EXPLAINABLE ARTIFICIAL INTELLIGENCE (XAI) AND ETHICAL DECISION-MAKING IN BUSINESS

Gabriel Marín Díaz, José Javier Galán Hernández, José Luis Galdón Salvador

Universidad Complutense de Madrid (Spain), Universidad Politécnica de Valencia (Spain)

gmarin03@ucm.es; josejgal@ucm.es; jogalsal@doe.upv.es

EXTENDED ABSTRACT

Introduction

The question of whether machines can think, proposed by Alan M. Turing (Turing, 1950), has become increasingly relevant in today's world. Artificial Intelligence (AI) has made remarkable progress, surpassing human capabilities in various tasks traditionally associated with intelligence and creativity. However, as AI becomes more pervasive in decision-making processes within businesses, concerns regarding transparency, interpretability, and ethics arise. This article explores the importance of Explainable Artificial Intelligence (XAI) in facilitating ethical decision-making within the business context.

The Rise of XAI and the Paradox of Unexplainable Algorithms

Advancements in AI algorithms, particularly "black box" algorithms, have enabled machines to make complex decisions without human intervention. While these algorithms can yield impressive results, their decision-making logic often remains opaque and incomprehensible to humans. This poses a paradox: decision-making increasingly relies on AI systems that we struggle to understand fully. This lack of transparency raises concerns about accountability, biases, and potential risks in business operations (Kliegr et al., 2021).

The Need for Interpretability and the Ethical Implications

Recognizing the need for interpretability in AI algorithms, the concept of XAI has emerged. XAI aims to provide transparency and understandability in the decision-making process, enabling humans to comprehend and validate AI-driven decisions. Within the business context, interpretability becomes crucial as it allows decision-makers to assess the fairness, accuracy, and ethical implications of AI-generated recommendations or actions. Furthermore, interpretability helps identify and address potential biases in algorithms (Vallor & Rewak, 2019).

Legislative and Regulatory Considerations

Some sectors, such as finance, have recognized the importance of interpretability and have implemented regulations that mandate explanations for algorithmic decisions. However, broader awareness and understanding of interpretability in society are still limited. It is imperative for businesses to proactively engage with regulators, industry experts, and policymakers to shape legislation that ensures transparency, fairness, and ethical AI practices.

Ethical guidelines, standards, and accountability frameworks should be established to govern AI-driven decision-making in businesses (Glauner, 2022).

Advances in XAI and Mitigating Algorithmic Biases

Research and technological advancements in XAI offer promising solutions for addressing the interpretability challenge. Techniques such as rule-based explanations, visualizations, and model-agnostic approaches enable stakeholders to understand how AI algorithms arrive at decisions. Additionally, interpretability can help identify and mitigate biases present in training data or algorithm design (Molnar, 2019).

Creating a Culture of Ethical Decision-Making

Incorporating XAI into business processes requires a cultural shift towards ethical decision-making. Organizations should prioritize transparency, accountability, and human oversight in AI-driven systems. Decision-makers must possess a comprehensive understanding of AI capabilities, limitations, and potential biases to ensure responsible use (Bibal et al., 2020).

Model-agnostic interpretability algorithms in the context of XAI

The Model-agnostic algorithms provide techniques that can provide explanations for the decision-making process of any machine learning model, regardless of its underlying architecture or complexity. These algorithms focus on understanding and interpreting the behaviour of AI systems without relying on specific knowledge about how the models are built (Lundberg & Lee, 2017). Here are a few examples of their use in business settings:

- Decision Support Systems (Ribeiro et al., 2016): Model-agnostic algorithms can assist decision-makers in understanding the factors that contribute to AI-driven decisions. By providing explanations for each prediction or recommendation, these algorithms enable business professionals to gain insights into the underlying rationale and factors influencing the outcomes. This helps decision-makers make more informed choices.
- Risk Assessment and Compliance (Goodman & Flaxman, 2017): In industries such as finance, insurance, and healthcare, regulatory requirements often demand transparent and explainable decision-making processes. Model-agnostic algorithms allow businesses to identify potential biases, discrimination, or errors in the AI systems' outputs. By understanding the variables and features that influence the decision-making process, organizations can ensure compliance with regulations and mitigate potential risks.
- Customer Experience and Personalization (Marín Díaz et al., 2022): Model-agnostic algorithms can aid businesses in understanding the preferences and behaviour of their customers. By providing explanations for recommendations or personalized offerings, these algorithms allow organizations to provide transparency and gain customer trust. Moreover, they can help identify instances where AI-driven personalization might lead to unintended consequences or biases, enabling businesses to refine their algorithms and ensure fair and ethical treatment of customers.

