



# NAVIGATING THE TENSION BETWEEN ALGORITHMIC AND HUMAN ETHICS: TOWARD A RESPONSIBLE BALANCE

## Tensión entre la ética algorítmica y la ética humana: encontrar un equilibrio responsable

Mario González-Arencibia 

Universidad de Ciencias Informáticas, La Habana, Cuba.   
[mgarencibia@uci.cu](mailto:mgarencibia@uci.cu)

Hugo-Armando Ordoñez-Erazo 

Universidad del Cauca, Popayán, Colombia.   
[hugordonez@unicauca.edu.co](mailto:hugordonez@unicauca.edu.co)

Juan-Sebastián González-Sanabria 

Universidad Tecnológica y Pedagógica de Colombia, Tunja, Colombia.   
[juansebastian.gonzalez@uptc.edu.co](mailto:juansebastian.gonzalez@uptc.edu.co)

Received: 16-04-2024

Accepted: 13-03-2025



### ABSTRACT

The increasing implementation of algorithmic systems in various areas of society has created a noticeable tension between algorithmic ethics and human ethics. This research examines this tension and proposes strategies to achieve a responsible balance between both ethical perspectives. The study focuses on three fundamental questions: What are the main points of conflict between algorithmic ethics and human ethics in automated decision-making processes? What approaches could harmonize these two ethical perspectives? How can human ethical principles be effectively incorporated into the design and operation of algorithmic systems? The main findings reveal that conflicts arise primarily in areas such as transparency, accountability, biases and discrimination, and the preservation of human autonomy. The research identifies key strategies to address these challenges, including the implementation of “ethics by design” frameworks, the development of ethical auditing processes, the promotion of diversity in development teams, and the advancement of digital ethics education. It is concluded that the effective integration of algorithmic ethics and human ethics requires a holistic approach that combines technical advancements, ethical reflection, ongoing education, and adaptive regulation. Only through this multifaceted effort can a technological ecosystem be created that enhances human capabilities and promotes a fairer and ethically conscious society in the digital age.

**Keywords:** algorithmic ethics; algorithmic responsibility; ethical balance; ethical tension; human values; social impact.

## RESUMEN

La creciente implementación de sistemas algorítmicos en diversos ámbitos de la sociedad ha generado una tensión notable entre la ética algorítmica y la ética humana. Esta investigación examina dicha tensión y propone estrategias para lograr un equilibrio responsable entre ambas perspectivas éticas. El estudio se centra en tres preguntas fundamentales: ¿Cuáles son los principales puntos de conflicto entre la ética algorítmica y la ética humana en los procesos de decisión automatizados? ¿Qué enfoques podrían armonizar estas dos perspectivas éticas? ¿Cómo se pueden incorporar efectivamente los principios éticos humanos en el diseño y funcionamiento de los sistemas algorítmicos? Los hallazgos principales revelan que los conflictos surgen principalmente en áreas como la transparencia, la responsabilidad, los sesgos y la discriminación, y la preservación de la autonomía humana. La investigación identifica estrategias clave para abordar estos desafíos, incluyendo la implementación de marcos de “ética por diseño”, el desarrollo de procesos de auditoría ética, el fomento de la diversidad en equipos de desarrollo, y la promoción de la educación en ética digital. Se concluye que la integración efectiva de la ética algorítmica y la ética humana requiere un enfoque holístico que combine avances técnicos, reflexión ética, educación continua y regulación adaptativa. Solo a través de este esfuerzo multifacético se podrá crear un ecosistema tecnológico que potencie las capacidades humanas y promueva una sociedad más justa y éticamente consciente en la era digital.

**Palabras clave:** equilibrio ético; ética algorítmica; impacto social; responsabilidad algorítmica; tensión ética; valores humanos.

## TENSÃO ENTRE A ÉTICA ALGORÍTMICA E A ÉTICA HUMANA: ENCONTRAR UM EQUILÍBRIO RESPONSÁVEL

### RESUMO

A crescente implementação de sistemas algorítmicos em diversos campos da sociedade tem gerado uma tensão notável entre a ética algorítmica e a ética humana. Esta pesquisa examina essa tensão e propõe estratégias para alcançar um equilíbrio responsável entre ambas as perspectivas éticas. O estudo se concentra em três questões fundamentais: Quais são os principais pontos de conflito entre a ética algorítmica e a ética humana nos processos de decisão automatizados? Quais abordagens poderiam harmonizar essas duas perspectivas éticas? Como os princípios éticos humanos podem ser efetivamente incorporados no design e funcionamento dos sistemas algorítmicos? Os principais resultados revelam que os conflitos surgem principalmente em áreas como transparência, responsabilidade, vieses e discriminação, e preservação da autonomia humana. A pesquisa identifica estratégias chave para lidar com esses desafios, incluindo a implementação de marcos de “ética por design”, o desenvolvimento de processos de auditoria ética, o incentivo à diversidade nas equipes de desenvolvimento e a promoção da educação em ética digital. Conclui-se que a integração eficaz da ética algorítmica e da ética humana exige uma abordagem holística que combine avanços técnicos, reflexão ética, educação contínua e regulamentação adaptativa. Somente por meio desse esforço multifacetado será possível criar um ecossistema tecnológico que potencialize as capacidades humanas e promova uma sociedade mais justa e eticamente consciente na era digital.

