Critical Citizens Regarding the Use of Artificial Intelligence:

An Approach from Social Engineering

Fernando Broncano[1], Aurora Cuartero[2], Mar Ávila[3], Andrés Caro[3], and Pablo G. Rodríguez[3]

[1] Department of Computer and Telematics Systems Engineering, School of Industrial Engineering, University of Extremadura, Badajoz, Spain
[2] Department of Graphic Design, Polytechnic School, University of Extremadura, Cáceres, Spain
[3] Department of Computer and Telematics Systems Engineering, School of Industrial Engineering, University of Extremadura, Cáceres, Spain
fbroncano@unex.es

Abstract. The impact of artificial intelligence (AI) on society must be examined from a critical and multidimensional perspective, addressing both pedagogical advances and the associated social, ethical, and environmental risks. Through a literature review and the analysis of relevant cases such as Cambridge Analytica, it becomes evident that the exponential growth of data and the development of new technologies have enabled unprecedented processes of control, surveillance, and social manipulation, grouped under the concept of social engineering. This study argues for the need to strengthen digital literacy and critical thinking among future professionals, integrating key competences to defend against digital disinformation, algorithmic manipulation, and cyberattacks that threaten democratic processes. It also presents didactic proposals aimed at fostering ethical training, environmental awareness, academic reading, the search and validation of new information sources, as well as encouraging slower and more reflective digital consumption habits. The coexistence of intelligent systems and citizens requires a population capable of understanding and questioning new forms of digital control and manipulation exerted by major digital influencers and multinational organizations.

Keywords: Artificial intelligence in education, Pedagogical innovation, Teaching innovation group, Inclusive and sustainable teaching, Emerging technologies.

1 Introduction

Computer science can be defined as the discipline concerned with the processing of information by automatic machines (Martí et al., 2008). This definition highlights its two fundamental pillars: information and its automated processing. It is also necessary

K. Becker et al. (eds.), *Proceedings of the Conference Growing up the Future: Education for Sustainability (GROUFES 2026)*, Atlantis Highlights in Social Sciences, Education and Humanities 54,
https://doi.org/10.2991/978-94-6239-760-6_13

to distinguish between information and data. Data represent raw facts, symbolic representations without inherent meaning, while information results from processing, organizing, and contextualizing data to generate a meaningful, purposeful, and useful message that supports decision-making and situational understanding.

Historically, the importance of computer science and automatic processing machines can be traced back to the two World Wars, when such machines were used to decipher messages (or raw data) transforming them into comprehensible and useful information that enabled concrete decision-making. Today, these transformations have become more efficient due to the integration of diverse information and communication devices into everyday life, such as mobile phones and personal computers.

The widespread introduction of consumer computing devices into daily life has led to a surge in data generation and an increased demand for computational capacity to distinguish between data that yield valuable information and those that must be discarded to avoid contaminating generated information. This exponential growth of data, produced daily by millions of devices, has given rise to the phenomenon known as *Big Data*. This concept refers to the management, storage, and processing of massive data volumes, characterized by their variety, velocity, and veracity. In the current digital ecosystem, the management of *Big Data* enables the discovery of patterns, the anticipation of behaviors, and the optimization of processes in diverse fields such as healthcare, education, economics, and public administration. It also extends to other areas, including the geolocation of devices through geodata and satellite information that allow climate and environmental prediction and observation. However, the abundance of data does not necessarily imply a deeper understanding of reality but rather introduces new risks associated with disinformation, algorithmic bias, and manipulation (Rumiche Chávarry et al., 2023).

In the data ecosystem, digital literacy emerges as an essential competence for critical citizenship, as it equips individuals with the tools required to interpret, analyze, and evaluate information ethically and responsibly. Promoting digital literacy entails not only improving access to technology but also fostering the capacity to question, verify, and contextualize received information, thereby safeguarding the integrity of democratic processes and the quality of public debate (Castañeda et al., 2024).

