating the "A"—the liberal arts—are relatively limited, as STEAM evolved from STEM education. In Japan, 51 technical colleges (KOSEN), including our institution, serve as regional hubs for STEAM education and provide support to junior high school students.

Since 2019, our college has implemented STEAM education as part of a specialized liberal arts course for fourth-year students. In this course, faculty members from various disciplines, including mathematics, debate, and economics, design unique themes based on their expertise, and students from five different departments select and attend lectures that align with their interests. Through collaborative learning among students from diverse academic backgrounds, we aim to foster deep learning that integrates knowledge and creativity. However, opportunities for students to apply and share their learning with society—particularly through teaching others—remain limited.

Meanwhile, the need for financial education in Japan has increased significantly. Factors such as prolonged low interest rates, diversified employment patterns, and changes in the pension system have made it increasingly necessary for individuals to manage their own financial assets. In addition, the lowering of the legal age of adulthood in 2022 has required younger people to make financial decisions and enter into contracts, contributing to the rise in financial troubles among youth. Under these circumstances, the development of financial literacy from an early age has become a pressing issue. Nevertheless, many people in Japan have had few opportunities to receive financial education either at school or at home, and international surveys have revealed that financial knowledge in Japan lags that of other countries.

In response to these challenges, financial education initiatives began in Japan in 2005. With the 2022 revision of the national curriculum guidelines, financial education was officially incorporated into high school home economics and civics classes. The curriculum covers four major areas: "life planning and household budgeting," "financial and economic systems," "consumer protection and prevention of financial problems," and "career education." These areas are designed to help students develop problem-solving skills by learning how to engage with financial matters in their personal lives and in relation to society. However, due to the lack of standardized teaching methods, differences in teachers' knowledge, limited instructional time, and the complexity of the content, financial education still faces numerous challenges.

Against this backdrop, the purpose of this study is to develop and implement financial education materials that integrate perspectives from economics (liberal arts) and mathematics, and to support STEAM education for junior high school students through an open lecture format. The goal is not only to stimulate students' interest in economics and mathematics but also to encourage multifaceted thinking. In addition, we involved students who had previously taken the liberal arts seminar as teaching assistants (TAs), aiming to improve their instructional skills and deepen their understanding of the material.

This initiative was realized in the form of an open lecture for junior high school students in 2024. The course began with the socially relevant and relatable topic of Japan's "20 million yen (approximately 118,000 euros) retirement issue," and included explanations on life planning and money management. This was followed by a smooth transition to mathematical content, introducing the concepts of simple and compound interest as methods of growing money, and demonstrating how to calculate the Rule of 72 as an application of compound interest. Survey results indicated that the materials effectively enhanced students' interest in both economics and mathematics. Furthermore, comments from TAs suggested that their participation improved their teaching skills and motivation to learn.

As a prior case, in 2022, we (Sakai et al., 2023) developed and implemented STEAM teaching materials on "financial economics" for a public lecture open to both junior high school students and the general public. Female students from the liberal arts seminar participated as instructors, applying their knowledge of economics and mathematical skills in a socially engaged context. Given the broad scope of financial economics, this lecture focused on "interest rates" as a central theme, integrating economics (A) and mathematics (M) in the instructional design. In the mathematics section, a female student from the author's mathematics class explained the differences between simple and compound interest, saving and repayment methods, and the long-term impact of compound interest. Exercises using calculators and financial simulators were incorporated to facilitate hands-on understanding, reflecting the simulation approach used in the economics section.

This paper is organized as follows. First, we introduce the background of our initiative, including the "20 million yen retirement issue" in Japan. Second, we outline the content of our practice and review the scenes from the open lecture, including photographs, questionnaire results, and participant feedback. Finally, we present our conclusions and discuss directions for future research.

20 million yen Retirement Issue

This issue gained widespread attention following a 2019 report by the Market Working Group of the Financial Services Agency. The report estimated that, in a model household consisting of a retired couple—specifically, a husband aged 65 or older and a wife aged 60 or older—a monthly shortfall of approximately 55,000 yen in living expenses would lead to a cumulative deficit of about 13 million yen over 20 years, and approximately 20 million yen over 30 years. Although this estimate was based on a specific model case and does not necessarily apply to all individuals, the figure was extensively reported in the media and triggered significant public concern. The issue highlighted the lack of financial preparedness for retirement and emphasized the need to enhance financial literacy among the Japanese population.

Construction of the Open Course

The course included 16 junior high school students and their parents. Participants were self-selected, and the questionnaire was designed to assess six STEAM competencies. Reliability was ensured through internal consistency checks, and statistical analysis (paired t-tests) was conducted to evaluate pre- and post-course changes.

In this class, we designed a financial literacy program as a theme for STEAM education, integrating perspectives from economics (liberal arts) and mathematics. Through activities in which participants consider future financial needs and planning—with guidance from instructors and teaching assistants (TAs)—the course aimed to: (1) help TAs improve their teaching skills and deepen their understanding, (2) stimulate participants' interest in economics and mathematics, and (3) provide students with insights for developing multifaceted thinking. For the TAs, this initiative also served as an opportunity to give back to society by applying their acquired knowledge of economics and mathematics in a real-world educational setting.

The remainder of this paper is structured as follows. First, as an introduction to financial education, we present the "20 million yen retirement issue" and its historical background. Next, we explain the three basic steps of financial planning—reviewing expenses, reducing expenditures, and increasing income—through a series of exercises. In particular, the section on increasing income uses a worksheet to help visualize money growth, accompanied by explanations of arithmetic and geometric sequences, including how to calculate their sums.

Finally, the course concludes with practical exercises that deepen understanding of wealth-building methods, such as simple and compound interest, the Rule of 72, and the effects of long-term saving with compound interest.

To evaluate the effectiveness of the course, pre- and post-class questionnaires were administered to examine changes in participants' awareness and understanding.

The structure of the class was as follows:

(Pre-class questionnaire)
(1) 20 million yen retirement issue
(2) A financial plan
(3) Problem exercises
(4) Methods to increase wealth
(5) Problem exercises
(Post-class questionnaire)

Content of the Practice

In this class, financial education was implemented using a STEAM-based approach for junior high school students and their parents. The program also provided an opportunity for our students—who had studied economics and mathematics through regular courses and liberal arts seminars—to serve as teaching assistants (TAs). The goal was not only to spark participants' interest in finance but also to encourage the development of multiple perspectives. At the same time, the TAs were given a chance to give back to society by applying their academic knowledge in a practical teaching context.

The class was structured as follows:

(a) Learning Content
Financial literacy from the perspectives of economics and mathematics
(b) Structure

- Participants: 16 junior high school students and their parents
- Facilitators: Three female students (TAs) and two instructors
- Duration: 120 minutes (60 minutes for economics, 60 minutes for mathematics)

The class began with a pre-class questionnaire. In this study, we defined six key competencies to be fostered through STEAM education and developed corresponding questionnaire items for each competency:

(1) Problem-Solving Skills
- When you didn't understand something, did you try to look it up yourself?
- When you faced a problem, did you try to solve it on your own?

(2) Critical Thinking
- When you encountered something you didn't understand or found strange, did you think, "Why is that?"
- When listening to your friends or teachers, did you ever think, "Why do they think that way?"

(3) Creativity
- Do you enjoy coming up with fun ideas or thinking of something new?
- When doing something, did you try to add your own creative twist or personal touch?

(4) Flexibility
- When plans or situations changed suddenly, were you able to stay calm and respond appropriately?
- When things didn't go as expected, did you try different approaches to solve the problem?

(5) Perspective-Taking
- When a problem occurred, did you try to consider the cause and solution from a broader point of view?

(6) Execution Ability
- Did you follow through with the plans you made?
- Even when things didn't go well, did you make efforts to complete the task?

-

Introduction to the Course

In this section, we present scenes from the open course along with photographs. Slides were displayed on the screen (Figure 1), and key points regarding the "20 million yen retirement issue" were explained. Using a newspaper article as a reference, we discussed how the amount of money required for retirement has changed over time. We also emphasized that this amount can vary significantly depending on individual living conditions and lifestyle choices. Through this discussion, participants were encouraged to recognize the importance of developing financial literacy that enables them to make informed decisions without being influenced by external information.

Participants took notes on their answer sheets before attempting the exercises, while TAs circulated the room and provided assistance (Figure 2). Through these interactions, the female students serving as TAs developed their ability to teach and support others.

Figure 1. *Lecture*

Figure 2. *Question answering by TAs*

Three Steps for creating a Financial Plan

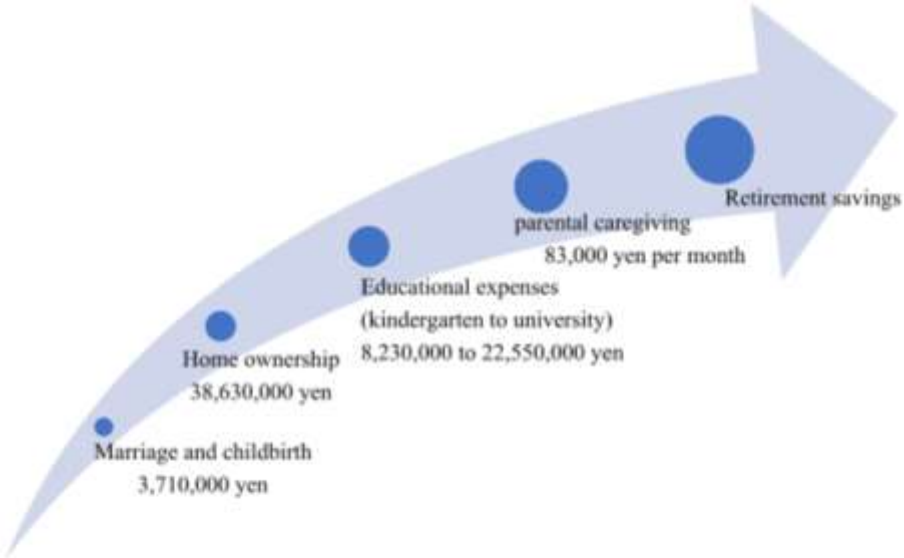
In the next part of the class, we introduced the three steps for creating a financial plan using presentation slides.

- Step 1 involves assessing future expenses by estimating the financial requirements for various life events and essential needs.
- Step 2 focuses on reducing expenditures by identifying opportunities to save through careful spending management and the elimination of unnecessary costs.
- Step 3 entails increasing income by developing strategies to accumulate the remaining required funds—such as through additional earnings, savings, or investment options.

Figure 3. *Three Steps for creating a Financial Plan*

In Step 1, we presented examples of major life events along with their estimated costs (Figure 4).

Figure 4. *Costs associated with Life Events*



Participants were then guided to complete a life planning worksheet to calculate the total amount of money needed through retirement (Table 1). Additionally, by estimating their annual income and expenditures during retirement, participants could project the total financial resources necessary to sustain post-retirement life (Figure 5).

Table 1. *A Sample of Life planning Sheet*

	Family members' ages				Life events, dreams, and goals	Budget
A.D.	Husband	Wife	First child	Second child		
2024	54	44				
2025	55	45			European trip	400,000 yen
2026	56	46			Buying a house (down payment)	4,500,000 yen
2027	57	47				
2028	58	48			Buying a massage chair	500,000 yen
2029	59	49				
2030	60	50				
2031	61	51			Asian trip	300,000 yen
2032	62	52				
2033	63	53				
2034	64	54			World trip	6,000,000 yen
2035	65	55				
2036	66	56			To get a dog	300,000 yen
2037	67	57				
2038	68	58				
					Total budget	12,000,000 yen

Figure 5. *Calculation of Retirement Funds*

Yearly retirement income	Yearly retirement expenses	Yearly deficit
① yen	② yen	③=①-② yen

Yearly deficit	Number of years after age 65	Nursing care costs	Retirement funds
③ yen	④ years	⑤ 8,000,000 yen	⑥=③×④+⑤ yen

In Step 2, participants used a structured worksheet to calculate the amount of savings required to meet their future needs (Figure 6). This was done by subtracting their current assets from the projected expenses identified in Step 1, helping them clarify the financial gap to be filled. After identifying the target savings amount, participants calculated their current monthly income and expenditures using another worksheet (Figure 7). This activity enabled them to better understand their potential monthly savings and evaluate the feasibility of achieving their financial goals. The worksheet also included recommended spending ratios for each expense category, allowing participants to compare these benchmarks with their own expenditures and visually identify areas for adjustment based on their priorities.

Figure 6. *Target savings Confirmation Sheet*

Projected financial needs	① yen

Current assets	
Cash	yen
Savings	yen
Stocks, bonds, and mutual funds	yen
Investment property	yen
cash surrender value	yen
Jewelry	yen
Others	yen
Required savings	
③=①-②	yen

Figure 7. *Household spending Ratio*

Income	
Item	Amount
Monthly income	yen

Expenses			
Items	Amount	Proportion income	Ideal Spending Ratio
Rent	yen	%	20～30%
Insurance premium	yen	%	3～5%
Communication expenses, utilities	yen	%	7～9%
Food expenses	yen	%	10～15%
Social expenses	yen	%	7～10%
Clothing expenses	yen	%	5～7%
Education expenses	yen	%	10%
Miscellaneous expenses	yen	%	3～5%
Savings	yen	%	15～20%
Total expenses	yen	100%	

In Step 3, we introduced various strategies for accumulating additional funds to address potential shortfalls. These included:

- Increasing income through employment (e.g., side jobs or post-retirement work)
- Enhancing pension benefits (e.g., deferring pension claims or using private pension plans)
- Utilizing tax-exempt accounts (e.g., defined contribution plans or NISA)
- Growing wealth through investments (e.g., bonds, stocks, real estate, mutual funds, gold, and foreign exchange)

To help participants understand how money can grow over time, two hands-on exercises using structured worksheets were conducted, as described in what follows:

Exercise 1. Let's stick five colors of round stickers on the board according to the following rules.

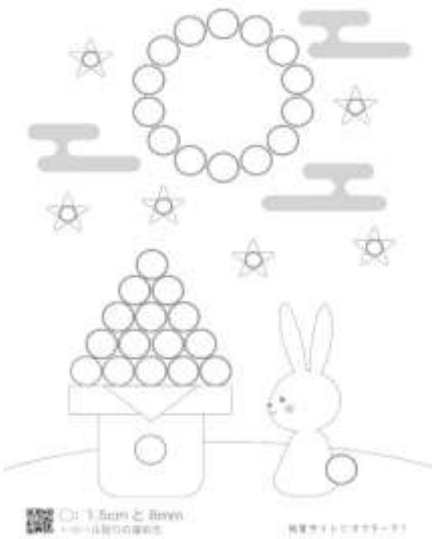
Figure 8. *Christmas. From Chiikumura, 2024 (https://chiiku-baby.jp/maru-sticker/)*

【Rules】
① Use one sticker of the first color.
② Use two more stickers than in ① for the second color.
③ Use two more stickers than in ② for the third color.
④ Use two more stickers than in ③ for the fourth color.
⑤ Use two more stickers than in ④ for the fifth color.

Question 1.
(1) How many stickers were used in total?
(2) If you continue the same rule and use a sixth color, how many stickers would you have in total?

Exercise 2. Let's stick five colors of stickers on the board according to the following rules.