**Palavras-chave:** equilíbrio ético; ética algorítmica; impacto social; responsabilidade algorítmica; tensão ética; valores humanos.

---

## 1. INTRODUCTION

The accelerated technological evolution and the increasing integration of algorithms in decision-making processes have generated a fundamental debate regarding the relationship between algorithmic ethics and human ethics. Currently, systems based on artificial intelligence (AI) and machine learning exert an increasingly significant influence in domains that directly affect individuals, ranging from credit assessment to personnel selection [1]. The phenomenon brings up fundamental questions about how to reconcile human ethical principles with automated decision mechanisms. It is necessary to address the relationship between algorithmic ethics and human ethics to ensure that the use of technology respects and promotes fundamental human values.

Previous research has explored different sides of this problem. For example, Hagendorff [2] examined the limitations of the current ethical approaches in AI, focusing on the need to develop more comprehensive frameworks that address the ethical complexities of these technologies. Tsamados et al. [3] analyzed the specific ethical challenges that arise when implementing AI systems in the health sector, highlighting the importance of balancing technological efficiency with fundamental human values [4].

Several international organizations have acknowledged the importance of addressing these ethical challenges. The Council of Europe, for instance, adopted in 2020 a recommendation on the impact of algorithmic systems on human rights, providing guidelines for the ethical development and use of these technologies [5]. Moreover, entities such as the Institute of Electrical and Electronics Engineers (IEEE) have developed ethical standards for the design of autonomous and intelligent systems, seeking to promote responsible practices in the technology industry [6].

Despite these efforts, a significant gap persists in the knowledge regarding how to effectively integrate human ethical principles into the architecture and operation of algorithmic systems, particularly in scenarios where conflicts may arise between both ethical perspectives [7]. The aim of this research is to address such gap, analyzing the current tensions between algorithmic ethics and human ethics, and proposing specific strategies to reach a responsible balance.

Thus, the following research questions will be addressed. What are the main conflict points between algorithmic ethics and human ethics in automated decision processes? Which approaches could be promoted to reconcile these two ethical perspectives? In what way can human ethical principles be effectively incorporated into the design and operation of algorithmic systems?

The relevance of addressing the relationship between algorithmic ethics and human ethics lies in the algorithms' potential to influence the critical decisions that affect individuals and entire societies. Since algorithms are the product of human decisions, they are not neutral and can perpetuate current biases and inequalities. For example, algorithmic biases may discriminate against certain groups of individuals based on characteristics such as race, gender, or socioeconomic background, which underscores the need for ethics that contemplates justice and equity in their design and application.

The harmonization of algorithmic ethics with human ethical principles is fundamental to preventing AI technologies from reproducing prejudices and threatening human rights. UNESCO [8] has emphasized the importance of an "ethical compass" for AI, underscoring that, without adequate ethical safeguards, these technologies may fuel divisions and perpetuate discrimination. Therefore, algorithmic ethics must integrate values such as justice, responsibility, and respect for human dignity to guarantee a technological development that benefits every sector of society.

## 2. ALGORITHMIC ETHICS AND HUMAN ETHICS

The study of the tension between algorithmic ethics and human ethics requires a clear understanding of the fundamental concepts that guide each of these areas. Both ethical frameworks have their own principles and values that influence the design and implementation of advanced technologies, particularly algorithms.

Algorithmic ethics refers to the principles and values that must guide the creation and use of algorithms to ensure their fair and transparent operation. This field of study emphasizes the need for integrating aspects such as equity, responsibility, and transparency in algorithms [9]. The ethical principles in this context include non-discrimination, equity in decision-making, and transparency in algorithmic processes

[10]. These principles seek to guarantee that the algorithms do not perpetuate biases or injustices and that the decisions made by algorithmic systems are understandable and explainable.

On the other hand, human ethics covers the values and moral principles that guide human behavior. It is based on concepts such as human dignity, justice, autonomy, and responsibility. These principles are essential for the development of a just and fair society and must be considered when designing and implementing advanced technologies. Human ethics focuses on promoting well-being, protecting human rights, and ensuring that technological decisions benefit society as a whole [11].

Technological advances significantly influence ethics, as they introduce new possibilities and challenges that require the reassessment of traditional ethical principles. Technology can amplify the ethical benefits and the risks. For example, algorithms can improve efficiency and precision in decision-making, but they can also perpetuate biases and inequalities if they are not designed and used ethically [12].

The relationship between ethics and technology is bidirectional: while technology poses new ethical challenges, ethical principles must also guide the development and use of technology to ensure that its benefits are distributed fairly and equitably. Technological advances pose new ethical dilemmas and challenge traditional conceptions of morality. At the same time, ethical principles influence the development and implementation of new technologies. Coeckelbergh [12] argues that this constant interaction between ethics and technology requires a reflective and adaptive approach to address emerging challenges.

In order to explore adequately the relationship between algorithmic ethics and human ethics, it is critical to consider several basic questions. One of the main questions is: Which are the main principles of the algorithmic ethics and the human ethics? This question seeks to identify and compare the values that guide both ethical frameworks and understand how they can be effectively integrated (Table 1).

