Revista Española de Pedagogía

Revista Española de Pedagogía (REP)
ISSN: 0034-9461
e-ISSN: 2174-0909

The ethics of generative artificial intelligence in education under debate. A perspective from the development of a theoretical-practical case study

La ética de la inteligencia artificial generativa en educación a debate. Perspectiva desde el desarrollo de un caso de estudio teórico-práctico

Francisco-José GARCÍA-PEÑALVO, PhD. Professor, Universidad de Salamanca *(fgarcia@usal.es).* (*)
María-José CASAÑ-GUERRERO, PhD. Associate Professor, UPC Universidad Politécnica de Cataluña *(ma.jose.casan@upc.edu).*

Marc ALIER-FORMENT, PhD. Associate Professor, UPC Universidad Politécnica de Cataluña *(marc.alier@upc.edu).*

Juan-Antonio PEREIRA-VARELA, PhD. Associate Professor, Universidad del País Vasco *(juanan.pereira@ehu.eus).*

(*) Corresponding author

Abstract:

This article examines the ethics of generative artificial intelligence (GenAI) in higher education from a theoretical and practical perspective. It reflects the growing importance of teaching ethics in information technology and its impact on society. Its aim is to develop a training model that integrates ethical reflection on GenAI into the education of computer engineering students. To achieve this, a case study is used based on the Social and Environmental Aspects of Computer Science module at the Universidad Politécnica de Cataluña, in which the theoretical principles of artificial intelligence (AI) ethics are applied through a case study of the use of GenAI within the course. Several examples of GenAI systems within AI applied to education are introduced. One of the examples, developed by the authors, involves an AI assistant built using the LAMB framework that enables students to analyse the proposed case using the PESTLE (political, economic, social, technological, legal, and environmental) method with the assistant acting as an expert. Students are then required to analyse a similar case study in another domain. The results suggest that this theoretical-practical approach, where abstract concepts of AI ethics and safety are grounded in specific decisions about application and concrete technological artefacts, effectively integrates ethical reflection into engineering education, highlighting the need for multidisciplinary approaches to address emerging ethical challenges in AI and education.

Date of receipt of the original: 2024-12-12.
Date of approval: 2025-02-03.
Please, cite this article as follows: García-Peñalvo, F.-J., Casañ-Guerrero, M.-J., Alier-Forment, M., & Pereira-Valera, J.-A. (2025). The ethics of generative artificial intelligence in education under debate. A perspective from the development of a theoretical-practical case study [La ética de la inteligencia artificial generativa en educación a debate. Perspectiva desde el desarrollo de un caso de estudio teórico-práctico]. *Revista Española de Pedagogía, 83*(291), 281-293 https://doi.org/10.22550/2174-0909.4577

Keywords: ethics of artificial intelligence, higher education, generative artificial intelligence, engineering education, PESTLE method, artificial intelligence in education, artificial intelligence regulation, ethical competencies.

Resumen:

El artículo analiza la ética de la inteligencia artificial generativa (IAGen) en la educación superior desde un enfoque teórico-práctico. Se enmarca en la creciente relevancia de la enseñanza de la ética en tecnologías de la información y en su impacto en la sociedad. El objetivo principal es desarrollar un modelo formativo que integre la reflexión ética sobre la IAGen en la formación de ingenieros informáticos. Para ello, se emplea un estudio de caso basado en la asignatura Aspectos sociales y medioambientales de la informática de la Universidad Politécnica de Cataluña, en la que se aplican los principios teóricos de la ética de la inteligencia artificial (IA) a partir de un caso de aplicación de IAGen en la propia asignatura. Se introducen varios ejemplos de sistemas de IAGen en el dominio de la IA aplicada a la educación. Uno de los ejemplos, desarrollado por los autores, utiliza un asistente de IA creado con el *framework* LAMB, que permite a los estudiantes analizar el caso propuesto mediante el método PESTLE (*political, economic, social, technological, legal, and environmental*) y el uso del asistente como experto. Con posterioridad, los estudiantes deben analizar un caso de estudio análogo en otro dominio. Los resultados sugieren que este enfoque teórico-práctico, en el que los conceptos abstractos de ética y seguridad de la IA se aterrizan en decisiones de aplicación específicas y en artefactos tecnológicos concretos, es efectivo para integrar la reflexión ética en la enseñanza de la ingeniería y subraya la necesidad de implementar enfoques multidisciplinares para abordar los desafíos éticos emergentes de la IA en la educación.

Palabras clave: ética de la inteligencia artificial, educación superior, inteligencia artificial generativa, educación en ingeniería, método PESTLE, inteligencia artificial en educación, regulación de la IA, competencias éticas.

1. Introduction

1.1. The ethics of information technology as a subject of study

Teaching ethics in information and communication technology (ICT) to engineering students is increasingly important owing to the profound impact of these technologies in contemporary society. As Casañ et al. (2020) observe, technology clearly influences our way of life, culture, economy, how we function in society, and how we relate to our surroundings, and ICT is no exception.