Figure 9. *Otsukimiusagi. From Chiikumura, 2024 (https://chiiku-baby.jp/maru-sticker/)*



【Rules】
① Use one sticker of the first color.
② Use twice as many stickers of the second color as in ①.
③ Use twice as many stickers of the second color as in ②.
④ Use twice as many stickers of the second color as in ③.
⑤ Use twice as many stickers of the second color as in ④.

Question 2.

(1) How many stickers were used in total?
(2) If you continue the same rule and use a sixth color, how many stickers would you have in total?

From a mathematical perspective, the number of stickers in Question 1 follows an arithmetic sequence with a first term of 1 and a common difference of 2 (i.e., 1, 3, 5, 7, 9, …). Part (1) asks for the sum of the first five terms, and part (2) asks for the sum of the first six terms.

In Question 2, the number of stickers forms a geometric sequence with a first term of 1 and a common ratio of 2 (i.e., 1, 2, 4, 8, 16, …). Part (1) asks for the sum of the first five terms, and part (2) asks for the sum of the first six terms.

These two questions use the familiar activity of placing stickers to help participants understand the difference between "adding 2" and "multiplying by 2" in terms of how quantities grow. This visual and hands-on approach supports the conceptual understanding of the fundamental distinction between arithmetic and geometric growth. Furthermore, the process of solving these problems naturally leads to the derivation of the formulas for the sums of arithmetic and geometric sequences (Figures 10 and 11).

Figure 10. *Solution to Problem 1*

(1)

Sticker	First Color	Second Color	Third Color	Fourth Color	Fifth Color
Number of Stickers	1	3	5	7	9

(2) Let's calculate $1 + 3 + 5 + 7 + 9 + 11$. Let this sum be denoted as x.
Then, $x = 1 + 3 + 5 + 7 + 9 + 11$. Since adding in reverse order doesn't change the result, we also have $x = 11 + 9 + 7 + 5 + 3 + 1$.

$$\begin{array}{rl} x = & \boxed{1} + \boxed{3} + \boxed{5} + \boxed{7} + \boxed{9} + \boxed{11} \\ +)\ x = & \boxed{11} + \boxed{9} + \boxed{7} + \boxed{5} + \boxed{3} + \boxed{1} \\ \hline \end{array}$$

Adding these two equations together gives:
$2x = 12 + 12 + 12 + 12 + 12 + 12 = 72$.
Dividing both sides by 2, we get $x = 36$.

Figure 11. *Solution to Problem 2*

(1)

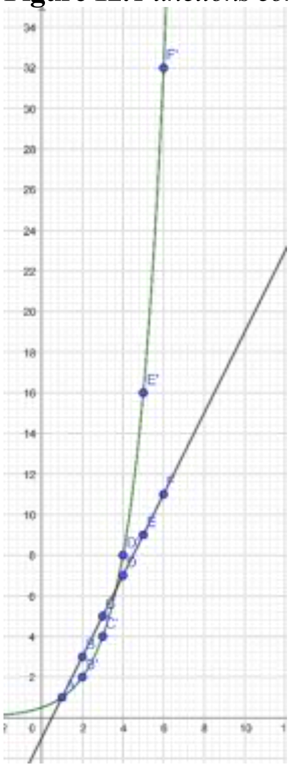
Sticker	First Color	Second Color	Third Color	Fourth Color	Fifth Color
Number of Stickers	1	2	4	8	16

(2) Let $x = 1 + 2 + 4 + 8 + 16 + 32$. If we multiply both sides by 2, we get: $2x = 2 + 4 + 8 + 16 + 32 + 64$. If we calculate $2x - x$, we have $x = 64 - 1 = 63$ (see the following calculation).

$$\begin{array}{rl} 2x = & \phantom{1+{}} 2 + 4 + 8 + 16 + 32 + 64 \\ -)\ \ x = & 1 + 2 + 4 + 8 + 16 + 32 \\ \hline x = & 64 - 1 = 63 \end{array}$$

Furthermore, GeoGebra was used to display graphs of the functions represented in Problems 1 and 2, thereby enhancing participants' visual understanding of the differences in the rate of increase in the number of stickers. In this context, the black graph represents the linear function $y = 2x - 1$ corresponding to Problem 1, while the green graph illustrates the exponential function $y = 2^{x-1}$ associated with Problem 2 (Figure 12).

Figure 12. *Functions corresponding to Problem 1 and 2*



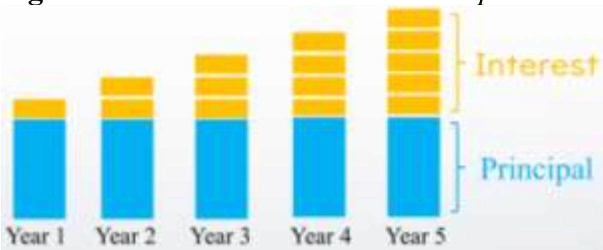
In this part of the class, we also explained the concepts of simple and compound interest, the Rule of 72 as a practical application of compound interest, the power of long-term accumulation, and basic methods for repaying borrowed money. These topics were presented to deepen participants' understanding of how money grows over time and the financial implications of interest in real-life scenarios.

Simple Interest and Compound Interest

In addition to simple and compound interest, financial literacy includes understanding inflation, time value of money, and money illusion. These concepts are essential for making informed financial decisions and should be considered in future curriculum development. In this section, we defined simple interest and compound interest, and provided practical examples to illustrate their applications.

Definition 1. Simple interest is defined as interest calculated solely on the initial principal, without incorporating the accrued interest into the principal.

Figure 13. *Investment Growth under Simple Interest*



Example 1. Suppose an initial investment of 1,000,000 yen is made at an annual simple interest rate of 3%. Since interest is calculated only on the principal, the amount increases linearly:

- After 1 year: $100 + 3 = 103$ (ten thousand yen)
- After 2 years: $100 + 3 + 3 = 106$ (ten thousand yen)
- After 3 years: $100 + 3 \times 3 = 109$ (ten thousand yen)

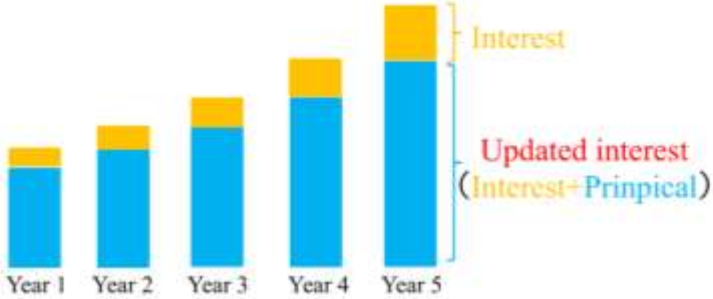
More generally, after n years, the total amount is given by:

$$A = a(1 + nr)$$

where a is the initial principal, r is the annual interest rate (as a decimal), and n is the number of years.

Definition 2. *Compound interest* is interest calculated on both the initial principal and the accumulated interest from previous periods.

Figure 14. *Investment Growth under Compound Interest*



Example 2. Suppose the same 1,000,000 yen is invested at an annual compound interest rate of 3%. The amount grows exponentially:

- After 1 year:
  $100 + 100 \times (3/100) = 100 \times (1 + 0.03) = 100 \times 1.03^1$ (ten thousand yen)
- After 2 years:
  $(100 \times 1.03^1) \times 1.03 = 100 \times 1.03^2$ (ten thousand yen)
- After 3 years:
  $(100 \times 1.03^2) \times 1.03 = 100 \times 1.03^3$ (ten thousand yen)

More generally, after n years, the total amount is given by:

$$A = a(1 + r)^n.$$

where a is the initial principal, r is the annual interest rate (as a decimal), and n is the number of years.

The differences between simple and compound interest, as well as their calculation methods, were explained. Participants then used calculators to solve an exercise designed to reinforce their understanding by comparing the outcomes of both interest calculations (Figures 15 and 16).

Exercise 3. Calculate the simple and compound interest for an investment of 1,000,000 yen at an annual interest rate of 3% using a calculator.

Figure 15. *Simple and Compound Interest Calculation Sheet*

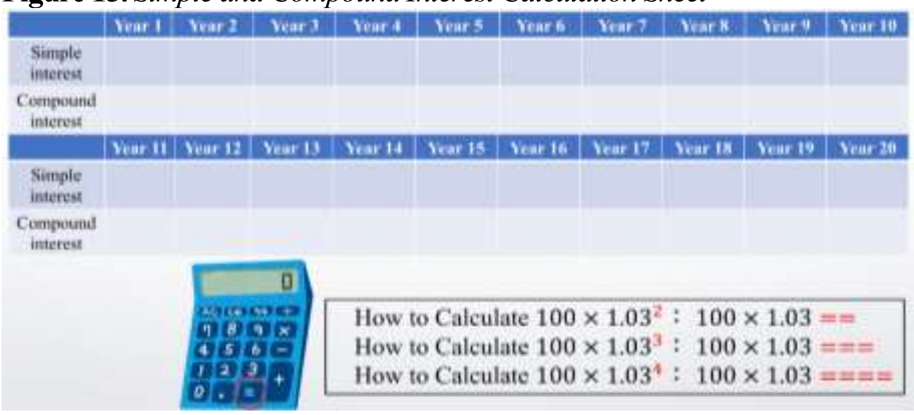
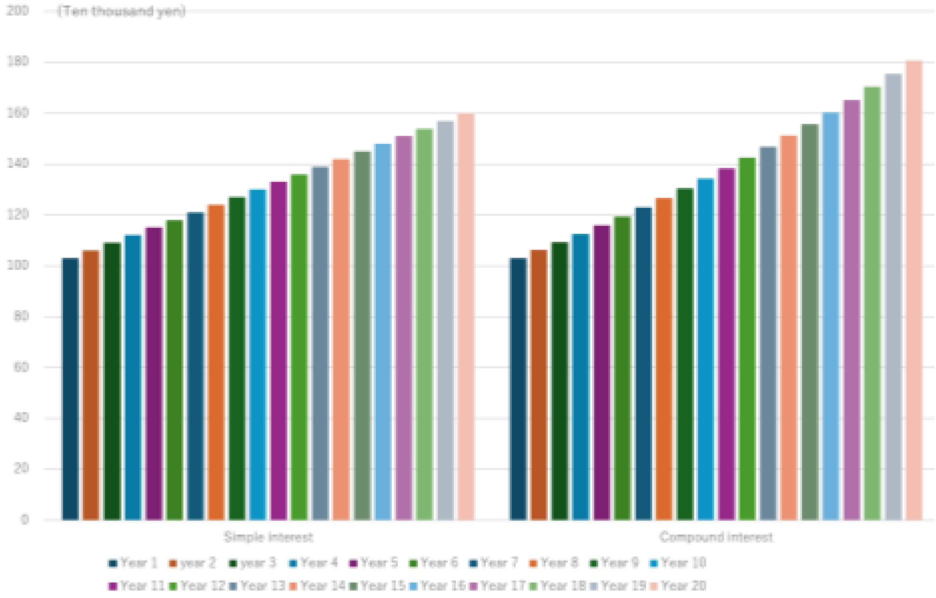


Figure 16. *Comparison between Simple and Compound Interest*



Rule of 72

Next, to illustrate the Rule of 72, the following question was presented to the participants:

Question 1: Under compound interest at an annual rate of 3%, how many years will it take for the principal to double?

As demonstrated in Example 2, we explained that the task was to find the smallest natural number *n* that satisfies the inequality $100 \times (1.03)^n \geqq 200$. After this explanation, participants were encouraged to use calculators to compute powers of 1.03 (such as $1.03^2, 1.03^3$, etc.) to determine the value of *n*.

After participants had estimated the value of *n*, we explained that, from a mathematical perspective, the exact value of *n* can be obtained using common (base-10) logarithms to solve the inequality: $100 \times (1.03)^n \geqq 200$. Dividing both sides by 100 gives:$(1.03)^n \geqq 2$. Taking the common logarithm of both sides results in:$\log_{10}(1.03)^n \geqq \log_{10} 2$.

Applying the logarithmic identity $\log_{10} a^p = p \log_{10} a$, we have:

$n \log_{10} 1.03 \geqq \log_{10} 2$. Solving for *n*, $n \geqq \frac{\log_{10} 2}{\log_{10} 1.03} \approx \frac{0.3010}{0.0128} \approx 23.5$.
Therefore, the smallest natural number satisfying the inequality is $n = 24$.

We then introduced the Rule of 72, an approximation method that estimates the doubling time under compound interest by dividing 72 by the interest rate: $n \approx \frac{72}{r}$. For an interest rate of 3%, this yields: $n \approx \frac{72}{3} = 24$. This illustrates that the Rule of 72 provides a close approximation to the exact logarithmic calculation (Table 2).

Table 2. *Calculation Examples using the Rule of 72*

(1) Interest rate (%)	(2) Doubling time under compound interest	Product of (1) and (2)
2	36	72
3	24	72
4	18	72
6	9	72

Long-term saving

To help participants understand the power of long-term saving through compound interest, the following problem was presented, followed by a detailed explanation:

Question 2: If you deposit 100,000 yen annually at an interest rate of 2% compounded annually, what will be the total amount at the end of the 5th year?

Answer to Question 2: The total amount accumulated by depositing 100,000 yen annually at a compound interest rate of 2% per year, with deposits made at the end of each year, is calculated step-by-step as follows. (Amounts are expressed in units of 10,000 yen.)

- Year 1:
  $10\times1.02 = 10.20$
- Year 2:
  $(10\times1.02+10)\times1.02=20.604$
- Year 3:
  $\{[(10\times1.02+10)\times1.02]+10\}\times1.02=31.216$
- Year 5 (after continuing the same process):
  $((((10\times1.02+10)\times1.02+10)\times1.02+10)\times1.02+10)\times1.02=53.0812$

Therefore, the total amount at the end of the 5th year is approximately 530,812 yen.

Remark 1. In general, the total amount accumulated after n years of annual deposits of a yen at a compound interest rate r is $\frac{a(1+r)\{(1+r)^n-1\}}{r}$.
In Question 2, the parameters are: the annual deposit is $a = 10$ (in ten thousand yen), the annual interest rate is $r = 2\%$, and the duration is $n = 5$ years.

To further support participants' understanding of how compound interest affects long-term savings—and to visualize how changes in interest rate and savings period influence the outcome—a savings simulator was used. Participants calculated the accumulated amounts under various conditions (Table 3 and Figure 17).

Exercise 4. Using the savings simulator, complete the following table.

Table 3. *Calculation examples using the Rule of 72*

Monthly deposit amount	Interest rate	10-year saving period	20-year saving period	30-year saving period
20,000 yen	0.001%			
	0.25%			
	3%			

Figure 17. *Savings Simulator. From FSA Japan, 2024 (https://www.fsa.go.jp/policy/nisa2/moneyplan_sim/index.html)*

Subsequently, a graph was presented to help participants visually understand the powerful effects of compound interest and long-term saving (Figure 18).

Figure 18. *Comparison of amounts by interest rate and saving period*

At the end of this section, methods of loan repayment were introduced. Specific loan amounts and interest rates were used to pose a realistic question and prompt participants to consider how different repayment methods affect total repayment amounts.

Question 3. I borrowed 1 million yen at an annual interest rate of 5%.

(1) If I repay 100,000 yen at the end of each year, how many years will it take to fully repay the loan?

(2) If I want to repay the loan in 10 years, how much should I repay at the end of each year?