- Fraud Detection and Cybersecurity (Zhang et al., 2022): Model-agnostic techniques can assist in identifying patterns and anomalies in large datasets, aiding in fraud detection and cybersecurity efforts. By explaining the features that contribute to suspicious activities or potential threats, these algorithms enhance the ability to interpret and validate AI systems' outputs, increasing the accuracy and effectiveness of fraud detection mechanisms.
- Process Optimization and Resource Allocation (Lakkaraju et al., 2016): Model-agnostic algorithms can uncover insights into complex business processes, enabling organizations to identify inefficiencies and optimize resource allocation. By providing explanations for the decisions made by AI models, businesses can pinpoint areas for improvement, streamline operations, and allocate resources more effectively.

Conclusions

By integrating ethics into AI practices, businesses can foster trust, maintain a positive reputation, and ensure the long-term viability and benefits of AI technologies within their operations. Embracing ethical AI not only aligns with societal expectations but also creates a competitive advantage by demonstrating responsible leadership in the ever-evolving landscape of AI-driven business practices.

- Transparency and accountability are paramount. Businesses should prioritize transparency in their AI systems and algorithms, ensuring stakeholders have a clear understanding of how decisions are made. This transparency builds trust and enables accountability for the outcomes produced by AI technologies.
- Fairness and non-discrimination should be prioritized. Businesses must actively identify and mitigate biases in their AI-driven processes that may result in discriminatory outcomes. Regular audits and evaluations are necessary to ensure equal opportunities and treatment for all individuals.
- Privacy and data protection are essential. Businesses must handle customer data responsibly, obtaining informed consent, implementing robust security measures, and adhering to relevant data protection regulations. Respecting privacy rights is critical for maintaining trust with customers and stakeholders.
- A human-centred approach is vital. AI should be designed to enhance human capabilities and improve decision-making, rather than replacing human workers. Businesses should prioritize the well-being of their employees and ensure that AI systems augment their skills and productivity.
- Ethical procurement and supply chain practices are necessary. Businesses should assess the ethical implications of AI technologies throughout their supply chains, ensuring that vendors and partners adhere to ethical standards and guidelines.
- Continuous monitoring and improvement are key. Ethical considerations should be an ongoing process, with businesses regularly evaluating the impact of AI on society, addressing emerging ethical challenges, and continuously improving their AI systems.

KEYWORDS: Decision-making, business processes, XAI, Ethics, AI practices.

REFERENCES

Bibal, A., Lognoul, M., de Streel, A., & Frénay, B. (2020). Legal requirements on explainability in machine learning. *Artificial Intelligence and Law*, *0123456789*. https://doi.org/10.1007/s10506-020-09270-4

Glauner, P. (2022). *An Assessment of the AI Regulation Proposed by the European Commission*. *April 2021*, 119–127. https://doi.org/10.1007/978-3-030-99838-7_7

Goodman, B., & Flaxman, S. (2017). European union regulations on algorithmic decision making and a “right to explanation.” *AI Magazine*, *38*(3), 50–57. https://doi.org/10.1609/aimag.v38i3.2741

Kliegr, T., Bahník, Š., & Fürnkranz, J. (2021). A review of possible effects of cognitive biases on interpretation of rule-based machine learning models. *Artificial Intelligence*, *295*. https://doi.org/10.1016/j.artint.2021.103458

Lakkaraju, H., Bach, S. H., & Leskovec, J. (2016). Interpretable decision sets: A joint framework for description and prediction. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, *13-17-August-2016*, 1675–1684. https://doi.org/10.1145/2939672.2939874

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, *2017-Decem*(Section 2), 4766–4775.

