Navigating the Ethical Challenges of Artificial Intelligence in Higher Education: An Analysis of Seven Global AI Ethics Policies

Zouhaier Slimi [1,2], Beatriz Villarejo Carballido [3]

[1]*Deusto University, Unibertsitate Etorb., 24, 48007 Bilbo, Bizkaia, Spain , Bilabo, Spain*
[2]*National University of Sciences and Technology Oman, Peripheral Rd, Liwa, Oman, Sohar, Oman*
[3]*Universitat Autonoma de Barcelona, Cerdanyola del Vallès, near the city of Barcelona in Catalonia, Barcelona, Spain*

***Abstract* – AI use in higher education raises ethical concerns that must be addressed. Biased algorithms pose a significant threat, especially if used in admission or grading processes, as they could have devastating effects on students. Another issue is the displacement of human educators by AI systems, and there are concerns about transparency and accountability as AI becomes more integrated into decision-making processes. This paper examined three AI objectives related higher education: biased algorithms, AI and decision-making, and human displacement. Discourse analysis of seven AI ethics policies was conducted, including those from UNESCO, China, the European Commission, Google, MIT, Sanford HAI, and Carnegie Mellon. The findings indicate that stakeholders must work together to address these challenges and ensure responsible AI deployment in higher education while maximizing its benefits. Fair use and protecting individuals, especially those with vulnerable characteristics, are crucial. Gender bias must be avoided in algorithm development, learning data sets, and AI decision-making.**

Data collection, labeling, and algorithm documentation must be of the highest quality to ensure traceability and openness. Universities must study the ethical, social, and policy implications of AI to ensure responsible development and deployment. The AI ethics policies stress responsible AI development and deployment, with a focus on transparency and accountability. Making AI systems more transparent and answerable may reduce the adverse effects of displacement. In conclusion, AI must be considered ethically in higher education, and stakeholders must ensure that AI is used responsibly, fairly, and in a way that maximizes its benefits while minimizing its risks.

***Keywords* – Artificial intelligence (AI), higher education, biased algorithms, decision-making, human displacement.**

DOI: 10.18421/TEM122-02
https://doi.org/10.18421/TEM122-02

Corresponding author: Zouhaier Slimi,
Deusto University, Unibertsitate Etorb., 24, 48007 Bilbo, Bizkaia, Spain , Bilabo, Spain
National University of Sciences and Technology Oman, Peripheral Rd, Liwa, Oman, Sohar, Oman
Email: slimizou@oopendeusto.es

Received: 17 February 2023.
Revised: 18 April 2023.
Accepted: 05 May 2023.
Published: 29 May 2023.

© 2023 Zouhaier Slimi & Beatriz Villarejo Carballido; published by UIKTEN. This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 License.

The article is published with open access at https://www.temjournal.com/

1. Introduction

Artificial intelligence (AI) has the potential to revolutionize numerous industries, including all facets of society, particularly education. As AI technologies continue to advance and become more widely adopted in higher education, the ethical issues surrounding their implementation must be addressed. This article explores the ethical challenges posed by AI in higher education, with a specific focus on biased algorithms, the decision-making process, and the potential displacement of human labour. The use of biased algorithms in AI poses a significant moral challenge in higher education, where decisions made by AI systems, such as admissions or grading, can significantly impact students' lives. Similarly, the potential displacement of human labour, including faculty and teaching assistants, presents concerns about the impact on employment and the need for individuals to stay competitive in the job market. As these systems become more integrated into decision-making processes, users must clearly understand how they work and how decisions are made to ensure transparency and accountability.

This article emphasizes the need for policymakers and stakeholders, including educators and administrators, to collaborate and ensure the responsible deployment of AI technologies in higher education. In conclusion, the ethical challenges posed by AI in higher education highlight the need for careful consideration and accountable implementation of these technologies to maximize their positive effects while minimizing any potential adverse effects.

1.1 AI and biased algorithms in higher education

Satterfield & Able [18] argue that emerging applications of artificial intelligence (AI), such as predictive software integrated into websites like Amazon Prime, autonomous features integrated into automobiles, or innovative home technologies like Alexa or Siri, have an increasing impact on business, industry, research, and higher education. New trends and innovations in applying AI technology to design, user experience, and behavioural psychology will fundamentally alter how humans interact with technology and design user experiences. Biased algorithms prioritise creators, empathizers, pattern-recognition experts, and meaning makers [18].

Shanklin et al. [19] argue that AI algorithms, even if designed to be neutral, may produce racially biased results if trained on data that reflect racial biases. In the context of medical appointment scheduling in the United States, their research found that algorithms predict that black patients are more likely to miss appointments than non-black patients. Although technically accurate based on available data, black patients are disproportionately scheduled in appointment slots with longer wait times, perpetuating racial inequalities and creating a lack of access to healthcare. This raises essential accuracy-fairness trade-offs, as policymakers and stakeholders must decide whether to prioritise efficiency or equity when using AI in these settings.

The potential for AI to exacerbate inequalities is not unique to medical appointment scheduling, but extends to other domains such as education, judicial systems, and public safety. As such, it is crucial to develop strategies to address these trade-offs. [19] propose a decoupling approach that separates an algorithm's machine learning and optimisation components, allowing for interventions at various stages to promote fairness. Specifically, the authors applied their method to medical appointment scheduling and identified four interventions that address disparities in different components of the algorithm.

While one approach eliminated disparities while maintaining comparable precision to state-of-the-art methods, other procedures resulted in varying accuracy and fairness trade-offs.

As such, policymakers and stakeholders must carefully consider the trades associated with each approach when using AI to avoid perpetuating racial and ethnic disparities in healthcare and other domains.

In conclusion, the study by [19] highlights the potential for AI algorithms to perpetuate racial and ethnic disparities in various domains, including healthcare. The research addresses these disparities by decoupling an algorithm's components and intervening at different stages. However, policymakers and stakeholders must carefully weigh the accuracy-fairness trade-offs associated with other interventions when deciding how to use AI in various settings.