To help participants solve this question, we first explained the concept of the present value of an annuity, using the following step-by-step analysis.

(1) Fixed Payment of 100,000 Yen Each Year
To calculate the present value of a payment of 100,000 yen to be made one year from now at an annual interest rate of 5%, we let x denote the present value. Then:
$1.05x = 100{,}000 \Longrightarrow x = \frac{100{,}000}{1.05} \fallingdotseq 95{,}238.$
Similarly, the present value of a payment of 100,000 yen two years from now is:
$(1.05)^2x = 100{,}000 \Longrightarrow x = \frac{100{,}000}{(1.05)^2} \fallingdotseq 90{,}702.$
For a payment made three years from now:
$(1.05)^3x = 100{,}000 \Longrightarrow x = \frac{100{,}000}{(1.05)^3} \fallingdotseq 86{,}383.$
In general, the present value of 100,000 yen to be paid n years from now is:
$x = \frac{100{,}000}{(1.05)^n}.$
(2) Fixed Repayment Over 10 Years
Suppose a fixed payment of x yen is made at the end of each year for 10 years. The total

present value of these payments must equal 1,000,000 yen.
The present value of these payments is given by the sum:

$$\frac{x}{1.05} + \frac{x}{(1.05)^2} + \frac{x}{(1.05)^3} + \cdots + \frac{x}{(1.05)^{10}} = 1{,}000{,}000$$

Factoring out x, we get:
$$x\left\{\frac{1}{1.05} + \frac{1}{(1.05)^2} + \frac{1}{(1.05)^3} + \cdots + \frac{1}{(1.05)^{10}}\right\} = 1{,}000{,}000$$
Let this sum of the geometric series be denoted by S, where
$S = \frac{1}{1.05} + \frac{1}{(1.05)^2} + \frac{1}{(1.05)^3} + \cdots + \frac{1}{(1.05)^{10}}.$
This is a finite geometric series, and its sum can be calculated using the standard formula: $S = \frac{\frac{1}{1.05}\left\{1-\frac{1}{(1.05)^{10}}\right\}}{1-\frac{1}{1.05}} = \frac{1-\frac{1}{(1.05)^{10}}}{1.05-1} = \frac{1-\frac{1}{(1.05)^{10}}}{0.05}$. Thus, the fixed annual payment x can be obtained by solving: $x = \frac{1{,}000{,}000}{S}$

This approach helps participants understand how interest rates, loan durations, and repayment schedules influence the total cost of borrowing.

Exercise 5. (1) Using the savings simulator, complete the table below and calculate the total present value of the loan repayments.

Table 4. *Calculation of Present Value*

	$\frac{100{,}000}{1.05}$	$\frac{100{,}000}{(1.05)^2}$	$\frac{100{,}000}{(1.05)^3}$	$\frac{100{,}000}{(1.05)^4}$	$\frac{100{,}000}{(1.05)^5}$	$\frac{100{,}000}{(1.05)^6}$	$\frac{100{,}000}{(1.05)^7}$	$\frac{100{,}000}{(1.05)^8}$	$\frac{100{,}000}{(1.05)^9}$	$\frac{100{,}000}{(1.05)^{10}}$
Present Value	95,238	90,703	86,384	82,270	78,352	74,621	71,067	67,683	64,460	61,390

	$\frac{100{,}000}{(1.05)^{11}}$	$\frac{100{,}000}{(1.05)^{12}}$	$\frac{100{,}000}{(1.05)^{13}}$	$\frac{100{,}000}{(1.05)^{14}}$	$\frac{100{,}000}{(1.05)^{15}}$
Present Value					

(2) Using the savings simulator, determine the values of S (the sum of the geometric series) and x (the fixed annual repayment amount), respectively.

Remark 2. The present value $\boldsymbol{PV}$ of an ordinary annuity, in which a fixed payment A is made at the end of each period over n periods with a constant interest rate r, is given by the following formula:

$$PV = A \times \frac{1-(1+r)^{-n}}{r}.$$

In Question 3(1), the present value PV is 1,000,000 yen, the annual payment A is 100,000 yen, and the interest rate r is 0.05. In Question 3(2), the present value $PV$is 1,000,000 yen, the annual payment A is 100,000 yen, and the number of periods n is 10.

Considerations Based on Questionnaire Results

To assess the extent to which the goals of this practice were achieved, we analyzed the percentage of correct responses on in-class exercises and the results of pre- and post-class questionnaires.

For the questionnaire analysis, each response item was scored on a 4-point Likert scale: "Agree" = 4 points, "Somewhat agree" = 3 points, "Somewhat disagree" = 2 points, and "Disagree" = 1 point. Table 1 presents the mean scores for each competency area, comparing responses before and after the class.

Table 5. *Average Score for Question Item Choices*

Competency	Question items	Before	After	Increase
Problem-Solving Skills	When you didn't understand something, did you try to look it up by yourself?	3.25	3.29	0.04
	When you faced a problem, did you try to solve it on your own?	3.19	3.24	0.05
Critical Thinking	When you encountered something you didn't understand or found strange, did you think, "Why is that?"	3.31	3.47	0.16
	When listening to your friends or teachers, did you ever think, "Why do they think that way?"	3.50	3.65	0.15
Creativity	Do you enjoy coming up with fun ideas or thinking of something new?	3.31	3.65	0.33
	When doing something, did you try to add your own creative twist or personal touch?	3.25	3.24	(0.01)
Flexibility	When plans or situations changed suddenly, were you able to stay calm and respond appropriately?	3.13	2.88	(0.24)
	When things didn't go as expected, did you try various approaches to solve the problem?	3.06	3.47	0.41
Perspective-Taking	When a problem occurred, did you try to consider the cause and solution from a broad point of view?	2.94	3.24	0.30
Execution Ability	Did you follow through with the plans you made?	3.19	3.24	0.05
	Even when things didn't go well, did you make efforts to complete the task to the end?	3.31	3.53	0.22

To evaluate the effectiveness of the financial literacy course, a paired t-test was conducted using the pre- and post-class questionnaire scores. The test compared the mean scores across six STEAM-related competencies: problem-solving, critical thinking, creativity, flexibility, perspective-taking, and execution ability. The analysis yielded a t-statistic of **−2.3679** and a p-value of **0.0394**. Because the p-value is below the conventional threshold of 0.05, the difference between the pre- and post- class scores before and after the course is statistically significant. These results suggests that the course had a meaningful impact on participants' development of STEAM competencies, particularly in areas such as critical thinking and perspective-taking. The results for each of the six competencies are summarized as follows:

1. Problem-Solving Skills

The results showed a slight improvement. Students already exhibited a moderately strong inclination toward solving problems independently, and the program modestly reinforced this tendency without significantly altering their overall behavior.

2. Critical Thinking

A clear improvement was observed. The program appears to have strengthened students' curiosity and their ability to critically reflect on their own and others' viewpoints. Activities involving discussion and inquiry likely contributed to this development.

3. Creativity

A notable increase was observed in students' enjoyment of generating new ideas, indicating heightened interest in creative thinking. However, there was little change in their ability to translate these ideas into practical improvements. This suggests that while students felt more creative, they did not necessarily take more creative action.

4. Flexibility

The results revealed contrasting tendencies. Students became more willing to try different approaches when encountering difficulties, indicating improved adaptive thinking. However, their confidence in staying calm during unexpected changes decreased slightly. This suggests that while their experimental mindset grew, their emotional resilience may require further support.

5. Perspective-Taking

This competency demonstrated one of the most substantial improvements, especially given its relatively low baseline. The program appears to have helped students broaden their outlook and integrate multiple viewpoints—an essential objective of interdisciplinary STEAM education.

6. Execution ability

A marked improvement was observed in students' perseverance and their capacity to follow through with tasks despite obstacles. The structure of the program, which emphasized project-based and goal-oriented activities, likely played a key role in fostering this competency.

Table 6. *Overall Summary*

Competency	Key Changes	Evaluation
Problem-Solving	Slight improvement	Stable
Critical Thinking	Consistent improvement	Positive
Creativity	Increased idea generation, but flat in execution	Partially effective
Flexibility	Increased problem-solving flexibility, decreased composure	Mixed results
Perspective-Taking	Major improvement	Highly effective
Execution Skills	Improvement in perseverance	Steady growth

The program effectively supported the development of several key STEAM-related competencies, particularly in critical thinking, creative ideation, perspective-taking, and resilience in execution. However, certain areas—such as emotional flexibility and the practical application of creative ideas—may benefit from more targeted instructional strategies in future implementations.

Overall, the results indicate a positive developmental impact, while also highlighting areas for further refinement and growth.

We also received the following comments from participants regarding the course theme and instructional materials:

- “In the economics part, I wanted more concrete examples. In the mathematics section, working with actual numbers helped me realize how significantly interest rates affect outcomes. Overall, I’m glad I attended the course.”
- “The female students' explanations were calm and easy to follow.”
- “The content was interesting, but I was disappointed that there were so few participants.”
- “I plan to try the simulation at home. The course was very informative. I wish more people had the chance to hear it.”

Limitations of this study include a small sample size and potential biases such as self-selection and social desirability. Additionally, the applicability of this model to other educational contexts may be constrained by differences in resources and curricula.

Conclusion and Future Directions

Future research should include longitudinal studies to assess the retention of financial concepts and compare outcomes with control groups. The STEAM approach could also be adapted to other domains such as physics and chemistry. We received the following comments from female teaching assistants regarding the course theme and instructional materials:

- "I was nervous, but it was a valuable experience."
- "I found the simulations useful and could easily understand the outcomes of long-term savings."
- "I should have practiced the simulation a bit more. I'm glad I learned about interest rates through this course. I hope to apply this knowledge when borrowing money in the future."

The results of the questionnaire indicate that the course was effective for learners and demonstrated promise as a STEAM education tool.

Looking ahead, we propose the development of two new teaching tools for future implementation:

1. Knot Theory
   Knot theory can be introduced to beginners without requiring prior background knowledge, making it highly accessible. It is supported by a range of visual teaching materials and is linked to various disciplines such as quantum field theory in physics, molecular design in chemistry, and DNA structure in biology. In future work, we aim to create STEAM teaching materials that connect mathematics with physics and chemistry.

2. Graph Theory
   Graph theory is another promising subject area. Like knot theory, it is accessible to beginners and does not require deep prior knowledge. It has broad interdisciplinary applications across fields such as physics, chemistry, computer science, linguistics, and the social sciences. We plan to explore its potential for integration into STEAM education through the development of engaging and visually intuitive materials.

Acknowledgments

The authors would like to express their sincere gratitude to the teaching assistants — Kokoro Kawaida, Funa Miyake, and Mizuki Yoshida — for their outstanding contributions and dedicated support throughout the project.

References

Collins WJ (2003) *The Greenwood Dictionary of Education*. Greenwood Publishing Group.

Kawashima K, Matsuda Y, Okita M, Sakai M & Tanaka T (2018) *A report on open courses in mathematics for junior high school students assisted by upper-class students at NIT, Kurume College.* Proceedings of ISATE 2018, 12th International Symposium on Advances in Technology Education, pp. 134–139.

Kawashima K, Sakai M & Matsuda Y (2021) *A practice of deep learning by geometry.* Proceedings of ISATE 2021, 14th International Symposium on Advances in Technology Education, pp. 124–128.

Kawashima K, Sakai M & Tanaka T (2014) An effort of mathematics teaching materials for secondary education using matrix and knot theory (in Japanese). *Gifu Educational Studies in Mathematics* 13: 1–11.

Kawashima K, Sakai M, Tanaka T & Matsuda Y (2015) *A practice of deep learning about mathematics using knot theory.* Proceedings of the 9th International Symposium on Advances in Technology Education, pp. 377–382.

Sakai M & Miyaji T (2013) *Various efforts to improve motivation to learn mathematics.* ATINER'S Conference Paper Series, MAT2013-0572, pp. 5–13.

Sakai M, Miyaji T & Nakabo S (2013) *A report of a mathematics open class for junior high school students at Kurume National College of Technology.* Proceedings of the 7th International Symposium on Advances in Technology Education, pp. 241–244.

Sakai M & Tanaka T (2023) *A practice of deep learning by knot theory and DNA.* Proceedings of ISATE 2022, 15th International Symposium on Advances in Technology Education, pp. 306–310.

Sakai M & Tanaka T (2014) *Development of teaching materials aimed for relevance on learning mathematics by the medium of complex numbers.* Proceedings of the 8th International Symposium on Advances in Technology Education, C1, 5 pages.

Sakai M, Aizawa M, Kawashima K, Okita M & Matsuda Y (2023) *A report on open courses in finance for citizens, taught by female students (RIKEJO) at NIT, Kurume College.* Proceedings of the 16th International Symposium on Advances in Technology Education, 5 pages.

Sakai M, Hagiwara Y & Aizawa M (2024) Introduction of a project to support budding girls in science and engineering at National Institute of Technology, Kurume College. *Journal of JSEE* 72(5), 5 pages, to appear.

Sakai M, Kawashima K & Matsuda Y (2016) *A report of a supplementary lesson in mathematics using teaching assistants at NIT, Kurume College*. Proceedings of the 10th International Symposium on Advances in Technology Education, pp. 164–168.

Sakai M, Nakamura S & Miki H (2024) Early STEAM education practice: Application of graph theory through teaching assistants. *Athens Journal of Technology & Engineering* 11(2): 145–166.

Sakai M & Tanaka T (2024) *Development and practice of mathematics teaching materials for early STEAM education of junior high school students using tiling*. Proceedings of the 19th International Symposium on Advances in Technology Education, 5 pages, to appear.

Sakai M & Tanaka T (2024) Development and practice of mathematics teaching materials for early STEAM education of junior high school students using molecular graphs. *Journal of JSEE* 72(3): 36–41.

Sakai M, Tokunaga M, Oiki H, Tanaka R, Tsuneki K & Okuyama T (2024) *Early STEAM education initiative targeting elementary school and junior high school students at National Institute of Technology, Kurume College – With NIT students as role models.* Proceedings of the Annual Conference of the Japanese Society for Engineering Education, pp. 368–369.

Van Rooij MCJ, Lusardi A & Alessie RJM (2011) Financial literacy and stock market participation. *Journal of Financial Economics* 101(2): 449–472.

Van Rooij MCJ, Lusardi A & Alessie RJM (2012) Financial literacy, retirement planning and household wealth. *The Economic Journal* 122(560): 449–478.

Yakman G (2008) *STEAM education: An overview of creating a model of integrative education.* In: Pupils' Attitudes Towards Technology (PATT-19) Conference: Research on Technology, Innovation, Design and Engineering Teaching, Salt Lake City, UT.

Athens Journal of Technology and Engineering - Volume 13, Issue 3,
September 2026 – Pages 195-210

Enhancing Multimodal Systems using Azure AI Image Embeddings and GPT-4.x Vision

By Mihail Mateev[*]

Image embeddings have become foundational in AI, enabling machines to transform visual data into structured numerical vectors for applications from predictive analytics to interactive user experiences. This paper presents a comprehensive study of Azure AI's image embedding technologies and their integration with GPT-4.x Vision to enhance multimodal retrieval-augmented generation (RAG) systems. We analyze the comparative strengths of Azure Machine Learning custom pipelines, the Azure AI Model Inference API, and Computer Vision v4.0 multimodal embeddings. The integration of CLIP embeddings and hybrid decomposition strategies is explored, with empirical results from a predictive maintenance case study. Theoretical frameworks, solution architectures, and experimental results are illustrated with six original schemas and diagrams, providing practical guidance for deploying scalable, accurate, and cost-efficient multimodal AI systems.

