ESI Preprints **Not Peer-reviewed**

Between Efficiency and Empathy: Evidence of Ethical Ambivalence and the Residual Moral Subject in AI-Mediated Decision-Making

Aljula Gjeloshi, PhD
Anila Boshnjaku, Associated Professor
Ledia Thoma, Professor
Agricultural University of Tirana, Albania

Approved: 08 March 2026
Posted: 10 March 2026

Cite As:
Gjeloshi, A., Boshnjaku, A. & Thoma, L. (2026). *Between Efficiency and Empathy: Evidence of Ethical Ambivalence and the Residual Moral Subject in AI-Mediated Decision-Making.* ESI Preprints. https://doi.org/10.19044/esipreprint.3.2026.p95

Abstract

The rapid integration of artificial intelligence (AI) into domains involving ethical judgment has generated renewed concern about the status of human moral agency. While much of the existing literature focuses either on technical optimization or normative prescriptions, fewer studies empirically investigate how human subjects themselves perceive moral responsibility under algorithmic mediation. This article presents an empirical–theoretical analysis of moral ambivalence in the age of AI, drawing on survey data (N = 146) collected among university students. The findings suggest that contemporary subjects neither fully reject nor uncritically accept AI as a moral authority. Instead, they occupy an intermediate position characterized by ethical ambivalence: technical efficiency increasingly competes with empathy, contextual judgment, and personal responsibility. We argue that this condition gives rise to a "residual moral subject," whose agency is not eliminated but progressively reshaped under algorithmic governance.

Keywords: Moral agency, decision making, artificial intelligence, algorithmic governance, ethical ambivalence

Introduction

The Crisis of the Moral Subject

Artificial intelligence is no longer merely a computational tool; it increasingly functions as a normative infrastructure shaping decisions about justice, responsibility, and human well-being. From algorithmic recommendations to automated assessments, AI systems intervene in domains traditionally governed by human ethical judgment. This transformation raises a fundamental philosophical question: **what happens to the human moral subject when ethical decision-making is partially delegated to algorithms?**

Rather than asking whether AI is "good" or "bad," this article shifts the focus toward the subjective dimension of moral authority. Specifically, it investigates how individuals perceive empathy, fairness, and responsibility in relation to AI-mediated decisions. While theoretical debates on algorithmic governance abound, empirical grounding remains limited. This study seeks to bridge that gap by examining how moral agency is experienced and negotiated in practice.

Literature Review

Algorithmic Governance and Moral Reconfiguration

Artificial intelligence is increasingly embedded in decision-making processes that were traditionally the domain of human moral judgment. From predictive policing and healthcare triage to autonomous vehicles, AI systems mediate decisions with significant ethical consequences. Scholars have conceptualized this shift as algorithmic governance, in which algorithms do not merely perform technical functions but actively shape norms, priorities, and perceptions of what constitutes a "good" decision (Beer, 2017; Yeung, 2018).

While proponents highlight efficiency, consistency, and objectivity, critics stress that algorithms embed normative assumptions through data curation, modeling choices, and optimization goals (O'Neil, 2016; Eubanks, 2018). This implies that delegation of decision-making to algorithms is never ethically neutral. It produces moral distancing, where humans are involved in outcomes without fully owning their moral implications (Introna, 2016). Despite extensive discussion of these structural transformations, empirical investigations into how individuals perceive AI-mediated moral authority remain limited.

Moral Delegation and Limits of Ethical Formalization

Moral delegation refers to the transfer of ethically significant judgment from humans to technical systems (Coeckelbergh, 2020). Machine ethics research suggests that ethical reasoning can be partially formalized

through rules, constraints, and objective functions (Moor, 2006; Wallach & Allen, 2009). Yet, philosophers emphasize that morality is irreducible to formal rules. Moral judgment relies on context, relational understanding, and emotional sensitivity (Dreyfus, 1992; Nussbaum, 2001; Tronto, 1993).

Empirical studies corroborate this critique. Bigman and Gray (2018) found that participants resist delegating morally sensitive decisions to machines, particularly those involving human suffering or social consequences. Gogoll and Müller (2017) similarly report that while AI is acceptable as an advisory tool, participants prefer human final authority in ethical dilemmas. These findings suggest that humans maintain boundaries of moral delegation, guided by empathy and contextual judgment.

Empirical Studies on Moral Trust in AI

Recent research expands our understanding of trust in AI beyond purely technical competence.