Huang et al. [11] argue that the rapid proliferation of artificial intelligence (AI) technologies has led to significant changes in the landscape of higher education, fundamentally altering traditional teaching and learning norms. However, these changes also raise critical ethical concerns about surveillance, social inequality, and job security. To address these concerns, the authors conducted an in-depth examination of the discourse surrounding the integration of AI in higher education, with a particular focus on the field of library and information science (LIS) and the role of librarianship in shaping the trajectory of AI in learning and teaching. They also examined the ethical implications of the use of AI in higher education and the role of professional LIS ethics in confronting these transformations.

While the work of [11] is a valuable contribution to the growing literature on the intersection of AI and higher education, it is essential to acknowledge some potential limitations of their study. For example, their focus on LIS and librarianship may limit the generalisability of their findings to other disciplines and fields. Furthermore, their examination of the ethical implications of AI in higher education is mainly theoretical and could benefit from more empirical research. Despite these limitations, their work represents an essential step towards a more nuanced understanding of the impact of AI on higher education and the ethical challenges it poses.

Yolder Himes et al. [21] opine that student with different skin colour, particularly women of colour, face significant barriers in STEM fields in higher education due to social isolation and various biases, such as interpersonal, technological, and institutional biases. The authors identify a bias in online exam proctoring software, which frequently uses facial detection technology to identify potential instances of cheating. However, facial detection algorithms utilised by exam proctoring software may be biased against students based on skin tones or gender, depending on each company's images used as training sets.

This phenomenon has not yet been quantified, nor is it readily available from software manufacturers.

Yolder Himes et al. [21] assessed instructor outputs of 357 students from four courses to determine if the automated proctoring software adopted by their institution and used by at least 1,500 universities in the United States was biased based on race, skin colour, or gender. The authors manually classified the skin tone of each student's self-reported race and gender using a high-resolution photograph. There was a significant increase in the likelihood that students with darker skin tones and black students would be marked as requiring more instructors review due to the possibility of cheating compared to students with lighter skin tones. In addition, women with the darkest skin tones were significantly more likely to be selected for review than men with darker skin or men and women with lighter skin tones.

While the authors do not observe any statistically significant differences between male and female students in the aggregate, their findings suggest that a prominent automated proctoring software may use AI algorithms that are biased against certain student groups. This study is the first quantitative examination of biases in facial recognition software at the intersection of race and gender. It has implications for multiple fields, including education, social justice, equity and diversity, and psychology. However, it is essential to note that the study was limited to a single institution and a small sample size. The generalisability of the findings to other institutions and populations requires further investigation.

In conclusion, [21] provide important insights into the potential biases in facial detection technology used in online exam proctoring software, particularly against students of colour and women of colour. Although this study raises significant concerns, more research is needed to better understand the extent of the problem and develop appropriate solutions. Nonetheless, this research highlights the need for greater attention to AI technologies' social and ethical implications in higher education. It underscores the importance of promoting equity and inclusion for all students in STEM fields.

Cornacchia et al. [4] have argued that artificial intelligence (AI) has increasingly become a popular solution for making critical judgments in various life-altering decisions. However, they cautioned that biased AI tools could cause significant harm and that these systems may improve or diminish individuals' well-being. Government regulations prohibit using sensitive features such as gender, race, and religion in algorithmic decision-making to avoid unfair outcomes.

Despite these regulations, [4] contend that these restrictions may not safeguard individuals from unfair decisions, since algorithms may continue to exhibit discriminatory behaviour, even when sensitive features are omitted.

Cornacchia et al. [4] proposed an end-to-end method for detecting bias in black-box models that comply with regulations. The method uses a module for counterfactual reasoning and an external classifier for sensitive features. The counterfactual analysis identifies minimum cost variations that result in a positive outcome. In contrast, the classifier identifies nonlinear patterns of nonsensitive features that act as surrogates for sensitive characteristics. The experimental evaluation demonstrates the effectiveness of the proposed technique to detect classifiers that learn from proxy features.

It is noteworthy that [4] conducted further research to explore the impact of cutting-edge debiasing algorithms on the proxy feature problem. However, a critical stance on this issue is necessary as the effectiveness of debiasing algorithms may still be limited due to the use of proxy features. It is, therefore, crucial to acknowledge that the proposed method is not a panacea for detecting bias in AI systems. Nonetheless, the proposed method is a significant contribution to the literature and paves the way for future research to improve algorithms' effectiveness in detecting bias in AI systems.

According to [1], the ubiquitous deployment of Artificial intelligence (AI) at the periphery has the potential to revolutionize various aspects of human life. However, the authors warn that the success of AI should be measured by its ability to benefit humanity. They argue that deep learning-based edge AI algorithms are intricately linked with human interests and must be viewed through a human-centric lens. Nevertheless, the authors suggest that the security and trustworthiness of AI applications are far from foolproof or ethical, despite their significant impact on human interests. Butt et al. [1] contend that social norms are often disregarded during the design, implementation, and deployment of edge AI systems, making it essential to analyze the application of AI at the edge from a human-centred standpoint.

Butt et al. [1] make two contributions in their paper. First, they present a development pipeline for human-centric embedded machine learning (HC-EML) applications using a generic human-centric artificial intelligence (HCAI) framework. The authors then analyse and discuss the privacy, dependability, robustness, and security aspects of HC-EML applications, offering an insider's perspective on their challenges and potential solutions.

The authors illustrate the gravity of these issues with a case study of human facial emotion recognition (FER) based on the AffectNet data set.

The case study by [1] analysed the effects of commonly used input quantisation on an EML model's security, robustness, fairness, and reliability. The findings revealed that input quantisation reduced the effectiveness of adversarial and backdoor attacks at the expense of a slight reduction in accuracy compared to clean input. The authors determined that the eyes, alar crease, lips, and jaws significantly impacted a FER model's decision, as per the explanations generated by SHAP. The authors also observed that input quantisation showed a significant bias against dark-skinned faces and hypothesised that the low contrast characteristics of dark-skinned faces might be responsible for the observed tendencies.

Finally, [1] concluded with cautionary comments and recommendations for future researchers. Despite the potential of AI at the periphery, they warn that the ethical implications of these technologies cannot be ignored. The authors recommend that researchers use human-centric approaches to design, implement and deploy AI systems to ensure that they benefit humans and adhere to ethical and social norms. The study underscores the importance of considering the ethical implications of AI at the periphery and provides information on potential solutions to mitigate the challenges associated with its deployment.