Against this backdrop, the present contribution focuses on analyzing educational environments, emphasizing the role of Teaching Innovation Groups in developing strategies and methodologies that strengthen a critical and informed citizenry, prepared to face the challenges and opportunities posed by artificial intelligence in the data-driven society.

2 Emerging technologies and socio-environmental footprint

The revolution of artificial intelligence (AI) has substantially transformed data processing and process automation across multiple sectors, including higher education. The ability of machine learning and generative models to personalize experiences, generate content, and optimize resources has opened new pedagogical opportunities. However, the impact of these technologies must be examined not only from the standpoint of innovation, but also in terms of their environmental and cybersecurity implications.

From a sustainability perspective, large-scale AI systems demand enormous computational capacity, which translates into a significant increase in energy consumption and the carbon footprint associated with their training and execution. Data centers, essential to sustain these infrastructures, consume growing amounts of electricity and water, creating tensions in terms of environmental sustainability and resource efficiency (Ramírez Morán, 2025). Training advanced models, such as those in generative AI, can equal the annual energy consumption of dozens of households and emit tens of tons of CO_2, not to mention the rising levels of electronic waste caused by the accelerated obsolescence of hardware required to support these technologies. It has been estimated that training GPT-3 consumed approximately 1,287 MWh of energy (the annual equivalent of over 350 households in Spain) and generated more than 500 tons of CO_2 equivalent. Similarly, training PaLM model, developed by Google, incurred an even higher energy cost, close to 2,500 MWh, with an estimated carbon footprint of over 1,000 tons of CO2 equivalent (Patterson et al., 2021; Strubell et al., 2019).

This scenario directly connects with the Sustainable Development Goals (SDGs), particularly SDG 7 (Affordable and Clean Energy), SDG 9 (Industry, Innovation, and Infrastructure), and SDG 12 (Responsible Consumption and Production), which promote a transition toward cleaner, more inclusive, and more efficient production models. It is therefore essential to integrate sustainability criteria into the adoption of these technologies, evaluating both the origin of the energy used and the efficiency of digital service providers, while encouraging the deployment of infrastructures that mitigate their environmental impact. In this sense, the integration of AI into academia is not only a technological challenge but also an opportunity to advance global commitments toward a more sustainable and equitable society.

On the cybersecurity front, the adoption of emerging technologies entails significant risks for citizens. The population is exposed daily to attacks such as *phishing*, *vishing*, and *smishing*, designed to obtain credentials, personal information, or access to sensitive systems, often leading to identity theft and serious economic and social consequences. With the rise of AI, these attacks have become increasingly sophisticated: generative models enable the creation of more convincing phishing content, the production of synthetic audio or video (*deepfakes*), and the automated scaling of campaigns, thereby increasing their reach and effectiveness (Carley, 2020).

Nonetheless, the main cybersecurity challenge lies in its social dimension. So-called *social cybersecurity* represents the weakest link in organizational protection, since the human factor is especially prone to error. For instance, opening a malicious email can facilitate the installation of *ransomware*, severely compromising the integrity of organizational information systems. Moreover, in many cases, affected users hesitate to report incidents promptly to IT teams due to fear of sanctions or personal exposure, delaying response and aggravating the consequences of the attack. This reality underscores the need to foster a culture of transparency and trust within organizations, where social cybersecurity is prioritized by expanding training for members to strengthen protection against threats and attacks.

This scenario highlights the need for a critical and responsible approach to the design, use, and evaluation of AI-based technologies, addressing in a cross-cutting manner their potential social and ecological impacts, while emphasizing the importance of regulation, transparency, and the ethical training of those who develop and use them (Fernandez Nieto, 2024). A comprehensive and sustainable vision of AI requires the active participation of all social and educational actors, along with the establishment of public policies aimed at minimizing its environmental footprint and maximizing its social benefits[1]. One of the public policies implemented by the Government of Spain, through the National Cybersecurity Institute (INCIBE) and with European Recovery Funds (Recovery and Resilience Facility – Next Generation[2]), involved the creation of university-sponsored chairs aimed at promoting and disseminating technological challenges in cybersecurity and artificial intelligence to society and higher education students[3].