- Studies in *Cognitive Research: Principles and Implications* (2024) reveal that individuals distinguish between AI's competence and its moral judgment, demonstrating selective trust in efficiency but skepticism toward ethical decision-making.
- Large-scale surveys (MDPI Social Sciences, 2024) show that trust in AI emerges from perceptions of impartiality and consistency, rather than intrinsic moral understanding.
- Experimental research such as "Zombies in the Loop" (2021) demonstrates that participants may rely on AI advice even when aware of its ethical unreliability, indicating conditional and context-dependent trust.
- Other studies (ArXiv, 2025) find that trust in AI can result from distrust in human actors, highlighting a compensatory dynamic in human–AI interactions.

Collectively, these findings indicate that moral trust in AI is multi-dimensional: it depends on task type, emotional stakes, transparency, and human oversight. Trust is conditional, often ambivalent, and never equivalent to full moral endorsement.

The Residual Moral Subject

Building on critical social theory and philosophy of technology, this article introduces the concept of the residual moral subject (Adorno & Horkheimer, 1947; Verbeek, 2011). This concept captures the form of moral agency that persists under algorithmic governance:

- Ethical awareness persists, but is mediated by algorithmic structures.
- Moral agency is fragmented and conditional, exercised selectively according to context, stakes, and perceived trustworthiness of AI.

- Individuals negotiate responsibility in a liminal space between delegation and autonomy, reflecting ethical ambivalence rather than abdication.

The residual moral subject provides a lens for understanding contemporary patterns of moral engagement: humans retain agency but operate under conditions that encourage partial outsourcing, conditional trust, and negotiated responsibility.

Unlike distributed agency frameworks that dissolve agency across human–technical assemblages, the residual moral subject preserves normative accountability within the human domain, even under structural algorithmic mediation.

Integration: Theory Meets Empiricism

By combining philosophical critique, social theory, and empirical evidence, the literature converges on several key insights:

1. AI systems reshape, but do not erase, moral responsibility.
2. Trust in AI is multi-faceted and context-dependent, particularly sensitive to empathy, transparency, and emotional salience.
3. Humans demonstrate ethical ambivalence: they accept AI as a supportive advisor but resist granting it full moral authority.
4. Moral judgment remains tied to human relational capacities, indicating that AI complements rather than replaces the residual moral subject.

This synthesis lays a strong conceptual foundation for the present study. It situates the survey findings within ongoing debates about algorithmic governance, moral delegation, and the persistence of human agency, providing both theoretical grounding and empirical resonance.

Methodology

Research Design and Sampling

This study employs an exploratory quantitative research design to examine perceptions of moral agency and algorithmic delegation in AI-mediated decision-making. Data were collected through a structured survey administered to 146 university students. Participants were recruited through convenience sampling in classroom settings. Participation was voluntary and anonymous. Respondents were informed about the academic purpose of the study, assured that no identifying information would be collected, and reminded that they could withdraw at any time without consequence. The study adhered to standard ethical guidelines for social research involving human subjects.

The sample consists primarily of early-stage undergraduate students (Mean age = 19.48, SD = 2.50; range 17–46), with most respondents enrolled in their first or second year of study. This population is analytically relevant because it reflects individuals in a formative stage of civic and moral identity development, navigating increasing exposure to digital infrastructures and AI systems. While the sample is not statistically representative of the general population, it provides insight into emerging ethical orientations among digitally socialized young adults. Sample characterists are shown in Table 1

Table 1: Sample Characteristics (N = 146)

Variable	Mean	SD	Min	Max
Age	19.48	2.50	17	46
Year of Study	1.86	0.83	1	3

Measurement and Instrument

The survey instrument consisted of Likert-scale items (1 = strongly disagree to 5 = strongly agree) designed to measure: a) Perceived moral centrality of empathy, b) Beliefs about contextual and non-programmable aspects of ethical judgment, c) Trust in AI's ethical and decision-making capacity, d) Willingness to delegate morally significant decisions to AI, e) Conditional trust and boundaries of delegation.

Items were developed based on themes identified in the literature on algorithmic governance, moral delegation, and trust in AI (Bigman & Gray, 2018; Coeckelbergh, 2020; Gogoll & Müller, 2017; Yeung, 2018).