Gardner [8] argues that the use of biased algorithms in education systems, as evidenced by the controversial A-level results in the UK in August 2020, highlights the need for greater awareness and accountability in algorithmic decision-making. While the transparency of the algorithm used by Ofqual is commendable, the design of the data set and the broader societal biases it reflects resulted in unfair outcomes that were difficult to deny or dismiss. However, [8] notes that similar biases and harmful consequences exist in many other algorithmic systems. Still, its impact is often less visible and more challenging to challenge, particularly for those without the privilege and resources to do so. This raises concerns about the ethics and accountability of algorithmic decision-making and the need for a more rigorous evaluation of the datasets and algorithms used in such systems. As Gardner [8] emphasises, it is crucial to ensure that algorithms are designed with sensitivity to potential biases and that those affected are informed of their existence and have mechanisms to challenge their outcomes. Further research and awareness of these issues are essential to ensure that algorithmic systems are deployed equitably and ethically.

1.2 AI and decision-making processes

The definition and categorization of AI and Machine Learning provided by [13] provide a valuable framework for understanding the scope and application of AI in various industries. They may have oversimplified things when they promote XAI as the solution to the problems with transparency and interpretability in AI systems. Although XAI techniques have been developed to explain AI algorithms' decision-making processes, their effectiveness remains significant limitations, particularly in more complex and opaque models. Additionally, the reliance on expert human interpretation of XAI explanations raises concerns about the potential for bias and the limitations of human understanding in assessing AI systems. More research is needed to evaluate the effectiveness of XAI in addressing the ethical and social implications of AI use. Thus, while [13] overview of XAI is a valuable contribution to the field, it should be considered in conjunction with a critical assessment of the limitations and challenges of implementing XAI in practise.

Although integrating AI systems into decision-making tasks aims to improve task performance, it is essential to recognise that AI is not infallible, and its recommendations may not always align with human values and ethics. Therefore, it is crucial to understand how humans behave when confronted with the challenge of knowledge imbalance, particularly when they lack the necessary knowledge to complete the task accurately. [9] provide valuable insights into this issue, highlighting the importance of involving users in the AI recommendation generation process. This approach increases the likelihood of users accepting the AI's suggestions and enhances their perception of collaboration with the AI agent. However, it is essential to note that such findings may not be generalisable to all AI-assisted decision-making tasks, and it is necessary to consider the context and nature of the study when implementing these insights. Furthermore, more research is essential to explore the ethical implications of integrating AI into decision-making tasks and the potential risks associated with overreliance on AI recommendations.

Cornacchia et al. [4] suggest that while artificial intelligence (AI) is increasingly being relied upon to inform critical judgments that impact people's lives, biased AI systems can negatively affect individuals' well-being. While laws ban sensitive qualities such as gender, ethnicity, and religion from influencing decisions, algorithms may employ proxy variables that are only distantly related to sensitive aspects, suggesting that these restrictions may not be enough to avoid discrimination.

Cornacchia et al. [4] propose an end-to-end method for detecting bias in black-box models to address this issue. This approach utilizes a module for counterfactual reasoning, which identifies the minimum cost variations that result in a positive outcome, and an external classifier for sensitive features that identify non-linear patterns of non-sensitive features serving as surrogates for sensitive characteristics.

However, while the experimental evaluation of the proposed technique indicates its effectiveness in detecting classifiers that learn from proxy features, the authors also acknowledge that the use of cutting-edge debiasing algorithms may have a limited effect on the problem of proxy features.

In light of these findings, it is crucial to develop more robust and comprehensive strategies to detect and prevent bias in AI systems, particularly as they continue to integrate into critical decision-making processes. Moreover, it is essential to critically evaluate and improve existing regulations to ensure that they effectively address the potential for discriminatory behaviour in AI systems.

Cornacchia et al. [4] suggest that while artificial intelligence (AI) is increasingly being relied upon to inform critical judgments that impact people's lives, biased AI systems can negatively affect individuals' well-being. The authors claim that algorithms can continue to use proxy traits that are only distantly connected to sensitive factors such as gender, ethnicity, and religion, despite laws that prohibit them from influencing judgments.

Cornacchia et al. [4] propose an end-to-end method for detecting bias in black-box models to address this issue. This approach utilizes a module for counterfactual reasoning, which identifies the minimum cost variations that result in a positive outcome, and an external classifier for sensitive features that identify non-linear patterns of non-sensitive features serving as surrogates for sensitive characteristics.

However, while the experimental evaluation of the proposed technique indicates its effectiveness in detecting classifiers that learn from proxy features, the authors also acknowledge that the use of cutting-edge debiasing algorithms may have a limited effect on the problem of proxy features.

Considering these findings, it is crucial to develop more robust and comprehensive strategies to detect and prevent bias in AI systems, particularly as they continue to integrate into critical decision-making processes. Moreover, it is essential to critically evaluate and improve existing regulations to ensure that they effectively address the potential for discriminatory behaviour in AI systems.

1.3 AI and Human Displacement

Artificial intelligence (AI) in the hiring process has become increasingly popular, despite growing concerns about the potential for biased evaluations. Zhang & Yencha [22] study aimed to explore the public's perceptions of resume and video interview screening algorithms. The authors used a nationally representative sample to investigate the effectiveness and fairness of hiring algorithms.

The study's results revealed that the public has a generally negative view of using algorithms in the hiring process, with most respondents considering them unfair and ineffective. Interestingly, the authors noted individual differences in algorithmic perceptions, with males having a higher level of education and income expressing more favourable views towards hiring algorithms than their counterparts. These findings are significant as they challenge the widespread assumption that AI-driven recruitment methods are universally accepted.

Although the study sheds light on the public's perceptions of hiring algorithms, it has several limitations. Firstly, it only focused on resume and video interview screening algorithms; thus, it may not apply to other hiring algorithms. Second, the study did not investigate the reasons for the public's negative perceptions of AI-driven recruitment methods. Future research could explore the factors influencing these perceptions to provide a better understanding of the public's attitudes towards hiring algorithms.