Marín Díaz, G., Galán, J. J., & Carrasco, R. A. (2022). XAI for Churn Prediction in B2B Models: A Use Case in an Enterprise Software Company. *Mathematics*, *10*(20). https://doi.org/10.3390/math10203896

Molnar, C. (2019). Interpretable Machine Learning. A Guide for Making Black Box Models Explainable. *Book*, 247. https://christophm.github.io/interpretable-ml-book

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why Should I Trust You?” Explaining the Predictions of Any Classifier. *NAACL-HLT 2016 - 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Demonstrations Session*, 97–101. https://doi.org/10.18653/v1/n16-3020

Turing, A. M. (1950). Computing Machinery and Intelligence Author (s): A . M . Turing Source : Mind , New Series , Vol . 59 , No . 236 (Oct ., 1950), pp . 433-460 Published by : Oxford University Press on behalf of the Mind Association Stable URL : http://www.jstor.org/sta. *Mind*, *59*(236), 433–460.

Vallor, S., & Rewak, W. J. (2019). *An Introduction to Data Ethics*. 63. https://www.scu.edu/media/ethics-center/technology-ethics/IntroToDataEthics.pdf%0Ahttps://www.accenture.com/t20160629T012639Z__w__/us-en/_acnmedia/PDF-24/Accenture-Universal-Principles-Data-Ethics.pdf#zoom=50

Zhang, Z., Hamadi, H. Al, Damiani, E., Yeun, C. Y., & Taher, F. (2022). Explainable Artificial Intelligence Applications in Cyber Security: State-of-the-Art in Research. *IEEE Access*, *10*(September), 93104–93139. https://doi.org/10.1109/ACCESS.2022.3204051

AI EXPLAINABILITY, TEMPORALITY, AND CIVIC VIRTUE

Wessel Reijers, Tobias Matzner, Suzana Alpsancar, Martina Philippi

Paderborn University (Germany)

Wessel.reijers@uni-paderborn.de; tobias.matzner@uni-paderborn.de; suzana.alpsancar@uni-paderborn.de; martina.philippi@uni-paderborn.de

EXTENDED ABSTRACT

The notion that artificial intelligence (AI) has to be explainable has become entrenched in the public discourse concerning the ethical impacts of this emerging technology (Mittelstadt et al., 2016). Most notably, the stated reason for this concern is the property of neural networks to function as 'black box' models (Pasquale, 2015) that nonetheless perform certain modalities of reasoning. That is to say, these models 'reason' from particular inputs, which may consist of characters, pixels, or digital information in other modalities, to particular outputs, without transparently disclosing the process of this reasoning. This is often contrasted with 'good old fashioned AI' (GOFAI) models that use decision trees which – in principle – can be followed by a human expert from input to output. The problem with neural nets, implemented in programs like ChatGPT and Dall-E, is that they can potentially influence or even autonomously make decisions about human affairs that cannot ex-post be explained by human interpreters – even if these are experts. At most, humans may figure out the particular artificial neurons that had an important influence on a decision.

Yet, the feasibility and relevance of the principle of explainability has been questioned. Robbins (2019) has argued that in fact, people are not required to explain every decision they make. Instead, explainability only becomes an issue in exceptional circumstances when the outcome of a particular decision requires explanation. It would therefore be unreasonable and unhelpful to insist on a standard for AI systems that does not apply to human decision-making. Moreover, meaningful human control over AI decision-making, which is arguably one of the aims of explainability, can be achieved by other means – for instance through proper legislation. Others have argued that explainability should not be reduced to explicability (i.e., accounting for the explanandum) but should involve the social context, considering it as a set of social practices (Rohlfing et al., 2021). Indeed, explaining takes place in a social context, and moreover has different modalities.

From this perspective, explainability as such is neither a mere technical matter, nor is it in any case relevant, nor is it a singular phenomenon. This paper proposes an initial way to grapple with these difficulties, by considering – first of all – the role of temporality in different modalities of explaining, and – secondly – the normative perspective of civic virtue to evaluate these different modalities, which then raises distinct requirements for explainability given distinct social contexts.

Let us start with the consideration of temporality, as it offers a ground to consider different modalities of explanation. In the *Rhetoric*, Aristotle set out the idea that argumentation occurs in different temporal modalities. It can be past-oriented, in which case it is *forensic*, explaining what has happened by reference to memory and traces. It can be present-oriented, in which case it is *epideictic*, explaining why a person or act deserves blame or honor, or the assignment

of virtue or vice. It can, furthermore, be future-oriented, in which case it is *deliberative*, explaining why particular future outcomes should or should not be supported. AI systems can, in principle, be involved in all three of these modalities of explaining, but they confront us with different normative requirements when they do. Forensic explanations, for instance, put forward requirements concerning historical proof, whereas deliberative explanations put forward requirements concerning (political) vision and conviction.