Keywords: *ChatGPT, AI, Generative AI, AI Foundry, Vector Search, Open AI, Azure AI Vision, Microsoft Azure, Image Embeddings, Predictive Analysis, IoT, Power Platform*

Introduction

The development of large language models (LLMs) and computer vision has changed how organizations utilize visual data. Image embeddings—dense vector representations of images—enable efficient search, classification, and cross-modal analytics. The ability to align image and text data in a unified vector space underpins advanced multimodal systems, enabling new applications in predictive maintenance, security, medical diagnostics, and e-commerce (Mateev 2025, Mateev 2024, Mateev 2024).

Azure AI, with its suite of embedding and vision services, offers a robust platform for deploying such systems at scale. This paper extends previous analyses by providing a detailed comparison of Azure's embedding services, integrating theoretical and practical perspectives, and presenting empirical results from real-world deployments.

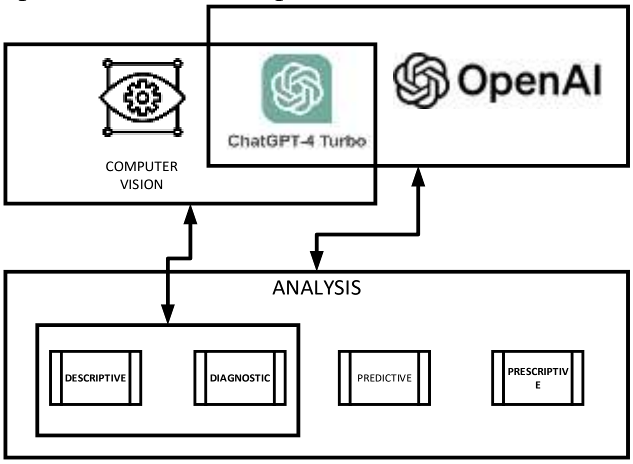
This study explores image analysis with Generative AI and GPT models via OpenAI API. OpenAI is a relatively new organization and laboratory for artificial intelligence research. It was established in 2015 by several tech leaders, such as Elon Musk and Sam Altman. Artificial Intelligence is the research field that OpenAI works in across different areas such as natural language processing (NLP), computer vision, reinforcement learning, and robotics.

[*]Chief Assistant Professor, UACEG, Faculty of Structural Engineering, CAE Department, Bulgaria.

The current research focuses on developing a reference architecture for leading image and video analysis platforms that utilize image embeddings. The focus is on various LLMs, which provide image analysis and image embeddings. The research uses preliminary OpenAI and Cohere models, GPT models, enhancing the analysis with computer vision services - especially GPT-4.x with Vision (GPT4-TV) to implement with a short time to market solutions, able to be implemented fast, to provide accurate results and to be available also for low code /no code technologies (Mateev 2025).

The paper extends the research from (Mateev, 2025), where options to use together GenAI and Computer Vision. The overall schema explaining relations between different types of analysis using Computer Vision and Open AI is demonstrated in Figure 1.

Figure 1. *Implementation of Different Types of analysis using Computer Vision, Open AI – General Dependencies*



The next chapter covers details on the research background and related work.

Methodology

Concept, Technologies, and Methodology Description

The Role of Image Embeddings

Image embeddings are high-dimensional vectors generated by deep learning models, capturing the semantic and contextual features of images. These vectors enable:

- **Similarity search**: Finding visually or semantically similar images.
- **Multimodal retrieval**: Linking images and text for cross-modal search.
- **Clustering/classification**: Grouping images by content or features.
- **RAG systems**: Enhancing LLMs with contextual image data for more accurate generation (Mateev 2025, Mateev 2024).

Multimodal Systems and RAG

Multimodal systems integrate data from multiple sources (e.g., images, text, audio) to provide richer, context-aware outputs. Retrieval-augmented generation (RAG) leverages embeddings to retrieve relevant content, which is then used by LLMs like GPT-4.x Vision for reasoning and generation (Mateev 2025, Mateev 2024).

Azure AI Image Embedding Solutions

In this section, options for Image Embeddings implementations are considered, available on the cloud platform used for this research: Microsoft Azure.

Azure Machine Learning Custom Embedding Pipelines

It is possible to create a custom AI service (based on AML or another AI platform) that implements custom AI algorithms. Some open-source projects and libraries can be utilized for developing custom embedding modules.

- **Customizability**: Allows domain-specific tuning.
- **Batch processing**: Efficient for large datasets.
- **Integration**: Seamless with Azure ML workflows (Mateev, 2025).

Azure AI Model Inference API

This option is the preferred one for research, offering simplicity and access to the latest LLMs as a service.

- **Pre-trained models**: Rapid deployment for general-purpose use.
- **Unified API**: Simplifies integration and model switching.
- **Scalability**: Supports serverless and managed compute endpoints (Mateev 2025).

Computer Vision v4.0 Multimodal Embeddings

- **Unified vector space**: Integrates images and text for cross-modal search.
- **Multilingual support:** Over 100 languages for text queries.
- **Optimized for RAG**: Enhances retrieval and generation tasks (Mateev 2025, Mateev 2024).

Integration with CLIP

CLIP (Contrastive Language-Image Pre-training) aligns images and text in a shared space, enabling zero-shot learning and flexible retrieval. Integration with Azure enables enhanced multimodal RAG systems (Mateev 2025, Mateev 2024).

The Theoretical and Technical Framework

The experimental research environment uses ChatGPT-4. x with Vision (ChatGPT-4-TV) for image analysis. This model has the following specifics:

- Open AI-based
- LLM-based.
- Good at solving not defined requirements in both conversational and completion modes
- Option to use Retrieval Augmented Generation (RAG)

In the case related to analyzing the video stream (input stream from drones and robots' cameras during construction observation), Azure AI Vision empowers ChatGPT-4-TV.

Azure AI Vision is a distinct solution for computer vision analysis, utilizing cognitive analysis to comprehend videos and images. It has the following characteristics:

- It's not LLM-based.
- Good at clearly defined tasks, object recognition, etc.
- Working with video streaming.

AI Vision extracts frames from the stream and sends them to the OpenAI service using a ChatGPT-4. x deployment.

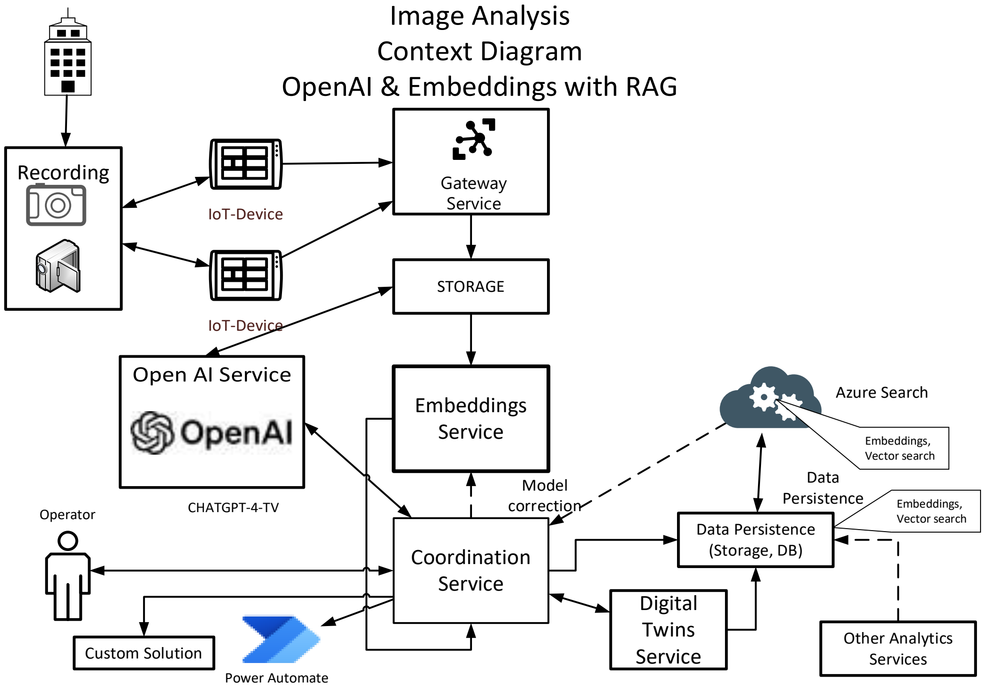
This study is focused on exploring the impact of various innovative approaches to improving efficiency and reducing costs in the realm of image analysis, particularly through the application of OpenAI's advanced multimodal language models, such as GPT-4-Turbo with Vision (ChatGPT-4-TV) and GPT-4. These models are deployed to perform a range of complex analytical tasks that go beyond traditional methods.

The experimental setup includes a sophisticated module for case decomposition, designed with the capabilities of Azure Digital Twins (ADT). This module is integral to the process, as it organizes and supervises the contextual framework of the solutions under analysis. It effectively breaks down intricate cases into smaller, manageable components, ensuring seamless integration of analyses. The decomposition process transforms domain-specific problems into general subcases, which are then analyzed using the robust tools provided by these large language models. By leveraging this systematic approach, the study aims to unlock new potential in image analysis, enhancing its applicability and precision across diverse scenarios.

One high-level schema of the solution is presented in the schema below.

Figure 2 illustrates the contextual schema of a solution utilizing OpenAI and AI Vision for digital content analysis.

Figure 2. *Predictive Analysis using Image Analysis with OpenAI and Embeddings*



The experimental framework employed for this study is meticulously designed using Microsoft Azure, Azure OpenAI Services, and Power Automate. This setup incorporates a diverse dataset, which includes 2,000 high-resolution images alongside 10 concise video clips, each with a duration of up to 20 seconds. The visual data primarily focuses on corroded steel construction elements sourced from actual structures within the United States. Figure 3 illustrates this dataset, providing rich visual references that form the foundation of the analysis process.

Figure 3. *Images from the Experimental Test Set: Steel Construction with Corrosion*

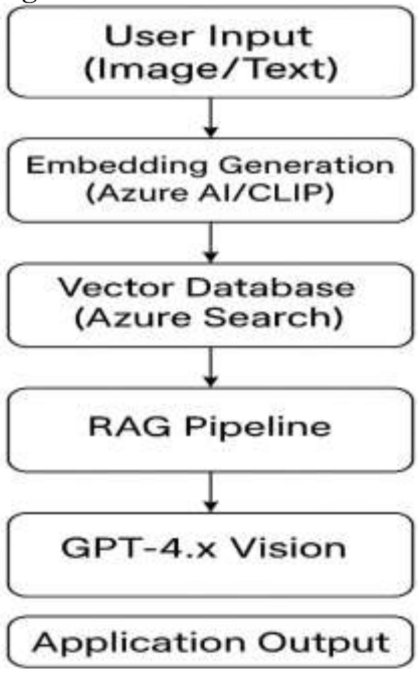
Hybrid Decomposition Architecture

The image analysis usually requires several steps, including:

- Collect the content (images)
- Generate embeddings
- Persist the embeddings in a vector database
- Search the distance with other embeddings in the database (RAG pipeline)
- Request a solution from AI (could be in several steps using LLMs and computer vision in different steps)
- Output results (for the user and persist them in a database)

Complex tasks are decomposed into smaller subtasks using the principles of Separation of Concerns (SoC) and Digital Twins (DT). This enables parallel processing and efficient resource utilization (Mateev, 2025; Mateev, 2024).

Figure 4. *High-Level Multimodal Solution Architecture (main flow)*



Processing chain details: The main flow begins with image/video ingestion, followed by on-the-fly frame extraction (when applicable), preprocessing (resize/ normalize), and embedding generation (Azure AI Vision or CLIP). Embeddings are persisted in a VectorDB (e.g., Azure AI Search/FAISS-compatible store) with metadata (timestamp, source, camera, asset id). Queries—text or image—undergo the same embedding function to ensure space alignment. Top-k candidates are retrieved via cosine similarity and passed as structured context into the LLM stage (GPT-4.x Vision) for reasoning, explanation, and decision support. The response and relevant matches are logged for traceability and future retraining.

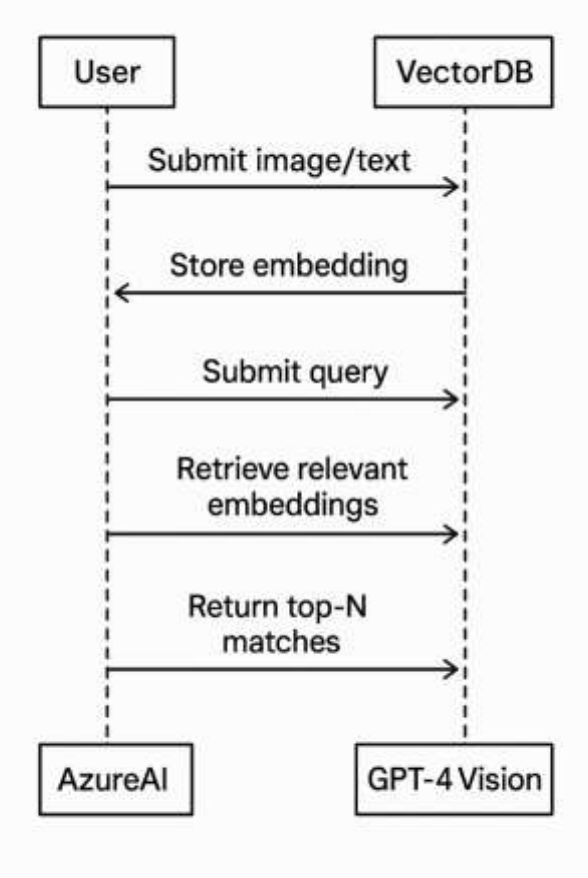
Embedding and Vector Search Workflow

RAG Pipeline Processing Details

The sequence diagram in Figure 5 below illustrates the interaction between four entities: User, AzureAI, VectorDB, and GPT-4 Vision. The user submits image or text data to AzureAI, which processes the input and stores its embedding in VectorDB. Later, the user sends a query to GPT-4 Vision, which retrieves the most relevant embeddings from VectorDB and returns the top-N matching results.

- Indexing: (i) Normalize images; (ii) produce embeddings; (iii) store vectors + metadata; (iv) build/update ANN indexes (HNSW/IVF).
- Retrieval: Convert user image/text to an embedding and perform vector search (cosine/L2) with filters (time/window, asset type, confidence).
- Augmentation: Package top-k results (thumbnails, captions, scores) as context, optionally merged with domain documents (manuals, SOPs) via hybrid (BM25+vector) search.
- Generation: GPT-4.x Vision composes an answer/explanation, citing retrieved items and highlighting uncertainty thresholds.
- Feedback & Logging: Capture user confirmations/corrections to update relevance signals and refresh the index.