ICT's potential to create ethical and social problems that differ from those posed by other technologies has been debated since the earliest days of digital computers (Wiener, 1950). This potential is amplified by the accelerated pace of innovation in the field, which, according to Moore's law, doubles approximately every 18 months (Moore, 1965). This rate of change led Ray Kurzweil (2005) to predict that the technological development we will experience during the 21st century will be equivalent to 20 000 years of progress at the current pace.

The importance of incorporating ethics into the Computer Engineering curriculum was formally recognised in 1991, when the study of ethics was first introduced into Computer Science study plans (Bynum et al., 1992). The concept *computer ethics* had been coined before then by Walter Maner (1980), who noted that ethical decisions become much more complex when computers are added to the equation, as Johnson (2009) explains. This idea is expressed

in a well-known quote from 1969, attributed to Paul Ehrlich or Bill Vaughan (Quote Investigator, 2010): "To err is human but to really foul things up requires a computer".

According to the guidelines of the IEEE/ACM Computer Science Curriculum 2023 (Kumar et al., 2024), the education that Computer Engineering students receive must incorporate all of the dimensions and functions of computing as a profession (technical, philosophical, and ethical) as well as understanding that these standards vary internationally (Casañ et al., 2020).

The emergence of generative artificial intelligence (GenAI) (García-Peñalvo & Vázquez-Ingelmo, 2023; Jovanović & Campbell, 2022) and the launch of ChatGPT on 30 November 2022 were a turning point marking a decisive moment that exemplified the transition of artificial intelligence (AI) from a "deceptive" phase to a "disruptive" phase, according to Diamandis's model of the 6Ds (Diamandis & Kotler, 2012). According to this model, a technology passes through six phases when it is digitised: digitisation, deceptive growth, disruptive growth, demonetisation, democratisation, and dematerialisation. After decades of deceptively slow development, GenAI has reached a turning point typical of exponential curves where its impact on society cannot be ignored.

This impact is present in the three principal dimensions of sustainability. In the social dimension, AI is revolutionising communication (Elmoudden & Wrench, 2024), work (Anurag et al., 2023), and decision making (Parra et al., 2024), most noticeably affecting the fields of art (Epstein et al., 2023), culture (Henriksen et al., 2025), language (Martínez-Arboleda, 2024), and education (García-Peñalvo et al., 2024), which poses a challenge for the rules of the game and existing laws and raises questions about the process of authorship when the creation of any type of content or knowledge is involved (González-Geraldo & Ortega-López, 2024). In the environmental dimension, training large language models (LLM) (Zhao et al., 2024) require significant energy consumption (Samsi et al., 2023) but AI can help optimise processes and reduce the environmental impact of various sectors (Kar et al., 2022). In the economic dimension, GenAI is creating new business models and jobs and transforming entire industries (Kanbach et al., 2024), promoting higher levels of productivity at the same time as threatening existing business models and jobs (García-Peñalvo & Vázquez-Ingelmo, 2023).

This emerging scenario means it is vital to incorporate the ethics of AI as a specific component in the education of future professionals in any branch of knowledge, but in particular those that relate to ICT. However, this presents unique challenges. The dizzying pace at which the state of the art in AI is evolving (with new models, capacities, and applications emerging almost daily) means that even specialists in the field find it hard to keep up to date. This poses an additional challenge for teaching the ethics of AI, as ethical aspects must be analysed on a constantly changing technological foundation (Alier et al., 2024).

As Flores-Vivar and García-Peñalvo (2023) note, it is important to develop ethical frameworks that not only address the technical aspects of AI, but that also consider its impact on society and the environment. Ethics education for AI must prepare students to confront dilemmas that do not yet exist and develop critical thinking that will allow them to evaluate the ethical implications of the applications of AI as these emerge.

1.2. Approaches to teaching the ethics of AI

The teaching of ethics in Software Engineering has its origins in Europe with pioneering initiatives like that of the Computing Faculty of the Universidad Politécnica de Cataluña, which in 1991 incorporated courses on professional ethics and the history of computing (Casañ & Alier, 2024). Another important initiative was the creation of the Centre for Computing and Social Responsibility at De Montfort University in 1995 (Gotterbarn et al., 1997).

There are various approaches to introducing ethics into ICT curricula. Some centre on the process of ethical decision making, while others emphasise practical applications through

deontological codes. Johnson (2017) suggests providing knowledge of codes and standards of behaviour, developing skills for practical application. However, an alternative perspective integrates both ethics and the social implications, exposing students to diverse cultural, social, and legal questions to expand their understanding. This focus was adopted by academics such as Barceló and Gordon (Spiekermann, 2015).

Education in engineering should include both technical and professional skills (Bowden, 2010). The criteria of the ABET Engineering Accreditation Commission (2021) emphasise competences such as communication, team work, and ethical comprehension, along with awareness of the overall context. In Spain, the white papers for adapting to the EHEA specifically recommended courses on ethics, following the guidelines of the ACM/IEEE and ABET (Miñano et al., 2019).

Teaching methodologies include using deontological codes, case studies, moral theory, and service learning. In order to teach this content effectively, teachers must combine competences in philosophy of science and in computing and technology.

1.3. Structure of the article

Section 2 of this article introduces the challenge of incorporating the ethics of AI into higher education in engineering. Section 3 addresses a possible introduction of this topic from both the theoretical and practical viewpoints. Finally, part 4 sets out the conclusions of the work.