Where conceptually appropriate, items were grouped into composite indices reflecting:

- Moral trust in AI
- Delegation willingness
- Conditional boundaries of trust

Internal consistency analysis was conducted to assess scale reliability. Composite indices demonstrated satisfactory internal consistency (Cronbach's $\alpha \geq .70$), indicating coherent underlying attitudinal constructs.

Analytical Strategy

The analysis proceeded in three stages:

1. **Descriptive Statistics**
   Means and standard deviations were calculated to assess central tendencies and dispersion of moral attitudes toward AI.
2. **Reliability Assessment**
   Cronbach's alpha coefficients were computed for composite scales to ensure internal consistency.

3. **Correlational Analysis**
   Pearson correlation coefficients (r) were calculated to examine associations between perceived ethical capacity of AI, perceived safety, collaboration willingness, and boundaries of moral delegation.

Effect sizes were interpreted using conventional benchmarks (small ≈ .10, moderate ≈ .30, strong ≥ .50). Statistical significance was assessed at $p < .05$. The purpose of the analysis is exploratory rather than predictive; therefore, emphasis is placed on relational patterns rather than causal inference.

Analytical Orientation

The study does not aim for statistical generalization. Instead, it seeks analytical insight into how moral agency is perceived under conditions of algorithmic mediation. The objective is interpretive: to identify structural patterns of ethical ambivalence and boundaries of moral delegation within a specific sociocultural context. The findings are situated within broader philosophical debates on moral agency and technological mediation, contributing empirical grounding to normative theory.

Limitations

First, the use of convenience sampling restricts generalizability beyond the university population studied. Second, the sample is demographically skewed toward young adults, whose attitudes may differ from older or professionally embedded populations. Third, self-reported perceptions of moral agency may not fully correspond to behavioral decisions in real-world contexts. Fourth, the cross-sectional design captures attitudes at a single moment in time and cannot assess longitudinal shifts in moral perception.

Despite these limitations, the study offers analytically valuable insight into emerging patterns of ethical ambivalence and conditional trust in AI, providing a foundation for future research employing more diverse samples, experimental designs, or longitudinal methods.

Findings and results

Empirical Indicators of Ethical Ambivalence Persistence of Empathy as a Moral Reference Point

Respondents attribute a significant moral value to empathy. The statement *"Empathy is an essential element in every decision I make"* yielded a mean score of **M = 3.12 (SD ≈ 0.91)**, indicating moderate but consistent agreement. Similarly, participants tended to agree that **humans are more empathetic than AI** (*M = 3.17, SD ≈ 1.21*), reinforcing the perception that empathy remains a distinctly human moral resource.

These findings suggest that, despite increasing exposure to AI systems, respondents do not perceive empathy as replaceable by algorithmic processes. Moral judgment is still closely associated with emotional and relational understanding rather than technical calculation alone.

Table 2: Descriptive Statistics of Core Ethical Attitudes Toward AI

Statement	Mean (M)	SD
Empathy is an essential element in every decision I make	**3.12**	0.91
Humans are more empathetic than AI	**3.17**	1.21
Ethical decisions must be context-based and cannot be fully programmed	**3.43**	1.10
AI can help humans make more ethical decisions	**2.74**	1.11
I feel safe relying on AI for complex decisions	**2.55**	1.12
I believe AI decisions can be more efficient but less fair	**3.31**	1.18
I would prefer decisions affecting me to be made by a human rather than AI	**3.48**	1.18
I am concerned that AI ignores human emotions in decisions	**3.54**	1.20

(Likert scale: 1 = Strongly disagree, 5 = Strongly agree)

The clustering of means around the midpoint (≈ 3) indicates **ethical ambivalence**, rather than trust or rejection. Empathy and contextual judgment score consistently higher than trust in AI's ethical competence.

Contextual and Emotional Judgment as Non-Programmable

A stronger consensus emerged around the contextual nature of ethical decisions. The statement asserting that *ethical decisions must be context-based and cannot be fully programmed* received a relatively high mean score (**M = 3.43, SD ≈ 1.10**). This indicates a widespread skepticism toward the idea that morality can be exhaustively formalized into algorithmic rules.

This perception directly challenges technocratic narratives that frame ethical decision-making as an optimization problem, and instead affirms the irreducibility of moral context.

Ambivalence Toward AI as an Ethical Agent

When asked whether *AI can help humans make more ethical decisions*, respondents' answers clustered near the midpoint of the scale (**M = 2.74, SD ≈ 1.11**). A similar ambivalence appears in responses to feeling *safe relying on AI for complex decisions* (**M = 2.55, SD ≈ 1.12**).