In conclusion, [22] study is essential for the emerging research on hiring algorithms. It highlights the need for businesses to address the public's negative perceptions of AI-driven recruitment methods and proposes strategies to increase their acceptance. However, the study's limitations call for caution when interpreting the results and further research in this area.

Data and algorithms are essential to develop data-driven and AI-driven economies. Users, data providers, and algorithm providers must interact to ensure the efficiency of exchange of data and algorithm effectiveness of recommender systems in connecting users and products in e-commerce environments. Their applicability for data and algorithm sharing has not been thoroughly investigated. To address this research gap, [15] conducted a study in which they identified six recommendation scenarios for supporting data and algorithm sharing, four of which differ significantly from traditional e-commerce recommendation scenarios.

These recommendation scenarios were evaluated by [15] using a novel interaction data set from the OpenML data and algorithm sharing platform. The authors examined three types of recommendation strategies: those based on popularity, collaboration, and content. The authors discovered that collaboration-based recommendations were the most accurate in every scenario, whereas the accuracy of other recommendations varied by scenario. For example, algorithm recommendations for users posed incredible difficulty than algorithm recommendations for datasets. In addition, the content-based strategy generated minor popularity-biased requests for the most critical datasets and algorithms.

While the study by [15] provides valuable insight into the effectiveness of recommender systems for data and algorithm sharing, it is essential to evaluate the study's findings critically. For instance, the scope and biases of the study's data set may limit the generalisability of the findings. Furthermore, the definition of accuracy used in the study should be scrutinised. Furthermore, the potential implications of popularity bias in recommendation systems should be considered.

In conclusion, the study by [15] contributes significantly to understanding recommendation systems for data and algorithm sharing. However, more research is needed to validate and extend the study findings, address potential biases, and develop context-specific recommendation approaches to support better data and algorithm sharing in various settings.

The ethical implications of decision-making in human resource management (HRM) have received significant attention in both academic and practitioner circles. On the contrary, research on the theoretical foundations of ethical positions and strategies in HRM decision-making and the accountability for these decisions after the fact has been scarce. Therefore, the present study proposes a Throughput Model framework that describes how perceptions, judgments, and information use influence individual decision-making processes in an algorithmic HRM context. Moreover, the model identifies algorithmic pathways that can facilitate diverse ethical decision-making strategies.

This study uses a variety of multidisciplinary theoretical lenses, including those related to AI-augmented HRM (HRM(AI)), HRM(AI) assimilation processes, AI-mediated social exchange, and the judgement and choice literature, to further explore the integration of artificial intelligence (AI) into HRM and its acceptance by stakeholders. Rodgers et al. [17] note that the use of algorithmic ethical positions in the adoption of AI has received limited exploration in the literature, despite its potential to enhance the intelligibility and accountability of AI-generated HRM decision-making. The authors argue that algorithmic ethical positions play a pivotal role in the selection of HRM strategies and highlight the importance of accounting for their use in HRM decision-making processes [17].

Overall, this study contributes to the existing literature by providing a theoretical framework that offers a better understanding of decision-making processes in algorithmic HRM contexts, while also shedding light on the crucial role of algorithmic ethical positions in the integration of AI in HRM. However, further research is necessary to test the efficacy of the proposed Throughput model framework in practical settings and explore algorithmic ethical positions' nuances in HRM decision-making processes.

Agent-based modelling is a powerful approach to understanding social phenomena by simulating individual behaviours and interactions. However, as modelling techniques continue to advance, the analysis of complex input factors in models can become more challenging, particularly when proposing specific policies for improving system outcomes. While traditional micro-dynamic analysis can be informative, it may also suffer from ambiguity and limited explanatory power. To address these limitations, [3] proposed a revised microdynamic analysis method that incorporates advanced artificial intelligence techniques to enhance model interpretation and facilitate group-specific policymaking. This modified method enables a more comprehensive causal understanding of a target phenomenon across subgroups, thereby reducing ambiguity and increasing the method's explanatory power. The authors applied this method to an agent-based model that evaluated the effects of a long-term care scheme on access to care. The findings showed that this revised method could suggest policies for improving access equity more effectively than conventional scenario analysis [3].

Despite the promising results, it is essential to note that further research is needed to validate the generalizability and applicability of this revised method in other contexts.

Fossen et al. [7] conducted an empirical investigation on the links between three types of patented technologies: artificial intelligence (AI), software and industrial robots, and wage fluctuations at the individual level in the United States over ten years (2011-2021). The study aimed to examine whether AI technologies are related to wage increments or decrements for individual workers and how this relationship compares to previous software and industrial robots’ innovations. The researchers used patient-derived indicators of occupational exposure to these three technological categories to conduct their analysis.

To investigate the impact of technology on wages, the authors merged individual wage data for the United States with novel technology measures and employed regression analysis to estimate the association between annual wage changes and technical measures while controlling for other variables. Research findings suggest that the advent of software and industrial robots is related to a decrease in wages, which may indicate a significant displacement effect on human labour.

However, it should be noted that the study has several limitations that should be considered. Firstly, the patent-derived indicators may not necessarily reflect occupational exposure to these technologies. Furthermore, the analysis did not consider the different skill sets required to work with these technologies, which could affect the wage impact of AI, software, and industrial robots. Therefore, while the findings are informative, caution should be exercised when interpreting them.

2. Methodology

Discourse analysis is valuable for analyzing written or spoken language in social or cultural contexts. It is commonly used to investigate how language is used to construct and negotiate meaning and to examine the social and political implications of language use. This approach seeks to expose and challenge how language is utilized to maintain and perpetuate inequalities and injustices while promoting a cognitive approach to focus on the mental processes involved in language use and how language shapes our understanding and interpretation of AI ethics policies. The current study applied discourse analysis to analyse seven central AI ethics policies based on three main themes. AI bias, decision making, and human labour displacement. The data were thematically coded according to these three objectives and analysed using the Fairclough dimensions based on the texts studied. The study employed a purposive sampling technique to select policies most relevant to the research questions and objectives and ensure representation from diverse regions and stakeholders. The ethical considerations included ensuring data accuracy and transparency, respecting intellectual property rights, maintaining objectivity and fairness throughout the research, avoiding negative consequences of recommendations, and adhering to relevant ethical guidelines and regulations. Purposive sampling was the appropriate strategy given the specific research questions and objectives, even though it would not accurately reflect the larger population and may limit the generalisability of the findings [6].