3 Cyberattacks on democracy and the case of *Cambridge Analytica*

Contemporary information systems have enabled new forms of population control and surveillance through the systematic analysis of everyday data captured by consumer devices (mobile phones, smart appliances, navigation systems, etc.) and processed to generate predictive inferences about individual behaviors, preferences, and decisions. This continuous monitoring capacity allows large organizations, states, and technology corporations to obtain detailed citizen profiles without their explicit knowledge, reshaping the traditional democratic ecosystem.

In this context, cyberattacks on democracy can be defined as the systematic disruption of political power acquisition processes through the interference of information systems and new technologies aimed at manipulating public opinion, distorting democratic debate, and illegitimately influencing electoral outcomes. These attacks exploit both technological and human vulnerabilities, employing techniques of social engineering and mass disinformation to erode the trust of citizens in democratic institutions (Hernández Peña, 2022).

3.1 The Cambridge Analytica case: architecture of electoral manipulation

The Cambridge Analytica–Facebook scandal (Greenfield, 2018) represents one of the most paradigmatic cases of cyberattacks against modern democracy. Between 2014 and 2015, the British consulting firm Cambridge Analytica harvested personal data from up

[1] Government of Spain, "Digital agenda – data economy and artificial intelligence" (2021), available at https://avance.digital.gob.es/es-es/Documents/210330-Agenda-Digital-9-Economia-dato-inteligencia-artificial.pdf.

[2] https://next-generation-eu.europa.eu/recovery-and-resilience-facility_es.

[3] The authors of this contribution are members of the INCIBE-UEx-EPCC sponsored chair (https://ciberseguridad.unex.es)

to 87 million Facebook users through a psychological application developed by Aleksandr Kogan, a researcher at the University of Cambridge. The application, called *thisismydigitallife*, was presented as an academic personality quiz but operated as a massive data-extraction mechanism that gathered information not only from voluntary respondents but also from their networks of contacts (González Montaño, 2024).

Cambridge Analytica used this information to design sophisticated psychographic campaigns, combining demographic data with psychological profiles to segment voters and deliver hyper-personalized messages tailored to influence electoral behavior. The company claimed to hold 5,000 specific data points for each voter in the United States, enabling the prediction and manipulation of voting patterns with unprecedented accuracy. These techniques were deployed during the 2016 presidential election in the United States, in support of the campaign of Donald Trump, and during the Brexit referendum in the United Kingdom.

The significance of this case lies in demonstrating how digital platforms, under the guise of free services, can become vehicles of large-scale democratic manipulation, allowing private actors with sufficient resources to alter electoral processes through the unauthorized use of personal data and algorithmic persuasion techniques.

3.2 Political influencers and new forms of digital populism

In the current context, various political actors, supported by digital influencers, have successfully entered European electoral processes by taking advantage of social media as primary channels of political communication (Boulianne, 2015). In the 2024 European elections, the case of Alvise Pérez in Spain illustrates this trend: his group “Se Acabó la Fiesta” obtained nearly 800,000 votes (4.6% of the total) and three European parliamentary seats within the Spanish delegation, becoming the sixth political force in the country without a formal electoral program, physical headquarters, or organized membership.

In parallel, in Cyprus, the YouTuber Fidias Panayiotou secured 19.4% of the vote and one European parliamentary seat as an independent candidate, receiving strong support particularly among voters aged 18–34. These phenomena reflect a common strategy: the intensive use of platforms such as Telegram, TikTok, and YouTube to build polarized digital communities, disseminate anti-establishment content, and monetize social discontent.

The impact of social networks on youth is particularly evident in the proliferation of extremist groups on Telegram that promote hate crimes and xenophobic discourse (Williams et al., 2019). The case of Torre Pacheco (Reina, 2025), which took place in a municipality of Murcia in 2025, illustrates this problem. After a local incident, more than 138,000 hate messages were detected against people of Maghrebi origin, coordinated through Telegram channels that encouraged “hunts” and violence against migrants. The Civil Guard identified organized patterns of hate speech dissemination, resulting in arrests and the closure of channels promoting mass deportations and xenophobic violence.