To make sense of these normative requirements, we may also draw from Aristotle. For in Aristotle, as Johnstone argues, (2023), ethics, rhetoric, and politics are fundamentally interrelated. Modalities of explanation, in other words, have a bearing on ethical and political life, in that they affect human virtues. Virtue is therefore a valid point of departure, as Vallor has forcefully argued (2016) in the context of technology ethics, in considering how AI affects explainability in a normative sense. Yet, virtue is also primarily grounded in the life of the individual, being anchored in *eudaimonia*, and does not yet offer the resources to bridge the gap between the ethics of the individual and the politics of the community. Civic virtue, developed in Aristotle's *Politics*, does offer this transitory concept, for it always mediates between the aim of the individual and the aim of the political community. As such, it is also inherently concerned with technology, as the technological infrastructure is a primary concern of the mode by which civic virtue is cultivated and enacted.

Strikingly, the distinct modalities of explanation and the distinct notions of civic virtue in political philosophy can each be grounded in a consideration of temporality. Like modalities of explanation, civic virtue can be past-, present-, and future-oriented. Past-oriented civic virtue finds its most vocal adherents in liberal and neo-republican thought, where it in an instrumental quality that draws from a history of reputational events, cultivating a sense of civility amongst a population (Pettit, 1997). Present-oriented civic virtue finds its footing in classical republican thought, where it requires institutional structures for the support of practices that aim at internal goods (MacIntyre, 2007). Future-oriented civic virtue finds its basis in existential republican thought, which puts forward the requirement of a durable public sphere that supports political action in concert (Arendt, 1958).

How do these different modalities of civic virtue help us to think through the modalities of explainable AI? First, they help us to consider the plurality of explanations insofar as they relate to different modalities of civic virtue. To give an example: when faced with a reputation-building AI (e.g., a credit scoring mechanism), the aim of such a system is to mediate past-oriented civic virtue; in that reputation building implies a historical record of reputational events. Such a mode of civic virtue put forward requirements deriving from forensic explanations. In other words, for such an AI to cultivate rather than to corrupt civic virtue, its explainability would need to safeguard requirements of – amongst others – historical proof. When faced with a more explicitly political AI (e.g., the use of AI in mass online deliberation), the aim of such a system is to mediate future-oriented civic virtue; in that it supports deliberative decision-making about alternative political pathways. Such a mode of civic virtue puts forward requirements deriving from deliberative explanations. Differently put, for such an AI to cultivate rather than to corrupt civic virtue, its explainability would need to respect requirements of – amongst others – political conviction. It goes without saying that the latter requirements would be rather more stringent and putting up a higher bar than the former.

What this tells us is, foremost, that not every explanation is equal. Whether an explanation is required at all, and what modality it should be in, depends on the temporal mode of the human activities that an AI system affects. In a shorthand manner, one could argue that the more AI infringes onto the political realm, the more stringent explainability requirements will be. At the same time, the modality of those requirements will also change, for instance shifting from forensic to deliberative requirements.

KEYWORDS: Explainability, AI, civic virtue, temporality.

REFERENCES

Arendt, H. (1958). *The Human Condition* (Vol. 24, Issue 1). University of Chicago Press. https://doi.org/10.2307/2089589

Johnstone, C. L. (2023). *An Aristotelian Trilogy: Ethics, Rhetoric, Politics, and the Search for Moral Truth*.

MacIntyre, A. (2007). *After Virtue: A Study in Moral Theory*. University of Notre Dame Press.

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data and Society*, *3*(2), 1–21. https://doi.org/10.1177/2053951716679679

Pasquale, F. (2015). *The Black Box Society*. Harvard University Press. https://doi.org/10.4159/harvard.9780674736061

Pettit, P. (1997). *Republicanism: A Theory of Freedom and Government*. Oxford University Press.

