AI Business Review

Journal homepage: https://theaibr.com/index.php/aibr

Volume 1| No. 1 |Spring 2025

When Machines Write: The Business and Ethical Impact of AI Text Automation

Hesham Allam *, Benjamin Kwasi Gyamfi, Lisa Makubvure, Kwadwo Nyarko Graham, Kehinde Akinwolere

[1] Center of Information and Communication Science (CICS), College of Communication, Information & Media, Ball State University, Muncie, IN, 47304; hesham.allam@bsu.edu, benjamin.gyamfi@bsu.edu, lisa.makubvure@bsu.edu, Kwadwo.nyarkograham@bsu.edu,kehinde.akinwolere@bsu.edu

* Correspondence: hesham.allam@bsu.edu

Abstract: Text automation powered by artificial intelligence is revolutionizing workflows and instructional materials by significantly increasing efficiency and productivity. But expeditious adoption has led to several ethical challenges that cannot wait to be addressed. This paper also explores critical moral issues, including bias and discrimination in AI-generated content, privacy breaches resulting from extensive data collection, and misinformation, which has the potential to erode public confidence, particularly in areas of great importance such as health and education. This paper examines the challenges of machine learning-based text classification, focusing on overfitting, underfitting, imbalanced datasets, training performance, and the size of the problem. It also underscores the challenge of grappling with linguistic nuances such as ambiguity, as many words have multiple meanings that can trip up text categorization. We outline practical solutions to these problems by calling for more transparent documentation, stronger tools for detecting and correcting discrimination, and models that can help explain how artificial intelligence arrives at its decisions. Finally, this study also highlights the necessity for AI developers, ethicists, policymakers, and end-users to work together to make advances in technology that are ethically consistent and beneficial to society as a whole.

Keywords: AI Ethics, Text Automation, Bias Mitigation, Data Privacy, Transparency in AI, Misinformation Detection, Accountability in AI

Academic Editor(s): Zahoor Khan
Received: 16 June 2025
Revised: 26 June 2025
Accepted: July 1, 2025
Published: date: 15 July
Citation: Allam et al. (2025). When Machines Write: The Business and Ethical Impact of AI Text Automation
AI Business Review, 1(1), x.
https://doi.org/10.64044/9sf5zg73

1. Introduction

The increased application of artificial intelligence in text automation has raised several critical ethical issues that require urgent attention. Bias, privacy, accountability, and misinformation speak to integrity and fairness within automated systems. This necessitates, most fundamentally, addressing ethical issues so that the deployment of AI technologies is responsible and does not cause harm (Hesham Allam et al., 2023; H. M. Allam, Gyamfi, & AlOmar, 2025). Some of the major concerns include bias and discrimination in AI-generated text. AI systems, since they are trained on large datasets, can reflect and reinforce existing societal biases, thereby leading to discriminatory outcomes in the generation of automated content. Research has shown that language models may incidentally favor specific demographics or viewpoints, raising questions about the fairness and inclusivity of AI-driven text automation (Illia, Colleoni, & Zyglidopoulos, 2023). Without proper interventions, biased outputs can lead to the perpetuation of stereotypes and social inequality. Other important ethical issues include those dealing with privacy and surveillance(H. Allam, Dempere, Akre, & Flores, 2023; Hesham Allam et al., 2023). A significant amount of private information is collected for training AI models, raising concerns about user consent, data protection, and surveillance. Users often unconsciously provide sensitive information to AI systems, which raises questions about how their data is stored, processed, and utilized. If left unchecked, AI-driven automation may infringe upon individuals' rights to privacy and leave them vulnerable to potential misuse of their information (Agrawal & Singh, 2025). Another relevant challenge is the accountability and transparency of AI decision-making. Many AI systems are "black boxes" whose inner workings are not explainable. Without knowledge of their inner workings, it is difficult to attribute responsibility in cases of errors or harmful outputs resulting from AI-generated content. Users and stakeholders might not understand how AI comes to conclusions, hence a lack of trust in the technology (Khan, 2023; Lockey, Gillespie, Holm, & Someh, 2021). Greater transparency in AI processes also means better prospects for the ethical and responsible deployment of AI. Additionally, there is a risk of spreading misinformation and violations of academic integrity. AI-generated text can sound very convincing, and this facility can be manipulated to create misleading or false information that affects public discourse and academic integrity. In an educational context, students may have an increased temptation to use generated content without adequate citation, which can be detrimental to original research work and intellectual growth (Lee & McLoughlin, 2007). Without ethics, AI could be used to deceive audiences and erode trust in credible sources. While these ethical concerns present a daunting set of challenges, they also underscore an urgent need for regulatory regimes and standards on the ethical use of AI in text automation. A nuanced, sliding-scale approach that addresses bias, secures data, promotes transparency, and minimizes information gaps is essential for the ethical implementation of AI. Robust governance protocols are also necessary to ensure that AI evolves in harmony with the changing needs and values of society.