3.3 Social engineering and social cyberattacks

Social engineering is academically defined as a set of psychological manipulation techniques that exploit human vulnerabilities to obtain confidential information, gain access to systems, or elicit specific behaviors from victims (Conteh & Schmick, 2016). In the context of cyberattacks against democracy, these techniques have become considerably more sophisticated through the use of generative AI, enabling the creation of hyper-realistic content (ranging from deepfakes to personalized messages) that is almost impossible to detect for users without specialized knowledge.

Social cyberattacks represent the evolution of traditional social engineering into coordinated campaigns that combine disinformation, emotional manipulation, and algorithmic exploitation to exert mass influence over democratic processes. These attacks operate on multiple levels, beginning with the creation of false narratives, followed by the artificial amplification of polarizing content, and extending to the psychographic segmentation of audiences and the personalization of persuasive messages.

The justification for labeling these phenomena as *social cyberattacks* lies in their systematic nature, their destabilizing intent, and their capacity to erode social cohesion and trust in democratic institutions. Unlike traditional cyberattacks directed at technological infrastructures, social cyberattacks primarily target the manipulation of popular will and the alteration of public opinion formation, representing a direct threat to democratic sovereignty.

This evolution of the threat landscape requires comprehensive educational responses that strengthen the critical competences of citizens and promote a culture of verification and digital responsibility, particularly within university environments where future social leaders and professionals are trained.

4 Didactic proposals from teaching innovation

According to the analyzed context, a set of recommendations can be proposed to ensure that students acquire knowledge about the current functioning of information systems (Domínguez Álvarez, 2025). The first key aspect is the incorporation of ethical, environmental, and social perspectives into formal education, encouraging students to advocate for the protection of human rights. A second line of action emphasizes training aimed at fostering awareness of digital disinformation and issues driven by self-interested agendas. The promotion of academic reading, together with the improvement of reading comprehension as a means of strengthening qualifications, constitutes a third proposal to counter manipulated and misleading information. The fourth proposal focuses on encouraging reflection on automated algorithmic behavior, which shapes the positioning of information presented to the general public and constitutes the seed of social engineering. Finally, a general recommendation is to slow down the pace of life and reconsider the culture of massive digital consumption.

4.1 Ethical, environmental, and social emphasis in education

One of the most significant challenges for subjects related to information management lies in integrating ethical, environmental, and social reflection across the curriculum. It is essential that students understand the implications of data processing and its impact on individual privacy, the environment, and society as a whole. Initiatives carried out by Teaching Innovation Groups should be committed to the incorporation of project-based methodologies and remote laboratories to address these challenges. Raising awareness about the environmental footprint of digital technologies (such as energy consumption or device obsolescence) must form part of academic training, promoting responsible and sustainable use of technological resources, as well as incorporating the principles of circular economy to enable the reuse of obsolete systems.

4.2 Awareness of hoaxes and digital disinformation

In the face of the proliferation of hoaxes and disinformation practices, it is a priority to train students in critical reading, source verification, and the identification of false narratives. Problem-based learning and the analysis of case studies allow students to contrast information, reflect on data integrity, and learn to avoid the dissemination of potentially harmful content. Tutorial activities and active participation in research projects foster digital competence and strengthen a culture of verification within the university context.

4.3 Promotion of Academic Reading and Recovery of Interest in Higher Education

The progressive decline in male enrollment in higher education (Zafra, 2025) underscores the need for intervention strategies aimed at strengthening motivation, providing role models, and revitalizing the practice of didactic and academic reading. It is essential to convey to students the importance of reading as a source of inspiration, research, and knowledge acquisition. Reading broadens knowledge by offering diverse perspectives and supporting the development of critical reflection and personal viewpoints. Such activities must guarantee sustained and sufficient efforts to ensure that students continue engaging in reading and maintain a genuine willingness to learn, thereby enabling them to develop independent opinions within a context increasingly shaped by manipulation and algorithmic bias.