Robbins, S. (2019). A Misdirected Principle with a Catch: Explicability for AI. *Minds and Machines*, *29*(4), 495–514. https://doi.org/10.1007/s11023-019-09509-3

Rohlfing, K. J., Cimiano, P., Scharlau, I., Matzner, T., Buhl, H. M., Buschmeier, H., Esposito, E., Grimminger, A., Hammer, B., Hab-Umbach, R., Horwath, I., Hullermeier, E., Kern, F., Kopp, S., Thommes, K., Ngonga Ngomo, A.-C., Schulte, C., Wachsmuth, H., Wagner, P., & Wrede, B. (2021). Explanation as a Social Practice: Toward a Conceptual Framework for the Social Design of AI Systems. *IEEE Transactions on Cognitive and Developmental Systems*, *13*(3), 717–728. https://doi.org/10.1109/TCDS.2020.3044366

Vallor, S. (2016). *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting*. Oxford University Press.

ETHICS UNVEILED: ILLUMINATING THE PATH OF AI INTEGRATION IN HIGHER EDUCATION

Jorge Pelegrín-Borondo, Cristina Olarte-Pascual, Luis Blanco-Pascual, Alba García-Milon

University of La Rioja (Spain)

jorge.pelegrin@unirioja.es; cristina.olarte@unirioja.es; luis.blanco@unirioja.es; alba.garciam@unirioja.es

EXTENDED ABSTRACT

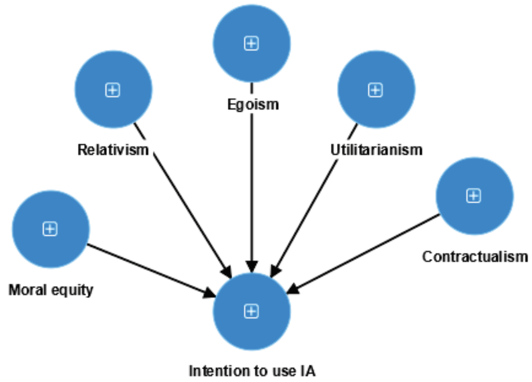
In the field of higher education, Artificial Intelligence (AI) presents both new possibilities and challenges (Silander & Stigmar, 2019). It offers opportunities to improve governance effectiveness and efficiency, benefiting students, teachers, administrative staff, and researchers (Nasrallah, 2014). Therefore, there is a need for integrating AI into higher education (Stefan & Sharon, 2017). However, the use of AI-based technologies for teaching and learning raises ethical issues (Celik, 2023). AI tools can exhibit systematic errors, leading to discrimination against students from diverse backgrounds and compromising inclusiveness in education (De Cremer & De Schutter, 2021; Dietvorst et al., 2018). Other ethical concerns associated with AI include content moderation, environmental impact, and the risk of copyright infringement (Cooper, 2023).

Currently, teachers face the dilemma of whether to encourage or discourage students from using AI. In this decision, teachers' ethical considerations regarding their students' use of this technology can be crucial in determining their role as integrators or opponents of AI. Ethics allows addressing the controversy between the potential benefits of technological progress and the duty not to jeopardize that progress (Olarte-Pascual, Pelegrín-Borondo, Reinares-Lara, Arias-Oliva, 2021). However, the impact of different dimensions of ethical judgment on this decision remains unexplored. This research aims to address this question, focusing on the widely recognized AI platform ChatGPT, which has gained global attention and public interest. Recent news in Spain indicates that university students are extensively using ChatGPT (Planas Bou, 2023).

Reidenbach and Robin (1990) developed the Multidimensional Ethical Scale (MES), which proposes that individuals use multiple reasons to make ethical judgments. Originally consisting of eight items measuring three subscales, the MES was distilled and validated from an initial inventory of 33 items (Reidenbach & Robin, 1990, p. 639). The MES (1990) and its modified versions (e.g. Kadić-Maglajlić et al., 2017; Mudrack & Mason, 2013; Pelegrín-Borondo et al., 2020) have been widely used to explain the influence of ethical judgment on behavior. Shawver and Sennetti (2009) proposed the Composite MES, a modification that incorporates items from the five major normative ethical theories. The Composite MES has been extensively used to explain the impact of ethical judgments on behavior (e.g., Kara et al., 2016; Manly et al., 2015; Mudrack & Mason, 2013). It includes the dimensions of moral equity, relativism, utilitarianism, egoism, and contractualism (Nguyen & Biderman, 2008; Reidenbach & Robin, 1990).