2. The challenge of incorporating the ethics of AI directly as content

The relationship between research into AI and its commercial application has undergone a significant change in recent years. The phrase "as soon as it works, nobody calls it AI any more" is attributed to John McCarthy (Meyer, 2011) (among others [Quote Investigator, 2024]), one of the parents of the concept *artificial intelligence* in the 1950s. This remark notes AI's nature as a field of research centred on giving computer systems capacities traditionally associated with human intelligence, and its achievements result in technological applications that we stop seeing as *intelligent*: logical inference, facial recognition, optimisation, etc.

However, we are currently seeing a blurring of the boundaries between academic research and commercial applications. This phenomenon is especially apparent in the field of GenAI, where the concept *artificial intelligence* has moved beyond its academic origin to become a commercial label widely used in products and services aimed at the general public.

One widely held position, including among experts in technology, views AI – especially the surprising GenAI technologies that have appeared in the last two years – as a technology that is beyond our control, resembling the metaphor of the *palantír* of J.R.R. Tolkien (1954): a powerful but dangerous magical tool, controlled by forces we do not fully understand (Alier, 2024). This analogy is particularly relevant when we observe how the very leaders of the technology industry contribute to this narrative.

Prominent figures such as Elon Musk, Ilya Sutskever, and Sam Altman have fed this perception when publicly discussing the existential risks of AI and the imminent arrival of artificial general intelligence (AGI) (Morrison, 2024), even going so far as to advocate artificial superintelligence (Altman et al., 2023). This vision has been amplified by influential philosophers such as Nick Bostrom (2014) and Yuval Noah Harari (2015), who have published extensively on the potential dangers of this technology.

Paradoxically, this idea of AI as an almost-magical tool is promoted both by its most ardent critics and by its principal defenders. This apparent contradiction could be explained if we consider that this type of narrative fosters a perception of inevitability of the adoption of these technologies, at the same time as justifying the need for their regulation.

The current situation recalls the *bootleggers and baptists* phenomenon described by Bruce Yandle (1983). As in the case of prohibition in the USA, when moralist preachers and alcohol smugglers alike supported prohibition for different motives, we currently see a similar alliance between people with an *apocalyptic* outlook who warn of the dangers of AI and the *integrationists* who benefit from these regulations (Smith & Yandle, 2014).

In view of this pressure, the European Union has passed specific legislation on AI without even having a clear and agreed definition of what it is (Morrison, 2024). Therefore, this legislative focus reflects an attempt to establish trust in AI systems through governance, albeit at the cost of the clarity of definitions (Bellogín et al., 2024). This haste contrasts with the slower development of other regulations such as the GDPR (General Data Protection Regulation) (European Parliament & Council of the European Union, 2016) and it could benefit large technology companies through *regulatory capture*, a phenomenon where regulation comes to favour established businesses instead of serving the public interest (Dal Bó, 2006; Saltelli et al., 2022; Wei et al., 2024), creating barriers to entry for new competitors.

The first step in approaching the study of the ethics of AI requires a dual focus. On the one hand, it is necessary to recognise and understand how the concept *artificial intelligence* has been positioned in popular culture (where the metaphor of the *palantír* illustrates the current perception of this technology). On the other hand, as the target population is mainly (but not exclusively) engineering students from fields relating to ICT (in areas such as data science, artificial intelligence, and computer engineering), AI should be presented from a rigorous technical perspective.

This technical perspective involves understanding AI as a group of technologies and tools with specific characteristics and functionalities shaped by conscious design decisions by researchers and developers. It is crucial to understand that these new technologies do not exist in isolation, but instead are integrated into pre-existing systems and technologies. This process of integration is not new in the field of technology; we see a parallel in how smartphones became integrated into existing web technology after their introduction in 2007, transforming, but not replacing, the previous technological ecosystem.

3. Introducing the ethics of AI in the curriculum

The Social and Environmental Aspects of Computing module will be used as a case study. This is a 6 ECTS (European Credit Transfer and accumulation System) credit optional module that is delivered in the final years of the degree in Computer Engineering of the Universidad Politécnica de Cataluña. Students usually take this module in their third or fourth year, when they already have a solid technical foundation in the theoretical-practical concepts of computer engineering.

This module has existed for more than 30 years, introduced into the study plans of the Faculty of Computing of Barcelona by Professor Miquel Barceló, a pioneer in the teaching of technoethics in software engineering courses (Casañ et al., 2020). The module has been part of numerous study plans over the decades and has constantly evolved, both in the content delivered and in the methodologies applied, adapting to the constant and rapid changes in ICT and its impact on society.

In its current version, the module's programme covers four principal areas:

- History of ICT
- Social impact of technology
- Environmental impact
- Ethics of ICT

The ethics block has an important weight in the module. The principal ethical theories are studied (Kantianism [Sevilla, 2014], utilitarianism [Zavala, 1999], virtue ethics [Hoyos, 2011], and the social contract [Leal, 1982]) and they are applied to case histories where students must analyse dilemmas and debate different moral positions relating to ICT and its social and environmental impact, integrating knowledge acquired in the course's other blocks. This practical focus is intended to develop critical thinking and the capacity for ethical analysis.

