Locating Responsibility in Data Science: Can Ethics Be Outsourced to AI Systems?

Introduction

This paper is grounded primarily in Leonelli's "Locating Ethics in Data Science", which argues that ethical responsibility in data science cannot be separated from everyday technical practices. Leonelli emphasizes that data science is inherently distributed, involving multiple actors, infrastructures and decisions, which makes responsibility and accountability difficult to locate (Leonelli, 2016). Ethics, she insists, is not an external constraint imposed after the fact, but an integral part of how data is collected, processed, interpreted and reused.

This concern resonates strongly with a recurring theme: the temptation to treat ethics as something that can be automated, delegated, or abstracted away. Across the readings and discussions, especially in Floridi and Taddeo's "What Is Data Ethics", ethics is framed as a macro-ethical problem, one that emerges from systems rather than isolated decisions (Floridi & Taddeo, 2016). Yet, as AI systems become more autonomous, opaque and persuasive, there is a growing tendency to assume that AI itself might resolve ethical problems, by auditing bias, enforcing fairness, or even reviewing human work.

This assumption is explicitly challenged in, initially your comment, my comment on Artificial Thinking and Doomsday Projections, where you noted that "AI will not solve AI problems", drawing a parallel to 2001: A Space Odyssey and the illusion of control through automation. That observation motivates the central question of this paper, which connects Leonelli's argument to concerns raised repeatedly in class discussions.

Leonelli writes that ethical reasoning should be "an integral part of data science" rather than an external add-on. Building on this claim, and on my concern about AI systems confidently performing tasks they are not qualified to do, such as reviewing academic manuscripts, the guiding question of this paper is:

If responsibility and accountability in data science are already distributed and fragile, what do we lose when ethical judgment itself is outsourced to AI systems?

This question matters because it addresses not only technical risk, but a deeper moral danger: the erosion of human ethical agency. The issue is not merely whether AI systems make mistakes, hallucinate, or reinforce bias, but whether reliance on them reshapes how humans understand responsibility, expertise and judgment.

A useful comparative text that can also be discussed is Floridi's "Five Risks of Being Unethical", which shows how ethical frameworks can be misused, diluted, or instrumentalized when ethics becomes performative rather than practiced (Floridi, 2019). Together with Leonelli, this text helps frame the problem not as a failure of regulation alone, but as a failure of ethical participation.

Discussion

Leonelli's central insight is that the distributed nature of data science makes accountability difficult but unavoidable. Data is produced across many institutions, countries and infrastructures, and ethical consequences often emerge far from the original site of decision-

making. Because of this, Leonelli argues for a participative and reflexive approach, where individuals remain responsible for their local contributions even within global systems (Leonelli, 2016).

However, the increasing use of AI systems to evaluate, monitor and even critique human work introduces a new layer of ethical displacement. Your comment about asking an AI to review a journal manuscript captures this tension vividly. While the AI was useful for detecting surface-level issues, its confidence in performing expert peer review was “deeply concerning”. This moment illustrates the core problem: AI systems do not know when they should not speak, yet they speak authoritatively.

This connects directly to the concern raised in Artificial Thinking and Doomsday Projections about hallucination, trust and automation. The danger is not only that AI produces false content, but that humans begin to trust AI judgment in domains that require lived expertise, interpretive depth and moral sensibility. When ethical oversight itself is automated, accountability becomes even harder to locate. Who is responsible when an AI-generated ethical judgment is wrong, the developer, the deployer, the user, or the system?

Floridi and Taddeo argue that data ethics must remain a human-centered macro-ethics, precisely because ethical harms arise from interactions among data, algorithms and social practices (Floridi & Taddeo, 2016). Outsourcing ethics to AI collapses this complexity into technical compliance. This concern is reinforced by Floridi’s discussion of “ethics bluewashing” or “ethics shirking”, where the appearance of ethical concern replaces genuine moral engagement (Floridi, 2019).

The gig economy case studies make these risks concrete. In The GIG Workers Who Fought an Algorithm, workers were subjected to opaque pay algorithms that claimed fairness while concealing exploitative outcomes (Calacci, 2024). Ethical responsibility was effectively displaced onto the system, allowing the company to deny accountability. Only when workers reclaimed data analysis for themselves did ethical agency re-emerge.

Similarly, in Did AI Win After All?, AI-driven platforms treat humans as “resources” to be optimized, reinforcing Heidegger’s notion of technological mode of thinking that prioritizes efficiency over meaning. In such systems, ethics becomes procedural rather than reflective. The danger, as noted in my comments on Jeff’s words, is not competition between AI and humans, but human evolution being placed on autopilot, guided by systems that lack understanding but simulate confidence.

Outsourcing ethical judgment to AI thus risks producing what is described as a “collective hallucination”, where decisions feel justified because they are automated, not because they are morally sound. This undermines Leonelli’s call for reflexive responsibility and replaces it with technological reassurance.

Conclusion

This paper has argued that outsourcing ethics to AI systems undermines the very conditions required for ethical responsibility in data science. Drawing on Leonelli’s account of distributed accountability, Floridi’s warnings about unethical practices, and my own critiques of AI hallucination and misplaced trust, the analysis shows that ethical judgment cannot be automated without significant loss.

AI systems can support ethical practice, but they cannot replace it. When ethics is delegated to

machines, humans risk disengaging from the moral labor that Leonelli identifies as essential to responsible data science. The result is not ethical neutrality, but ethical erosion: where accountability is diffused, judgment is simulated and responsibility quietly disappears.

My recurring concern, that AI confidently answers questions it does not understand, captures the core danger. Ethics requires not just answers, but the ability to recognize uncertainty, limits and responsibility. These are not technical properties but human ones.

In the end, the question is not whether AI can help with ethics, but whether humans are willing to remain ethically present in systems increasingly designed to absolve them. Leonelli's framework, reinforced by my critiques, suggests that the future of ethical data science depends less on smarter machines than on more reflexive humans.