The comparison and observations between algorithmic ethics and human ethics reveal significant similarities and differences. Both sets of principles share fundamental values, such as justice, beneficence, and non-maleficence. However, algorithmic ethics focuses on artificial systems, while human ethics applies to interactions among people. Algorithmic ethics presents unique challenges, such as algorithmic opacity and the difficulty of attributing responsibility, while indicators can be used to evaluate compliance with principles in both contexts [4].

The importance of the table lies in its capacity to present visually and concisely the key principles of both ethics, facilitating the identification of similarities and differences, and serving as a guide for the development and evaluation of ethical algorithmic systems. Moreover, this tool can be useful to educate about fundamental concepts of ethics in the digital era. It is crucial to take into account the cultural context and the constant evolution of algorithmic ethics, as well as the importance of co-creating ethical principles that involve diverse stakeholders.

It is critical to consider and apply the fundamental principles of algorithmic and human ethics in the development and use of AI systems. These principles must guide all stages of the algorithmic systems' life cycle to ensure their responsible and ethical development and use.

The integration of these principles in the context of algorithmic decision-making poses significant challenges. For example, the principle of human autonomy may conflict with the efficiency and precision of automated systems in certain contexts. Zerilli et al. [1] analyze this dilemma in the context of criminal justice, where algorithmic risk assessment systems can influence the decisions that affect directly the freedom of individuals.

**Table 1.** Comparison of the principles of algorithmic ethics and human ethics

| Principle                                | Algorithmic ethics                                                                          | Human ethics                                                    | Examples                                                                                              | Indicators                                                                                |
|------------------------------------------|---------------------------------------------------------------------------------------------|-----------------------------------------------------------------|-------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------|
| <b>Transparency</b>                      | Explainability of the algorithm, traceability of the decisions.                             | Clear and complete information on the options and consequences. | A product recommendation algorithm must be able to justify why it is suggesting a particular product. | Level of detail in explanations, access to information about the algorithm's functioning. |
| <b>Justice and non-discrimination</b>    | Avoiding biases in data and algorithms, guaranteeing equity in the results.                 | Treating everyone with equity and respect.                      | A personnel selection algorithm cannot discriminate due to race or gender.                            | Algorithmic impact assessment, equity metrics.                                            |
| <b>Responsibility and accountability</b> | Identifying those responsible for algorithmic decisions, establishing oversight mechanisms. | Accepting the consequences of our actions.                      | An AI developer must be responsible for the damage caused by their algorithm.                         | Responsibility frameworks, audit mechanisms.                                              |
| <b>Privacy and security</b>              | Protecting personal data, guaranteeing the security of the systems.                         | Respecting people's intimacy.                                   | A facial recognition system must protect biometric privacy.                                           | Complying with privacy regulations (GDPR, CCPA), cybersecurity measures.                  |
| <b>Beneficence</b>                       | Maximizing the benefit and minimizing the damage.                                           | Doing good, promoting welfare.                                  | A medical diagnosis algorithm must be precise and reliable.                                           | Social impact assessment, efficiency and efficacy metrics.                                |
| <b>Autonomy</b>                          | Respecting people's capacity to make informed decisions.                                    | Freedom to make one's own decisions.                            | A recommendation system should not manipulate the user's decisions.                                   | Customization options, transparency in the decision-making.                               |
| <b>Non-Maleficence</b>                   | Avoid doing harm.                                                                           | Do not harm intentionally.                                      | An autonomous driving algorithm should not cause accidents.                                           | Error rates, risk analysis.                                                               |

Likewise, the principle of transparency in algorithmic ethics may clash with the inherent complexity of certain machine learning models, such as deep neural networks. Burrell [13] examines this tension and proposes approaches to improve the interpretability of complex algorithms without sacrificing their performance.

Understanding the theoretical foundations of algorithmic ethics and human ethics, as well as their interrelationship, is essential for addressing the ethical challenges that arise from the increasing implementation of AI systems in society. This conceptual framework provides a solid foundation for analyzing and proposing solutions to conflicts that may arise between both ethical approaches.

### 3. MAIN AREAS OF TENSION

The analysis of the tension between algorithmic ethics and human ethics examines the conflicts that arise when automated systems make decisions that may contradict people's values and ethical principles. This issue has become increasingly relevant in the digital era, where algorithms influence multiple aspects of daily life.

An example of this is the use of algorithms in the criminal justice system. In the USA, the COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) system is used to assess the recidivism risk of criminals. Nevertheless, studies have demonstrated that this system tends to exhibit

racial biases, assigning higher risk scores to African American individuals compared to white individuals with similar records [14]. This conflict highlights how an algorithmic decision can contradict the values of equity and justice that are fundamental to human ethics.

In the context of hiring, companies have begun to use algorithms to filter candidates and make hiring decisions. Although these algorithms can increase efficiency, they have also shown gender and race biases. For example, a study conducted by Amazon revealed that its personnel selection tool tended to penalize curricula vitae that included the word “woman” or that belonged to women because the algorithm was trained with historical data dominated by men [15]. This case evidences how depending on algorithms can perpetuate current inequities and go against equity and non-discrimination principles.