This work contributes to the ongoing discussion about AI ethics and technical issues by clearly connecting the ethical implications of AI-generated text to the underlying technical limitations of large language models (LLMs). While prior research has examined these issues separately, our work combines them, illustrating how technological flaws—such as overfitting in text classification—can lead to biased outputs that exacerbate ethical concerns. By establishing direct links between machine learning constraints and AI-generated prejudice, this research gives a framework for understanding how strengthening model robustness can improve ethical outcomes. Furthermore, we propose multidisciplinary solutions, arguing for collaboration between AI ethicists and machine learning technologists to address these linked issues

2. Background

Although the technique of AI-based text automation is expected to become increasingly efficient and productive, it also presents specific ethical challenges and implications.

A. Bias and Fairness

Nevertheless, even trained on vast stores of data, AI models frequently replicate and may even intensify the biases present in the data they were trained on. One important issue is trust bias, in which decision makers defer to a computer-generated answer and do not actively search for opposing evidence, particularly under time constraints (H. Allam et al., 2025; Cummings, 2017).

B. Transparency and Explainability

Many AI systems, particularly those built on deep learning, operate as" black boxes," meaning that their internal decision-making processes are not easily understood. This lack of transparency extends to multilingual NLP models, where standard approaches usually involve learning separate language-specific embeddings. These embeddings are then mapped into a shared cross-lingual space, rather than learning a single unified embedding, because words naturally group by language (Aboagye, 2022).

C. Accountability and Responsibility

It is a complicated issue to hold someone accountable for the outputs produced by AI-powered text automation. The text, audio, images, and videos generated by AI are being misused to make nonconsensual intimate imagery, steal someone's identity, and spread disinformation or violent materials. Identification of responsible parties among developers, users, or deploying entities is often not straightforward, reiterating the necessity of such accountability mechanisms (Nightingale & Farid, 2022).

D. Privacy and Data Protection

Text automation systems using AI require a large amount of data, which impacts the privacy of data subjects. While these approaches can have the potential to provide significant benefits, the means of collecting, processing, and potentially misusing this data may pose risks to individuals that they are unaware of (Borenstein & Howard, 2021).

E. Impact on Employment

Automated tools that gather information, such as scraping a candidate's social media profile, can also result in job seekers being unfairly disqualified. Suppose employers are influenced by potentially flawed, outdated, or biased information in their hiring decisions. In that case, the privacy conundrum becomes more difficult to resolve, and the cloak-and-dagger nature of reputation systems makes it challenging for an individual to challenge or correct any possible errors (Borenstein & Howard, 2021).

3. Methodology

We pursue this project using a systematic literature review approach to investigate how limitations of AI technology intersect with concerns of a more general ethical nature. To cover a diverse set of sources, we queried multiple academic indexers with focused queries such as "AI ethics,"" bias in language models," "machine learning fairness," "text classification challenges," and" ethical issues in automation."

To maintain a balanced coverage, we included ethics-relevant PC magazines—e.g., AI and Ethics, Journal of AI & Society—as well as technical journals, such as IEEE Transactions on Neural Networks and JMLR. We also included industry white papers and policy reports to capture the progression of thinking from practitioners and regulators. While the sources were complementary, this combination overall allowed us to consider both the technological difficulties on the one hand and their ethical implications on the other in a grounded manner.

4. Challenges

A. Challenges in Machine Learning-Based Text Classification

In this section, we introduce the challenges in machine learning-based text classification and cover the crucial issues that influence the performance of these models. It considers overfitting and underfitting, which impact the generalization ability; class imbalance, which confounds classification accuracy; the complexity of the feature space, which adds to the complexity of training and interpretability; and linguistic challenges, including ambiguity and polysemy, that present formidable barriers in achieving accurate text understanding and categorization.

B. Overfitting and Underfitting in Text Classification

Overfitting and underfitting pose considerable obstacles to the efficacy of categorization models. Overfitting occurs when a model captures both the underlying patterns and the noise, resulting in high accuracy on the training data but inadequate generalization to new data. Both factors contribute to training errors that can significantly undermine the reliability of deep learning-based communication systems (Zhang, Zhang, & Jiang, 2019). Regularization, dropout layers, and data augmentation are efficacious methods for alleviating overfitting by regulating model complexity and diminishing sensitivity to specific parameters. The process of generalization is essential for mitigating overfitting and improving the model’s performance on unfamiliar data (Bu & Zhang, 2020). Underfitting occurs when a model is too simple to capture the underlying structure of the data, resulting in poor performance on both the training and test datasets. Underfitting can occur through the use of simple algorithms and inadequate feature extraction in text categorization, limiting the model's ability to understand language settings and theme variations on topics. To mitigate underfitting, model complexity must be increased through the use of complex architectures, such as transformers or pre-trained models, and the inclusion of a more diverse dataset. Moreover, regularization and dropout prevent overfitting, while it is possible to use more layers or pre-trained models to manage underfitting. These techniques enrich the generalizability of text classification models, enabling them to classify a broader range of texts more accurately (Dogra, 2022).