Since 2023, a new transversal block on AI has been added, including:

- The history and evolution of AI
- Its social impact (automation, the future of work, etc.)
- Its environmental impact (including the costs of AI and the possible benefits of its use to solve environmental problems)
- The ethics of AI

This update is in response to the need to address the unique challenges that AI poses, and it maintains the module's dual focus by involving theoretical study followed by a practical application using analysis of cases.

3.1. Theoretical aspects of the ethics of AI

Content that has previously been covered in the module is adapted to address the theoretical aspects of the ethics of AI, giving this content a specific focus in this area. For example, the social aspects of information technologies are analysed after discussing the questions of how automation has impacted the future of work, with similar phenomena being seen in previous industrial revolutions, where some occupations disappeared while others emerged. On this line, it is relevant to address specifically the influence of AI on this problem.

Consequently, in the theoretical aspect of the ethics of AI, the teaching team has designed a two-hour theory session addressing the following topics:

- Apocalyptic visions of AI: from Frankenstein (Shelley, 1831) and the Luddites (Prieto, 2016) to contemporary authors such as Bostrom (2014) and Harari (2015, 2024) who theorise on the real or hypothetical dangers of the imminent development of AI.
- A realistic and pragmatic analysis of the current ethical challenges and dilemmas of AI:
    - Algorithmic bias and discrimination
    - Privacy and data protection
    - Transformation of the job market
    - Control and autonomy of AI systems
    - Transparency and explainability of the algorithms
    - Alignment with human values
    - Implications for cybersecurity

It is deemed necessary that after the theoretical foundations of the ethics of AI have been presented, students will be able to apply the concepts acquired to specific contexts that are close to the reality of their lives. To this end, designing of a new practical activity is proposed. This is described in the following section.

3.2. Designing practical activities to work on the ethics of AI

The students are presented with a practical case for discussion so that they can design a practical activity where they apply the theoretical concepts relating to the ethics of AI that have previously been addressed.

The case study centres on the application of AI in the educational sphere. In this context, the authors developed an AI assistant using the LAMB (learning assistant manager and builder) framework (Alier et al., 2025).

An AI assistant with a specific knowledge base (Casañ et al., 2024) relating to anthropomorphic robots and their possible commercialisation in the years to come was created using LAMB. This knowledge base encompasses environmental, legal, economic, technological, ethical, and political aspects relating to anthropomorphic robots. Consequently, the assistant act as an expert consultant on the topic of anthropomorphic robots. In the thematic block of the module relating to social aspects of computing, the students do a case study using the AI assistant that was developed using LAMB, in which they can consult it to carry out an analysis of the PESTLE (political, economic, social, technological, legal, and environmental) factors (Casañ et al., 2025) of these robots.

In late 2024, the "Safe AI in education manifesto" (Alier et al., 2024a, 2024b) was launched. This document sets out a series of practical principles for the design and implementation of AI technologies in education, including a checklist to evaluate these technologies and strategies (García-Peñalvo et al., 2024).

The principles for safe AI in education proposed in the manifesto are (Alier et al., 2024a, 2024b):

- **Human oversight and accountability:** AI must be a support tool for teachers and not replace their role. Key decisions must be under human control.
- **Guaranteeing confidentiality:** the privacy of students' data must be protected with strict security measures.
- **Alignment with educational strategies:** AI tools must adapt to the policy objectives of each educational institution.
- **Alignment with didactic practices:** AI must follow the pedagogical parameters established to guarantee its effectiveness in the classroom.
- **Accuracy and explainability:** systems must offer clear and precise information to avoid errors and confusion.
- **Comprehensive interface and behaviour:** AI should be easy to understand for users and openly communicate its limitations.
- **Ethical training and transparency:** AI models should be trained ethically, ensuring transparency in the use of data and methodologies.

These principles present an interesting framework of analysis for cases where AI is used in education, but they can be also extrapolated to other domains. Therefore, the decision was taken to use the manifesto to analyse the AI assistant that teachers and students had used on the same course in the anthropomorphic robots case study.

How the AI-based assistant, which acted as an expert consultant on questions relating to anthropomorphic robots, aligned with the principles of the manifesto was analysed in class. The teachers presented the tool's technological design, its integration into the university's digital strategy, and the design of its didactic application, providing a practical example of application of the theoretical concepts relating to the ethics of AI in the educational context (Casañ et al., 2024).

Contrasting with the case of the previous AI assistant, the case of an English school that decided to carry out a pilot scheme with secondary students to replace teachers with ChatGPT in 2024 was discussed with the students (Escobar, 2024). The discussion centred on whether this complied with the principles of the manifesto. Applying the manifesto and the checklist showed that this strategy did not comply with most of the principles established.

To progress beyond the discussion, the students worked in small teams of three people on a practical case relating to another field of application of GenAI. The case used is shown in Figure 1.

FIGURE 1. Wording of the case study.