Implementing algorithms in credit decision-making is another field where ethical conflicts arise. Financial institutions use algorithms to determine the eligibility of prospective borrowers. However, research has shown that these systems can discriminate against ethnic minorities and low-income individuals, based on historical data that reflect socioeconomic prejudices [16]. This example underscores the need for reviewing and adjusting the algorithms to ensure that the automated decisions respect the human values of equity and social justice.

In India’s financial sector, credit scoring algorithms have shown biases against applicants in rural areas or with low income. A study revealed that these systems frequently assign lower scores to individuals from specific regions, despite their financial profiles being comparable to those of other candidates. This stands in direct opposition to the principles of financial inclusion and non-discrimination [17].

In the health sector in Brazil, an AI system was implemented to prioritize attention in public hospitals. However, it was discovered that the algorithm favored patients from certain neighborhoods, reproducing existing socioeconomic inequities. This stirred up a debate on equity in the access to health services and the need for adjusting these systems [18].

In Japan, a human resources company used a personnel selection algorithm that showed gender biases. The system penalized systematically female candidates for directive positions, based on historical hiring patterns. This demonstrated the tension between the efficiency promised by AI and the values of equal opportunity [19].

In China, the use of facial recognition systems for public surveillance has generated concerns about privacy and individual rights. Although these systems are justified for security reasons, their implementation raises ethical dilemmas about the balance between collective security and personal freedom [20].

These cases demonstrate how algorithmic decisions, despite seeking efficiency and objectivity, can conflict with fundamental human values such as equity, non-discrimination, and privacy. The challenge is to develop systems that are technically efficient and ethically aligned with human principles and rights.

[Table 2](#) provides a detailed vision of the fundamental conflicts that arise between human values and algorithmic decisions. A notable aspect is algorithmic opacity, which hinders understanding how decisions are made, generating distrust and impeding the detection of biases. Seeking efficiency may lead to the omission of important ethical values, such as justice and equity, which poses significant ethical dilemmas.

Another key point is the presence of inherent biases in algorithms, which can perpetuate and amplify the biases present in training data, giving rise to discriminatory consequences. The lack of context is also a challenge since algorithms can have issues considering individual contexts and exceptional situations, which may negatively affect decision-making.

Furthermore, the rigidity in rule application can lead to inflexible and inadequate results in complex situations, which highlights the need for flexibility and adaptability in algorithmic systems. These conflicts pose significant challenges to the development and implementation of algorithmic systems, underscoring the importance of addressing them proactively.

**Table 2.** Fundamental conflicts between algorithmic decisions and human values

| Conflict                              | Description                                                                                                                                                | Examples                                                                                                | Ethical implications                                                                |
|---------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|
| Transparency vs. technical complexity | Complex algorithms, especially those based on deep learning, can be difficult to understand. This makes it hard to assess how they work and detect biases. | Content recommendation algorithms that suggest information without a clear explanation of the criteria. | Lack of trust in algorithmic systems, difficulty in holding developers accountable. |
| Efficiency vs. ethical considerations | Optimizing algorithms to achieve specific goals can result in the omission of wider ethical considerations.                                                | Hiring algorithms that prioritize efficiency over diversity and inclusion.                              | Discrimination, perpetuation of inequities.                                         |
| Apparent objective vs. hidden biases  | Algorithms can include and amplify biases present in training data, which leads to discriminatory results.                                                 | Facial recognition algorithms that have difficulty identifying people of color.                         | Perpetuation of stereotypes, systemic discrimination.                               |
| Scale vs. individuality               | Algorithms can process vast amounts of data, but they may be unable to account for the particularities of each individual.                                 | Medical attention algorithms that do not consider the individual clinical context of a patient.         | Loss of personalized attention, inadequate medical decisions.                       |
| Consistency vs. flexibility           | Algorithms apply rules consistently but may lack the flexibility to adapt to exceptional circumstances.                                                    | Chatbots that cannot understand and answer open or ambiguous questions.                                 | Unsatisfactory user experiences, lack of empathy.                                   |

## 4. CHALLENGES AND LIMITATIONS

The implementation of algorithmic ethics faces significant challenges, particularly regarding transparency, accountability, biases, and discrimination. In parallel, human ethics shows limitations when facing technology, especially in terms of processing capacity and consistency in decision-making. These aspects configure a complex landscape at the intersection between artificial intelligence and human values.

### A. Challenges in the Implementation of Algorithmic Ethics

One of the most critical challenges is the lack of transparency in the algorithms. Most machine learning systems work as black boxes, which means that their internal processes and decisions are opaque to users and regulators. This opacity hinders understanding how decisions are made and poses accountability problems since it is difficult to assign blame when algorithms make errors or perpetuate injustices. In the financial sector, for example, automated credit models often fail to provide a clear explanation as to why an applicant has been rejected. This can lead to perceptions of discrimination and a lack of equity [16].

Responsibility is also a critical aspect. It is necessary to establish clear mechanisms that allow holding accountable the entities that design and use algorithms when they make errors or cause harm.

Without those mechanisms, victims of unfair algorithmic decisions may not have effective resources to seek reparations. This problem is exacerbated in contexts where automated decisions have serious consequences, such as in healthcare or the judicial system. For example, the use of algorithms to diagnose diseases or recommend treatments must be transparent and backed by solid scientific evidence to avoid errors that could have fatal consequences.