Building upon this theoretical framework, the authors propose to investigate how the different dimensions of ethical judgment influence university professors' intention to encourage their students to use AI in their tasks and academic activities. To achieve this, the following model is proposed (Figure 1).

Figure 1. Proposed model



A self-administered survey was conducted among university professors from Business Faculties in Spain to test the proposed model. An invitation was sent to all professors through a national association representing business faculties. A total of 270 valid surveys were collected, with 53% males and 47% females. The average age was 49.95 years (SD = 9.88). The MES Composite scale by Shawver and Sennetti (2009) was used to measure ethical judgment dimensions, employing an 11-point semantic differential scale. The professors' intention to encourage their students to use AI in academic activities was measured using a 2-item Likert scale based on Venkatesh and Davis's (2000) Technology Acceptance Model TAM2. The statistical analysis of the model was conducted using PLS (Partial Least Squares).

Regarding the results, the reliability and validity of the scales were examined. One item from the relativism dimension was removed due to convergent validity issues. The final scales demonstrated satisfactory reliability, convergent validity, and discriminant validity, as shown in Table 1.

Table 1. Composite reliability, Cronbach’s alpha, AVE (convergent validity) and discriminant validity.

Construct	**Composite reliability > 0.7**	**Cronbach’s *Alpha* > 0.7**	**AVE > 0.5**	**HTMT**				
				ME	**R**	**E**	**U**	**C**
Moral Equity (ME)	0.969	0.970	0.942					
Relativism (R)	0.863	0.863	0.880	0.898				
Egoism (E)	0.938	0.938	0.941	0.857	0.853			
Utilitarianism (U)	0.859	0.870	0.876	0.835	0.866	0.884		
Contractualism (C)	0.965	0.965	0.966	0.830	0.852	0.777	0.835	
Intention to use (IU)	0.958	0.958	0.960	0.764	0.707	0.743	0.669	0.699

Table 2 displays the values of R^2 and Q^2, the path coefficients (direct effects), and p-values for each antecedent variable of professors' intention for their students to use AI. The R^2 for the model of AI use intention was high (R^2 = 0.629), and the Q^2 provided by PLS Predict was greater than 0.5 (Q^2 = 0.565). This indicates that the dimensions of ethical judgment have explanatory and predictive power over professors' intention for their students to use AI. In Table 2, it is shown that the dimensions of moral equity, egoism, and contractualism positively influence the intention to use AI.

Table 2. Effect on the endogenous variables.

	R^2	Q^2	Path coefficient	*p*-value
INTENTION TO USE AI	0.585	0.565		
Moral Equity =>(+) Intention to use IA			0.401	0.000
Relativism =>(+) Intention to use IA			-0.014	0.850
Egoism =>(+) Intention to use IA			0.304	0.001
Utilitarianism =>(+) Intention to use IA			-0.076	0.310
Contractualism =>(+) Intention to use IA			0.195	0.013

The findings show that professors' ethical judgment dimensions have a differentiated impact on their intention to promote student use of AI in tasks and teaching activities. Three dimensions, namely moral equity, egoism, and contractualism, positively influence this intention. Among them, moral equity has the strongest explanatory power, indicating that perceiving AI use as fair motivates teachers to encourage it. Egoism is the second influential dimension, suggesting that personal benefits from student AI use increase teachers' inclination to promote it. Contractualism is the third influencing dimension, indicating that perceiving an implicit agreement within the university for AI use leads to greater encouragement. However, no evidence supports the impact of relativism and utilitarianism dimensions on professors' intention to promote student AI use. These conclusions emphasize the significance of considering professors' ethical perceptions when integrating AI in education and provide valuable insights for developing effective strategies for AI integration in teaching.

KEYWORDS: Artificial intelligence, ethical concerns, higher education, intention to use.

ACKNOWLEDGEMENTS: The authors gratefully acknowledge the support from the University of La Rioja for funding Teaching Innovation Projects during the academic year 2023-24, as well as the grant given to the COBEMADE Research Group at the University of La Rioja.

REFERENCES

Celik, I. (2023). Towards Intelligent-TPACK: An empirical study on teachers' professional knowledge to ethically integrate artificial intelligence (AI)-based tools into education. Computers in Human Behavior, 138, 107468.