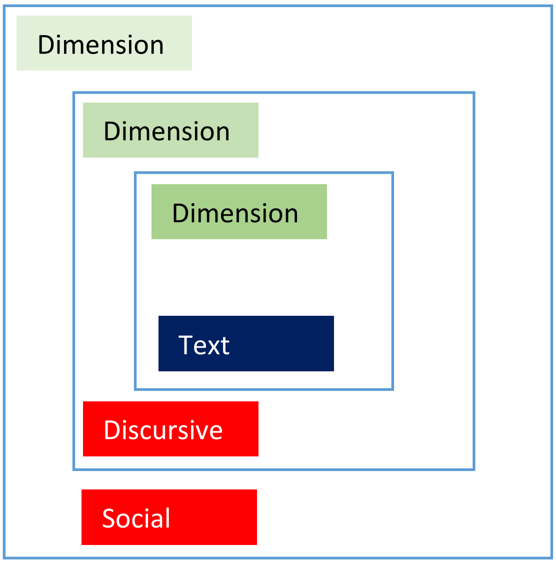

Figure 1. Fairclough discourse analysis model

3. Findings

These are the main findings based on the seven documents studied.

3.1. AI and biased algorithms in higher education

The research paper in question sheds light on the issues of biased algorithms and their potential impact on society. As demonstrated by the findings in table 1, all policies studied emphasize the importance of guiding AI towards unbiased algorithms. However, taking a critical stance on these policies is vital, as they may not fully address the issue of unwanted bias in AI.

While policies aim to ensure fair use of AI and avoid inequitable effects on individuals, they may not always succeed in distinguishing between fair and unfair biases, which can vary between cultures and societies. Moreover, the Intelligence Community Directive 203 requires objectivity and awareness of assumptions and risks when analysing, but it remains unclear whether these requirements are consistently met.

Furthermore, while the policies recognize the importance of avoiding biases based on sensitive characteristics such as race, ethnicity, gender, nationality, income, sexual orientation, ability, and political or religious belief, the research paper highlights that there are still instances where biased algorithms have negatively impacted individuals with these characteristics. As a result, it is essential to continue critically examining these policies and how they are implemented to ensure that they effectively reduce the risks of unwanted bias and promote fairness and equality in AI.

Table 1. AI and biased algorithms in higher education

[12]	The UNESCO intelligence analysis is objective. Intelligence Community Directive 203 requires objectivity and awareness of assumptions and risks when analysing. "We must use reasoning and practical mechanisms to expose and counteract bias". The Intelligence Community builds artificial intelligence with "biases" for legal, policy, and mission reasons. UNESCO's models and datasets filter out irrelevant data, focus on specific foreign intelligence targets, and minimise US person data collection and use. The Intelligence Community reduces bias by identifying and reducing unintended bias. "Unwanted bias" is a bias that could undermine the validity and reliability of the analysis, harm individuals, or affect civil liberties like freedom from excessive government intrusion into speech, religion, travel, or privacy. Data collection, feature extraction, curation, labelling, model selection and development, and user training may unintentionally introduce bias. Discover bias throughout an AI's lifecycle, mitigate unwanted bias, and document and communicate known preferences and how they were addressed to ensure long-term reliability, model reuse, and trustworthiness of outputs.
[16]	Algorithmic Bias Algorithms can be technically biased as algorithm accuracy may depend on training data. Algorithm bias harms society and consumers, so the Chinese government is trying to regulate algorithm use industrially and data protection-wise. Article 8.2: Biometrics the PIPL considers facial recognition and biometric data sensitive personal information. Processing such information requires separate consent for specific purposes and sufficient necessity. Public image collection or personal identification equipment can only be used for general security unless consent is given. Companies must determine if commercialized processing is necessary and find ways to obtain "separate consent" under the PIPL. Car cameras typically record pedestrians in the automotive industry. Automobile data processors must consider the PIPL and the recently issued Several Provisions on the Management of Automobile Data Security (for Trial Implementation), which believe videos and images with facial information are essential data when training their algorithms and providing relevant services. The Supreme People's Court also provides its judicial view on facial information processing. It defines legal liability for infringing the law while performing facial verification, recognition, or analysis in commercial settings and public areas; failing to disclose regulations on facial information processing; or failing to expressly state the aims, methods, and extent of such processing.
[5]	Human agency and oversight, including fundamental rights, human agency, and human oversight. Technical Robustness and Safety: Includes resistance to attack and security, a fallback plan and general safety, as well as precision, dependability, and reproducibility. Privacy and data governance, including respect for privacy, the quality and integrity of data, and data access. Traceability, explicability, and communication are all components of transparency. Diversity, non-discrimination, and fairness, including avoiding unfair bias, accessibility and universal design, and stakeholder involvement.
[10]	Inequitable biases can be reflected, reinforced, or reduced by AI algorithms and datasets. We acknowledge that distinguishing between fair and unfair preferences is not always straightforward and varies between cultures and societies. We will strive to avoid inequitable effects on individuals, especially those based on sensitive characteristics such as race, ethnicity, gender, nationality, income, sexual orientation, ability, and political or religious belief.
[14]	Massachusetts Institute of Technology (MIT): MIT has a set of guidelines for AI research, which includes a commitment to fairness, transparency, and accountability in AI development. The guidelines state that researchers should be aware of potential biases in data and algorithms and take steps to mitigate them. MIT also has a Centre for Responsible AI, which conducts research and education on the ethical, social, and policy implications of AI.
[20]	Stanford University has a set of principles for AI that state that researchers should be aware of potential biases in data and algorithms and take steps to mitigate them. Stanford also has a Centre for AI Policy and Governance, which conducts research and education on AI's ethical, social, and policy implications.
[2]	Carnegie Mellon University (CMU) has several initiatives and resources related to the ethical use of AI, including in higher education. For example, the CMU AI and Ethics Initiative aims to promote AI's responsible and ethical use through research, education, and engagement. In addition, the CMU Centre for Machine Learning and Health is dedicated to advancing the safe, effective, and ethical use of machine learning in healthcare. Regarding bias in AI algorithms, CMU has resources and research focusing on detecting and mitigating bias in algorithms. For example, researchers at the University have developed methods to detect bias in data sets and algorithms, as well as techniques to mitigate bias in machine learning models. Additionally, the university offers courses and resources for students and faculty to learn about ethical considerations in AI, including the potential for bias in algorithms.

