Optimizing Schools: An Ethical Analysis of AI Integration in Education

Department of Computer Science, Virginia Tech

CS 5024: Ethics and Professionalism in Computer Science

Adeyemi Aina

10/12/24

Optimizing Schools: An Ethical Analysis of AI Integration in Education

Imagine an institution with a crystal ball capable of predicting which students are at risk of dropping out. In 2012, Minerva High School in Pittsburgh, Pennsylvania, faced a critical challenge: soaring dropout rates and declining academic performance, with only 55% of students graduating on time compared to the state average of 76.4%. To address these issues, the school administration proposed leveraging artificial intelligence (AI) to analyze extensive student data, including academic records, attendance, and behavioral patterns, to predict at-risk students and recommend interventions. While promising to improve outcomes, this initiative also raised significant ethical concerns around privacy, fairness, and the potential overreach of technology in education.

This case highlights the intersection of technology, society, and ethics, where AI offers transformative opportunities to identify struggling students and enhance their well-being through tailored interventions. However, it also presents risks, including algorithmic bias, data misuse, and a shift away from human-centered education. The Minerva High School case underscores broader ethical challenges in integrating AI into public institutions, particularly those serving vulnerable populations, prompting critical questions about balancing innovation with respect for individual rights and whether technological efficiency should outweigh traditional educational values. This analysis explores these dilemmas through ethical frameworks, offering insights into the responsible deployment of technology in society.

A closer look at Minerva High School, in 2012, Minerva High School in Pittsburgh, Pennsylvania, was battling with a severe dropout crisis. With only 55% of its students graduating on time, which was far below the state average of 76.4%, the school was at risk of losing funding and potentially facing closure. Pressured by the school board to address these challenges, Principal Vulcani turned to an unconventional solution: leveraging the school's comprehensive datasets and artificial intelligence (AI). The school had been collecting a lot of data on student behavior, including academic performance, disciplinary actions, attendance records, internet logs and even activity and from scannable ID cards. These cards tracked everything from library visits and class attendance to cafeteria purchases. They employed a data science organization called "*Hephaestats*". The organization proposed an AI system, aimed to analyze the school's data collection and identify students at risk of dropping out, then recommend tailored interventions, such as counseling or adjustments to their academic pathways.

The plan held some promise for improving student outcomes, however it also raised ethical concerns. The key stakeholders, which are students, parents, teachers, and administrators,

expressed divergent views on the initiative, there were pushbacks from various stakeholders: teachers expressed concerns, students launched petitions, and parents sought to opt their children out of the AI system as they were worried about privacy and potential misuse of personal information. Teachers also questioned whether data-driven decisions might undermine their decisions and overlook the nuances of individual student circumstances. Administrators also faced some challenge of balancing the urgency to improve outcomes with the ethical responsibility to protect students' rights and dignity.

After implementing the system, the school identified eight key indicators that could predict dropout rates with 92% accuracy. The case does not explicitly list these indicators, but it provides some examples of the types of factors considered, such as administrative issues, external influences, and nutritional options in the school cafeteria, which influenced the system's recommendations. By the end of the 2016/2017 academic year, following the deployment of the AI system developed by Hephaestats, Minerva High School's four-year graduation rate increased significantly, from 55% to 85%.

The outcome was impressive, creating the impression that the AI system's recommendations solely drove the improvement, but it might be important to consider correlation might not always imply causation. Other unexamined factors, such as values, interpersonal dynamics, or broader human elements, may have also contributed to the results. This case illustrates a complex ethical landscape where competing interests collide: the need for innovation in addressing educational inequities, the need to safeguard privacy and autonomy, and the risk of embedding biases into systems that disproportionately affect vulnerable populations. The Minerva High School case provides a lens through which to explore the ethical dilemmas inherent in using AI for societal decision-making, particularly in public institutions. By applying some ethical frameworks, we can better understand the moral complexities and guide responsible decision-making in this context.

Utilitarianism, a consequentialist ethical framework, evaluates the morality of actions based on outcomes and aims to maximize happiness. In the case of Minerva High School, this framework offers a lens to assess whether implementing AI to predict at-risk students serves the broader societal good. From a utilitarian perspective, the benefits of leveraging AI in education are significant. By identifying struggling students early, the system can recommend targeted interventions, such as personalized counseling, tutoring programs, or modified curricula. These measures could improve academic outcomes, reduce dropout rates, and increase the likelihood of students achieving long-term success. The ripple effects of these improvements extend beyond individual students, to the society, as better-educated individuals contribute positively to their

communities, workforce, and economic stability. Studies, such as those by (Bates et al. 2020) support this view, demonstrating that AI-driven predictive systems enhance educational success by enabling schools to provide timely, tailored support to at-risk students.

However, utilitarianism also necessitates evaluation of potential harms. The extensive data collection required by AI raises concerns about privacy and the potential misuse of sensitive information. If this data is mishandled or breached, the harm to students' trust and personal security could outweigh the intended benefits mentioned. Furthermore, the effectiveness of the AI system depends on the quality of its algorithms and training data. Also biases within the algorithms could disproportionately flag students from marginalized communities as at-risk, exaggerating some systemic inequalities rather than addressing them. Studies by (Eubanks, 2019) highlights how data-driven decision-making in public systems often stigmatizes and marginalizes the groups it intends to support, underscoring the need for vigilance in the system's design and implementation, which might be a potential issue in Minerva case.