Another significant challenge is the presence of biases and discrimination in algorithms. Machine learning systems are often trained with historical data that may contain social and economic biases. These biases can be perpetuated and amplified in algorithmic decisions, negatively affecting minority and marginalized groups. For example, in the field of hiring, a study revealed that automated systems tended to penalize candidates belonging to racial or ethnic minorities, perpetuating existing inequalities [21].

Moreover, algorithms can exacerbate current inequalities if they are not designed carefully. In the housing sector, for example, algorithms used to determine mortgage eligibility may discriminate against certain populations based on historical data that reflects discriminatory housing practices. It is essential to implement measures to detect and mitigate these biases, making sure that algorithms operate fairly and equitably.

## ***B. Limitations of Human Ethics Against Technology***

Human ethics has significant limitations when confronted with advanced technology. One of them is the processing capacity. Human beings have limits when it comes to the amount of information they can process and how quickly they can make decisions. Algorithms, instead, can analyze vast data sets and make decisions in split seconds, which makes them more efficient in many tasks. For example, AI-based medical diagnostic systems can analyze thousands of medical images in much less time than a team of human radiologists, identifying patterns and anomalies with remarkable precision [22].

However, this superior processing capacity is not always translated into fairer or more coherent decisions. Algorithms can process data quickly, but if the data are biased or incomplete, the resulting decisions may be unfair or incorrect. Therefore, it is crucial to ensure that the data used to train algorithms is representative and free from bias.

Coherence when making decisions is another aspect where human ethics can be limited in comparison to technology. Due to cognitive and emotional biases, human beings can be inconsistent in their judgments. For example, judges may show significant variations in their sentences based on factors such as mood, fatigue, or even the time of day [23]. In contrast, algorithms, when well-designed, can offer consistency and uniformity in decisions that humans cannot match.

Nevertheless, these algorithms must be designed and supervised in such a way that their decisions reflect ethical values and do not perpetuate injustices. Human oversight is necessary to guarantee that algorithms operate within an adequate ethical framework. This includes revising algorithms regularly and updating their parameters to reflect changes in social values and norms.

## 5. ETHICAL CONSIDERATIONS

The implementation of algorithms in decision-making raises several ethical challenges that require careful analysis and a multidisciplinary approach. These challenges span areas such as transparency and accountability, bias and discrimination, privacy and data use, and the preservation of human autonomy and agency.

Transparency and accountability in algorithmic systems are essential to guarantee that automated decisions are understandable and justifiable. However, achieving this transparency presents significant technical challenges, especially in the case of complex algorithms such as those with deep learning. The opacity of these systems, often called “black boxes,” makes it difficult to explain how certain decisions are made, which poses problems of trust and accountability.

The question of who is responsible for decisions made by algorithms is also complex. Should responsibility fall on developers, organizations that implement systems, or end users? This question has deep legal and ethical implications for which there are no clear answers in many contexts yet.

Bias and discrimination represent another significant challenge. Algorithms can reproduce and amplify existing prejudices in society, either through biased training data or through the system’s design itself. This can lead to unfair decisions in critical areas such as criminal justice, medical attention, or access to economic opportunities. Ensuring equity and non-discrimination in algorithmic decision-making processes requires not only technical solutions but also a deep understanding of the social and cultural contexts in which these systems operate.

Privacy and data use present another set of ethical dilemmas. On the one hand, access to large amounts of personal data can significantly improve the efficiency of algorithms in many areas. On the other hand, this extensive use of personal data poses a series of concerns regarding individual privacy and potential abuse. Finding a balance between data usefulness and privacy protection is an ongoing challenge that requires both technical solutions and robust regulatory frameworks.

The question of who should have access to and control over the data used by algorithms is also controversial. Should companies, governments, or individuals control this data? Each option has different implications for privacy, innovation, and power in the digital society.

In addition, the preservation of human autonomy and agency in a world increasingly mediated by algorithms is a topic of growing importance. To what extent should we delegate important decisions to automated systems? How can we ensure that humans retain the ability to make informed and autonomous decisions in an environment where algorithms can process information at a scale and speed far beyond human capabilities?

These challenges do not have simple solutions and require an interdisciplinary approach that combines technical advances, ethical reflection, adequate regulatory frameworks, and broad social debate. It is essential to develop algorithmic systems that are not only efficient, but also fair, transparent, and respectful of fundamental human values. This involves designing AI interfaces and systems that are explainable and keep humans at the center of the decision-making process.

Ultimately, the goal must be to create a technological ecosystem that enhances human capabilities rather than replacing them, and that promotes a more just and equitable society. This will require a continuous effort of research, development, and dialog among multiple sectors of society.

## 6. PROPOSALS FOR A RESPONSIBLE BALANCE

The integration of algorithmic ethics and human ethics requires a multifaceted approach that addresses technical, social, and normative challenges. This process entails the development of specific strategies and a clear definition of the roles of the various actors involved.

One of the fundamental strategies for promoting the balance between algorithmic ethics and human ethics is the development of regulatory and normative frameworks. These frameworks must establish clear guidelines for the design, implementation, and use of algorithmic systems, ensuring that they respect human values and fundamental rights. For example, the European Union has proposed the Artificial Intelligence Regulation, which classifies AI systems according to their risk level and establishes specific requirements for each category [24]. This type of regulation can be a model for other countries and regions.