AI-Powered hiring system in the ACME company

Background:

The ACME company, a large multinational, has recently introduced an advanced system powered by Artificial intelligence (AI) in its recruitment processes. This AI system is designed to analyse CVs, carry out initial filtering with chatbots, and even score candidates according to their responses and qualifications. The company believes that this system will streamline the hiring process, reduce human bias, and help select the most qualified candidates.

Questions:

1. Identify the most relevant ethical questions from the case.
2. What measures can the company take to address the ethical questions?
3. Analyse the economic viability of the proposals mentioned in question 2.
4. Should the company continue with this system of selecting people to hire? Base the answer on the answers to the previous questions.

The case of the recruiting company described in Figure 1 is considered through team work so that in small groups students can discuss the different ethical aspects relating to the case. To analyse this situation, students have all of the previously mentioned theoretical material: the theory set out in class, the "Safe AI in education manifesto", as well as various articles about the British school that decided to replace the teachers with ChatGPT.

The results of this case study suggest that this focus enables the abstract concepts of ethics and security of AI to become a reality for specific decisions about the application and design of technology. Therefore, the authors consider that this theoretical-practical approach can be effective for integrating ethical reflection into the teaching of engineering, underlining the need for multidisciplinary focuses to tackle the ethical challenges in education posed by AI.

4. Conclusions

This article offers a methodological proposal for approaching the study of the ethics of AI from a theoretical-practical focus, principally intended for students in fields of engineering related to ICT. As this field is relatively new and is constantly evolving, studying it poses significant challenges regarding defining relevant content and its application in real contexts.

To provide a solid basis, the theoretical framework of the ethics of ICT has been adopted. This means that the ethical challenges and dilemmas of AI can be analysed from an interdisciplinary perspective.

Finally, structured theoretical content is proposed, as is a strategy of using case histories in which GenAI tools developed specifically to be part of the learning process are both the foundation of the educational programme and an object for analysis of their level of compliance with the SAFE-AI principles based on the "Safe AI in education manifesto". In this way, we can conclude that the experience acquired through studying a specific use case in the field of education underpins the viability of this methodology for reconciling theoretical concepts with students' professional environment.

To approach with some guarantee of success the challenge of teaching a subject area like the ethics of AI that is rapidly changing, the teaching team must have a multidisciplinary body of knowledge. On the one hand, a grounding in knowledge of ethics and philosophy of science is important. On the other hand, knowledge of the short but intense history of information technologies and their impact on society is needed. Furthermore, there is a need for knowledge of the different legal frameworks that affect ICT in general and AI in particular (some of these laws will be passed as the course is in progress) as well as knowledge of the foundational literature about AI's potential and dangers, including how it is presented in works of fiction, such as *Frankenstein* (Shelley, 1831), *Terminator* (Cameron, 1984), or *Her* (Jonze, 2013). This last part is very relevant given how it appeals to students and for the media coverage through which the collective imagination has influenced social perceptions of the concept of AI.

Authors' contributions

Francisco-José García-Peñalvo: Conceptualisation; Methodology; Supervision; Writing (review and editing).

María-José Casañ-Guerrero: Research; Validation; Writing (original draft).

Marc Alier-Forment: Conceptualisation; Software; Writing (original draft).

Juan-Antonio Pereira-Varela: Supervision; Software; Writing (review and editing).

Artificial Intelligence (AI) Policy

The authors do not claim to have made use of Artificial Intelligence (AI) in the preparation of their articles.

Funding

This research is part funded by the Spanish Ministry of Science and Innovation through the AvisSA project, (reference PID2020- 118345RB-I00), the Department of Research and Universities of the Regional Government of Catalonia through the 2021 SGR 01412 support for research groups, and the Universidad del País Vasco/Euskal Herriko Unibertsitatea through the GIU21/037 contract within the "Call for Funding for Research Groups in the Universidad del País Vasco/Euskal Herriko Unibertsitatea (2021)" programme.

References

ABET - Engineering Accreditation Commission. (2021). *2022-2023 Criteria for Accrediting Engineering Programs*. https://d66z.short.gy/WvENCg.

Alier, M. (2024, October 24). *Beyond the palantír. Safe AI in education from ethics to tools* [Keynote Address]. Technological Ecosystems for Enhancing Multiculturality 2024 (TEEM 2024), Alicante, España.

Alier, M., Pereira, J., Garcia-Peñalvo, F. J., Casañ, M. J., & Cabré, J. (2025). LAMB: An open-source software framework to create artificial intelligence assistants deployed and integrated into learning management systems. *Computer Standards y Interfaces*, *92*, 103940. https://doi.org/10.1016/j.csi.2024.103940

Alier, M., García-Peñalvo, F. J., & Camba, J. D. (2024a). Generative artificial intelligence in education: From deceptive to disruptive. *International Journal of Interactive Multimedia and Artificial Intelligence*, *8*(5), 5-14. https://doi.org/10.9781/ijimai.2024.02.011

Alier, M., García-Peñalvo, F. J., Casañ, M. J., Pereira, J. A., & Llorens-Largo, F. (2024b). *Manifiesto para una IA Segura en la Educación (versión en español). Version 0.4.0 [Safe Ai in Education Manifesto]*. https://manifesto.safeaieducation.org/index_es.html