Figure 5. *Embedding and Vector Search Pipeline*

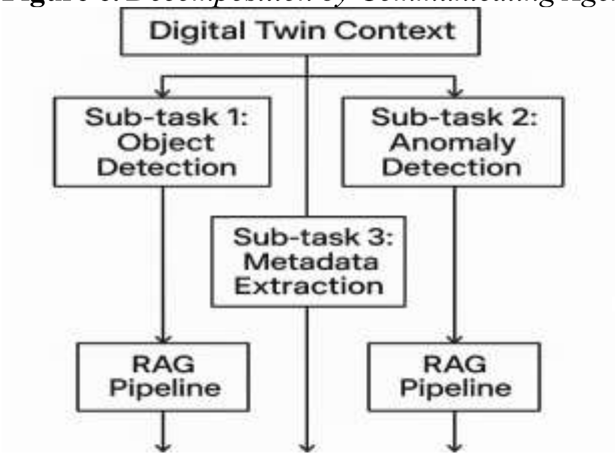


Often, complex flows need to be split into a system of simpler flows, and for such scenarios, the approach used is based on decomposition using the DT (Digital Twins) concept. The design, based on DT, simplifies the logic of the solution and allows for applying more formal analysis (decoupled from the context) for part of the flows.

Decomposition for Complex Image Analysis

The diagram illustrates a decomposition of tasks within a Digital Twin Context using communicating agents. The central context is broken down into three sub-tasks: Object Detection, Anomaly Detection, and Metadata Extraction. Each sub-task acts as an independent agent that processes specific information. These agents communicate their results to corresponding RAG Pipelines for further reasoning and retrieval-augmented generation. This modular design enables scalable, context-aware decision-making throughout the digital twin system.

Figure 6. *Decomposition by Communicating Agents*



Within the realm of image analysis and image embeddings, this workflow illustrates a modular and task-specific architecture designed for processing and comprehending intricate visual data in the context of a "Digital Twin system":

1. **Digital Twin Context**: Represents the virtual replica of a physical system (e.g., a smart building or industrial environment) where real-time data from sensors or cameras is continuously fed.
2. **Sub-Task Agents**:

   - **Object Detection**: Identifies and classifies entities (e.g., equipment, people, vehicles) within images. It generates localized features and object labels.
   - **Anomaly Detection**: Detects deviations from normal patterns using visual embeddings—useful in predictive maintenance or safety monitoring.

- **Metadata Extraction**: Captures contextual information from the image (e.g., timestamps, location, visual descriptors) and converts it into structured data or embeddings.

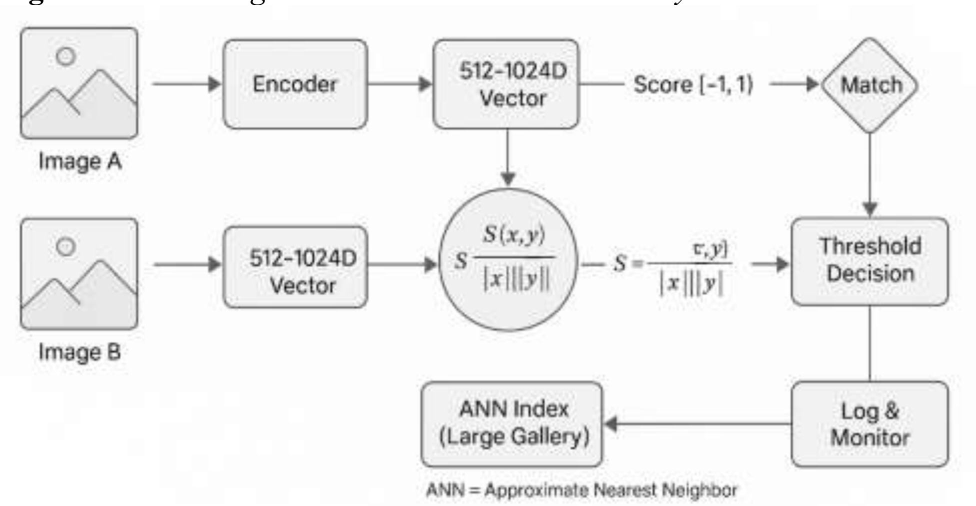
3. **RAG Pipelines** (Retrieval-Augmented Generation): Each sub-task sends its embeddings to a dedicated RAG pipeline, where relevant knowledge (e.g., documents, technical manuals, or prior cases) is retrieved from a vector database. This allows the system to generate rich, informed responses or alerts based on both the visual input and external knowledge.

The application of such a solution could be in various domains: for example, in a factory setting, this architecture could analyze surveillance footage in real-time, detect machinery, flag overheating components, extract operational context, and query technical documentation to provide instant insights or automated reports. The use of embeddings and retrieval makes it scalable and explainable.

Embedding Generation and Similarity Computation

The schema shows how two images are converted into embeddings, which are then compared using cosine similarity. This comparison calculates a similarity score indicating how visually or semantically alike the two images are. It's a key process in image matching, retrieval, and clustering tasks.

Figure 7. *Embedding Generation and Cosine Similarity*



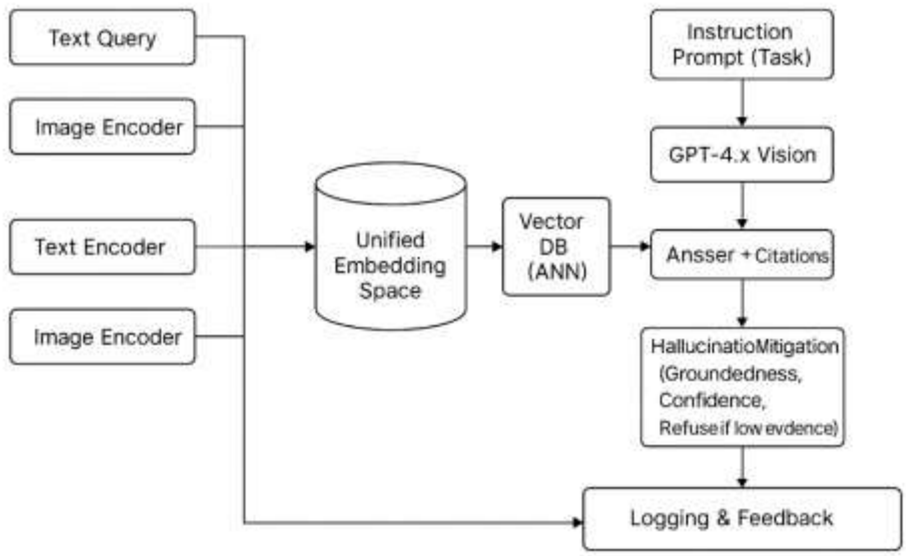
Processing chain details: Two images are independently encoded to 512–1024-D vectors. Cosine similarity $S(x,y)=\langle x,y\rangle/(\|x\|\|y\|)$ yields a score in $[-1,1]$, thresholded for match decisions. Batch normalization of vectors and approximate-nearest-neighbor (ANN) indexes improve both stability and latency for large galleries.

RAG with Multimodal Embeddings

Nowadays, most of the LLMs are multi-content. The same trend is promoted for models offering embeddings, such as those that embed different types of content or even models that generate content and provide embeddings.

The diagram illustrates a RAG (Retrieval-Augmented Generation) workflow using multimodal embeddings. A user submits a query—image or text—which is processed through Azure AI Search to perform a vector similarity search. Relevant embeddings are retrieved and sent to GPT-4.x Vision. The model uses them as context to generate a rich, informed output. This approach enhances answer accuracy and relevance by combining LLM generation with retrieved visual or textual knowledge.

Figure 8. *Retrieval-Augmented Generation with Multimodal Embeddings*



Processing chain details: Multimodal RAG first retrieves visually and textually relevant items using a unified vector space, then conditions GPT-4.x Vision with (a) the query, (b) retrieved image/text snippets, and (c) task instructions (e.g., defect classification). This reduces hallucinations and improves faithfulness compared with generation-only baselines.

Research

The experimental environment for this study is based on Microsoft Azure and Azure AI Foundry (a part of the platform that offers AI services)

Implementation of Image Analysis based on Computer Vision and OpenAI/ChatGPT

The research setup employs Azure AI, OpenAI GPT-4: x Vision, and Power Automate for workflow orchestration. The test dataset consists of 10,000 images and 10 short videos of corroded steel structures. Digital Twins are utilized for context management and decomposition, enabling domain-agnostic subcases for Large Language Model (LLM) analysis (Mateev 2025, Mateev 2024, Mateev 2024).

The experimental environment uses ChatGPT 4o with Vision (ChatGPT-4-TV) for image analysis. The solution includes a Digital Twin (DT) module based on Azure Digital Twins (ADT). ADT is used to extract the context from the observed solution, decompose the case, and unify the analysis case, converting it from domain-specific to domain-agnostic subcases suitable for LLM analysis.

One high-level scheme of the solution is presented in Figure 2 (Mateev, 2023).]

Findings/Results

In this paper, there are added metrics related to the experimental project PoC using the following KPI:

1. Embedding Performance
    a. Azure AI Vision and GPT-4o
    b. Hugging Face
    c. Replicate
    d. ChatGPT-4+CLIP
2. Vector Search Performance
3. Predictive Analysis Accuracy

Embedding Performance

The summarization of results is shown in the Table 1 below.

Table 1. *Embedding Performance*

Service	Embedding Quality	Dimensionality	Inference Speed	Cost Efficiency	Flexibility
Azure AI Vision (GPT-4o)	9/10	512	Moderate	Moderate	High
Hugging Face	8/10	768	Fastest	Best	Highest
Replicate	7/10	1024	Slowest	Low	Moderate
ChatGPT-4o (CLIP)	9/10	512	Moderate	Moderate	High

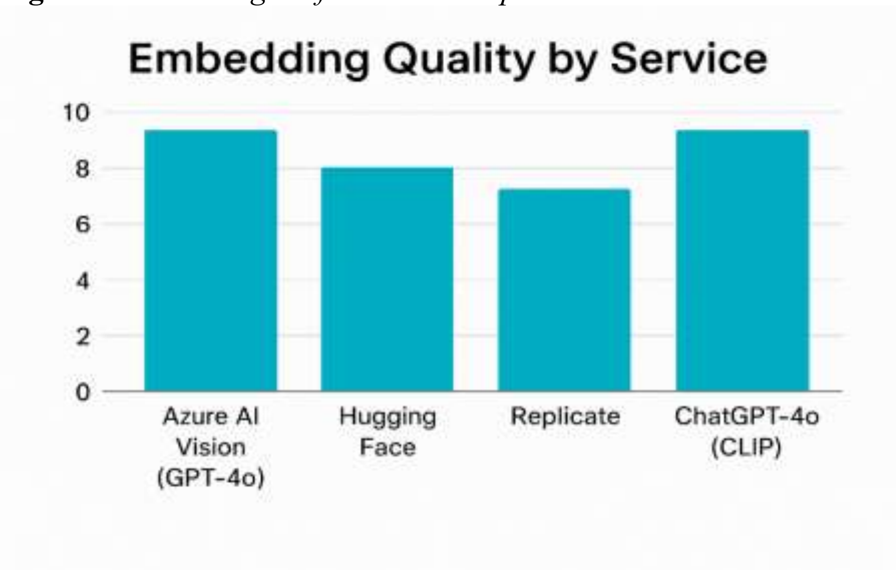
Embedding quality (rated 0–10) reflects how well a model converts inputs (images or text) into vector representations that capture semantic meaning and content similarity. This score is typically based on the following metrics:

Discussion: Hugging Face and Replicate refer to curated model hubs used in our PoC. Hugging Face experiments relied primarily on open-source CLIP variants with ViT backbones that offered strong zero-shot retrieval and fast inference on commodity GPUs. Replicate runs leveraged larger embedding models with 1024-D outputs, which increased storage and latency but provided richer feature representations for difficult edge cases. Vector dimensionality directly impacts index memory and query time; thus, 512-D models tended to be more cost-efficient at scale, while 768–1024-D models were beneficial when maximum recall was required. These trade-offs informed the design recommendations.

- Metrics and protocol: We report Recall@k ($k \in \{1,5,10\}$), MRR, and Precision for retrieval; p50/p95 latency and QPS for serving; and index footprint (GB) at various gallery sizes. Predictive maintenance accuracy is computed per-image and per-case, with confidence-threshold analysis; we additionally track false-positive/negative rates for anomaly detection sub-tasks. These measurement definitions are now included to support reproducibility.
- Zero-Shot Performance: Ability to generalize to unseen queries or tasks, often tested on benchmark datasets (like ImageNet or LAION).
- Cross-Modal Alignment: For multimodal models (e.g., GPT-4o, CLIP), how accurately do image and text embeddings align in a shared space?
- Clustering Consistency: How tightly similar items cluster and how distinctly different items are separated in embedding space (visualized via t-SNE or UMAP).
- Cosine Similarity Tests: Controlled comparisons between image pairs to measure the precision of similarity scores.

The 0–10 scale is often derived from normalized benchmark results or expert evaluations across these categories. A score of 9–10 implies state-of-the-art performance in aligning embeddings with real-world meaning.
The embedding quality comparison is described in Figure 9.

Figure 9. *Embedding Performance Comparison*



Vector Search Performance

The Vector Search performance, split by components, is described below:

- Relevance: Azure AI Vision and ChatGPT-4o (CLIP) yield the most relevant results.
- Latency: Hugging Face is fastest; Replicate is slowest.
- Storage: Lower dimensionality (512) is more efficient for storage and computation.
- Cost: Hugging Face is the most cost-efficient; Replicate is the least (Mateev 2025, Mateev 2024).

Predictive Analysis Accuracy

The relevance of the services compared by accuracy is described below:

- Azure AI Vision (GPT-4o) and CLIP (ChatGPT-4o) achieve the highest accuracy for predictive maintenance (up to 99.4% in defect detection).
- Hybrid approaches (using SoC and DTs) further optimize performance and cost, especially for complex or large-scale deployments (Mateev 2025, Mateev 2024).

Conclusions

Azure AI's embedding solutions, combined with GPT-4.x Vision and CLIP enable scalable, accurate, and cost-efficient multimodal systems. The choice of embedding service should be tailored to the application's quality, speed, and budget requirements. Future research will focus on automated decomposition, expanding hybrid frameworks, and maximizing API aggregation platforms for embedding generation (Mateev 2025, Mateev 2024).

Key Insights

The central key insight from this research is based on a comparison between the analyzed methods for image embedding generation:

- **Best Quality**: Azure AI Vision and ChatGPT-4o (CLIP) for high-stakes applications.
- **Fastest**: Hugging Face for real-time, large-scale deployments.
- **Most Cost-Efficient**: Hugging Face for startups and budget-conscious projects.
- **Richest Embeddings**: Replicate for research and feature-rich scenarios.

Design Recommendations

A critical takeaway from this study is about the design recommendations when implementing image embedding.

- Use custom pipelines for domain-specific tasks.
- Leverage pre-trained APIs for rapid deployment.
- Employ multimodal embeddings for unified search and retrieval.
- Integrate hybrid decomposition for complex workflows.
- Aggregate APIs (e.g., Eden AI) for cost and flexibility optimization (Mateev, 2025; Mateev, 2024).

References

Biswas D (2023) *Contextualizing large language models (LLMs) with enterprise data.* LinkedIn Pulse. Available at: https://www.linkedin.com/pulse/contextualizing-large-language-models-llms-enterprise-debmalya-biswas/ [Accessed 16 July 2023].

D'Amico RD, Erkoyuncu JA, Addepalli S & Penver S (2022) *Cognitive digital twin: An approach to improve the maintenance management.* CIRP Journal of Manufacturing Science and Technology 38: 123–134.