C. Class Imbalance in Text Classification

The problem of class imbalance in text classification arises when some categories are overrepresented in a dataset, leading to a model's preference for majority classes and, consequently, biased predictions. This challenge is especially relevant in tasks such as spam and sentiment analysis, where minority classes are most critical. Therefore, it is essential to mitigate this class imbalance to achieve the robustness and fairness of text classification models. In the domain of tropical cyclone (TC) intensity forecasting, one of the main obstacles is multi-class imbalance. Classes such as Rapidly Intensifying (RI) and Extraordinarily Intensifying (EI) often have significantly fewer training samples than more dominant stages like Neutral and Weakening. This imbalance leads to biased model performance and hinders accurate prediction of rare but critical intensity shifts (Hachiya, 2024).

The words Rapidly Intensifying (RI) and Extraordinarily Intensifying (EI) are used to describe changes in the intensity of the system. RI refers to a rapid and significant change in intensity over a relatively short period, while EI denotes a more unusual and extreme shift in strength beyond the typical change factor. Neutral is when conditions are about the same or nearly the same, while weakening applies when systems become weaker due to whatever is in their way. Such classifications offer a necessary perspective on system behavior and contribute to the precision of analysis and prediction.

Imbalance between the classes is a common occurrence due to the nature of the data. For example, sports or politics-based topics will have more data compared to niche domains of environmental news, particularly in the case of user-generated content or real-time applications. This imbalance 'pulls' models towards the majority class, resulting in poor generalization to a wide variety of scenarios. To mitigate this issue, techniques like SMOTE (Synthetic Minority Over-sampling Technique) aim to balance the dataset by oversampling the minority category and undersampling the majority category. On the other hand, models trained and evaluated on imbalanced datasets are likely to be overfitted to the majority class, yielding artificially inflated classification accuracy, thus, unreliable performance estimation (Hachiya, 2024). This class imbalance problem is typically addressed by an algorithmic method that modifies the learning algorithm to assign relatively heavier weights to classes that occur less frequently. Similar to the boosting and bagging methods, a better characterization of small categories can help increase the accuracy of the model. Participant model: Advanced models (such as BERT and GPT) also contribute to recognizing minority classes in extremely imbalanced datasets, aided by fine-tuning and cost-sensitive learning, which helps achieve more balanced and accurate classification results (Liu, Loh, & Sun, 2009).

D. Complexity in Feature Space

Pattern distribution analysis over feature spaces provides an in-depth understanding of the complexity and difficulty in various classification problems. The study of such distributions may help in discovering patterns or variations and in explaining ambiguities, thereby supporting better interpretability and performance of the model (Nagy & Zhang, 2006). The feature space in text classification –The structured dimensions or variables used to make the text data amenable to machine learning. The text is structurally unbound, with words, fragments, and syntax that are not already mapped into a numerical representation. Texts are raw and unstructured, and unlike structured data in tables or spreadsheets, they require preprocessing, such as tokenization and embedding, to make sense semantically and grammatically, resulting in a high-dimensional feature space. Such complexity brings difficulties in model training, increases computation cost, and easily causes the risk of overfitting. For this purpose, they attempted to manage the complexity of the feature space using feature selection and dimensionality reduction, which retained essential information (Gasparetto, 2022).

