Navigating the non-coherent landscape of digital ethics and law. Exploring the OREO mandate

Expanded notes from the presentation delivered at the ICICTE Conference on 09 July 2025

Mart Susi

I. Introduction. The digital literacy imperative in higher education

1.1 The existential question is to Chatbot or not to Chatbot?

The integration of Artificial Intelligence into the fabric of higher education represents not merely a technological upgrade, but a fundamental epistemological rupture. The OREO project, titled "To chatbot in higher education or not to chatbot?", positions itself at the epicenter of this disruption. The titular question, while seemingly binary, serves as a rhetorical gateway into a complex labyrinth of ethical, legal, and pedagogical challenges. As the project documentation articulates, the question is effectively moot: the technology is present, powerful, and accessible. The choice is not whether to engage with it, but how to survive its integration without sacrificing the core tenets of critical thinking and academic integrity.

The prevalence of Large Language Models like ChatGPT has created a paradox in academia. On one hand, these tools offer wide access to information synthesis; on the other, they threaten to erode the very cognitive processes, that is, research, writing, and analysis, that constitute the learning experience. The OREO project identifies a critical vulnerability: the potential for abuse by students who may utilize these tools to bypass the intellectual labor of assignments. This "cheating paradox" forces a re-evaluation of assessment methods. If a machine can generate a passing essay, the assessment measures the machine's capability, not the student's. However, the OREO vision transcends the immediate concern of plagiarism. It posits that the antidote to misuse is not prohibition, which is futile, but a robust and sophisticated form of "digital literacy". This literacy involves a deep, technical, and ethical understanding of how AI models function. Students and educators must learn to critically evaluate the probabilistic nature of AI-generated information, recognize the hallucinations and biases inherent in the training data, and understand the ethical ramifications of data privacy. The project’s mission is to support the academic community in shaping these tools into assets that improve education while significantly reducing their risks, moving from a posture of defensive prohibition to one of informed, critical engagement.

1.2 The core hypothesis of the OREO vision

The intellectual foundation of the OREO project is built upon a skeptical, precautionary principle regarding AI ethics. The distinct "OREO Vision" can be articulated as this: the ethical aspect of AI must always be approached with extreme caution.
This vision challenges the anthropomorphic tendency to attribute moral agency to machines. The prevailing narrative in the tech industry often suggests that with enough data and "alignment" training, AI can become an ethical agent. The OREO vision rejects this. It posits that the assumption that AI has an ethical component should be

questioned or outright rejected. The guiding principle which I have proposed is stark: "Better assume that AI has no ethics".

This assumption dictates the pedagogical strategy. If AI is viewed as an amoral, high-functioning statistical engine rather than a "thinking" partner, the responsibility for ethical judgment remains firmly with the human user. This shifts the focus of digital literacy from "collaborating with AI" to "auditing and supervising AI."

II. The capabilities approach. Deconstructing the myth of the ethical machine

2.1 The question of capability vs. output

In assessing whether AI solutions can provide ethical decisions, the discourse often focuses on the *output*: essentially, does the machine produce a statement that aligns with our moral intuition? However, the OREO project's research argues that this focus is misplaced. To determine if an entity is ethical, we must look not at its output, but at its *capability*.

This theoretical shift draws upon the "Capabilities Approach" pioneered by economist Amartya Sen and philosopher Martha Nussbaum. Sen argued that assessments of well-being and justice should be based on what a person is capable of being and doing, that is, their "functionings", rather than merely on resources or utility. Martha Nussbaum expanded this by identifying a list of central human capabilities, including bodily health, emotional attachment, and practical reason, which are prerequisites for a life of dignity. When applied to the domain of Artificial Intelligence, this framework forces a reformulation of the core question. Before we can adjudicate whether an AI's decision is "ethical," we must determine "whether AI is in the first place capable for this". Does the machine possess the internal constitution, the "capabilities," required for moral reasoning?