El Mokhtari K, Panushev I & McArthur JJ (2022) *Development of a cognitive digital twin for building management and operations.* Frontiers in Built Environment 8. doi:10.3389/fbuil.2022.856873.

Fernandez T (2023) *How to choose the best OpenAI model for your AI application.* Semaphore CI. Available at: https://semaphoreci.com/blog/openai-models [Accessed 10 August 2023].

Lin KL & Suen JB (2019) *A vision-based method for determining degradation level of a road marking.* Presented at the ATINER Conference Presentation Series No. CIV2019-0138, Athens, Greece.

Mateev M (2023) *Design and implementation of cognitive digital twins with generative AI and ChatGPT.* Presented at the 4th Annual International Conference on Computer and Software Engineering (ATINER SFW2023-0344), Athens, Greece, 17–20 July 2023.

Mateev M (2023) *Predictive analytics based on digital twins, generative AI, and ChatGPT.* Presented at the 27th World Multi-Conference on Systemics, Cybernetics and Informatics (WMSCI 2023), pp. 168–174. doi:10.54808/WMSCI2023.01.168.

Mateev M (2023) *Using digital twins, IoT, and anomaly detection for predictive analysis in construction industry.* Presented at the 19th Annual International Conference on Information Technology & Computer Science, Athens, Greece, 15–18 May 2023, Abstract Book, p. 68.

Mateev M (2024) *Evolution of predictive analysis using GPT OpenAI models.* Presented at Industry 4.0 2024 Winter Session Proceedings, Scientific-Technical Union of Mechanical Engineering "Industry 4.0", Borovets, Bulgaria, 11–14 December 2024, vol. 2, pp. 78–81.

Mateev M (2024) *Implementing hybrid (AI and data analytics) solutions for optimal performance and cost optimization for image analysis with GPT-4 Turbo with Vision for predictive analysis.* Presented at World Multi-Conference on Systemics, Cybernetics and Informatics (WMSCI 2024), pp. 74–80. doi:10.54808/wmsci202.[01.74.

Mateev MA (2025) *Comparative analysis on implementing embeddings for image analysis.* Journal of Information Systems Engineering and Management 10(17s): 89–102. doi: 10.52783/jisem.v10i17s.2710

Ramaji I (2024) *Advancing building management: Digital twins for sustainable HVAC efficiency.* Presented at the 12th Annual International Conference on Industrial, Systems & Design Engineering, Athens, Greece, 24–27 June 2024, Abstract Book

Ünal P (2023) *Cognitive digital twins: Digital twins that learn by themselves, foresee the future, and act accordingly.* Digital Twin Consortium. Available at: https://www.digitaltwinconsortium.org/2022/09/cognitive-digital-twins-digital-twins-that-learn-by-themselves-foresee-the-future-and-act-accordingly/ [Accessed 9 June 2023]

Athens Journal of Technology and Engineering - Volume 13, Issue 3, September 2026 – Pages 211-228

GPT-only vs. GPT with RAG: A Study on Accuracy in Handling University-Specific Queries

*By Meltem Cakar**

While Large Language Models (LLMs) have demonstrated impressive capabilities in general natural language processing, their accuracy often diminishes in domain-specific contexts where precise, factual responses are crucial. This study addresses this limitation within the higher education sector by comparing two approaches to handling university-specific queries. We evaluate a Generative Pre-trained Transformers (GPT)-only model that relies on prompt engineering against a Retrieval-Augmented Generation (RAG) model that incorporates external university documents, specifically program flyers and a module handbook, integrated using Langchain. We benchmark both systems using 90 academic queries categorized by the question difficulty and assess their performance through automatic metrics and blind expert ratings. Our results demonstrate that RAG significantly outperforms the GPT-only approach, particularly for complex questions concerning curriculum and program structure. This research offers valuable insights for higher education institutions seeking to implement reliable and effective AI-powered solutions for student support and information provision.

Introduction

The field of natural language processing has seen significant advancements with the development of large neural network models trained on vast datasets, commonly referred to as Large Language Models (LLMs) (Wu et al. 2023). These models are designed to generate human-like text and have demonstrated a capacity for various language understanding and generation tasks, including question answering and summarization (Naveed et al. 2025). GPT models, such as those developed by OpenAI, represent a prominent family of LLMs known for their ability to generate coherent and contextually relevant text based on extensive pre-training (Kalyan 2025).

Despite their capabilities, LLMs have inherent limitations. A notable concern is 'hallucination', where the models generate content that is factually incorrect, irrelevant or inconsistent with the input data (Bang et al. 2023, Ji et al. 2023). This issue poses risks in environments like academia and education, where factual accuracy and reliability are important. Even if the information is plausible, it can still be misleading and lead to incorrect conclusions or decisions, particularly when influencing academic advising or information retrieval.

In response to the issue of LLM hallucinations (Waldo & Boussar 2024) and to improve factual accuracy, RAG has emerged as a promising framework (Liang et al. 2025). RAG combines the generative capabilities of LLMs with a retrieval mechanism that retrieves relevant information from verifiable external data

*Academic Assistant, DHBW Heidenheim, Duale Hochschule Baden-Württemberg Heidenheim - Cooperative State University, Germany.

sources prior to generating a response. This enables models to base their outputs on actual, human-authored data, thereby reducing the occurrence of inaccuracies in the generated content (Gao et al. 2024). In educational contexts, structured content like university program flyers and module handbooks represent a rich source of context-specific information that LLMs often ignore. Integrating such retrieval mechanisms enables LLMs to adapt more effectively to specific organizational contexts while mitigating the risk of factual errors.

The need for accurate and context-specific information is particularly pronounced in higher education institutions, where students and staff frequently seek detailed answers on curricula, program structures, and administrative procedures. Current LLM-based chatbots, when deployed without a robust retrieval mechanism, may struggle to provide the necessary precision academic advising and support systems. This highlights a critical gap in the reliable application of LLMs for domain-specific information retrieval within universities.

This paper investigates the application of RAG in the context of academic information retrieval within a university setting. Specifically, we benchmark a RAG-based system against a GPT-only model, evaluating their respective performances in answering program-related questions derived from authentic university documents. Our methodology involves a structured approach outlined in three key steps.

1. System preparation and data ingestion: Initially, a Retrieval-Augmented Generation (RAG) system was implemented using the Langchain[1] framework. This involved ingesting documents specific to DHBW Heidenheim, specifically flyers for the Business Informatics and Computer Science programs, as well as the official Business Informatics module handbook. These documents form the knowledge base for the RAG system.
2. Structured benchmarking with automatic metrics: A structured benchmark was then conducted, comparing the RAG system (developed with Langchain) with the GPT-only model. For this purpose, a curated dataset of 90 university-specific questions was generated from the mentioned DHBW documents. These questions were categorized into three difficulty levels: easy, medium and difficult (detailed breakdown in Document Sources and Question Development). Reference answers (gold standards) were established and validated for accuracy by two independent domain experts for all questions. The responses generated by both the RAG and GPT-only systems were then evaluated quantitatively against these gold standards using the following automatic metrics: F1-score, BLEU and METEOR. These metrics were selected due to their widespread acceptance and complementary strengths in evaluating the quality of text generation.
3. Complementary Human Expert Evaluation: In addition to the automatic metrics, a qualitative assessment was performed by three human experts. These experts independently evaluated the answers from both systems against the described gold standards, unaware of which system generated which response (a detailed description is provided in Human Evaluation and

[1]LangChain Developers: Building Chatbots. LangChain v0.x Documentation. Accessed on October 28, 2025 from https://python.langchain.com/v0.2/docs/tutorials/chatbot/

Qualitative insights). This step provided crucial qualitative insights into aspects such as factual accuracy and completeness.

This comprehensive Langchain-RAG framework and evaluation process established a structured benchmark, combining automatic metrics with invaluable human expert assessment. This multidimensional approach enabled a thorough analysis of answer quality from both systems, considering various perspectives.

Related Work

The field of natural language processing (NLP) has achieved significant progress, largely due to the development of large language models (LLMs). Trained on extensive text corpora, these models have demonstrated their ability to perform various NLP tasks, such as text generation, summarization and answering questions (Brown et al. 2020, Wu et al. 2023). Early iterations, such as GPT-3, demonstrated the advantages of few-shot learning, allowing models to perform new tasks with few examples or even through prompt engineering alone (Brown et al. 2020, Beltagy 2022). Despite their versatility, LLMs have limitations. The most notable of these is the phenomenon of hallucination, whereby models generate content that is factually incorrect, irrelevant or fabricated (Ji et al. 2023, Bang et al. 2023). This limitation poses significant challenges in fields requiring high levels of factual accuracy and reliability, such as education, healthcare and legal advice. In academic settings, for example, the generation of misleading or unverified information can undermine trust and hinder effective knowledge transfer.

To address these challenges, the RAG approach has emerged as a robust solution. RAG systems combine the generative capabilities of LLMs with a dynamic information retrieval component, enabling models to base their responses on external, verifiable data sources (Lewis et al. 2020, Gao et al. 2024). This approach reduces hallucinations and enhances factual consistency by anchoring outputs in real, human-authored content (Barnett et al. 2024). Research specifically on RAG in education highlights its potential to address these accuracy concerns. For instance, Calfoforo et al. (2024) investigated the integration of RAG and the Langchain framework to develop a question-answering system using the Llama-2 model. Their study aimed to improve information retrieval accuracy and relevance for policy-related questions based on a faculty handbook and FAQs in PDF format, demonstrating enhanced information accessibility and support efficiency within academic institutions. Similarly, Khan et al. (2025) developed an educational virtual assistant that leverages RAG with Llama-2 and Mistral models to provide university- related information. They evaluated the assistant's responses using BLEU scores and emphasized its potential for automating administrative support. Other studies have examined the use of RAG in generating precise answers from course materials. Furthermore, studies like Soygazi and Oguz (2023) have analyzed the performance of LLMs and Langchain-based models in mathematics education, highlighting the challenges LLMs face in deterministic fields and the role of frameworks like Langchain in integrating specialized knowledge or tools to improve accuracy.

Building upon existing literature, including recent work by Calfoforo et al. (2024) and Khan et al. (2025) on RAG in educational contexts, this study addresses a critical research gap by providing structured benchmark. Previous studies have often focused on single RAG system implementations or general LLM comparisons. Our study compares a GPT-only model with a RAG system to identify their respective strengths and weaknesses when processing university-specific queries. Furthermore, we use a variety of university documents to assess performance and information complexity. We also use different quantitative metrics alongside a detailed, blind human expert evaluation.

Materials and Methods

Model Foundation

This study evaluates two LLM configurations: a GPT-only baseline model and RAG system. Both configurations are based on OpenAI's GPT-3.5 Turbo model. GPT-3.5 Turbo was chosen due to its optimal combination of linguistic capabilities in question-answering tasks. This model provides a robust foundation for both experimental setups, offering a cost-effective yet high-performing solution that is well-suited to research projects.

GPT-only Model Configuration

The GPT-only baseline model uses the GPT-3.5-turbo model provided by OpenAI directly. In this configuration, responses are generated based solely on the model's pre-trained internal knowledge. Prompt engineering was applied to guide the model's output, using a system prompt to direct its behavior and response style. The system prompt used was:

Figure 1. *System Prompt GPT-only*

> *You are an assistant for question-answering tasks. Use the following pieces of retrieved context to answer the question. If you don't know the answer, say that you don't know. Use three sentences maximum and keep the answer concise.*

RAG System Configuration

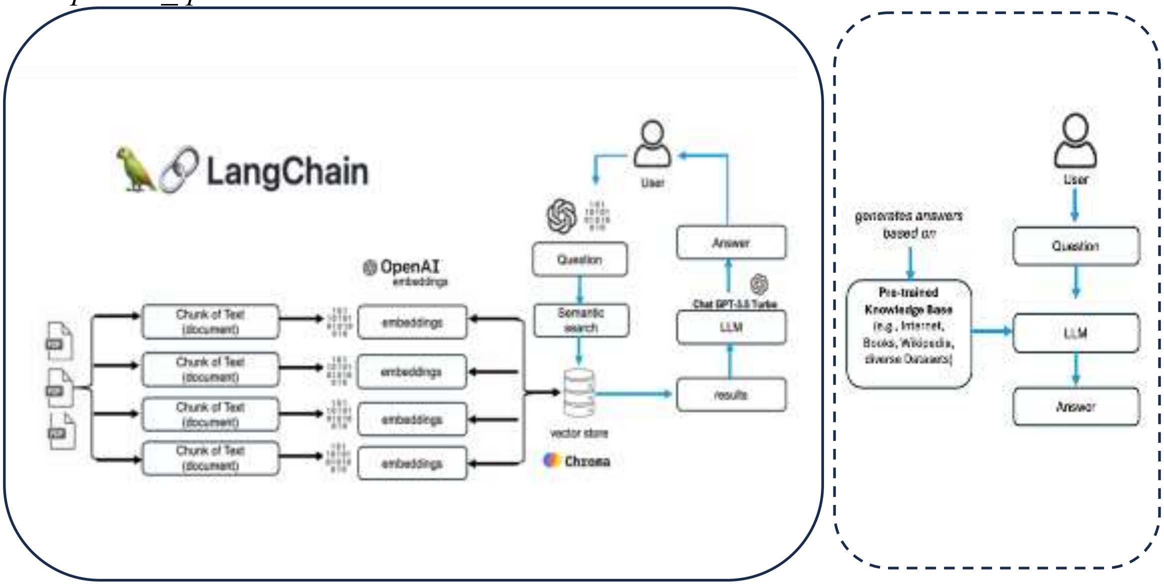
The RAG system also employs OpenAI's GPT-3.5-turbo but augments its capabilities with an information retrieval component. This system was implemented using the Langchain framework. The core components of the RAG (Langchain[2]) are detailed below.

[2]LangChain Developers: Building Chatbots. LangChain v0.x Documentation. Accessed on October 28, 2025 from https://python.langchain.com/v0.2/docs/tutorials/chatbot/

- Document ingestion and pre-processing (OCR): University-specific documents, including program flyers for Business Informatics and Computer Science and the official Business Informatics module handbook, were used as the knowledge base. As these flyers contained images and text, an Optical Character Recognition (OCR) process was applied to them first using the *Tesseract OCR* engine to accurately extract text from image based content within the PDFs. This ensured that all textual information, regardless of its original format, was available for processing. The extracted text was then prepared for the next step.
- Chunking: The pre-processed text from all documents was divided into smaller, semantically meaningful units, or 'chunks', to optimize retrieval accuracy. For this purpose, the *RecursiveCharacterTextSplitter* from Langchain was implemented with the following parameters: chunk_size=500 characters and chunk_overlap= 200 characters. This overlap helps maintain contextual continuity across consecutive chunks and avoids critical information being split.
- Embedding Model: To convert the text chunks into numerical vector representations (embeddings), an embedding model is essential. These embeddings allow for efficient semantic search. We utilized OpenAI's *text-embedding-ada-00* model for generating these embeddings.
- Vector Storage: The generated embeddings of the document chunks were then stored in a vector database. For vector storage and efficient similarity search, ChromaDB was chosen due to its ease of integration within the Langchain framework This vector store facilitates rapid retrieval of relevant document segments based on the semantic similarity to a user's query.
- Retrieval Method: When a user query is submitted, its embedding is generated and used to perform a similarity search within the vector store. The retrieval method identifies the most relevant chunks based on vector similarity. These retrieved chunks are then passed as context to the LLM.
- System Prompt for RAG: The same system prompt as in the GPT-only configuration was used to ensure consistent conversation style and response constraints (Figure 1).