The psychological impact of surveillance is another factor to consider. Students who feel closely monitored by AI may experience some anxiety, which could affect their mental well-being. Teachers, too, may feel undermined if AI recommendations override their professional judgment, creating some tension with actual human expertise. For the AI system at Minerva High School to align with utilitarian principles, the institution must implement robust safeguards, which includes transparency about data usage, regular algorithmic audits to mitigate biases, and strong privacy protections, which is an undoubtedly complex and challenging endeavor, even by 2024 standards.

Deontological ethics, rooted in the philosophy of Immanuel Kant, focuses on adherence to moral duties and principles rather than outcomes. This framework emphasizes respect for individual rights, autonomy, and the inherent dignity of each person, making it particularly relevant to the ethical concerns surrounding AI implementation at Minerva High School. Deploying AI systems to track and predict student behavior raises critical questions about whether students and their families fully understand and agree to how their personal data is used. According to this framework, failing to obtain explicit consent violates the moral duty to respect individuals' autonomy. Transparency is also an essential factor, students and parents must be informed about the nature of the data collected, how it is analyzed, and the purposes it serves. Without such transparency, the system undermines trust and autonomy.

Deontological ethics also stresses the importance of treating individuals as ends in themselves, not as means to institutional goals. The AI may help achieve objectives like boosting graduation rates or securing funding, however it risks reducing students to data points in a broader analysis process.

This system by Minerva overlooks the unique needs, aspirations, and contexts of each student, leading to one-size-fits-all solutions that fail to account for individual circumstances. For example, relying solely on AI-generated recommendations might sideline human judgment, diminishing the moral obligation to treat students with dignity and care.
Fairness is another critical consideration under this framework. If the AI system disproportionately impacts certain groups, such as students from lower socioeconomic backgrounds, it fails to uphold the principle of equality. (Garrett, 2019) reinforces this point, highlighting how algorithmic systems often replicate societal biases and must be carefully designed to avoid perpetuating inequities. Ensuring that the system is free from bias and equitable in its recommendations is not just a practical necessity but a moral imperative.

Finally, deontological ethics highlights the duty of accountability. Those who design, implement, and oversee the AI system bear a responsibility to ensure its ethical use. If the system causes harm or unfair treatment, mechanisms must be in place to address grievances. Accountability extends to regular audits, ensuring that the system adheres to ethical principles and remains transparent in its operations. The Minerva High School case also reveals broader societal tensions, such as privacy versus optimization. Schools must strike a balance between respecting privacy and deriving actionable insights. Similarly, equity versus efficiency arises as a significant clash. AI's potential to optimize processes must not come at the expense of fairness. As described in the previous paragraph, regular algorithmic audits and the use of inclusive datasets are crucial to mitigating these risks. Lastly, the tension between human judgment and automation underscores the importance of leveraging AI as a supportive tool rather than replacing Teacher's expertise. Teachers' nuanced understanding of individual students must remain central to decision-making.

The studies from (Sweeney, 2013) highlights the risks of deploying AI systems without adequate safeguards, particularly in contexts involving vulnerable populations. Their findings underscore the importance of transparency and accountability in building trust and ensuring ethical AI deployment. By integrating these insights, Minerva High School can design a system that aligns with both utilitarian and deontological principles, addressing concerns about privacy, fairness, and accountability.

The ethical analysis of Minerva High School's AI initiative demonstrates that while the technology holds significant promise, its success depends on navigating complex moral and societal challenges. Utilitarianism underscores the importance of maximizing benefits, urging the school to ensure transparency and equity. Deontological ethics highlights the need to respect individual rights, prioritize fairness, and maintain accountability.

Ultimately, this case serves as an example for broader debates about the role of AI in public institutions. By addressing privacy concerns, mitigating bias, and preserving human oversight, Minerva High School can harness the power of AI responsibly, creating a model for ethical innovation in education.

References

Princeton University. (2018). Case Study 3: Optimizing Schools. Princeton Dialogues on AI and Ethics. https://aiethics.princeton.edu/wp-content/uploads/sites/587/2018/10/Princeton-AI-Ethics-Case-Study-3.pdf

Bates, L. (2018). Justice for victims of sexual abuse and harassment: Lessons for westminster? Journal of Gender-Based Violence, 2(2), 397–402. https://doi.org/10.1332/239868018X15263882926256

Eubanks, Virginia. (2019). Automating inequality: how high-tech tools profile, police, and punish the poor. Picador, St. Martin's Press.

Garrett, Y. C. (2019). Review of Algorithms of Oppression: How Search Engines Reinforce Racism. In Journal of Contemporary Archival Studies (Vol. 6).

Sweeney, L. (2013). Discrimination in Online Ad Delivery. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.2208240

Bates, T., Cobo, C., Mariño, O., & Wheeler, S. (2020). Can artificial intelligence transform higher education? International Journal of Educational Technology in Higher Education, 17(1). https://doi.org/10.1186/s41239-020-00218-x