Alier, M., García-Peñalvo, F. J., Casañ, M. J., Pereira, J. A., & Llorens-Largo, F. (2024c). *Safe AI in Education Manifesto. Version 0.4.0*. https://manifesto.safeaieducation.org

Altman, S., Brockman, G., & Sutskever, I. (2023, 22 de mayo). *Governance of superintelligence*. OpenAI. https://openai.com/index/governance-of-superintelligence/

Anurag, Vyas, N., & Lilhore, U. K. (2023). Transforming work: The impact of artificial intelligence (AI) on modern workplace. In N. Chaudhary (Ed.), *Proceedings of the 2023 3rd International Conference on Technological Advancements in Computational Sciences (ICTACS)* (pp. 602-607). IEEE. https://doi.org/10.1109/ICTACS59847.2023.10390258

Bellogín, A., Grau, O., Larsson, S., Schimpf, G., Sengupta, B., & Solmaz, G. (2024). The EU AI act and the wager on trustworthy AI. *Communications of the ACM*, *67*(12), 58-65. https://doi.org/10.1145/3665322

Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford University Press.

Bowden, P. (2010). Teaching ethics to engineers - A research-based perspective. *European Journal of Engineering Education*, *35*(5), 563-572. https://doi.org/10.1080/03043797.2010.497549

Bynum, T. W., Maner, W., & Fodor, J. L. (Eds.). (1992). *Teaching computer ethics*. Southern Connecticut State University.

Cameron, J. (Director) (1984). *The Terminator* [Film]. Orion Pictures.

Casañ, M. J., & Alier, M. (2024). Case studies about moral dilemas to apply ethical theories in engineering education. *IEEE Revista Iberoamericana de Tecnologias del Aprendizaje*, *19*, 1-6. https://doi.org/10.1109/RITA.2024.3368609

Casañ, M. J., Alier, M., & Llorens, A. (2020). Teaching ethics and sustainability to informatics engineering students, an almost 30 years' experience. *Sustainability*, *12*(14), 5499. https://doi.org/10.3390/su12145499

Casañ, M. J., Alier, M., Pereira, J., & García-Peñalvo, F. J. (2024). Asistentes de aprendizaje basados en inteligencia artificial: principios de seguridad y experiencias de implementación en educación superior [Artificial intelligence-based learning assistants: Security principles and implementation experiences in higher education]. In M. Navarro, J. J. Sánchez, P. Berbel, & C. Rodríguez-Jiménez (Eds.), *Investigación y conocimientos en la educación actual [Research and knowledge in education today]* (pp. 13-35). Dykinson.

Casañ, M. J., Llorens, A., Alier, M., & Pereira, J. (2025). Using an AI based learning assistant for a PESTLE case study learning activity. In *Proceedings of the 12th International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM) 2024*. Springer.

Dal Bó, E. (2006). Regulatory capture: A review. *Oxford Review of Economic Policy*, *22*(2), 203-225. https://doi.org/10.1093/oxrep/grj013

Diamandis, P. H., & Kotler, S. (2012). *Abundance: The future is better than you think*. The Free Press.

Elmoudden, S., & Wrench, J. S. (Eds.). (2024). *The role of generative AI in the communication classroom*. IGI Global. https://doi.org/10.4018/979-8-3693-0831-8

Epstein, Z., Hertzmann, A., Akten, M., Farid, H., Fjeld, J., Frank, M. R., Groh, M., Herman, L., Leach, N., Mahari, R., Pentland, A., Russakovsky, O., Schroeder, H., & Smith, A. (2023). Art and the science of generative AI. *Science, 380*(6650), 1110-1111. https://doi.org/10.1126/science.adh4451

Escobar, D. (2024, September 10). Reemplazan profesores por inteligencia artificial: la educación de los niños con realidad virtual y mucho más [Replacing teachers with artificial intelligence: Educating children with virtual reality and more]. *Infobae*. https://d66z.short.gy/vvFgTW

European Parliament, & Council of the European Union. (2016). *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)*. https://bit.ly/2O2juE9

Flores-Vivar, J. M., & García-Peñalvo, F. J. (2023). Reflections on the ethics, potential, and challenges of artificial intelligence in the framework of quality education (SDG4). *Comunicar*, *31*(74), 35-44. https://doi.org/10.3916/C74-2023-03

García-Peñalvo, F. J., & Vázquez-Ingelmo, A. (2023). What do we mean by GenAI? A systematic mapping of the evolution, trends, and techniques involved in generative AI. *International Journal of Interactive Multimedia and Artificial Intelligence*, *8*(4), 7-16. https://doi.org/10.9781/ijimai.2023.07.006

García-Peñalvo, F. J., Llorens-Largo, F., & Vidal, J. (2024). The new reality of education in the face of advances in generative artificial intelligence. *RIED: revista iberoamericana de educación a distancia*, *27*(1), 9-39. https://doi.org/10.5944/ried.27.1.37716

González-Geraldo, J. L., & Ortega-López, L. (2024). Can AI fool us? University students' lack of ability to detect ChatGPT. *Education in the Knowledge Society*, *25*, e31760. https://doi.org/10.14201/eks.31760

Gotterbarn, D., Miller, K., & Rogerson, S. (1997). Software engineering code of ethics. *Communications of the ACM*, *40*(11), 110-118. https://doi.org/10.1145/265684.265699

Harari, Y. N. (2015). *Homo Deus: A brief history of tomorrow*. Harvill Secker.