Figure 2 provides a visual overview of both the GPT-only and the RAG system architectures, illustrating the distinct workflow of each approach. The diagram highlights how the RAG system augments the LLM's capabilities by integrating external document retrieval and a dedicated vector store, contrasting with the GPT-only model's reliance only on its internal pre-trained knowledge.

Figure 2. *Comparative Architecture of GPT-only and RAG System based on Langchain, own illustration based on Langchain: https://python.langchain.com/docs/concepts/text_splitters/*



Document Sources and Question Development

To simulate realistic student query scenarios and ensure the practical relevance of our evaluation, we compiled a comprehensive dataset of questions based on authentic university documents from DHBW Heidenheim. The primary sources of these documents were two program flyers (Business Informatics and Computer Science) and the official Business Informatics module handbook. These documents are typical sources of information for students seeking details on study options.

A total of 90 questions were manually developed from this content. To ensure a structured assessment of the models' capabilities across different cognitive tasks, these questions were systematically classified into three difficulty levels: Easy, Medium and Difficult.

- Easy level (remembering): Questions at this level target factual information that is stated directly within the source documents. They align with the 'Remembering' cognitive level of Bloom's Taxonomy (Bloom et al. 1956). Examples include questions about the length of studies, the number of credits, and the basic structure of dual study courses.
- Medium level (understanding): Questions categorized as 'medium' require a broader understanding and interpretation of the content of the document. These correspond to the 'Understanding' cognitive level of Bloom's Taxonomy (Bloom et al. 1956). Examples of such questions include identifying overarching program goals or highlighting the differences between similar study program based on their key characteristics.
- Difficult level (challenging questions) These questions were designed to go beyond direct recall or simple interpretation. Drawing upon information

within the documents, they required comparison or a deeper understanding of interrelationships. Examples include inquiring about learning outcomes across modules, the role of specific modules in competence development and drawing detailed comparisons between different study options. They were designed to evaluate the models' capacity to address more complex, academic questions that transcend the explicit levels of 'Remembering' and 'Understanding' as defined by Bloom's Taxonomy.

This multi-level classification system allowed us to assess how accurately each system could retrieve explicit answers and reflecting the diverse informational needs of a university setting.

Results

Quantitative Results by Document Type and Question Difficulty

The quantitative evaluation was conducted using a dataset comprising 90 questions, which were structured to simulate the diverse queries that students might bring to a university environment. The dataset was derived from three distinct document types representative of DHBW Heidenheim.

- Business Informatics Flyer (n=30 questions),
- Computer Science/Software Engineering Flyer (n=30 questions),
- Business Informatics Module Handbook (n=30 questions).

Each set of 30 questions for each document type was divided equally into three difficulty levels, based on the cognitive demands posed to the models.

- Easy (10 questions per document type),
- Medium (n = 10 questions per document type),
- Difficult (n = 10 questions per document type).

A precise reference answer (gold standard) was established for every question, ensuring a reliable basis for automated evaluation. The responses generated by the GPT-only and RAG-enhanced models were then assessed quantitatively using the F1-score, BLEU and METEOR metrics. Table 1 shows the performance metrics achieved by both models for different document types and question difficulty levels.

Table 1. *Model Performance in structured Benchmarks by Document Type and Question Difficulty*

GPT- only vs. RAG mean scores

Document Type	Model	Question Difficulty	F1-Score	Bleu-Score	Meteor-Score
Business Informatics Flyer	GPT-only	easy	0.72	0.56	0.64
	RAG	easy	0.68	0.42	0.61
	GPT-only	medium	0.74	0.56	0.65
	RAG	medium	0.63	0.39	0.55
	GPT-only	difficult	0.66	0.39	0.53
	RAG	difficult	0.60	0.34	0.49
Computer Science Flyer	GPT-only	easy	0.70	0.62	0.63
	RAG	easy	0.96	0.94	0.95
	GPT-only	medium	0.66	0.52	0.58
	RAG	medium	0.94	0.88	0.90
	GPT-only	difficult	0.59	0.52	0.55
	RAG	difficult	0.90	0.84	0.86
Module Handbook	GPT-only	easy	0.62	0.60	0.64
	RAG	easy	0.97	0.89	0.92
	GPT-only	medium	0.58	0.41	0.48
	RAG	medium	0.93	0.87	0.94
	GPT-only	difficult	0.52	0.43	0.49
	RAG	difficult	0.88	0.81	0.88

Interpretation of Automatic Metric Scores

In order to interpret the results presented in Table 1, it is important to understand what high or low scores for each metric indicate in the context of question answering.

F1-score: Ranges from 0 to 1, with 1 indicating a perfect match between the generated answer and the reference answer (Chauhan & Daniel 2023). A higher score suggests that the model's response is both complete and concise.

BLEU: Also ranges from 0 to 1, indicating the n-gram overlap with the reference answer. A higher BLEU score indicates greater similarity in phrasing and word order to the ground truth, and better fluency. BLEU is particularly good at capturing exact matches of phrases (Chauhan & Daniel 2023).

METEOR (Metric for Evaluation of Translation with Explicit Ordering): It typically ranges from 0 to 1 and measures semantic similarity by considering synonyms, as well as exact word matches. A higher METEOR score indicates that the generated answer expresses the same meaning as the reference, even if different words are used (Lavie & Denkowski 2009).

Analysis of Results by Document Type and Difficulty

The quantitative results reveal varying performance differences between the GPT-only and RAG systems, depending on the document type and question difficulty.

Performance on the Business Informatics Flyer

For questions based on the Business Informatics Flyer, the GPT-only model outperformed the RAG system consistently across all difficulty levels. As shown in Table 1, the GPT-only model achieved higher F1, BLEU and METEOR scores for easy (F1: 0.72 vs. 0.68), medium (F1: 0.74 vs. 0.63) and difficult (F1: 0.66 vs. 0.60) questions. This unexpected outcome suggests that, for this specific document which is likely designed to contain widely known information (e.g. standard dual study advantages and general program descriptions), GPT-only's extensive pre-training enabled it to provide semantically similar answers without the need for external retrieval. The information required for these questions may have been sufficiently covered by GPT-only's pre-existing knowledge base, rendering the retrieval step less beneficial and potentially introducing minor noise, given that some of the information in the flyer is common knowledge.

Performance on the Computer Science/Software Engineering Flyer

By contrast, when questions from the Computer Science/Software Engineering Flyer were evaluated, the RAG system demonstrated a clear and significant advantage. RAG showed competitive performance for easy questions, but its strength became more apparent as the questions became more difficult. RAG achieved an F1 score of 0.94 for medium questions, compared to 0.66 for GPT-only. For difficult questions, RAG scored 0.90 (F1), compared to 0.59 for GPT-only. Similar trends were observed for BLEU and METEOR scores. This suggests that the Computer Science flyer, despite also being a 'flyer', probably contained more specialized or less commonly covered information that GPT-only found difficult to recall precisely from its general training data. This meant that it significantly profited from RAG's targeted retrieval.

Performance on the Module Handbook

The most significant performance difference of the RAG system was observed with questions from the Business Informatics Module Handbook. Across all difficulty levels, RAG consistently outperformed GPT-only. RAG achieved an F1 score of 0.97 for easy questions (compared to GPT-only's 0.62), 0.93 for medium questions (compared to GPT-only's 0.58), and 0.88 for difficult questions (compared to GPT-only's 0.52). These results strongly confirm the hypothesis that retrieval augmentation is critical for highly specific, detailed and voluminous domain knowledge. The module handbook contains intricate and unique curriculum details, learning outcomes and module descriptions that are unlikely to be fully or accurately represented in the general training data of GPT-only, thereby highlighting the ability of RAG to ground responses in precise information.

Examples of response Generation

To illustrate the quantitative findings and provide qualitative context for the performance differences, specific examples of model responses at different difficulty levels are presented below. These examples demonstrate how the RAG mechanism enables more accurate and contextually relevant answers, especially when dealing with domain-specific or complex information.

Example 1: Easy question (Business Informatics flyer)
Prompt: What are the advantages of alternating between theoretical and practical phases?
GPT-only completion: 'Theory is applied directly in practice, which increases learning success.'
Quantitative analysis: While this answer is semantically correct, it is also generic. GPT-only's F1, BLEU and METEOR scores for the easy Business Informatics flyer questions are comparatively high, (F1: 0.72, BLEU: 0.56, METEOR: 0.64) but reflect this general nature. A reference answer from the flyer might use more specific phrasing that is not precisely captured by GPT-only.
RAG completion: 'The combination of theoretical and practical elements ensures the course is both varied and practical.'
Quantitative analysis: The RAG-based response utilizes adjectives ('varied', 'practical') that are aligned with the specific terminology found in DHBW's official materials. Although the F1, BLEU and METEOR scores are slightly lower than those for GPT-only in this particular category, they demonstrate RAG's ability to ground answers in document-specific phrasing.
Example 2: Difficult question (computer science/software engineering flyer).
Prompt: In which field of study does the development of secure software belong?
GPT-only completion: 'The development of secure software is usually part of the degree program.'
Quantitative analysis: This response is highly unspecific and lacks the precise institutional detail necessary. The lower F1, BLEU and METEOR scores for GPT-only on difficult questions from the Computer Science Flyer (e.g. F1: 0.59) clearly reflect this generalization and the inability to provide accurate and specific information from its internal knowledge base alone.
RAG Completion: 'The Software Engineering course provides instruction on the development of secure software.'
Quantitative analysis: In stark contrast, the RAG-based response precisely attributes the topic to the 'Software Engineering' course, directly referencing the structure of DHBW's curriculum. This accuracy is a key driver for RAG's substantially higher scores (e.g., F1: 0.90) in this difficult category, showcasing its ability to retrieve and integrate specific domain knowledge.
These examples reinforce the quantitative findings, showing that while GPT-only can provide generic answers, RAG is essential for providing specific, accurate and contextually relevant information derived directly from university documents — especially for complex, domain-specific queries.

Human Evaluation and Qualitative insights

To complement the quantitative results presented in Quantitative Results by Document Type and Question Difficulty, a structured evaluation by human experts was conducted. This qualitative assessment aimed to evaluate the practical usefulness and factual accuracy of the systems' responses from the perspective of experienced academic advisors, identifying nuances that automated metrics might overlook.

Evaluation Setup

Three academic experts, all holding leadership positions as program directors in dual study programs (Business Informatics and Software Engineering) at DHBW, participated in a blind review process. Their deep familiarity with the university's curricula and official documents ensured a highly relevant assessment.

The evaluation focused exclusively on student query scenarios derived from the Business Informatics program flyer. This specific focus enabled an in-depth qualitative analysis, allowing direct comparisons to be made between system responses and the experts' domain knowledge. For each pre-determined query, the experts were presented with two anonymized answers: one from the RAG-based model and one from the GPT-only model. They were instructed to:

- compare each response against a human-annotated gold standard (previously defined based on official study documents).
- Assign a score from 0 to 2 for factual correctness and completeness (*0 = not fulfilled, 1 = partially fulfilled, 2 = fully fulfilled*).
- They were also asked to provide a concise written justification for each score, highlighting any missing information, inaccuracies, or overly generic language.

Findings and Expert Perspectives

The human expert evaluation of the questions on the Business Informatics program flyer revealed critical insights that complement and sometimes nuance the automatic metric results. Table 2 provides an overview of the average human ratings alongside the automated F1, BLEU and METEOR scores for this category of document.

Table 2. *Combined Automatic Metrics and Average Human Ratings for Business Informatics Flyer*

GPT- only vs. RAG mean scores

Document type	Model	Question Difficulty	human rating	F1-Score	Bleu-Score	Meteor-Score
Business Informatics Fyler	GPT-only	easy	0.45	0.72	0.56	0.64
	RAG	easy	0.71	0.68	0.42	0.61
	GPT-only	medium	0.55	0.74	0.56	0.65
	RAG	medium	0.57	0.63	0.39	0.55
	GPT-only	difficult	0.47	0.66	0.39	0.53
	RAG	difficult	0.48	0.60	0.34	0.49

Analysis of Combined Metrics for Business Informatics Flyer

The results of questions based on the Business Informatics Flyer indicate a dynamic performance landscape, with varying strengths for GPT-only and RAG depending on the difficulty of the question. This section critically examines the interplay between automatic metrics and human expert judgements.

Easy questions: For straightforward queries, the RAG system achieved a notably higher average human rating (0.71) than GPT-only (0.45). Despite GPT-only achieving superior scores in automatic metrics (F1: 0.72 vs RAG: 0.68; BLEU: 0.56 vs RAG: 0.42; METEOR: 0.64 vs RAG: 0.61), human experts perceived RAG's responses as more accurate and reliable, or more grounded in the content of the flyer. This discrepancy highlights that, for easy factual questions, human experts prioritize precise factual alignment over broader semantic coverage or stylistic fluency, which automatic metrics might favour.

Medium questions: For questions of medium difficulty, both models performed competitively in terms of human ratings. RAG achieved an average rating of 0.57, just above GPT-only's 0.55. However, GPT-only maintained a clear lead in all automatic metrics: F1 score (0.74 vs. 0.63 for RAG); BLEU score (0.56 vs. 0.39 for RAG); and METEOR score (0.65 vs. 0.55 for RAG). This suggests that, for moderately complex questions on this particular flyer, GPT-only's vast pre-trained knowledge enabled it to generate accurate answers that aligned well with general linguistic patterns.

Difficult questions: When it came to the most challenging questions related to the Business Informatics Flyer, the RAG system marginally outperformed GPT-only in terms of human ratings (0.48 versus 0.47). This is a notable finding, given that GPT-only had a consistent lead in terms of automatic metrics (F1: 0.66 versus RAG's 0.60; BLEU: 0.39 versus RAG's 0.34; METEOR: 0.53 versus RAG's 0.49). This suggests that, despite the concise nature of the flyer, RAG's ability to retrieve limited relevant context provided a slight edge in perceived accuracy and grounding by human experts for highly complex questions. Conversely, GPT-only's tendency to generalize or infer from its broader knowledge base, while achieving higher automatic metric scores for linguistic similarity, may have been perceived as less accurate human evaluators for these specific, demanding queries from such a concise document.

Figure 3 visualizes the individual expert scores for each question, providing a granular understanding of the expert agreement and discrepancies. This level of detail is essential for understanding the accuracy and consistency of the responses, showing where the models consistently met expectations and where they struggled in the eyes of the experts.

For *easy* questions, RAG responses tend to be highly consistent across experts, often receiving a score of 2, which indicates strong agreement on full fulfilment. In contrast, GPT-only responses to *easy* questions demonstrate greater variability and lower scores, particularly from Experts 1 and 3. This confirms the lower average human rating for GPT-only responses in this category.

While the average human ratings for *medium* questions are very close, Figure 3 reveals mixed expert opinions for both models. This indicates that neither model consistently achieved full agreement on high scores.