4.4 Pauses in Life Rhythm and Rethinking Digital Consumption

In light of the constant acceleration of life rhythms and the overwhelming presence of digital stimuli, it is advisable to incorporate structured pauses and reflective spaces into daily routines, promoted and supported within the classroom. Fostering slow thinking, allocating time for academic reflection, and encouraging knowledge transfer through written assignments, case analyses, and argumentation exercises contribute to enhancing both the quality of learning and the mental well-being of students. This approach enables temporary disengagement from rapid digital consumption and reinforces the capacity for critical analysis when confronted with the immediacy of digital content.

4.5 Unmasking Automated Content and Digital Social Engineering

Raising awareness among students regarding the origin and algorithmic nature of online content is of critical importance. The concept commonly referred to as the *dead internet theory* (Muzumdar et al., 2025; Walter, 2025) highlights that a considerable proportion of messages and entries indexed by search engines or disseminated through social media are generated or amplified by automated bots serving commercial and political purposes. In this context, formal education should provide insights into algorithmic positioning mechanisms and the functioning of social networks, enabling students to recognize that these platforms prioritize the delivery of content designed to maximize user engagement, as a result of digital social engineering. The analysis of patterns and the search for supplementary information on a given topic facilitate the detection of biases, manipulations, and vulnerabilities within digital platforms, thereby strengthening the ability of students to develop critical self-defense strategies against such phenomena.

The implementation of these teaching proposals led to the establishment of a teaching innovation group called the Innovation Group on Socio-Environmental Methodologies and Inclusive Engineering at the University of Extremadura, within the framework of the EUGREEN European Universities Alliance. The objective of this group is to transfer knowledge to the regions where the member universities of the EUGREEN Alliance are located, providing information on technological challenges, their socio-environmental footprint, and teaching proposals designed to foster a critical and engaged society, as well as to enhance protection against disinformation.

5 Conclusions

Emerging technologies, including artificial intelligence, undoubtedly provide significant advances for the improvement of educational and social processes. However, their potential benefits should not overshadow the necessity of cultivating a citizenry capable of recognizing inherent threats, particularly those that enable interference in domains as sensitive as democracy, ethics, and the environment.

The proposed measures are not intended to constitute a universal solution but rather aim to encourage a critical and reflective coexistence between citizens and technology. The objective is to enable society to coexist with intelligent systems from a perspective rooted in critical reasoning, informed reading, and up-to-date knowledge of new techniques of digital control and manipulation, thereby avoiding naïve positions regarding technological progress. Only through the integration of these competencies will it be possible to respond effectively to contemporary challenges and to preserve both the quality of public debate and the integrity of social processes in contemporary society.

Acknowledgements

This work has been carried out by the teaching innovation group “Grupo de Innovación en Metodologías Socioambientales e Ingeniería Inclusiva” (GIMII) at the Universidad de Extremadura, which has received a grant for groups under the 2025/2026 call for proposals.