Classifiers trained and tested on highly imbalanced datasets may yield a falsely elevated classification accuracy estimation, which can be misleading. This happens due to the tendency of the model to bias towards the majority with a good overall accuracy, but results in poor minority class performance. Accordingly, accuracy is not a sufficient performance measure for models in the imbalanced setting, which calls for more reliable assessment criteria, such as precision, recall, F1-score, and AUC-ROC (Nagy & Zhang, 2006). The high-dimensionality and sparsity of text features in the process of overlapping computation and pattern recognition necessitate good feature selection to efficiently maximize model performance and accuracy in the study (Nagy & Zhang, 2006). The sparsity in high-dimensional feature space restricts its generalization performance and makes the training process longer. Simple representations such as bag-of-words fail to account for language phenomena, especially polysemy and synonyms. To address such a complex issue, a multi-dimensional feature space can be introduced, allowing models to represent more complex language constructs with greater power, and thereby improving classification accuracy (Le, 2012). Feature engineering is crucial for making text data readable and comprehensible. Techniques such as TF-IDF and n-grams help models pick up essential words and phrase structures. In contrast, word embeddings, including Word2Vec, GloVe, and fastText, provide condensed feature vectors to facilitate generalization. State-of-the-art models, such as BERT and GPT, produce contextually coded representations, encoding word meanings depending on their context for enhanced text comprehension (Mars, 2022). Dimensionality reduction techniques like PCA, SVD, and autoencoders maximize feature space, retain important features, improve model interpretability, and reduce training time. State-of-the-art embeddings, such as BERT and GPT, enhance text classification by leveraging contextual information, thereby improving performance on complex languages. However, these developments give rise to interpretability issues, because deep learning models are often treated as "black boxes", which represent an indispensable challenge for areas such as healthcare and finance in which interpretability is crucial. Attention mechanisms and XAI tools aim to resolve this problem by emphasizing essential features while maintaining a trade-off between complexity and interpretability. These strategies enable well-informed decisions in complex language processing tasks without compromising the model's efficiency and accuracy (Sinjanka, Musa, & Malate, 2024).

E. Ambiguity and Polysemy in Language

Ambiguity and polysemy pose significant challenges for most NLP tasks, particularly in text categorization. Ambiguity refers to polysemy, where a word or short phrase has multiple meanings, such as the word "bank," which can be interpreted as a financial institution or the land alongside a river. Polysemy, a type of ambiguity, is the property in which a single word describes diverse concepts. For example, as applied in English, "run" means both a physical movement and the execution of a program. The difficulties associated with these complexities make it challenging to achieve good model accuracy, as nuance must be understood contextually for accurate interpretation, and this is an area where general models often struggle. In text classification, ambiguity arising from word sense can result in misclassification and may require context-aware solutions (Bashiri & Naderi, 2024). "Local bank raises money" is a headline that requires local knowledge to disambiguate between financial and non-financial implications. Simple models often misclassify these examples; even sophisticated neural architectures, such as the transformer, can fail to recognize these types of examples with subtle or culturally inflected cues. This also highlights the need for more advanced strategies to handle context for improved classification (Yadav, Patel, & Shah, 2021).

Polysemy is a key challenge in this work, as static word embeddings fail to capture contextual homonyms. For example, the word "light" can mean either brightness or weight, depending on the context. Contextual embeddings, such as those produced by BERT and GPT, dynamically shift meaning according to the words they accompany, but they still struggle to comprehend complex expressions and subtle meanings. With multilingual NLP, text categorization is further complicated by variations in ambiguity and polysemy between languages. Some languages use morphology to disambiguate between the senses of a polysemous word, while others rely more heavily on context, thus complicating the performance of, e.g., machine translation tasks. Multilingual bi-lingual models, such as mBERT, trained on a wide range of data, are deployed to address these problems, but linguistic diversity continues to limit coverage (Seneviratne, 2024).

There are several methods for handling ambiguity and polysemy in NLP. Far-reaching domain-specific models enhance contextual comprehension, minimizing misclassification, and auxiliary tasks, such as part-of-speech tagging, assist in making meaning explicit. For example, ensembles of models that aggregate estimations from different models achieve better accuracy. However, such methods are computationally expensive, demonstrating the continued difficulty that ambiguity and polysemy present to NLP (Garg, 2021).

5. Ethical Issues Related to AI-Based Text Automation

This article highlights the symbiotic relationship between the technical limitations of AI systems and the ethical issues they give rise to. More specifically, problems such as class distribution in training data and linguistic ambiguity, particularly polysemy, have been shown to play a crucial role in the spread of disinformation and to lead to biased or discriminatory outcomes. For example, imbalanced dataset-trained LLMs could induce biased societal bias learning since they might learn biased representations that overemphasize or leverage some demographic groups, while ignoring others. Similarly, the imprecise nature of natural language can lead to misunderstandings and yield incorrect or contextually irrelevant answers.

Resolving these ethical concerns demands a dedicated focus on solving the underlying technical limitations. Better feature selection policies and the incorporation of fairness-aware learning frameworks can lead to a more systematic and complex model, thereby minimizing the potential for overfitting and unfair bias. Tackling and adequately addressing these technical ethical dependencies is key to the emergence of strong and responsible AI governance frameworks. AI-powered text automation technology has significantly enhanced what can be done – and needs to be done – in various fields, including content creation and customer service, among others. Employing AI to automate text-oriented work may enhance efficiency, save money, and provide creative new options; nevertheless, it also raises serious ethical issues that must be addressed to ensure ethical deployment. These ethical dilemmas concern transparency, accountability, bias, data privacy, job displacement, and the misuse of AI, among others (Kumar, Verma, & Mirza, 2024). In this paper, we examine the primary ethical implications associated with AI-powered textual automation.