These neutral-to-slightly-negative means do not indicate outright rejection, but rather **a lack of full moral trust**. AI is not perceived as ethically illegitimate, but neither is it granted moral authority. This empirical ambivalence supports the central thesis that AI currently occupies a **liminal ethical position** rather than a dominant one.

Moral Resistance to Algorithmic Efficiency

A notable pattern emerges when efficiency is contrasted with human values. Respondents expressed a clear preference for **human-centered judgment over algorithmic efficiency**. The statement *"I would prefer to trust a person, even if AI suggests a more 'efficient' solution"* scored **M = 3.48 (SD ≈ 1.18)**.

Furthermore, participants strongly agreed that they would trust AI **only if it explicitly considers human aspects such as emotions and well-being** (**M = 3.68, SD ≈ 1.15**). This conditionally framed trust suggests that efficiency alone is insufficient to legitimize moral delegation.

Table 3: Conditional Trust and Moral Delegation Boundaries

Statement	Mean (M)	SD
I would trust AI only if it considers human emotions and well-being	**3.68**	1.15
I would collaborate with AI in complex decision-making	**3.08**	0.98
I would accept AI suggestions even if I partially disagree	**2.99**	1.07
I would trust AI for decisions, but not those with major emotional impact	**3.41**	1.30
I would trust a human even if AI suggests a more "efficient" solution	**3.48**	1.18
AI decisions should always be transparent and explainable	**3.68**	1.15

Trust in AI is **explicitly conditional**, bounded by emotional impact, transparency, and human oversight. This empirically substantiates the notion of **partial moral delegation**, not abdication.

Limited Willingness to Delegate High-Stakes Moral Decisions

Respondents demonstrated particular caution regarding emotionally impactful decisions. Agreement with the statement *"I would trust AI to make decisions, but not those with major emotional consequences for people"* reached **M = 3.41 (SD ≈ 1.30)**.

This result indicates a clear **boundary of moral delegation**: while AI may be acceptable in instrumental or advisory roles, respondents resist transferring authority in ethically sensitive contexts. Moral responsibility thus remains anchored to human agency.

Table 4: Correlation Matrix: Trust in AI and Moral Delegation

Variable	1	2	3	4	5
1. AI can help humans make more ethical decisions	1				
2. I feel safe relying on AI for complex decisions	**.54**	1			
3. I would collaborate with AI in decision-making	**.48**	**.52**	1		
4. I would trust AI only if it considers human aspects	.21	.18	.26	1	
5. I would trust AI, but not for emotionally impactful decisions	.09	.14	.19	**.41**	1

(Pearson's r, N = 146)

Correlation analysis (Table 4) reveals that trust in AI's ethical capacity is strongly associated with perceived safety and willingness to collaborate, while remaining weakly related to delegation in emotionally consequential decisions. This pattern indicates a selective normalization of moral outsourcing rather than full ethical displacement."

Summary Pattern: Ethical Ambivalence Rather Than Moral Abdication

Taken together, the findings reveal a consistent pattern:

- Empathy and contextual judgment remain central moral criteria.
- AI is viewed as potentially useful but ethically incomplete.
- Efficiency competes with, rather than replaces, human responsibility.
- Delegation is conditional and limited, not absolute.

Empirically, this constellation of attitudes supports the notion of a **"residual moral subject"**: a subject who has not relinquished ethical agency, but whose moral self-understanding is increasingly negotiated within algorithmic environments.

Discussion

Ethical Ambivalence and the Reconfiguration of Moral Agency

The findings of this study provide empirical support for a central claim in contemporary philosophy of technology: artificial intelligence does not simply replace human moral agency, but reconfigures the conditions under which moral judgment is exercised. Rather than indicating either moral abdication or technophobic resistance, respondents' attitudes reveal a persistent state of ethical ambivalence.

This ambivalence is most clearly expressed in the simultaneous endorsement of empathy and contextual judgment as essential moral criteria, alongside a conditional openness to AI-assisted decision-making. Such a pattern challenges deterministic narratives according to which algorithmic systems inevitably erode human responsibility. Instead, moral agency appears neither fully sovereign nor fully displaced, but **re-situated within algorithmically mediated environments**.