Harari, Y. N. (2024). *Nexus: A brief history of information networks from the Stone Age to AI*. Random House.

Henriksen, D., Oster, N., Mishra, P., & McCaleb, L. (2025). Generative AI, creativity, culture, and the future of learning: A conversation with Mairéad Pratschke. *TechTrends*, *69*(1), 3-9. https://doi.org/10.1007/s11528-024-01036-y

Hoyos, D. (2011). Revisiones de la ética de la virtud [Some revisions of virtue ethics]. *Estudios de Filosofía*, (44), 61-75.

Johnson, D. G. (2009). *Computer ethics: Analyzing information technology* (4th ed.). Prentice Hall.

Johnson, D. G. (2017). Can engineering ethics be taught? *The Bridge*, *47*, 59-64.

Jonze, S. (director) (2013). *Her* [Film]. Warner Bros. Pictures.

Jovanović, M., & Campbell, M. (2022). Generative artificial intelligence: Trends and prospects. *Computer*, *55*(10), 107-112. https://doi.org/10.1109/MC.2022.3192720

Kanbach, D. K., Heiduk, L., Blueher, G., Schreiter, M., & Lahmann, A. (2024). The GenAI is out of the bottle: Generative artificial intelligence from a business model innovation perspective. *Review of Managerial Science*, *18*(4), 1189-1220. https://doi.org/10.1007/s11846-023-00696-z

Kar, A. K., Choudhary, S. K., & Singh, V. K. (2022). How can artificial intelligence impact sustainability: A systematic literature review. *Journal of Cleaner Production*, *376*, 134120. https://doi.org/10.1016/j.jclepro.2022.134120

Kumar, A. N., Raj, R. K., Aly, S. G., Anderson, M. D., Becker, B. A., Blumenthal, R. L., Eaton, E., Epstein, S. L., Goldweber, M., Jalote, P., Lea, D., Oudshoorn, M., Pias, M., Reiser, S., Servin, C., Simha, R., Winters, T., & Xiang, Q. (2024). *Computer Science Curricula 2023*. ACM Press, IEEE Computer Society Press and AAAI Press. https://doi.org/10.1145/3664191

Kurzweil, R. (2005). *The singularity is near: When humans transcend biology*. Viking Penguin.

Leal, J. G. (1982). La teoría del contrato social: Spinoza frente a Hobbes [The theory of the social contract: Spinoza vs. Hobbes]. *Revista de estudios políticos*, (28), 125-194.

Maner, W. (1980). *Starter kit in computer ethics*. Helvetia Press; National Information and Resource Center for Teaching Philosophy.

Martínez-Arboleda, A. (2024). Language sustainability in the age of artificial intelligence. *Alfinge, 36*(36), 1-37. https://doi.org/10.21071/arf.v36i.17761

Meyer, B. (2011, October 28). John McCarthy. *BLOG@CACM*. https://d66z.short.gy/P91Wo6

Miñano, R., Uribe, D., Moreno-Romero, A., & Yáñez, S. (2019). Embedding sustainability competences into engineering education. The case of informatics engineering and

industrial engineering degree programs at Spanish universities. *Sustainability*, *11*(20), 5832. https://doi.org/10.3390/su11205832

Moore, G. (1965). Cramming more components onto integrated circuits. *Electronics Magazine*, *38*(8), 114-117.

Morrison, R. (2024, November 12). Sam Altman claims AGI is coming in 2025 and machines will be able to *think like humans* when it happens. *Tom's guide*. https://d66z.short.gy/dJBnLv

Parra, Y. J., Theran, C., & Aló, R. (2024). Ethical considerations of generative AI: A survey exploring the role of decision makers in the loop. *Proceedings of the AAAI Symposium Series*, *3*(1), 391-398. https://doi.org/10.1609/aaaiss.v3i1.31243

Prieto, E. (2016). *La ley del reloj. Arquitectura, máquinas y cultura moderna [The law of the clock. Architecture, machines and modern culture]*. Ediciones Cátedra.

Quote Investigator. (2010, December 7). *Quote origin: To err is human; to really foul things up requires a computer*. Quote Investigator. https://d66z.short.gy/wYawuu

Quote Investigator. (2024, June 20). *Quote origin: As soon as it works, no one calls it AI anymore*. Quote Investigator. https://d66z.short.gy/dTNx8W

Saltelli, A., Dankel, D. J., Di Fiore, M., Holland, N., & Pigeon, M. (2022). Science, the endless frontier of regulatory capture *Futures*, *135*, 102860. https://doi.org/10.1016/j.futures.2021.102860

Samsi, S., Zhao, D., McDonald, J., Li, B., Michaleas, A., Jones, M., Bergeron, W., Kepner, J., Tiwari, D., & Gadepally, V. (2023). From words to watts: Benchmarking the energy costs of large language model inference. In *2023 IEEE High Performance Extreme Computing Conference (HPEC)*. IEEE. https://doi.org/10.1109/HPEC58863.2023.10363447

Sevilla, S. (2014). La actualidad de la ética kantiana como metacrítica [The relevance of Kantian ethics as meta-criticism]. In J. J. García, R. Rodriguez, & M. J. Callejo (Eds.), *De la libertad del mundo: homenaje a Juan Manuel Navarro Cordón [From the freedom of the world: Homage to Juan Manuel Navarro Cordón]* (pp. 375-391). Escolar y Mayo.