A. Transparency and Accountability

Text automation Systems based on AI, particularly ML-based systems, can often act as "black boxes," meaning the rationale behind the outputs is not easily understandable to human beings (Chaudhary, 2024). The opacity of AI systems makes it even more challenging for users to understand how AI content is generated and how automated decisions are made. It is challenging to determine accountability for AI-generated content when the decision-making processes are opaque and poorly understood. For instance, in the context of AI-generated content, it may be ambiguous whether the issue lies with the AI software, the training data, or even the individuals and organizations that deploy the tool. And this uncertainty in accountability adds another layer of ethical murk too: Who is responsible for any damage or violation caused directly by AI-powered text automation?

B. Bias and Discrimination

A long-standing ethical challenge of AI-enabled text automation that people have been concerned about is the risk of bias and discrimination (Hanna, 2024). AI systems require extensive data for training, and if this data is biased in any way — either intentionally or unintentionally —that bias will be reflected in the AI's output and may even be amplified. If an AI model learns from textual data that reflects societal biases on gender, ethnicity, or social status, the model can produce results that essentially reinforce or exacerbate those biases. In content creation, this can result in biased or harmful representations in auto-generated articles, news stories, or advertisements. Similarly, in Customer Service, biased language models have the potential to perpetuate or exacerbate unfair treatment or discrimination against specific user groups, thereby reinforcing social biases (Gallegos, 2024).

To mitigate these risks, AI models should be trained on diverse and representative data that reasonably reflects the demographics they serve. What's more, continuous audits and analysis of AI-generated content to pinpoint and fix biases that would cause harm to people or groups are needed.

C. Data Privacy and Security

It takes enormous amounts of personal data, preferences, interactions, etc, for text-based AI automation models to work well. So, security and privacy are the biggest concerns. Automated text-based systems (e.g., chatbots, recommendation systems) routinely collect sensitive user data to personalize interactions or enhance system performance (Chen, 2024). Without adequate protection of this information, unauthorized access and disclosure are possible.

Moreover, AI systems can sometimes inadvertently collect an excessive amount of data, raising potential concerns about data privacy and surveillance. For instance, text autoreply systems used in customer service may collect sensitive personal information during the chat process, which could be regarded as an infringement of user privacy rights. Organizations developing and deploying AI systems must put in place strong data protection policies, gain informed consent from users, and, where relevant, ensure that data is anonymized or encrypted.

D. Workforce Disruption and Financial Disparities

And it’s not just disrupting labor markets; AI text automation is going to do for jobs what previous technical advances have done for job structures. For instance, the deployment of telephone switchboards in the early 20th century created thousands of operator positions, all of which are now redundant thanks to automated switching. Similarly, gas station attendants were commonly found offering fill-up services; however, automation has significantly reduced the presence of such jobs (Hesham Allam et al., 2025). These historical examples suggest that while automation can remove classes of jobs, it also often generates new jobs in areas that are closely related. Though traditional writing and editing jobs may fade away as AI takes over automated text, the future could see demand swell for AI auditors, prompt engineers, and oversight professionals. This paper suggests proactive workforce-resilience strategies, such as mass reeducation drives and ethical AI lessons, to mitigate the havoc of automation-driven turmoil.

Road to further improvements. The widespread use of AI-powered text generation could have effects on work and jobs in the future. With AI systems catching up to and overtaking jobs that people have traditionally done (writing, answering customer inquiries, and transcribing), we've been afraid that specific jobs might be gone forever. While AI could be effective and efficient, it may also lead to reduced demand for human labor, particularly in industries that rely heavily on text-based activities.

This economic shift could exacerbate economic inequality, as those with less education or fewer tech skills may have difficulty adapting to an AI-fueled economy. Additionally, the benefits of AI-based automation can primarily benefit large industries and tech companies, which have the capital to invest in advanced AI technologies, leaving smaller businesses and individuals behind (Challoumis, 2024). Policy makers need to ensure that workers have the skills and training they need to transition into new jobs and that social safety nets are in place for those who are displaced by technology.

E. Intellectual Property and Ownership

AI-generated text raises important questions regarding intellectual property (IP) and authorship (Gaffar & Albarashdi, 2024).

F. Misinformation and Fake Content

AI text generation has the potential to create highly realistic fake content, such as fake news stories, disingenuous social media posts, and fabricated reviews or endorsements. As generated text by AI can be so close to what is written by humans, this is a serious issue for the trustworthiness of the information on the web (Hayawi, Shahriar, & Mathew, 2024). Malicious parties can use this ability to disseminate false information, influence public opinion, or meddle in political and social affairs. AI-based fake news and propaganda can be highly realistic in situations like an election or a public health crisis, leading people to believe false stories and even have a real-world impact. The way to address this problem is to develop tools that can recognize AI-generated content and distinguish it from real human writing. Organizations and governments must also ramp up public education on the risks of AI-driven misinformation and how to use AI technologies ethically.