References

1. Boulianne, S. (2015). Social media use and participation: A meta-analysis of current research. Information, Communication & Society, 18, 524-538. https://doi.org/10.1080/1369118X.2015.1008542
2. Carley, K. M. (2020). Social cybersecurity: An emerging science. Computational and Mathematical Organization Theory, 26(4), 365–381. https://doi.org/10.1007/s10588-020-09322-9
3. Castañeda, L., Haba-Ortuño, I., Villar-Onrubia, D., Marín, V. I., Tur, G., Ruipérez-Valiente, J. A., & Wasson, B. (2024). Desarrollando el marco DALI de alfabetización en datos para la ciudadanía. RIED-Revista Iberoamericana de Educación a Distancia, 27(1), 289–318. https://doi.org/10.5944/ried.27.1.37773
4. Conteh, N., & Schmick, P. (2016). Cybersecurity: Risks, vulnerabilities and countermeasures to prevent social engineering attacks. International Journal of Advanced Computer Research, 6, 31–38. https://doi.org/10.19101/IJACR.2016.623006
5. Domínguez Álvarez, J. L. (2025). La innovación docente como herramienta indispensable para fomentar la adquisición de competencias digitales y el establecimiento de una auténtica cultura de privacidad y ciberseguridad en la sociedad española. In Nuevos desafíos de la educación superior: Transformando la docencia universitaria. Colex. https://dialnet.unirioja.es/servlet/articulo?codigo=10368000
6. Fernandez Nieto, B. (2024). Toward a Human-Centered Framework for Trustworthy, Safe and Ethical Generative Artificial Intelligence: A Multi-Level Analysis of Large Language Models Social Impact. Proceedings of the 28th International Conference on Evaluation and Assessment in Software Engineering, EASE ’24, 505-509. https://doi.org/10.1145/3661167.3661177
7. González Montaño, A. (2024). When machines dream for us. Democracy in the world of algorithmic capitalism: Case study of Cambridge Analytica. Revista Científica Del Centro de Investigación Para La Comunicación Aplicada. https://dialnet.unirioja.es/servlet/articulo?codigo=9974028
8. Greenfield, P. (2018). The Cambridge Analytica files: The story so far. The Guardian. https://www.theguardian.com/news/2018/mar/26/the-cambridge-analytica-files-the-story-so-far
9. Hernández Peña, J. C. (2022). Electoral campaigns, Big Data and ideological profiling. Addressing their problematic from the perspective of the fundamental right to data protection. Revista Española de Derecho Constitucional. https://dialnet.unirioja.es/servlet/articulo?codigo=8416119
10. Martí, M. C., D’Agostino, M. J., Cabo, J. V. de, & Sanz-Valero, J. (2008). Alfabetización Digital: Un peldaño hacia la sociedad de la información. Medicina y Seguridad Del Trabajo, 54, 11-15.

11. Muzumdar, P., Cheemalapati, S., Ramireddy, S., Singh, K., Kurian, G., & Muley, A. (2025). The Dead Internet Theory: A Survey on Artificial Interactions and the Future of Social Media. Asian Journal of Research in Computer Science, 18. https://doi.org/10.9734/ajrcos/2025/v18i1549
12. Patterson, D., Gonzalez, J., Le, Q., Liang, C., Munguia, L.-M., Rothchild, D., So, D., Texier, M., & Dean, J. (2021). Carbon Emissions and Large Neural Network Training. https://arxiv.org/abs/2104.10350
13. Ramírez Morán, D. (2025). La huella medioambiental de la IA. Boletín IEEE (Instituto Español de Estudios Estratégicos).
14. Reina, E. (2025). Torre Pacheco, la tierra donde el melón se comió a la ultraderecha. El País. https://uex.be/5zo1y
15. Rumiche Chávarry, R. del P., Matas Terrón, A., Ríos Ariza, J. M., & Chunga Chinguel, G. R. (2023). Implicaciones sociales de la irrupción del Big Data y la robótica: Un análisis prospectivo en docentes hispanoamericanos. Revista Electrónica Interuniversitaria de Formación Del Profesorado, 26(1), 115–127. https://doi.org/10.6018/reifop.543871
16. Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and Policy Considerations for Deep Learning in NLP. In A. Korhonen, D. Traum, & L. Màrquez (Eds), Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (pp. 3645–3650). Association for Computational Linguistics. https://doi.org/10.18653/v1/P19-1355
17. Walter, Y. (2025). Artificial influencers and the dead internet theory. AI & SOCIETY. https://doi.org/10.1007/s00146-023-01857-0
18. Williams, M., Burnap, P., Javed, A., Liu, H., & Ozalp, S. (2019). Hate in the Machine: Anti-Black and Anti-Muslim Social Media Posts as Predictors of Offline Racially and Religiously Aggravated Crime. British Journal of Criminology. https://doi.org/10.1093/bjc/azz049
19. Zafra, I. (2025). Las chicas sacan cada vez más ventaja a los chicos en educación: "Para algunos, estudiar resta masculinidad". El País. https://uex.be/3ukqp