For *difficult* questions, the average human rating shows that RAG has a minimal lead over GPT-only. This is supported by instances where RAG's responses achieved a higher agreement on "partially fulfilled" or "fully fulfilled" scores, compared to GPT-only which received more "not fulfilled" scores (0) from certain experts. This suggests that, despite the challenges of concise summaries, RAG occasionally provided the critical information that GPT-only missed entirely for difficult queries.

Figure 3. *Expert Evaluation Scores by Question Difficulty for Business Informatics Flyer (0 = not fulfilled, 1 = partially fulfilled, 2 = fully fulfilled)*

Difficulty level	Question (Q)	Reference (R)	Answer Model	Expert 1	Expert 2	Expert 3	Answer Model	Expert 1	Expert 2	Expert 3
easy	Q1	R1	RAG Answer	2	2	2	GPT-only Answer	2	2	2
	Q2	R2	RAG Answer	2	2	2	GPT-only Answer	2	2	2
	Q3	R3	RAG Answer	2	2	2	GPT-only Answer	2	2	2
	Q4	R4	RAG Answer	2	2	2	GPT-only Answer	2	2	2
	Q5	R5	RAG Answer	2	2	2	GPT-only Answer	1	1	1
	Q6	R6	RAG Answer	1	1	1	GPT-only Answer	1	0	0
	Q7	R7	RAG Answer	1	1	1	GPT-only Answer	0	0	0
	Q8	R8	RAG Answer	1	1	1	GPT-only Answer	2	2	0
	Q9	R9	RAG Answer	1	1	2	GPT-only Answer	2	2	0
	Q10	R10	RAG Answer	0	0	2	GPT-only Answer	0	0	0

Difficulty level	Question (Q)	Reference (R)	Answer Model	Expert 1	Expert 2	Expert 3	Answer Model	Expert 1	Expert 2	Expert 3
medium	Q1	R1	RAG Answer	2	2	2	GPT-only Answer	2	2	1
	Q2	R2	RAG Answer	1	1	1	GPT-only Answer	2	2	2
	Q3	R3	RAG Answer	1	1	1	GPT-only Answer	0	0	0
	Q4	R4	RAG Answer	2	2	2	GPT-only Answer	1	1	1
	Q5	R5	RAG Answer	2	2	2	GPT-only Answer	1	1	1
	Q6	R6	RAG Answer	2	2	2	GPT-only Answer	2	2	2
	Q7	R7	RAG Answer	2	2	2	GPT-only Answer	1	2	1
	Q8	R8	RAG Answer	2	2	2	GPT-only Answer	2	2	2
	Q9	R9	RAG Answer	2	2	2	GPT-only Answer	2	2	2
	Q10	R10	RAG Answer	2	2	2	GPT-only Answer	2	2	2

Difficulty level	Question (Q)	Reference (R)	Answer Model	Expert 1	Expert 2	Expert 3	Answer Model	Expert 1	Expert 2	Expert 3
difficult	Q1	R1	RAG Answer	1	1	1	GPT-only Answer	2	2	2
	Q2	R2	RAG Answer	1	1	1	GPT-only Answer	2	2	2
	Q3	R3	RAG Answer	2	2	2	GPT-only Answer	0	0	0
	Q4	R4	RAG Answer	2	2	2	GPT-only Answer	2	2	0
	Q5	R5	RAG Answer	0	0	0	GPT-only Answer	0	0	0
	Q6	R6	RAG Answer	2	2	2	GPT-only Answer	0	0	0
	Q7	R7	RAG Answer	1	1	1	GPT-only Answer	1	0	0
	Q8	R8	RAG Answer	0	0	0	GPT-only Answer	2	2	0
	Q9	R9	RAG Answer	0	0	0	GPT-only Answer	2	2	0
	Q10	R10	RAG Answer	2	2	2	GPT-only Answer	1	1	1

The qualitative justifications provided by the experts further elaborate on these observations:

Expert 1 (Head of Degree Program Business Informatics): ‘RAG provided answers that mirrored the structure of the flyer, particularly with regard to questions about learning outcomes and course flow. GPT-Base provided plausible, but occasionally inaccurate, information.’

Interpretation: This expert's comment reinforces RAG's strength in mirroring document structure. The human ratings suggest this structural alignment is highly valued for easy questions.

Expert 2 (Head of the Informatics Degree Program): 'For complex or curriculum-related queries in particular, RAG was clearly more reliable. GPT tended to generalize.'

Interpretation: This observation from Expert 2 aligns with the general benefits of RAG for specific, complex content, as confirmed by the Module Handbook results in Quantitative Results by Document Type and Question Difficulty. Even for the concise Business Informatics flyer, RAG's lead in human ratings for difficult questions suggests that its clarity for specific points are perceived as more valuable than GPT's generalizations.

Expert 3 (Head of Degree Program Business Informatics): 'The answers from RAG felt more grounded — they referenced specific content from the documents we use for guidance. GPT alone guessed too much.'

Interpretation: The emphasis on 'grounded' answers is crucial. For easy questions, this grounding gives RAG a clear human rating advantage. For medium and difficult questions, both models struggle to achieve consistently high human scores on this concise flyer. However, RAG's marginal lead for difficult questions suggests that its 'grounded' approach, even when yielding limited information, is slightly preferred to GPT's potentially less verifiable inferences or predictions.

These qualitative insights, when combined with the detailed visual representation of the individual expert scores, show that the performance differences between RAG and GPT-only depend highly on the nature of the source document and the specific demands of the query. This emphasizes the importance of analyzing specific use cases and document characteristics when designing LLM-based advisory systems.

Discussion & Educational Implications

This section provides a holistic interpretation of the quantitative and qualitative results presented in the Results Section, discussing their wider implications for the use of Large Language Models (LLMs) in higher education. The aim is to summarize the findings, explain the reasons behind the observed results, and derive practical recommendations for academic information systems.

Results and Interpretation of Key Findings

Performance on the Business Informatics Flyer: For questions derived from the Business Informatics Flyer, the GPT-only model demonstrated competitive performance and in some areas (automatic metrics and medium human ratings), superior performance compared to RAG. This suggests that, for documents containing broadly accessible information or relating to general concepts (e.g. the advantages of dual studies), GPT-only's extensive pre-trained knowledge base is often sufficient. The information required for these queries may be adequately represented in its training data, enabling it to generate accurate responses without external retrieval. The slight divergence, whereby RAG achieved higher human ratings for easy questions despite lower automatic scores, suggests that human evaluators prioritize direct factual alignment, even if the fluency of the response is not optimized.

Performance on Computer Science/Software Engineering Flyer: In contrast, the evaluation of questions from the Computer Science/Software Engineering Flyer showed a clear and significant superiority of the RAG system across all metrics and difficulty levels. This difference is particularly pronounced for medium and difficult questions. This suggests that the content of this flyer likely contained more specialized or less commonly encountered information that was not as robustly represented in GPT-only's general training data. In this case, RAG's capacity to accurately retrieve specific details from the target document became critical, enabling it to provide accurate and contextually relevant answers where GPT-only struggled to recall correctly. Performance on the Business Informatics Module Handbook: The RAG system demonstrated the most significant performance improvements with questions derived from the Business Informatics Module Handbook. RAG performed better than GPT-only across all difficulty levels and metrics. This finding strongly supports the idea that retrieval augmentation is essential for highly specific, detailed and extensive domain knowledge.

Added Value and Educational Implications

The results of this study show that adding RAG to academic information systems is valuable, especially in higher education. Addressing LLM limitations and enhancing trust: A key challenge with GPT-only models is that they are liable to hallucinations and rely on static training data, which makes it difficult to control the recency and source of information. In dynamic educational environments where curricula and policies frequently change, this can result in the generation of outdated or inaccurate advice. RAG addresses this directly by enabling the integration of current institutional documents, such as up-to-date module handbooks or recently revised degree plans.

This on-demand update capability gives university administrators greater transparency and control over content, ensuring that AI-powered services provide reliable information. From a university's perspective, the ability to base responses on verified documents fundamentally enhances the quality of answers, fostering greater trust among students and staff. The study shows that a one-size approach to LLM deployment is not ideal. Instead, a differentiated deployment strategy is efficient. GPT-only models can be suitable for simple queries (e.g. contact details or general study facts) or questions related to widely available information (as demonstrated by the Business Informatics flyer). Their extensive pre-trained knowledge enables them to provide plausible and accurate responses. However, for complex, context-dependent queries that require high levels of precision and specific institutional knowledge (e.g. detailed curriculum planning, module competencies or unique program comparisons), RAG is essential. Its superior performance on the Computer Science flyer and the Module Handbook demonstrates its critical role in providing accurate responses in such scenarios. Qualitative expert feedback consistently reinforces this, highlighting RAG's ability to mirror the structure of the flyer and provide accurate answers.

Limitations and Future Work

Although this study provides valuable insights, it is subject to certain limitations. The evaluation was conducted using documents from a single university and focused on specific program types. Future work could involve scaling the system to support additional degree program by integrating a wider range of institutional content, including dynamic online sources (e.g. official web pages and news feeds), to ensure real-time content alignment. Additionally, it would be valuable to explore the impact of different chunking strategies or embedding models for diverse document types. Investigating user satisfaction and the long-term impact of RAG-based chatbots on administrative efficiency through different studies would also provide further practical insights.

Conclusion

This study systematically evaluated the accuracy of RAG against a GPT-only in handling university-specific queries, demonstrating the critical role of retrieval augmentation in educational question-and-answer scenarios. Through a structured benchmark involving 90 categorized questions derived from various university documents, our findings reveal the nuances of performance. While the GPT-only model proved sufficient for surface-level inquiries and questions related to broadly accessible information, RAG consistently demonstrated superior performance when deeper comprehension was required.

Our results confirm that integrating a document retrieval mechanism directly improves the accuracy of responses. This provides clear evidence that augmenting LLMs with retrieval capabilities is a significant technical advancement for educational institutions. By basing responses on verified official documents, RAG systems effectively reduce the risk of misinformation and promote transparent communication.

This paper provides a replicable blueprint for creating domain-specific chatbots for academic contexts. It illustrates how openly available documents can be systematically integrated into AI systems to create context-aware educational assistants. Future work will expand the system's knowledge base to include a wider range of institutional content, such as dynamic online sources. It will also explore integrating multimodal inputs to enhance the chatbot's capabilities as a comprehensive academic assistant.

Acknowledgements

The author would like to express their sincere gratitude to Prof. Dr. Anke Hutzschenreuter (DHBW Heidenheim) for her invaluable guidance and thoughtful suggestions during the development of this manuscript. Her support and critical feedback were instrumental in refining the work.

References

Asai A, Min S, Zhong Z & Chen D (n.d.) *Tutorial proposal: Retrieval-based language models and applications.*

Barnett S, Kurniawan S, Thudumu S, Brannelly Z & Abdelrazek M (2024) Seven failure points when engineering a retrieval augmented generation system. *Proceedings of the IEEE/ACM 3rd International Conference on AI Engineering – Software Engineering for AI*, 14 April 2024, ACM, pp. 194–199. doi:10.1145/3644815.3644945.

Bloom BS, Engelhart MD, Furst EJ, Hill WH & Krathwohl DR (1956) *Taxonomy of educational objectives: The classification of educational goals. Handbook I: The cognitive domain.* Philadelphia.

Brown TB, Mann B, Ryder N & et al. (2020) Language models are few-shot learners. *arXiv preprint* arXiv:2005.14165. doi:10.48550/arXiv.2005.14165.

Calfoforo JC & Raga RC (2024) Unleashing AI in education: A pre-trained LLMs for accurate and efficient question-answering systems. *2024 21st International Conference on Information Technology Based Higher Education and Training (ITHET)*, 6 November 2024, IEEE, pp. 1–6. doi:10.1109/ITHET61869.2024.10837606.

Chauhan S & Daniel P (2023) A comprehensive survey on various fully automatic machine translation evaluation metrics. *Neural Processing Letters* 55(9): 12663–126717. doi:10. 1007/s11063-022-10835-4.

Fan W, Ding Y, Ning L & et al. (2024) A survey on RAG meeting LLMs: Towards retrieval-augmented large language models. *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 25 August 2024, ACM, pp. 6491–6501. doi:10.1145/3637528.3671470.

Gao Y, Xiong Y, Gao X & et al. (2024) Retrieval-augmented generation for large language models: A survey. *arXiv preprint* arXiv:2312.10997. doi:10.48550/arXiv.2312.10997.

Huang Y & Huang J (2024) A survey on retrieval-augmented text generation for large language models. *arXiv preprint* arXiv:2404.10981. doi:10.48550/arXiv.2404.10981.

Kalyan KS (2024) A survey of GPT-3 family large language models including ChatGPT and GPT-4. *Natural Language Processing Journal* 6: 100048. doi:10.1016/j.nlp.2023.100048.

Khan UH, Khan MH & Ali R (2025) Large language model based educational virtual assistant using RAG framework. *Procedia Computer Science* 252: 905–911. doi:10.10 16/j.procs.2025.01.051.

Lavie A & Denkowski MJ (2009) The METEOR metric for automatic evaluation of machine translation. *Machine Translation* 23(2–3): 105–115. doi:10.1007/s10590-009-9059-4.

Lewis P, Perez E, Piktus A, Petroni F, Karpukhin V, Goyal N, Küttler H, Lewis M, Yih W-T, Rocktäschel T & et al. (2020) Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems* 33: 9459–9474.

Li H, Su Y, Cai D, Wang Y & Liu L (2022) A survey on retrieval-augmented text generation. *arXiv preprint* arXiv:2202.01110. doi:10.48550/arXiv.2202.01110.

Liang H, Zhou Y & Gurbani VK (2024) Efficient and verifiable responses using retrieval augmented generation (RAG). *Proceedings of the 4th International Conference on AI-ML Systems*, 8 October 2024, ACM, pp. 1–6. doi:10.1145/3703412.3703431.

Naveed H, Khan AU, Qiu S & et al. (2025) A comprehensive overview of large language models. *ACM Transactions on Intelligent Systems and Technology* (online first). doi:10.1145/3744746.

Quidwai MA & Lagana A (2024) A RAG chatbot for precision medicine of multiple myeloma. *Genetic and Genomic Medicine*, preprint, 18 March 2024. doi:10.1101/2024. 03.14.24304293.

Soygazi F & Oguz D (2023) An analysis of large language models and LangChain in mathematics education. *Proceedings of the 2023 7th International Conference on Advances in Artificial Intelligence*, 13 October 2023, ACM, pp. 92–97. doi:10.1145/3633598.3633614.

Vakayil S, Sujitha Juliet D, Anitha J & Vakayil S (2024) RAG-based LLM chatbot using Llama-2. *2024 7th International Conference on Devices, Circuits and Systems (ICDCS)*, 23 April 2024, IEEE, pp. 1–5. doi:10.1109/ICDCS59278.2024.10561020.

Waldo J & Boussard S (2025) GPTs and hallucination. *Communications of the ACM* 68(1): 40–45. doi:10.1145/3703757.

Zhao P, Zhang H, Yu Q, Wang Z, Geng Y, Fu F, Yang L, Zhang W & Cui B (2024) Retrieval-augmented generation for AI-generated content: A survey. *arXiv preprint* arXiv:2402.19473.

Zhong N, Qian Z & Zhang X (2021) Deep neural network retrieval. *Proceedings of the 29th ACM International Conference on Multimedia*, 17 October 2021, ACM, pp. 3455–3463. doi:10.1145/3474085.3475505.