G. Ethical Use in Sensitive Contexts

AI-driven text automation is also utilized in areas where sensitive needs exist, such as mental health counseling, legal advice, and medical diagnosis. Although AI systems can provide efficient and convenient services to consumers, there is concern about their ability to respond appropriately to sensitive and complex emotional or ethical matters. However, an AI-driven mental health chatbot that cannot fully understand human emotional status or exhibit genuine empathy in its communication may cause unexpected side effects. In law and medicine, for example, AI systems may lack the necessary human experience to make ethical decisions or comprehend complex situations (Nightingale & Farid, 2022). Depending entirely on automation in these domains may compromise the quality of care, support, or advice that humans receive, as AI systems may struggle to resolve ethical quandaries that arise in these complex and sensitive areas.

6. Discussion

The growing presence of AI-powered text generation, such as GPT-2, has proven to be a gold mine for creating game-changing workflows and educational resources. But the feverish pace at which it’s infiltrating different professions poses enormous ethical challenges that need to be addressed. One obvious worry is bias and discrimination. Generative AI models are developed on large datasets, which may be biased, and this bias can be transferred to the developed model, further producing biased outputs and amplifying societal bias reflected in the model outputs (Rao, 2024). This is especially troubling as biased data can compromise the fairness and diversity of automated systems.

The issue of privacy is also a primary ethical concern. Unfortunately, AI in educational and other sensitive domains requires the collection and processing of large datasets. The collection of such rich data can be prone to privacy violations if adequate precautions are not taken, putting sensitive information at stake (Sargiotis, 2024; Torres, 2024). At the same time, misinformation poses a significant risk as well. The ability of AI to create content that is believable but untrue threatens trust and informed discussion, particularly about important issues such as health (Williamson & Prybutok, 2024).

To mitigate these challenges, several strategies have been proposed:

Transparency and Documentation: Clear documentation of how data has been gathered, and an explanation of the AI's contribution are required to maintain integrity, especially in academic writing (Williamson & Prybutok, 2024). This is beneficial because it ensures that every party involved is aware of the foundation of AI-created content.

Bias Detection and Correction: Building and deploying frameworks that can detect and mitigate biases in the outputs of AI systems, thereby increasing their fairness. Such strategies could help to reduce the perpetuation of harmful stereotypes and to ensure fair outcomes (Rao, 2024).

Explainability Tools: Explanation tools can foster increased trust in the model by showing users how decisions are made within the AI model, especially in situations where transparency is crucial (Ranjan, 2024; Williamson & Prybutok, 2024). These instruments enable the “man on the street” to unpack the headroom in AI to inform better, more informed decisions.

These remedies emphasize the need for collaboration among AI developers, ethicists, policymakers, and end-users. An interdisciplinary approach of this sort is essential if we are to establish strong frameworks that guarantee transparency and accountability in the use of AI-driven text automation.

6. Conclusion

In conclusion, AI-based text automation holds the promise of transformational efficiency and innovation in workflows and education, yet it faces a set of ethical challenges. The emergence of biased outcomes, infringement of privacy, and the propagation of false information serve to underscore the imperative of robust and participatory ethical guidelines. Highlighting transparency, bias correction methods, and the deployability of explainability tools can further help participants in coping with these challenges. The question of considering ethics in technological development is all the more difficult, yet necessary. Further close examination and intervention will be required for the emerging AI-based text automation to be beneficial to societies without risk.

Acknowledgement

The authors acknowledge that all research, analysis, and writing were conducted independently without external assistance. Grammarly, an AI-powered writing assistant, was used solely during the revision stage to support minor improvements in grammar,

clarity, and readability, without contributing to the intellectual content or structure of the manuscript.