3.2. AI and decision-making processes

In the AI and decision-making process, findings reveal that gender bias should be avoided or minimised in algorithm development, learning data sets, and AI decision making, as detailed in Table 2. Policies indicate that automated decisions using personal information must be transparent, fair, impartial and not discriminate against the trading price or other trading conditions. Data collection, labelling, and algorithm documentation should be top-notch to ensure traceability and transparency. Explainable AI decisions should impact lives; therefore, all university AI ethics policies must emphasise fairness, transparency, and accountability in AI development. Similarly, the ethical, social, and policy implications of AI should be studied at each university's centre.

Table 2. decision-making process and AI ethics

[12]	Google Images' "schoolgirl" search will likely show women and girls in sexualized costumes. Schoolboys dominate "schoolboy" results. There are either no men or very few men dressed in sexualised costumes. Society's gender stereotypes AI. The technology behind search engines is not impartial because it processes large amounts of data and ranks results based on which ones have received the most clicks, which is determined by the user's preferences and location. Therefore, a search engine has the potential to become an echo chamber that reinforces biases that exist in the real world and further solidifies the beliefs associated with these prejudices and stereotypes online. How can we ensure that the results are more accurate and evenly distributed? Can we report search results that have bias? How should women be accurately represented in search results, and what would such a representation look like? Gender bias should be avoided or minimised in algorithm development, learning data sets, and AI decision-making. UNESCO aims to eliminate gender bias in AI.
[16]	The PIPL governs automated decision-making. First, automated decision-making using personal information must be transparent, fair, impartial, and not discriminate in trading price or other trading conditions. Automated decision making in information feeds or commercial marketing to individuals must provide options not specific to the individual's characteristics or easy opt-out options. Individuals whose interests are materially affected by an automated decision have the right to request explanations from the relevant service provider/processor and to refuse automated decisions.
[5]	Explicability depends on AI system data, system, and business model transparency. Traceability. Data collection, labelling, and algorithm documentation should be top-notch to ensure traceability and transparency. AI decisions follow. Identifying why an AI decision was wrong prevents future errors. Traceability allows for audibility and explanation. Explainability. AI systems can explain their technical processes and human decisions (e.g., system application areas). Technical explainability requires human understanding and tracing AI system decisions. System explicability and precision may also need to be prioritised (at the cost of explainability). Explainable AI decisions should impact lives. Explain quickly and to the stakeholder's expertise (e.g., layperson, regulator or researcher). Explain how an AI system affects the organisation's decision-making, design, and deployment (thus ensuring business model transparency). AI should inform, not impersonate. AI must be recognisable. Optional interaction protects fundamental rights. The use case should inform AI practitioners and end-users of the AI system's capabilities and limitations. Communicate the accuracy and boundaries of the AI system.
[10]	Google will design AI systems with feedback mechanisms, relevant explanations, and appeal mechanisms. Our AI technologies will be overseen and directed by humans.
[14]	A small amount of reckoning is upon technology: The Impact of Algorithms on Free Speech, Privacy, and Autonomy AI, or the datasets on which AI is trained, are frequently biased or misused to manipulate individuals.
[20]	Stanford University has a set of AI research principles, including a commitment to "fairness, non-discrimination, transparency, and accountability.
[2]	Carnegie Mellon University AI research principles include "fairness, transparency, and accountability." The codes also require researchers to mitigate data and algorithm biases. All university AI ethics policies emphasise fairness, transparency, and accountability in AI development and the need to identify and mitigate data and algorithm biases. The ethical, social, and policy implications of AI are studied in each university's centre.

Consequently, AI ethics policies in higher education are an emerging concern as universities and colleges increasingly adopt and integrate AI systems into their operations and decision-making processes. These policies can address a wide range of issues related to AI and decision-making processes in higher education, such as:

- Fairness: Ensuring that the AI systems used in admissions, financial aid, and other student services do not perpetuate or exacerbate existing biases and discrimination based on race, gender, and socioeconomic status.
- Transparency: Making sure that the decision-making processes of AI systems used in grading, student evaluations, and other academic decisions are explainable and understandable so that students and faculty can trust and have confidence in the systems.
- Accountability: Holding universities and colleges responsible for the actions and decisions of AI systems and ensuring that there are mechanisms in place for redress and remediation in case things go wrong.
- Safety: Minimising the potential negative impacts of AI systems on students and faculty, such as privacy violations and physical harm.
- Human autonomy: Ensuring that the decisions of AI systems are consistent with human values and do not undermine human independence.

It is important to note that, while these policies are essential, they are still a work in progress and subject to change as technology, society and laws continue to evolve. Some universities or colleges may have specific AI ethics policies or guidelines, but the level of implementation and enforcement of these policies may vary.

3.3. AI and Human Displacement

Concerning AI and human displacement, policies studied in this paper argue that when AI decisions affect human life, they should be explainable, as detailed in Table 3. Human interaction with AI should be optional and not impersonated. Design choices and rationale for deployment should also be explained to ensure business model transparency. Traceability aids audibility and an explanation of AI decision making is required. A system's explainability and accuracy may need to be balanced (at the cost of explainability). The European Commission's AI ethics policy emphasizes the need for responsible development and deployment of AI. The policy calls for the promotion of transparency and explainability in AI systems to help mitigate the adverse effects of displacement. It also encourages research into ways to minimise AI's potential negative impacts on employment and create new opportunities for workers.