Education and training on algorithmic ethics and human ethics is another essential strategy. Technology developers and ethics experts must be trained in ethical principles and technological concepts, respectively. A study made by Morley et al. [25], revealed that only 35% of AI developers had been trained in ethics, which underscores the need for improving education in this field. Interdisciplinary programs that combine computer science, ethics, and social sciences can help train professionals capable of addressing the ethical challenges of technology in a comprehensive manner.

Regarding the roles of different actors, technology developers have the primary responsibility of incorporating ethical considerations from the earliest stages of algorithmic design. This involves adopting approaches such as “ethical design by default” and conducting ethical impact assessments throughout the entire technological development lifecycle [26].

Technology users, for their part, must be aware of the ethical implications of the systems they use and demand transparency and accountability from developers and providers. Auxier & Rainie [27] found that 68% of Americans believe that technology companies have too much power and influence in the economy, which suggests a growing public awareness of the need for oversight.

Regulators play a fundamental role in the creation and implementation of regulatory frameworks. They must be updated on technological advances and collaborate with ethics experts, developers, and civil society representatives to create effective and adaptable regulations. The USA Federal Trade Commission, for example, has increased its focus on AI, issuing guidelines on the use of algorithms and warning companies about the legal consequences of biased or misleading AI systems [28].

Academic and research institutions are tasked with advancing knowledge of algorithmic ethics and developing methodologies to assess and mitigate ethical risks. The Oxford University Institute for Ethics in AI, for example, conducts interdisciplinary research to address the ethical challenges of AI [29].

Civil society, including non-governmental organizations and advocacy groups, must actively participate in public debate on algorithmic ethics, representing the interests of affected communities and promoting transparency and accountability.

To achieve an effective balance between algorithmic ethics and human ethics, close collaboration between all these actors is necessary. Multisectoral forums, such as the UN Internet Governance Forum, can provide platforms for dialog and cooperation [30].

In particular, the integration of algorithmic ethics and human ethics requires a holistic approach that combines regulation, education, shared responsibility, and collaboration among diverse stakeholders.

Only through these coordinated efforts can we ensure that technological progress aligns with human values and contributes to the well-being of society as a whole.

The effective incorporation of human ethical principles into the design and operation of algorithmic systems is a complex challenge that requires a multidisciplinary approach and careful consideration of various technical, social, and ethical factors. This task involves not only technological advances but also a profound reflection on human values and how these can be translated into the digital realm.

A fundamental strategy for incorporating ethical principles into algorithmic systems is the “ethics by design” approach. This concept, proposed by Dignum [31], suggests that ethical considerations should be integrated from the initial stages of technological development, rather than being an afterthought. This means that developers and designers of algorithmic systems must consider the potential ethical and social impacts of their creations from the outset of the design process.

To implement ethics by design, it is necessary to create interdisciplinary teams that include not only engineers and data scientists, but also ethicists, sociologists, psychologists, and public policy experts. This collaboration can help anticipate and address the ethical implications of algorithmic systems from multiple perspectives. For example, in the development of personnel selection algorithms, the inclusion of experts in human resources and organizational psychology can help identify and mitigate potential discriminatory biases.

Another key strategy is the implementation of ethical auditing processes for algorithmic systems. Raji et al. [32] propose a framework for auditing algorithms that includes ethical and social impact assessments. These audits can help identify potential ethical issues before systems are implemented on a large scale. For example, an ethical audit of a credit scoring algorithm could reveal whether the system inadvertently discriminates against certain demographic groups, allowing for corrections before its implementation.

The transparency and explainability of algorithmic systems are fundamental to incorporating human ethical principles. Techniques such as LIME (Local Interpretable Model-agnostic Explanations) and SHAP (SHapley Additive exPlanations) can help make algorithmic decisions more understandable to users and regulators [33], [34]. For example, in the context of an AI-based medical diagnostic system, these techniques could provide clear explanations about how a particular diagnosis was reached, allowing physicians and patients to understand and evaluate the system’s decision.

The incorporation of human feedback mechanisms into algorithmic systems can also be an effective strategy. This allows systems to learn and adapt to the users’ ethical values and preferences. For example, in content recommendation systems, options could be implemented for users to express their ethical concerns about certain types of content, allowing the system to adjust its recommendations accordingly.

It is also important to consider diversity and inclusion in the process of designing algorithmic systems. Buolamwini and Gebru [35] demonstrated how biases in training data can lead to facial recognition systems that function unequally for different demographic groups. Ensuring diversity in development teams and in the data used to train algorithms can help mitigate these problems.

Continuing education and training in digital ethics for developers, users, and regulators is essential. Universities such as MIT have incorporated courses on AI ethics into their engineering and computer science programs [36]. These initiatives must extend beyond academic institutions, encompassing continuing education programs for practicing professionals and public awareness campaigns.

The effective incorporation of human ethical principles into algorithmic systems requires a holistic approach that combines technical, educational, and regulatory strategies. Ethics by design, ethical audits, transparency and explainability, human feedback mechanisms, diversity in development, and continuing education are key components of this approach. Only through these combined efforts can we create algorithmic systems that are not only efficient but also respect and promote fundamental human ethical values.