References

Aboagye, P. O. (2022). *Normalization of language embeddings for crosslingual alignment.* Paper presented at the International Conference on Learning Representations. https://openreview.net/forum?id=Nh7CtbyoqV5

Agrawal, A., & Singh, J. (2025). The Dark Side of AI: Risks, Ethics, and Safeguarding Human Interests. In *Ethical AI Solutions for Addressing Social Media Influence and Hate Speech* (pp. 131-162): IGI Global Scientific Publishing. DOI: 10.4018/979-8-3693-9904-0.ch007

Allam, H., Dempere, J., Akre, V., Parakash, D., Mazher, N., & Ahamed, J. (2023). *Artificial intelligence in education: an argument of Chat-GPT use in education.* Paper presented at the 2023 9th International Conference on Information Technology Trends (ITT). https://ieeexplore.ieee.org/abstract/document/10184267

Allam, H., Dempere, J., Lazaros, E., Davison, C., Kalota, F., & Hua, D. (2025). *Unleashing educational potential: Integrating chatgpt in the classroom*. Paper presented at the HCT International General Education Conference (HCTIGEC 2024). https://www.atlantis-press.com/proceedings/hctigec-24/126008621

Allam, H., Makubvure, L., Gyamfi, B., Graham, K. N., & Akinwolere, K. (2025). Text Classification: How Machine Learning Is Revolutionizing Text Categorization. *Information, 16*(2), 130. **https://doi.org/10.3390/info16020130**

Allam, H. M., Gyamfi, B., & AlOmar, B. (2025). Sustainable innovation: Harnessing AI and living intelligence to transform higher education. *Education Sciences, 15*(4), 398. **https://doi.org/10.3390/educsci15040398**

Bashiri, H., & Naderi, H. (2024). Comprehensive review and comparative analysis of transformer models in sentiment analysis. *Knowledge and Information Systems*, 1-57. https://link.springer.com/article/10.1007/s10115-024-02214-3

Borenstein, J., & Howard, A. (2021). Emerging challenges in ai and the need for ai ethics education. *AI and Ethics, 1*, 61-65. https://link.springer.com/article/10.1007/s43681-020-00002-7

Bu, C., & Zhang, Z. (2020). *Research on overfitting problem and correction in machine learning.* Paper presented at the Journal of Physics: Conference Series. **DOI** 10.1088/1742-6596/1693/1/012100

Challoumis, C. (2024). *Building a sustainable economy-how ai can optimize resource allocation.* Paper presented at the XVI International Scientific Conference. https://conference-w.com/wp-content/uploads/2024/10/USA.P-0304102024.pdf#page=191

Chaudhary, G. (2024). Unveiling the black box: Bringing algorithmic transparency to ai. *Masaryk University Journal of Law and Technology, 18*(1), 93-122. https://www.ceeol.com/search/article-detail?id=1287954

Chen, J. (2024). When large language models meet personalization: Perspectives of challenges and opportunities. *World Wide Web, 27*(4), 42. https://link.springer.com/article/10.1007/s11280-024-01276-1

Cummings, M. L. (2017). Automation bias in intelligent time critical decision support systems. In (pp. 289-294): Routledge. https://www.taylorfrancis.com/chapters/edit/10.4324/9781315095080-17/automation-bias-intelligent-time-critical-decision-support-systems-cummings

Dempere, J., Modugu, K., Allam, H., & Ramasamy, L. K. (2023). The impact of chatgpt on higher education. *Frontiers in Education, 8*, 1206936. https://doi.org/10.3389/feduc.2023.1206936

Dogra, V. (2022). A complete process of text classification system using state-of-the-art nlp models. *Computational Intelligence and Neuroscience, 2022*(1), 1883698. **https://doi.org/10.1155/2022/1883698**

Gaffar, H., & Albarashdi, S. (2024). Copyright protection for ai-generated works: Exploring originality and ownership in a digital landscape. *Asian Journal of International Law*, 1-24. DOI: https://doi.org/10.1017/S2044251323000735

Gallegos, I. O. (2024). Bias and fairness in large language models: A survey. *Computational Linguistics*, 1-79. https://doi.org/10.1162/coli_a_00524

Garg, R. (2021). Potential use-cases of natural language processing for a logistics organization. In (pp. 157-191): Springer. https://link.springer.com/chapter/10.1007/978-3-030-68291-0_13

Gasparetto, A. (2022). A survey on text classification algorithms: From text to predictions. *Information, 13*(2), 83. **https://doi.org/10.3390/info13020083**

Hachiya, H. (2024). Multi-class auc maximization for imbalanced ordinal multi-stage tropical cyclone intensity change forecast. *Machine Learning with applications, 17*, 100569. https://doi.org/10.1016/j.mlwa.2024.100569

Hanna, M. (2024). Ethical and bias considerations in artificial intelligence (ai)/machine learning. *Modern Pathology*, 100686. https://doi.org/10.1016/j.modpat.2024.100686

Hayawi, K., Shahriar, S., & Mathew, S. S. (2024). The imitation game: Detecting human and ai-generated texts in the era of chatgpt and bard. *Journal of Information Science*, 01655515241227531. https://doi.org/10.1177/01655515241227531

Illia, L., Colleoni, E., & Zyglidopoulos, S. (2023). Ethical implications of text generation in the age of artificial intelligence. *Business Ethics, the Environment & Responsibility, 32*(1), 201-210. **https://doi.org/10.1111/beer.12479**