2.2 Vulnerability as the source of ethics

Martha Nussbaum's contribution to the capabilities approach is particularly salient here. She posits that human reasoning and ethics are deeply rooted in our vulnerability. We understand the value of life because we are mortal; we understand the pain of injustice because we can suffer; we understand the necessity of care because we are fragile.

AI systems, by definition, lack this vulnerability. They operate in a state of digital invulnerability, which means that they cannot die, they cannot feel pain, they cannot be humiliated. They mimic the syntax of ethical language without accessing its semantic emotional core. An AI can process the sentence "torture is wrong" based on statistical correlations in its training data, but it has no capability to comprehend the visceral reality of torture.

The OREO project highlights this ontological gap. If ethical judgment relies on the capacity to empathize with the vulnerability of others, and if empathy requires a shared experience of vulnerability, then AI is structurally excluded from the realm of ethical agents. It remains a simulator of ethics, not a participant. This "non-capability" is the first pillar of the OREO vision: we cannot trust the machine to be ethical because it lacks the biological and existential prerequisites for morality.

2.3 The anthropomorphic fallacy in education

The implications of this for higher education are profound. Students often succumb to the "anthropomorphic fallacy," projecting human intent and understanding onto the chatbot. When a chatbot apologizes ("I'm sorry, I cannot fulfill that request"), the student perceives a social interaction.

The OREO project's mandate to improve digital literacy requires deconstructing this fallacy. The "Trivia" board game and the curriculum developed in WP2 are designed to expose the mechanical nature of the system. By understanding that the "apology" is merely a predicted string of text with the highest probability of concluding a sensitive interaction, students can learn to treat the AI as a tool rather than an authority. This aligns with the project's finding that "digital literacy" must include understanding how AI models work and recognizing their limitations.

III. The positivist paradox. Why AI cannot be trusted

3.1 AI as the ultimate positivist

The limitations of AI in the realm of ethics lead to a provocative hypothesis regarding its legal character. Elsewhere I have argued: "You cannot find a better positivist than artificial intelligence".[1] Legal positivism is the theory that the only legitimate sources of law are those written rules, regulations, and principles that have been expressly enacted, adopted, or recognized by a governmental entity or political institution. Validity is a matter of source, not merit. AI systems are the ultimate positivists because they take "positive law" (their code, their training data, their prompt instructions) as absolute granted truth. They do not question the validity of their instructions. If an AI is programmed to prioritize engagement over truth, it will do so with ruthless efficiency, because it lacks the capacity to question the "law" of its code.

3.2 The Radbruch Formula and the capacity for distrust

To understand why this positivism is dangerous, it is useful to invoke Gustav Radbruch, the German legal philosopher. Post-WWII, Radbruch argued that while positive law should generally be obeyed, there is a threshold of injustice, which is "intolerable injustice", where the law must yield to justice. This was a direct response to the "legal" atrocities committed under the Nazi regime, which were technically lawful under the positivist framework of the time. Radbruch's insight implies that a judge must possess the capacity to *distrust* the law. On the basis of this approach, it can be argued that our trust in physical judges is paradoxically related to our expectation that they are capable of *distrusting* the law when it violates fundamental human rights. We trust them to be the "circuit breaker" against tyranny. Our trust in automated systems is weaker because we assume they *completely trust* the law. We cannot imagine an AI judge spontaneously questioning the constitutionality of a legal provision due to its unethical nature.

3.3 Fairness and algorithmic bias

---

[1] M. Susi. *The Non-Coherence Theory of Digital Human Rights*. Cambridge University Press; 2024.

The "Ethics Guidelines for Trustworthy AI"[2] emphasize fairness and the prevention of bias. However, viewing AI through the lens of the positivist paradox reveals why "fairness" is so elusive. Fairness often requires correcting for historical injustices or recognizing context that is not present in the literal data. If an AI is trained on historical data (which contains historical bias), it accepts that data as "positive law." It optimizes based on that reality. It cannot "intuit" that the data reflects an unjust history that ought to be changed. Therefore, the OREO project concludes that the capability of AI to apply the principle of fairness is unproven and likely non-existent.