## 7. CONCLUSIONS

The relationship between algorithmic ethics and human ethics is a complex and dynamic field that poses significant challenges for contemporary society. As algorithmic systems become omnipresent in various aspects of life, the need to align these systems with human ethical values becomes increasingly urgent. This interaction is not about separate entities, but rather a deep intertwining where algorithmic systems, being human creations, inevitably reflect the values, biases, and limitations of their creators.

A fundamental conclusion that emerges from this analysis is the critical importance of transparency and explainability in algorithmic systems. The ability to understand and explain the decisions made by these systems is not merely a matter of user trust, but rather an essential requirement for legal and ethical accountability. However, achieving this transparency presents considerable technical challenges, especially in the case of complex deep-learning algorithms.

Bias and discrimination in algorithmic systems are persistent problems that require continuous and rigorous attention. Research has shown how biases present in training data can lead to discriminatory results in artificial intelligence applications. This underscores the need to address ethical issues not only in the design of algorithms but also in the collection and selection of the data that feeds them.

Looking toward the future, several recommendations emerge for the development and implementation of ethical algorithmic systems. The adoption of “ethics by design” frameworks emerges as a fundamental strategy, integrating ethical considerations from the initial stages of technological development. This involves forming interdisciplinary teams that include not only technical experts, but also ethicists, sociologists, and public policy specialists.

The development and implementation of robust ethical auditing processes is emerging as another key recommendation. These processes should include comprehensive assessments of the ethical and social impact of algorithmic systems prior to their large-scale implementation. Concurrently, fostering diversity in technological development teams can help mitigate unintentional biases and promote broader consideration of the ethical implications of these systems.

Continuing education and training in digital ethics is emerging as a fundamental pillar for addressing these challenges. It is essential to expand education in AI and technology ethics beyond academic institutions, including training programs for practicing professionals and public awareness campaigns. This will help create a more informed and critical society regarding the use and implications of algorithmic systems.

In the regulatory sphere, it is recommended that flexible and adaptable frameworks that can evolve at the same pace as technological advances be developed. These frameworks must be sufficiently robust to protect fundamental rights, yet sufficiently agile so as not to impede innovation. Furthermore, given the global nature of these challenges, the need to promote closer international collaboration in formulating ethical principles and regulatory standards for AI and algorithmic systems becomes evident.

In conclusion, the effective integration of algorithmic ethics and human ethics requires an approach that combines technical advances, profound ethical reflection, continuous education, and adaptive regulation. Only through this multifaceted and sustained effort can we aspire to create a technological ecosystem that not only enhances human capabilities, but also promotes a more just, equitable, and ethically conscious society in the digital age.

## AUTHORS' CONTRIBUTION

**Mario González-Arencia:** Conceptualization, Methodology, Investigation, Writing-Original Draft.

**Hugo-Armando Ordoñez-Erazo:** Investigation, Writing-Original Draft.

**Juan-Sebastián González-Sanabria:** Investigation, Writing-Review & editing.

## REFERENCES

- [1] J. Zerilli, A. Knott, J. Maclaurin, C. Gavaghan, "Algorithmic decision-making and the control problem," *Minds and Machines*, vol. 31, pp. 555-578, 2019. <https://doi.org/10.1007/s11023-019-09513-7>
- [2] T. Hagendorff, "The ethics of AI ethics: An evaluation of guidelines," *Minds and Machines*, vol. 30, pp. 99-120, 2020. <https://doi.org/10.1007/s11023-020-09517-8>
- [3] A. Tsamados, N. Aggarwal, J. Cows, J. Morley, H. Roberts, M. Taddeo, L. Floridi, "The ethics of algorithms: key problems and solutions," *AI & Society*, vol. 37, pp. 215-230, 2022. <https://doi.org/10.1007/s00146-021-01154-8>
- [4] Computing, *La ética de la IA: 5 principios fundamentales*, 2024. <https://www.computing.es/mundo-digital/los-5-principios-eticos-de-la-inteligencia-artificial/>
- [5] Council of Europe, *Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems*. Strasbourg: Council of Europe, 2020.
- [6] IEEE, *Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems*, 2021.
- [7] M. A. Méndez, J. M. Palomares, *Caso de ética: Algoritmos ocultos*, 2024. [https://www.anahuac.mx/mexico/biblioteca/sites/default/files/2024-06/Algoritmos\\_ocu.pdf](https://www.anahuac.mx/mexico/biblioteca/sites/default/files/2024-06/Algoritmos_ocu.pdf)
- [8] UNESCO, *Recommendation on the Ethics of Artificial Intelligence*, 2019. <https://unesdoc.unesco.org/ark:/48223/pf0000373434>
- [9] UNESCO, *Ethics of Artificial Intelligence*, 2024. <https://www.unesco.org/en/artificial-intelligence/recommendation-ethics>.
- [10] L. Floridi, M. Taddeo, "What is Data Ethics?," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 374, no. 2083, e20160360, 2016. <https://doi.org/10.1098/rsta.2016.0360>
- [11] T. W. Bynum, "Computer and Information Ethics," in *The Stanford Encyclopedia of Philosophy*, 2018. <https://plato.stanford.edu/archives/win2018/entries/ethics-computer/>
- [12] M. Coeckelbergh, "AI Ethics," in *The MIT Press*, 2020.