Khan, S. (2023). The Ethical Imperative: Addressing bias and discrimination in AI-driven education. *Social Sciences Spectrum, 2*(1), 89-96. https://sss.org.pk/index.php/sss/article/view/23

Kumar, S., Verma, A. K., & Mirza, A. (2024). Digital revolution, artificial intelligence, and ethical challenges. In (pp. 161-177): Springer. https://link.springer.com/chapter/10.1007/978-981-97-5656-8_11

Le, P. Q. (2012). Representing visual complexity of images using a 3d feature space based on structure, noise, and diversity. *Journal of Advanced Computational Intelligence, 16*(5). https://pure.bit.edu.cn/en/publications/representing-visual-complexity-of-images-using-a-3d-feature-space

Lee, M. J., & McLoughlin, C. (2007). Teaching and learning in the Web 2.0 era: Empowering students through learner-generated content. *International journal of instructional technology and distance learning, 4*(10), 1-17. https://researchoutput.csu.edu.au/en/publications/teaching-and-learning-in-the-web-20-era-empowering-students-throu-2

Liu, Y., Loh, H. T., & Sun, A. (2009). Imbalanced text classification: A term weighting approach. *Expert systems with applications, 36*(1), 690-701. https://doi.org/10.1016/j.eswa.2007.10.042

Lockey, S., Gillespie, N., Holm, D., & Someh, I. A. (2021). *A review of trust in artificial intelligence: Challenges, vulnerabilities and future directions*. https://aisel.aisnet.org/hicss-54/os/trust/2/

Mars, M. (2022). From word embeddings to pre-trained language models: A state-of-the-art walkthrough. *Applied Sciences, 12*(17), 8805. **https://doi.org/10.3390/app12178805**

Nagy, G., & Zhang, X. (2006). Simple statistics for complex feature spaces. In (pp. 173-195): Springer.https://link.springer.com/chapter/10.1007/978-1-84628-172-3_9

Nightingale, S. J., & Farid, H. (2022). Ai-synthesized faces are indistinguishable from real faces and more trustworthy. *Proceedings of the National Academy of Sciences, 119*(8), e2120481119. https://doi.org/10.1073/pnas.2120481119

Rao, T. V. N., Stephen, M., Manoj, E., & Sangers, B. (2025). Exploring Bias and Fairness in Machine Learning Algorithms. *Innovations in Optimization and Machine Learning*, 369-398. DOI: 10.4018/979-8-3693-5231-1.ch014

Sargiotis, D. (2024). Data security and privacy: Protecting sensitive information. In *Data governance: a guide* (pp. 217-245): Springer Nature Switzerland. https://link.springer.com/chapter/10.1007/978-3-031-67268-2_6

Seneviratne, I. S. (2024). *Text simplification using natural language processing and machine learning for better language understandability*. The Australian National University, https://web.archive.org/web/20240403234313id_/https://openresearch-repository.anu.edu.au/bitstream/1885/313769/1/Sandaru_Seneviratne_PhD_Thesis_2024.pdf

Shanmugam, L., Tillu, R., & Jangoan, S. (2023). Privacy-preserving ai/ml application architectures: Techniques, trade-offs, and case studies. *Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online), 2*(2), 398-420. **DOI:** https://doi.org/10.60087/jklst.vol2.n2.p420

Singh, A., Kathait, S., Kothari, A., Joshi, S., Agarwal, Y., Badoni, S., . . . Mishra, P. (2024). Beyond the Black Box: XAI strategies for safeguarding critical infrastructure. In *Data Protection: The Wake of AI and Machine Learning* (pp. 129-154): Springer. https://link.springer.com/chapter/10.1007/978-3-031-76473-8_7

Sinjanka, Y., Musa, U. I., & Malate, F. M. (2024). *Text analytics and natural language processing for business insights: A comprehensive review*. https://doi.org/10.22214/ijraset.2023.55893

Williamson, S. M., & Prybutok, V. (2024). The Era of Artificial Intelligence Deception: Unraveling the Complexities of False Realities and Emerging Threats of Misinformation. *Information, 15*(6), 299. **https://doi.org/10.3390/info15060299**

Yadav, A., Patel, A., & Shah, M. (2021). A comprehensive review on resolving ambiguities in natural language processing. *AI Open, 2*, 85-92. https://doi.org/10.1016/j.aiopen.2021.05.001

Zhang, H., Zhang, L., & Jiang, Y. (2019). *Overfitting and underfitting analysis for deep learning based end-to-end communication systems.* Paper presented at the 2019 11th International Conference on Wireless Communications and Signal Processing (WCSP).**DOI:** 10.1109/WCSP.2019.8927876