Cooper, G. (2023). Examining science education in ChatGPT: An exploratory study of generative artificial intelligence. Journal of Science Education and Technology, 32(3), 444-452.

De Cremer, D., & De Schutter, L. (2021). How to use algorithmic decision-making to promote inclusiveness in organizations. *AI and Ethics*, 1(4), 563–567. https://doi. org/10.1007/s43681-021-00073-0

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2018). Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management Science*, 64(3), 1155–1170. https://doi.org/10.1287/mnsc.2016.2643

Kadi´c-Maglajli´c, S., Arslanagi´c-Kalajdˇzi´c, M., Micevski, M., Michaelidou, N., & Nemkova, E. (2017). Controversial advert perceptions in SNS advertising: The role of ethical judgement and religious commitment. *Journal of Business Ethics*, 141(2), 249–265. https://doi.org/10.1007/s1055.1-015-2755-5

Kara, A., Rojas-M´endez, J. I., & Turan, M. (2016). Ethical evaluations of business students in an emerging market: Effects of ethical sensitivity, cultural values, personality, and religiosity. *Journal of Academic Ethics*, 14(4), 297–325. https://doi. org/10.1007/s10805-016-9263-9

Manly, T. S., Leonard, L. N., & Riemenschneider, C. K. (2015). Academic integrity in the information age: Virtues of respect and responsibility. *Journal of Business Ethics*, 127 (3), 579–590.

Mudrack, P. E., & Mason, E. S. (2013). Ethical judgments: What do we know, where do we go? Journal of Business Ethics, 115(3), 575–597. https://doi.org/10.1007/s10551-012-1426-z

Nasrallah, R. (2014). Learning outcomes role in higher education teaching. *Education, Business and Society*, 7(4), 257–276. https://doi.org/10.1108/EBS-03-2014-0016

Nguyen, N. T., & Biderman, M. D. (2008). Studying ethical judgments and behavioral intentions using structural equations: Evidence from the multidimensional ethics scale. *Journal of Business Ethics,* 83(4), 627–640. https://doi.org/10.1007/s10551-007-9644-5

Olarte-Pascual, C., Pelegrín-Borondo, J., Reinares-Lara, E. Arias-Oliva, M. (2021). From wearable to insideable: Is ethical judgment key to the acceptance of human capacity-enhancing intelligent technologies? *Computers in Human Behavior*, 114, 106559.

Pelegrín-Borondo, J., Arias-Oliva, M., Murata, K., & Souto-Romero, M. (2020). Does ethical judgment determine the decision to become a cyborg? *Journal of Business Ethics*, 161(1), 5–17. https://doi.org/10.1007/s10551-018-3970-7

Planas Bou, C (2023). Universitarios y adolescentes se pasan en masa a ChatGPT para hacer trabajos (y exámenes). El Periódico (9-05-2023). https://www.elperiodico.com/es/sociedad/20230508/chatgpt-universidad-escuelas-inteligencia-artificial-estudiantes-deberes-examenes-86837251

Reidenbach, R. E., & Robin, D. P. (1988). Some initial steps toward improving the measurement of ethical evaluations of marketing activities. *Journal of Business Ethics*, 7, 871–879. https://doi.org/10.1007/BF00383050

Reidenbach, R. E., & Robin, D. P. (1990). Toward the development of a multidimensional scale for improving evaluations of business ethics. *Journal of Business Ethics*, 9(8), 639–653. https://doi.org/10.1007/BF00383391

Shawver, T. J., & Sennetti, J. T. (2009). Measuring ethical sensitivity and evaluation. *Journal of Business Ethics*, 88(4), 663–678. https://doi.org/10.1007/s10551-008-9973-z

Silander, C., & Stigmar, M. (2019). Individual growth or institutional development? Ideological perspectives on motives behind Swedish higher education teacher training. *Higher Education: The International Journal of Higher Education Research*, 77, 265–281. https://doi.org/10.1007/s10734-018-0272-z

Stefan, A. D. P., & Sharon, K. (2017). Exploring the impact of artificial intelligence on teaching and learning in higher education. *Research and Practice in Technology Enhanced Learning*, 1, 3–13. https://doi.org/10.1186/s41039-017-0062-8

Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the Technology Acceptance Model: Four longitudinal field studies. *Management Science*, 46, 186–204. https://doi.org/10.1287/mnsc.46.2.186.11926