- [13] J. Burrell, "How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms," *Big Data & Society*, vol. 3, no. 1, pp. 1-12, <https://doi.org/10.1177/2053951715622512>
- [14] J. Angwin, J. Larson, S. Mattu, L. & Kirchner, *Machine Bias*, 2016. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>
- [15] J. Destin, *Amazon scraps secret AI recruiting tool that showed bias against women*, 2018. <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G>
- [16] M. Hurley, J. Adebayo, "Credit scoring in the era of big data," *Yale Journal of Law and Technology*, vol. 18, no. 1, pp. 148-216, 2016. <http://hdl.handle.net/20.500.13051/7808>
- [17] T. Bono, K. Croxson, A. Giles, "Algorithmic fairness in credit scoring," *Oxford Review of Economic Policy*, vol. 37, no. 3, pp. 585-617, 2021. <https://doi.org/10.1093/oxrep/grab020>
- [18] A. D. Chiavegatto, et al., "Artificial intelligence in public health in Brazil: Opportunities and challenges," *The Lancet Digital Health*, vol. 2, no. 12, e658, 2020.
- [19] International Labour Organization, *World Employment and Social Outlook 2021: The role of digital labour platforms in transforming the world of work*, 2021.
- [20] F. Liang, "Surveillance in China's social credit system: Ethical considerations and public perceptions," *Journal of Information, Communication and Ethics in Society*, vol. 17, no. 4, pp. 375-392, 2020.
- [21] A. Datta, M. Tschantz, A. Datta, "Automated Experiments on Ad Privacy Settings: A Tale of Opacity, Choice, and Discrimination," *Proceedings on Privacy Enhancing Technologies*, vol. 2015, no. 1, pp. 92-112, 2015. <https://doi.org/10.1515/popets-2015-0007>
- [22] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. Swetter, H. Blau, S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks." *Nature*, vol. 542, no. 7639, pp. 115-118, 2017. <https://doi.org/10.1038/nature21056>
- [23] S. Danziger, J. Levav, L. Avnaim-Pesso, "Extraneous factors in judicial decisions," *Proceedings of the National Academy of Sciences*, vol. 108, no. 17, pp. 6889-6892, 2011. <https://doi.org/10.1073/pnas.1018033108>
- [24] Comisión Europea, *Proposal for a Regulation laying down harmonised rules on artificial intelligence*, 2021. <https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence>
- [25] J. Morley, L. Floridi, L. Kinsey, A. Elhalal, "From what to how: An initial review of publicly available AI ethics tools, methods and research to translate principles into practices," *Science and Engineering Ethics*, vol. 26, pp. 2141-2168, 2020. <https://doi.org/10.1007/s11948-019-00165-5>
- [26] L. Floridi, J. Cows, M. Beltrametti, R. Chatila, P. Chazerand, V. Dignum, E. Vayena, "AI4People—An ethical framework for a good AI society: opportunities, risks, principles, and recommendations," *Minds and Machines*, vol. 28, pp. 689-707, 2018. <https://doi.org/10.1007/s11023-018-9482-5>
- [27] B. Auxier, L. Rainie, *Key takeaways on Americans' views about privacy, surveillance and data-sharing*, 2019. <https://www.pewresearch.org/fact-tank/2019/11/15/key-takeaways-on-americans-views-about-privacy-surveillance-and-data-sharing/>
- [28] Federal Trade Commission, *Report Warns About Using Artificial Intelligence to Combat Online Problems*, 2018. <https://www.ftc.gov/news-events/news/press-releases/2023/06/ftc-report-warns-about->

using-artificial-intelligence-combat-online-problems

- [29] University of Oxford, *Institute for Ethics in AI*, 2023. <https://www.oxford-aiethics.ox.ac.uk/>
- [30] Internet Governance Forum, *About IGF*, 2023.  
<https://www.intgovforum.org/en/content/about-igf-0>
- [31] V. Dignum, "Ethics in artificial intelligence," *Ethics and Information Technology*, vol. 20, no. 1, pp. 1-3, 2018.
- [32] I. D. Raji, A. Smart, R. N. White, M. Mitchell, T. Gebru, B. Hutchinson, et al., "Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing," in *Proceedings of the 2020 conference on fairness, accountability, and transparency*, 2020, pp. 33-44.
- [33] M. T. Ribeiro, S. Singh, C. Guestrin, "Why should I trust you?" Explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 2016, 1135-1144.
- [34] S. M. Lundberg, S. I. Lee, "A unified approach to interpreting model predictions," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 2017, pp. 4768-4777.  
<https://doi.org/10.5555/3295222.3295230>
- [35] J. Buolamwini, T. Gebru, "Gender shades: Intersectional accuracy disparities in commercial gender classification," in *Conference on Fairness, Accountability and Transparency*, 2018, pp. 77-91.
- [36] MIT News, *MIT reshapes itself to shape the future*, 2019.  
<https://news.mit.edu/2019/mit-reshapes-itself-stephen-schwarzman-college-of-computing-0904>