IV. The Non-Coherence Theory of Digital Human Rights

4.1 The fundamental rupture

One of the theoretical backbones of the OREO project's legal analysis is the "Non-Coherence Theory of Digital Human Rights".[3] This theory challenges the comfortable assumption that human rights are universal and constant across all domains.
The theory asserts that "the digital and non-digital human rights normative and practice landscapes appear non-coherent". The meaning, scope, and content of well-established rights (like privacy or free speech) undergo significant variance when transposed from the physical (offline) to the digital (online) world. These are "incompatible images."

4.2 Scenarios of normative transposition

How does this non-coherence arise? The theory identifies three scenarios of legal transposition. The first is socio-geographic transposition, which is analogous to colonization. When a legal framework is carried to a new territory, it encounters "indigenous" resistance. The internet, with its "indigenous" culture of anonymity and speed, resists the "colonial" imposition of offline state laws. The second is ideological transposition (regime change). When a new regime replaces an old one (e.g., the Bolshevik revolution), laws change. The internet represents a "regime change" in human interaction, fundamentally altering the ideology of governance from vertical (state-citizen) to horizontal (private platform-user). And the thirs is transposition into normative *carte blanche*. The early internet was seen as an empty sheet (terra nullius) where laws could be written from scratch. However, this view failed to account for the "inadequacy of protection thesis", namely the idea that established human rights might simply fail to function in the new environment due to the novelty of the threat.

4.3 The failure of multistakeholderism

A key insight of the Non-Coherence theory is the critique of "Multistakeholderism"[4], which is the governance model involving states, tech companies, and civil society. The theory argues that this model acts as a "veil". It was an attempt to incrementally transpose offline frameworks into the digital domain through dialogue. However, the

---

[2] High-Level Expert Group on Artificial Intelligence. (2019). *Ethics Guidelines for Trustworthy AI*. European Commission. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai
[3] Supra 1
[4] A. Kovacs, 'Moving multistakeholderism forward: lessons from the NETmundial', Internet Policy Review: Journal on internet regulation, 12 May 2014, https://policyreview.info/articles/news/moving-multistakeholderism-forward-lessons-netmundial/281

result has been a "whitewashing" of the interests of the most powerful stakeholders (the tech giants). The theory concludes that multistakeholderism has failed to create a coherent legal landscape. Instead, non-coherence is a "permanent condition" of the digital age. We must accept that we live in two legal realities that do not match.

4.4 The "sameness of rights" myth

International bodies, including the UN and the Council of Europe, have championed the doctrine that "the same rights that people have offline must also be protected online".[5] The OREO research characterizes this as an "instrument of faith" rather than a fact. My theory questions the "universality" of rights. If rights change meaning based on the domain, they are relative, not universal. As noted by the advocacy group Article 19, without effective remedies, the statement of sameness is "no more than words just written on paper".[6] The Non-Coherence theory suggests that the "sameness" doctrine evaporates the moment one looks at the details. It is a comforting political slogan that obscures the radical legal fragmentation actually occurring.

V. Privacy, identity, and the equilibrium of rights

5.1 The dual nature of privacy

The most striking example of non-coherence is the transformation of the right to privacy. Historically, privacy was defined by Warren and Brandeis (1890) as the "right to be let alone."[7] It is existential. As humans, we are "thrown" into the world involuntarily (Sartre's existentialism), and privacy is our retreat. In the digital domain, the dynamic is inverted. We often enter the digital world *voluntarily*(social media) with the specific intent *not* to be let alone. We enter to be seen, to share, to exhibit.

5.2 Voluntary vs. involuntary entry

My theory distinguishes between two modes of digital existence. Here, privacy is about *control* over one's image. It is about "curating" the self. The user creates a "digital identity" that may be distinct from their physical reality. The interest is not seclusion, but managed exhibition. Yet, citizens are increasingly forced to enter the digital domain to access banking, taxes, and health services (e.g., Blockchain records). In this sphere, the "right to be let alone" is meaningless. You cannot be "let alone" by a blockchain. Here, privacy shifts to a right to *accuracy*(rectification) or *erasure* (the right to be forgotten). The variance is so extreme that it might be more appropriate to refrain from the usage of the expression 'privacy' for online environment. Using the same word for two opposite concepts creates confusion.