Empathy as a Boundary of Moral Delegation

One of the most consistent empirical findings concerns the moral centrality of empathy. Respondents overwhelmingly regard empathy as an essential component of ethical decision-making and express concern that AI systems fail to adequately account for emotional and relational dimensions. This result resonates with long-standing critiques of ethical formalism, particularly those rooted in care ethics and virtue ethics, which emphasize moral sensitivity, contextual understanding, and emotional engagement (Gilligan, 1982; Nussbaum, 2001; Tronto, 1993).

From this perspective, AI's perceived deficiency is not merely technical but structural: algorithmic systems operate through abstraction and generalization, whereas moral judgment often depends on situated understanding. The resistance to delegating emotionally consequential decisions to AI therefore reflects not irrational distrust, but an implicit recognition of the limits of ethical formalization.

Conditional Trust and Algorithmic Authority

The data further indicate that trust in AI is explicitly conditional. Respondents express moderate willingness to collaborate with AI and acknowledge its potential utility, yet consistently reject unconditional reliance, particularly in high-stakes moral contexts. Transparency, explainability, and human oversight emerge as non-negotiable requirements for ethical legitimacy.

This aligns with prior empirical research showing that people distinguish between instrumental efficiency and moral authority when evaluating automated systems (Bigman & Gray, 2018; Gogoll & Müller, 2017). AI is accepted as an advisory or supportive agent, but not as a final moral arbiter. Algorithmic efficiency, while valued, does not override concerns about fairness, empathy, and responsibility.

Trust is **conditional, context-dependent, and relational**, supporting findings from MDPI Social Sciences (2024), AI and Ethics (2025), and ArXiv (2025). High technical trust does not equate to willingness to delegate morally sensitive decisions.

Importantly, the correlation analysis suggests that increased trust in AI's technical competence does not translate into greater willingness to delegate morally sensitive decisions. This decoupling of efficiency and moral legitimacy underscores the persistence of human-centered ethical boundaries.

The Residual Moral Subject in Algorithmic Environments

These findings substantiate the theoretical concept of the **residual moral subject**. Contrary to narratives proclaiming the "death of the moral subject" under technological rationality, respondents demonstrate continued ethical awareness and resistance to full moral outsourcing. However, this agency is no longer exercised under conditions of full autonomy.

Instead, moral subjectivity becomes:

- fragmented across human–machine assemblages,
- negotiated through conditional trust,
- and constrained by algorithmic infrastructures that prioritize optimization and scalability.

This condition reflects what scholars of algorithmic governance describe as a redistribution rather than an elimination of responsibility (Beer,

2017; Yeung, 2018). Moral agency persists, but in a diminished and mediated form, shaped by institutional, technical, and epistemic constraints. The ethical ambivalence observed in this study should therefore be interpreted not as indecision or confusion, but as an adaptive response to structurally ambiguous moral environments. Individuals navigate between reliance and resistance, efficiency and empathy, delegation and accountability.

Implications for AI Ethics and Governance

The results carry important implications for AI ethics and governance. First, they suggest that ethical frameworks focused exclusively on technical compliance or algorithmic fairness risk overlooking the lived moral experiences of users. Second, they highlight the necessity of preserving spaces for human judgment in AI-assisted systems, particularly where emotional and contextual considerations are central.

Rather than striving to replace moral agents with moral machines, ethical AI design should support **human moral reflection**, acknowledge moral limits of automation, and resist the reduction of ethics to optimization problems. Recognizing and sustaining the residual moral subject may thus be a key normative challenge in the age of algorithmic governance.

Conclusion

Moral Agency After Automation

This article set out to examine how moral agency is experienced and negotiated under conditions of increasing algorithmic mediation. Rather than approaching artificial intelligence as either a moral solution or a moral threat, the study focused on the subjective perceptions through which ethical authority is accepted, resisted, or conditionally delegated.

The empirical findings indicate that human moral agency has not been supplanted by AI. Empathy, contextual judgment, and responsibility remain central moral reference points for respondents, particularly in decisions with emotional or ethical significance. At the same time, participants express a cautious openness toward AI as a supportive tool, especially where efficiency, consistency, and informational support are concerned. This coexistence of reliance and resistance does not reflect confusion, but a structurally induced ethical ambivalence.