Shelley, M. (1831). *Frankenstein, or the modern Prometheus*. H. Colburn and R. Bentley.

Smith, A., & Yandle, B. (2014). *Bootleggers and baptists: How economic forces and moral persuasion interact to shape regulatory politics*. Cato Institute.

Spiekermann, S. (2015). *Ethical IT innovation: A value-based system design approach*. Auerbach Publications.

Tolkien, J. R. R. (1954). *The two towers*. George Allen y Unwin.

Wei, K., Ezell, C., Gabrieli, N., & Deshpande, C. (2024). How do AI companies "fine-tune" policy? Examining regulatory capture in AI governance. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, *7*(1), 1539-1555. https://doi.org/10.1609/aies.v7i1.31745

Wiener, N. (1950). *The human use of human beings. Cybernetics and societ.* Houghton Mifflin.

Yandle, B. (1983). Bootleggers and Baptists: The education of a regulatory economist. *Regulation*, *7*(3), 12-16.

Zavala, N. L. (1999). *Tecnología y ética [Technology and ethics]*. Universidad Nacional Autónoma de México.

Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z., Du, Y., Yang, C., Chen, Y., Chen, Z., Jiang, J., Ren, R., Li, Y., Tang, X., Liu, Z., Liu, P., Nie, J.-Y., & Wen, J.-R. (2024). A survey of large language models. *arXiv*, Article arXiv:2303.18223v15. https://doi.org/10.48550/arXiv.2303.18223

Authors' biographies

Francisco-José García-Peñalvo. Doctorate in Computing from the Universidad de Salamanca. He is a university professor in the Department of Computing and Automation at the Universidad de Salamanca (USAL), with four recognised six-year research blocks, one six-year transfer period, and five-year teaching blocks. He was awarded the Gloria Begué

prize for teaching excellence in 2019 and the María de Maeztu prize for research excellence in 2023. Since 2006 he has been director of the GRIAL (Interaction and eLearning research group), a USAL-recognised research group that is a consolidated research unit of the regional government of Castilla y León (UIC 81). He is currently the deputy head of the University Institute of Educational Sciences (IUCE) and the coordinator of the Education in the Knowledge Society Doctoral Programme at the Universidad de Salamanca.

https://orcid.org/0000-0001-9987-5584

María-José Casañ-Guerrero. Doctor of Sciences (2013) from the Universidad Politécnica de Cataluña (UPC) (2013), she also has a degree in Computer Engineering from the same university (1997). Since 2004 she has worked as a researcher and teacher, giving classes in the UPC's Faculty of Computing. She has also been an instructor on courses with the Universidad Abierta de Cataluña (UOC). She has one recognised six-year research block and four five-year teaching blocks (two of them with a special mention for teaching quality). She currently teaches modules relating to software engineering projects, databases, social and environmental aspects of computing, and history of computing.

https://orcid.org/0000-0002-5072-6745

Marc Alier-Forment. Associate professor at the Universidad Politécnica de Cataluña (UPC) since 2002. He has a doctorate in Sustainability and a degree in Computer Engineering from the UPC, and has two recognised six-year research blocks as well as four recognised five-year teaching blocks (three of them with a special mention for teaching quality). His areas of work include computing, information systems, online education, and the ethics of technology. He has participated actively in the Moodle community, collaborating in the creation of modules in the form of wikis, online service layers, and implementing the IMS LTI standard. With more than 25 years' experience in research and development of educational software, he has published more than 190 scientific articles and is a member of the EduSTEAM research group. He is currently head of the Engineering and Technology Education doctoral programme at the UPC and teaches various modules in the Faculty of Computing in Barcelona. Outside the academic sphere, he is a guitar luthier and has produced various podcasts since 2007.

https://orcid.org/0000-0003-3922-1516

Juan-Antonio Pereira-Varela. Associate professor in the Computing Faculty of the Universidad del País Vasco (UPV/EHU), where he has been teaching for 20 years. He has a doctorate in Software Engineering (2014) and his research centres on generative AI applied to software engineering, as well as on studying the development of open-source software in this area. He has published numerous articles in both areas, participating in various national research projects and a European project. He has written a book on HTML5 and JavaScript APIs (2021). He was recognised for excellence in teaching based on the student evaluations in four consecutive years (2016–2020). He has been joint head of the ZIUR cybersecurity classroom, and has trained various teams for the SWERC programming competition. He is currently the lead developer on two projects relating to generative AI: RepoSearch, a semantic search engine of computing final degree project dissertations at a national level and LAMB (learning assistant manager and builder), an open-code project to create AI assistants to help learning.

https://orcid.org/0000-0002-7935-3612