Table 3. AI and Human Displacement

[12]	This criterion includes data, system, and business models relevant to an AI system's transparency and is closely related to explicability. AI data collection, labelling, and algorithms should be meticulously documented to ensure traceability and transparency. This also applies to AI decision making. Such regulation helps identify an AI's wrong decision and prevent future errors. Traceability aids audibility and explanation. Machine learning systems can explain their technical processes and human decisions (e.g., system application areas). To be technically explicable, humans must understand and reconstruct AI system decisions. Additionally, a system's accuracy and predictability may suffer a considerable trade-off as it improves (at the cost of explainability). When AI decisions affect human life, they should be able to explain them. Explain this quickly and to the stakeholder's level of understanding (e.g., layperson, regulator, or researcher). The extent to which an AI system influences and shapes an organisation's decision-making process, design choices, and rationale for deployment should also be explained to ensure business model transparency (thus ensuring business model transparency). Humans interacting with AI should be informed and not impersonated. AI systems must be identifiable. AI practitioners and end-users should be informed of the system's capabilities and limitations in a way that fits the use case. Communicating the system's precision and regulations may help protect fundamental rights. However, the interaction should be optional.
[16]	The 2017 State Council New-Generation Artificial Intelligence Development Plan: An intelligent court data infrastructure was proposed was proposed that incorporates trials, staff, data applications, judicial disclosure, and active surveillance to promote the use of artificial intelligence in evidence collection, case analysis, and reading of reading of legal documents. Several jurisdictions have begun fruitful research into the legal system's use of AI. Speech recognition technology aids court recording in many domestic courts. Locally developed intelligent assistant case-handling systems for criminal cases unify evidence standards, rules, and models. Some local civil courts use a smart trial platform that lets parties participate in trials remotely. The AI assistant judge could preside. The AI assistant will guide parties through evidence presentation, cross-examination, and other courtroom procedures if they are online. More legal cases will use AI. Artificial intelligence technology will help unify case trial standards and other areas due to its training on a massive amount of case data.

[5]	The European Commission's AI ethics policy includes guidelines for addressing the potential displacement of human labour caused by the deployment of AI systems. The approach emphasizes the need for responsible development and deployment of AI, including considering potential impacts on employment and the need for retraining and social safety nets for workers affected by automation. Additionally, the policy calls for the promotion of transparency and explainability in AI systems to help mitigate the adverse effects of displacement.
[10]	The scientific method, open enquiry, intellectual rigour, honesty, and collaboration underpin technological innovation. Therefore, AI tools could advance biology, chemistry, medicine, and environmental sciences. We pursue scientific excellence in AI development. Designers will work with stakeholders to promote thoughtful leadership in this field using scientifically rigorous and multidisciplinary methods. Researchers will publish educational materials, best practices, and research to help more people create practical AI applications.
[14]	MIT's AI ethics policy states that the development and deployment of AI should be guided by ethical principles, including the responsible use of AI to avoid the displacement of human labour. The policy also encourages research into ways to mitigate AI's potential negative impacts on employment and create new opportunities for workers. Additionally, the procedure calls for collaboration between researchers, policymakers, and industry to ensure that the benefits of AI are widely shared, and its potential downsides are minimised.
[20]	Technology companies also face resistance due to their massive impact on people and democracy. Policymakers must address these issues. Stanford University's associate chair for education in computer science, Mehran Sahami, believes universities also prepare future computer scientists. "Computer scientists must consider ethical issues from the start, rather than developing technology and waiting for problems."
[2]	Carnegie Mellon's Centre for AI and Policy Research researches and teaches about AI's ethical, social, and policy implications.

To sum up, AI ethics policies in higher education concerning AI and the displacement of human labour are essential areas of concern as universities and colleges increasingly adopt and integrate AI systems into their operations. These policies aim to ensure that the use of artificial intelligence in higher education does not lead to the displacement of human labour in unfair or harmful ways. One key aspect of these policies is to ensure that AI systems are used in ways that complement and enhance human delivery rather than replace it.

This can involve providing training and support for workers to develop new skills that will enable them to work effectively alongside AI systems and create new job opportunities that take advantage of the capabilities of AI. Another critical aspect of these policies is to ensure that the displacement of human labour caused by AI is done fairly and responsibly. This can involve providing support and assistance to workers affected by AI adoption, such as retraining programmes and financial aid. It can also include ensuring that AI systems are not used in ways that perpetuate existing biases or discrimination in the workforce.

Hence, it is essential to note that implementing these policies can be challenging and requires a multidisciplinary approach involving collaboration between different departments and stakeholders. Universities and colleges may also need to adopt a proactive approach to identifying and addressing the potential labour-related impacts of AI, such as conducting impact assessments, engaging with workers and other stakeholders, and monitoring and evaluating the effects of AI on the workforce.

4. Conclusion

Transparency and explainability are essential components of responsible AI development. This criterion includes data, system, and business models relevant to an AI system's transparency and is closely related to explicability. AI data collection, labelling, and algorithms should be meticulously documented to ensure traceability and transparency, which helps identify incorrect decisions and prevent future errors. When AI decisions affect human life, they should be explainable and communicated quickly and at the stakeholder's level of understanding.

Guidelines for dealing with the potential displacement of human labour brought on using AI systems are included in the AI ethics policy of the European Commission. This policy emphasizes the need for responsible AI development and deployment, including considering potential effects on employment and the necessity of retraining and social safety nets for workers who may be affected by automation.

MIT and Stanford University's AI ethics policies state that the development and deployment of AI should be guided by ethical principles, including the responsible use of AI to avoid the displacement of human labour. They also encourage research into ways to mitigate AI's potential negative impacts on employment and create new opportunities for workers.

Carnegie Mellon University's Centre for AI and Policy Research researches and teaches about AI's ethical, social, and policy implications.

In addition, there is fruitful research on the legal system's use of AI, such as an intelligent court data infrastructure that incorporates trials, staff, data applications, judicial disclosure, and active surveillance. AI technology will help unify case trial standards and other areas due to its training on a massive amount of case data.

Finally, AI tools can advance various fields, such as biology, chemistry, medicine, and environmental sciences. Scientific excellence in AI development can be promoted by working with stakeholders to pursue thoughtful leadership in this field, using scientifically rigorous and multidisciplinary methods. Educational materials, best practises, and research can help more people create practical AI applications.