To account for this condition, the article introduced the concept of the residual moral subject: a subject whose ethical awareness persists, yet whose agency is increasingly exercised within algorithmically structured environments. Moral responsibility is neither fully retained nor fully outsourced; it becomes fragmented, conditional, and negotiated across human–machine assemblages. This reconfiguration challenges both techno-

deterministic narratives of moral displacement and normative ambitions to fully formalize ethics within technical systems.

The implications of these findings extend beyond individual attitudes. They suggest that ethical AI governance cannot be reduced to technical optimization, compliance frameworks, or abstract principles alone. If moral legitimacy depends on empathy, contextual sensitivity, and accountability, then preserving meaningful spaces for human judgment is not a temporary safeguard but a normative necessity. The empirical findings carry important implications for AI governance and regulatory design. The observed pattern of ethical ambivalence suggests that citizens do not reject AI outright, nor do they accept full moral automation. Instead, they expect structured safeguards that preserve human judgment in morally consequential contexts. Governance frameworks must therefore move beyond technical compliance toward preserving meaningful human moral agency.

While this study is exploratory and limited in scope, it contributes empirical grounding to philosophical debates on moral agency in the age of AI. Future research should expand this inquiry across diverse populations, institutional contexts, and applied domains, examining how different forms of algorithmic mediation reshape moral self-understanding over time. Understanding not only what AI can do, but how humans remain moral agents alongside it, remains a critical task for ethics in the algorithmic age.

Conflict of Interest: The authors reported no conflict of interest.

Data Availability: All data are included in the content of the paper.

Funding Statement: The authors did not obtain any funding for this research.

References:

1. Adorno, T. W., & Horkheimer, M. (1947). *Dialectic of enlightenment*. Herder and Herder.
2. Beer, D. (2017). The social power of algorithms. *Information, Communication & Society, 20*(1), 1–13. https://doi.org/10.1080/1369118X.2016.1216147
3. Bigman, Y. E., & Gray, K. (2018). People are averse to machines making moral decisions. *Cognition, 181*, 21–34. https://doi.org/10.1016/j.cognition.2018.08.003
4. Coeckelbergh, M. (2020). *AI ethics*. MIT Press.
5. Cognitive Research: Principles and Implications. (2024). Perceptions of AI's moral and competence judgment.

https://cognitiveresearchjournal.springeropen.com/articles/10.1186/s41235-024-00573-7
6. Dreyfus, H. L. (1992). *What computers still can't do: A critique of artificial reason*. MIT Press.
7. Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.
8. Gilligan, C. (1982). *In a different voice: Psychological theory and women's development*. Harvard University Press.
9. Gogoll, J., & Müller, J. F. (2017). Autonomous cars: In favor of a mandatory ethics setting. *Science and Engineering Ethics, 23*(3), 681–700. https://doi.org/10.1007/s11948-016-9806-x
10. Human Trust in AI: A Relationship Beyond Reliance. (*AI and Ethics*, 2025). https://link.springer.com/article/10.1007/s43681-025-00690-z
11. Introna, L. D. (2016). Algorithms, governance, and governmentality. *Science, Technology, & Human Values, 41*(1), 17–49. https://doi.org/10.1177/0162243915587360
12. MDPI Social Sciences. (2024). Exploring motivators for trust in the dichotomy of human–AI trust dynamics. https://www.mdpi.com/2076-0760/13/5/251
13. Moor, J. H. (2006). The nature, importance, and difficulty of machine ethics. *IEEE Intelligent Systems, 21*(4), 18–21. https://doi.org/10.1109/MIS.2006.80
14. Nussbaum, M. C. (2001). *Upheavals of thought: The intelligence of emotions*. Cambridge University Press.
15. O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.
16. Trust in AI emerges from distrust in humans. (2025). arXiv:2511.16769. https://arxiv.org/abs/2511.16769
17. Tronto, J. C. (1993). *Moral boundaries: A political argument for an ethic of care*. Routledge.
18. Verbeek, P.-P. (2011). *Moralizing technology: Understanding and designing the morality of things*. University of Chicago Press.
19. Wallach, W., & Allen, C. (2009). *Moral machines: Teaching robots right from wrong*. Oxford University Press.
20. Yeung, K. (2018). Algorithmic regulation: A critical interrogation. *Regulation & Governance, 12*(4), 505–523. https://doi.org/10.1111/rego.12158
21. "Zombies in the Loop? Humans trust untrustworthy AI-advisors for ethical decisions." (2021). arXiv:2106.16122. https://arxiv.org/abs/2106.16122