5.3 The equilibrium of relative rights thesis

---

[5] U.N. Human Rights Council, 'The promotion, protection and enjoyment of human rights on the Internet', HRC 20th Session, UN Doc A/HRC/20/L.13 (2012); World Summit on the Information Society, Geneva 2003 – Tunis 2005, Document WSIS-03/GEBEVA7DOC74-E, 12 December 2003, Declaration of Principles, Building the Information Society: a global challenge in the new Millennium, https://www.itu.int/net/wsis/docs/geneva/official/dop.htm

[6] ARTICLE 19 statement at the 35th Session of the UN Human Rights Council on the 14th June 2017, as part of the Item 3 General Debate, see: https://www.article19.org/resources/article-19-at-the-unhrc-the-same-rights-that-people-have-offline-must-also-be-protected-online/

[7] S.D. Warren and L.D. Brandeis, 'The Right to Privacy' (1890) 4 *Harvard Law Review* 5, 193.

The Non-Coherence theory introduces a novel concept: the "Equilibrium of Relative Rights Thesis," drawing an analogy to quantum mechanics. It suggests that certain rights exist in a "shared state." For example, privacy and freedom of expression are inextricably linked. In the digital domain, if the scope of the right to privacy broadens (e.g., strict data controls, right to be forgotten), the scope of freedom of expression must necessarily narrow (content removal, censorship). We can measure the narrowing of one right by observing the broadening of the other.

VI. The "Death of Law" debate. Code vs. norms

6.1 Is law dying?

The OREO project engages with the provocative thesis of the "Death of Law," the subject of a major conference at Tallinn University. Professor Francisco J. Ansuátegui Roig argued that the interaction between technology and law causes a rupture so severe that traditional legal concepts become "unrecognizable". Roig suggested that the distinction between "code" (what is technically possible) and "law" (what is permissible) is collapsing. In the digital world, the "rule" is enforced not by a judge, but by the architecture itself. If the code does not permit an action, it cannot happen. Traditional law operates on "ought", for instance you *ought* not to steal. Digital architecture operates on "can", for instance you *cannot* steal because the system prevents it. This bypasses the moral agency of the subject and the enforcement mechanism of the state.

My Non-Coherence theory offers a middle path. It agrees that technology changes law to an unrecognizable extent and rejects the "continuity thesis". However, instead of the "Death of Law," I propose a "Dual Nature" thesis. The offline legal system and the online "techno-legal" system coexist in parallel. They use similar words (rights, privacy, speech) but mean different things. The "Death of Law" is not the end of regulation, but the end of the unity of law. We are entering an era of legal schizophrenia, where we must navigate two non-coherent normative universes simultaneously.

VII. Conclusion: The OREO Vision for a non-coherent future

The OREO project, through its rigorous exploration of the "To chatbot" question, has unveiled a digital landscape defined by non-coherence. We are not moving toward a seamless integration of human rights into the digital sphere, but rather navigating two parallel realities with distinct normative gravities.

The ultimate contribution of the OREO project is a call for Digital Realism. This has the following elements. First, reject the Myth of the ethical machine. Educators must teach students that AI is a powerful, amoral positivist. It has no ethics, no empathy, and no capacity for justice. Second, embrace non-coherence. We must stop pretending that offline rights apply seamlessly online. We need new frameworks that acknowledge the unique "physics" of the digital domain. And third, human-centric oversight must be secured. Because the machine cannot "distrust" the law or feel the weight of a right, the human user must remain the ultimate arbiter. Digital literacy is not about coding; it is about cultivating the critical, ethical, and empathetic capabilities that the machine will never possess.

In this light, the answer to "To chatbot or not to chatbot?" is neither a simple yes nor no. It is a conditional "Yes" - but only if the user is equipped with the literacy to see the machine for what it truly is: a statistical mirror of our data, devoid of the vulnerability that makes us moral.