References

[1]. Butt, M. A., Qayyum, A., Ali, H., Al-Fuqaha, A. & Qadir, J. (2023). Towards secure private and trustworthy human-centric embedded machine learning: An emotion-aware facial recognition case study. *Computers and Security*, *125*. Doi: 10.1016/j.cose.2022.103058

[2]. Carnegie Mellon University. (2023). *Ethics & Artificial Intelligence Department, Dietrich College of Humanities and Social Sciences.* Carnegie Mellon University. Retrieved from: https://www.cmu.edu/dietrich/philosophy/research/areas/ethics-value-theory/ethics-ai.html [accessed: 02 February 2023]

[3]. Chang, S., Asai, T., Koyanagi, Y., Uemura, K., Maruhashi, K., & Ohori, K. (2023). Incorporating AI Methods in Micro-dynamic Analysis to Support Group-Specific Policymaking. In Aydoğan, R., Criado, N., Lang, J., Sanchez-Anguix, V., Serramia, M. (eds) *PRIMA 2022: Principles and Practice of Multi-Agent Systems.* Springer, Cham. Doi: 10.1007/978-3-031-21203-1_8

[4]. Cornacchia, G., Anelli, V. W., Biancofiore, G. M., Narducci, F., Pomo, C., Ragone, A. & di Sciascio, E. (2023). Auditing fairness under unawareness through counterfactual reasoning. *Information Processing and Management*, *60*(2). Doi: 10.1016/j.ipm.2022.103224

[5]. European Commission. (2021). *Ethics Guidelines for Trustworthy AI.* European Commission. Retrieved from: https://ec.europa.eu/futurium/en/ai-alliance-consultation.1.html [accessed: 5 March 2023]

[6]. Fairclough, N. (2009). *Critical discourse analysis is the critical study of language in London Longman*. Cambridge.

[7]. Fossen, F. M., Samaan, D., & Sorgner, A. (2022). How Are Patented AI, Software and Robot Technologies Related to Wage Changes in the United States? *Frontiers in Artificial Intelligence, 5*. Doi: 10.3389/frai.2022.869282

[8]. Gardner, A. (2022). Responsibility, Recourse, and Redress: A Focus on the Three Rs of AI Ethics. *IEEE Technology and Society Magazine*, *41*(2), 84–89. Doi: 10.1109/MTS.2022.3173342

[9]. Gomez, C., Unberath, M. & Huang, C.-M. (2023). Mitigating knowledge imbalance in AI-advised decision-making through collaborative user involvement. *International Journal of Human-Computer Studies*, *172*. Doi: 10.1016/j.ijhcs.2022.102977

[10]. Google. (2023). *Our principles – Google AI.* Retrieved from: https://ai.google/principles/ [accessed: 06 February 2023].

[11]. Huang, S. & Fang, N. (2013). Predicting student academic performance in an engineering dynamics course: A comparison of four types of predictive mathematical models. *Computers & Education*, *61*(1), 133–145. Doi: 10.1016/j.compedu.2012.08.015

[12]. INTEL.gov. (2022). *INTEL - Artificial Intelligence Ethics Framework for the Intelligence Community*. INTEL.gov. Retrieved from: https://www.intelligence.gov/artificial-intelligence-ethics-framework-for-the-intelligence-community [accessed: 07 February 2023]

[13]. Kangra, K., & Singh, J. (2023). Explainable Artificial Intelligence: Concepts and Current Progression. In Hassanien, A.E., Gupta, D., Singh, A.K., Garg, A. (eds) *Explainable Edge AI: A Futuristic Computing Perspective. Studies in Computational Intelligence*. Springer, Cham. Doi: 10.1007/978-3-031-18292-1_1

[14]. MIT Media Lab. (2020). *Overview ‹ Ethics and Governance of Artificial Intelligence.* MIT Media Lab. Retrieved from: https://www.media.mit.edu/groups/ethics-and-governance/overview/ [accessed: 06 March 2023]

[15]. Müllner, P., Schmerda, S., Theiler, D., Lindstaedt, S., & Kowald, D. (2022). Towards employing recommender systems for supporting data and algorithm sharing. *DE 2022 - Proceedings of the 1st International Workshop on Data Economy, Part of CoNEXT 2022*, 8–14. Doi: 10.1145/3565011.3569055

[16]. Ning, S., & Wu, H. (2022). *Artificial Intelligence 2022 – China.* Chambers and Partners. Retrieved from: https://practiceguides.chambers.com/practice-guides/artificial-intelligence-2022/china [accessed: 10 March 2023].

[17]. Rodgers, W., Murray, J. M., Stefanidis, A., Degbey, W. Y., & Tarba, S. Y. (2023). An algorithmic artificial intelligence approach to ethical decision making in human resource management processes. *Human Resource Management Review*, *33*(1). Doi: 10.1016/j.hrmr.2022.100925

[18]. Satterfield, D., & Abel, T. D. (2020). Ai is the new UX: Emerging research innovations in ai, user experience, and design as they apply to industry, business, education, and ethics. In Spohrer, J., Leitner, C. (eds) *Advances in the Human Side of Service Engineering.* Doi: 10.1007/978-3-030-51057-2_26

[19]. Shanklin, R., Samorani, M., Harris, S., & Santoro, M. A. (2022). Ethical Redress of Racial Inequities in AI: Lessons from Decoupling Machine Learning from Optimization in Medical Appointment Scheduling. *Philosophy and Technology*, *35*(4).
Doi: 10.1007/s13347-022-00590-8

[20]. Stanford HAI. (2023). *Ethics and Artificial Intelligence.* Stanford HAI. Retrieved from: https://hai.stanford.edu/ethics-and-artificial-intelligence [accessed: 03 February 2023].

[21]. Yoder-Himes, D. R., Asif, A., Kinney, K., Brandt, T. J., Cecil, R. E., Himes, P. R., Cashon, C., Hopp, R.M. P. & Ross, E. (2022). Racial, skin tone, and sex disparities in automated proctoring software. *Frontiers in Education*, *7*. Doi: 10.3389/feduc.2022.881449

[22]. Zhang, L., & Yencha, C. (2022). Examining perceptions towards hiring algorithms. *Technology in Society*, *68*. Doi: 10.1016/j.techsoc.2021.101848