Oksana Chaika
Associate Professor, National University of Life and Environmental Sciences of Ukraine (NUBiP of Ukraine), D. Sc. (Pedagogy), Ph.D. (Linguistics),
e-mail: oxana.chaika@yahoo.es

AI ETHICS IN HUMANITIES EDUCATION AS A TOOL OF NATIONAL INFORMATION RESILIENCE AND IDENTITY PRESERVATION

Artificial intelligence (AI) has become structurally embedded in knowledge production, educational administration and classroom practice across higher education systems in Europe and beyond. Its rapid diffusion across the humanities, not traditionally driven by automation, creates a new policy and epistemic situation in which questions of reliability, governance and human agency move from a rhetorical concern to a structural condition of teaching and learning. Unlike STEM domains, where algorithmic optimisation is integrated into the epistemology of the field, the humanities derive their legitimacy from interpretive plurality, normative reasoning and cultural continuity. AI adoption in this sector along with technical or procedural questions raises the questions about the stability of interpretive authority, the preservation of cultural and linguistic identity, and the resilience of democratic information spaces in which human judgement historically functions as a filter against manipulation.

European, UNESCO and OECD policy frameworks, though issued by distinct institutions, converge on a foundational principle: AI in education is permissible only as long as it remains human-centred, ethically governed and epistemically accountable. The European Commission's *Ethics Guidelines for Trustworthy AI* define trustworthy AI as "lawful, ethical and robust" and emphasise that "AI systems should empower human beings, allowing them to make informed decisions and fostering their fundamental rights" (European Commission, 2019, pp. 3–4). This two-part injunction, highlighting lawfulness and empowerment, anchors the entire European normative architecture on AI, including education-specific documents. In the *Ethical Guidelines on the Use of Artificial Intelligence (AI) and Data in Teaching*

and Learning for Educators, the Commission states that educators must be enabled to "engage positively, critically and ethically with AI systems and exploit their full potential" (European Commission, 2022, p. 6). Thus, AI is framed not as an autonomous cognitive agent but as a supervised pedagogical instrument subordinate to human oversight. Next, UNESCO extends this principle to the global normative level. In its *Recommendation on the Ethics of Artificial Intelligence* (2021), adopted by 193 Member States, AI is required to operate "in the service of humanity, peace and the common good" (UNESCO, 2021, Art. 23). In education, UNESCO's *Guidance for Generative AI in Education and Research* warns that unregulated adoption may "amplify bias, misinformation and inequity" (UNESCO, 2023, p. 7) and calls for explicit teacher training and institutional governance before classroom use. Besides, the OECD's *Recommendation of the Council on Artificial Intelligence* similarly mandates that governments ensure that AI is "robust, secure and safe throughout its entire life cycle" and that its benefits are distributed inclusively (OECD, 2019, pp. 5–7). Across these instruments, AI in education is conceptualised as a socio-technical object whose legitimacy does not derive from capability but from compliance with ethical constraints.

The humanities represent a privileged testing ground for these policies because they depend on precisely those human capacities, e.g. context recognition, value judgement, textual criticism, that AI can imitate syntactically but not assume responsibility for. Automated summarisation, translation, topical clustering and generative paraphrase can assist interpretation, but they also risk replacing first-order reading with second-hand abstraction, weakening the development of epistemic vigilance. UNESCO's 2023 guidance notes that generative AI "blurs the distinction between original and synthetic content" (p. 9), a risk that is epistemic rather than technical: students may begin to accept plausible language for verified knowledge. The OECD (2023) similarly cautions that AI-driven platforms can "narrow learning to measurable outputs rather than conceptual understanding" (p. 15), potentially flattening interpretive complexity into computational convenience.

A concrete illustration of these risks can be observed in environments exposed to active information warfare. In such contexts, generative AI trained on mixed-provenance corpora may reproduce narrative frames that reflect geopolitical bias rather than verified historical fact. The phenomenon is not confined to any one state: similar distortions have been documented in AI output related to territorial disputes in Asia, conflicts in the Middle East, and secessionist movements in Europe. These cases show that AI can inadvertently function as a vector of discursive asymmetry when adopted uncritically in educational settings. However, the same cases also demonstrate that when AI is used under ethical supervision, for example, to contrast competing narratives or annotate ideological markers — its presence can strengthen rather than weaken interpretive competence. This is aligned with UNESCO's requirement that "humans, not machines, should remain accountable for decisions that affect people's lives" (UNESCO, 2021, Art. 13), implying that AI is permissible only when its epistemic outputs remain subordinated to critical scrutiny.

The policy convergence across the EU, UNESCO and the OECD has practical implications for higher education governance. First, educator capacity building is explicitly mandated where both the EU (2022) and UNESCO (2023) call for structured teacher training in AI literacy and ethics prior to classroom deployment. Second, institutional oversight is not optional, and the UNESCO Recommendation (2021, Art. 18) requires Member States and institutions to establish mechanisms for ethical review and risk assessment of AI deployment. Third, transparency obligations apply not only to developers but also to adopters, which means that the AI Act obligates that learners receive "clear information about the capabilities and limitations" of systems used in educational contexts (European Commission). Fourth, fairness is not an aspirational value but a regulatory criterion, the OECD (2019) requires AI use to promote "inclusive growth and sustainable development" (p. 5), rendering discriminatory or culturally erasing implementations normatively incompatible with international guidance.

For the humanities, these obligations translate into methodological rather than merely managerial consequences. If AI is to be used without eroding interpretive

agency, assignments, assessment and classroom practice must be redesigned to require justification, comparison and source-verification rather than passive reproduction of AI-generated text. AI can support transliteration (for instance, *Kyiv* versus *Kiev* or *Lviv* versus *Lvov* as proper names used for the Ukrainian cities), corpus search, concordance analysis, intertextual mapping and reference discovery, but the interpretive act must remain human. When AI provides candidates and humans adjudicate them, the alignment with ethical policy remains intact. When AI outputs are accepted without challenge, the conditions identified by UNESCO (2023) as high-risk — misinformation, inequity, epistemic dependence, are activated.

The combined reading of EU, UNESCO and OECD documents shows that ethical control of AI in humanities education is not an aesthetic or moral supplement to innovation but a structural precondition for legitimate adoption. The common denominator is constant: AI may extend human cognition but cannot replace human judgement without violating international ethical standards. Under these frameworks, humanities instructors are repositioned as ethical mediators between algorithmic affordances and civic epistemic needs. The resulting alignment of AI literacy, interpretive responsibility and information resilience suggests that the humanities are not displaced by AI but re-tasked, moving from defending tradition against technology to governing technology on behalf of tradition and democratic knowledge.

References

1. European Commission. (2019). *Ethics guidelines for trustworthy AI.* High-Level Expert Group on Artificial Intelligence. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai
2. European Commission. (2022). *Ethical guidelines on the use of artificial intelligence (AI) and data in teaching and learning for educators*. Publications Office of the European Union. https://op.europa.eu/en/publication-detail/-/publication/d81a0d54-5348-11ed-92ed-01aa75ed71a1/language-en
3. European Commission. (n.d.-a). *AI Act — Regulatory framework for AI.* https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

4. OECD (2019). *OECD recommendation of the Council on artificial intelligence*. OECD Publishing. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

5. OECD (2023). *The OECD.AI Policy Navigator is a central resource for tracking public AI policies*. https://oecd.ai/en/dashboards/overview

6. UNESCO. (2021). *Recommendation on the ethics of artificial intelligence*. UNESCO Publishing. https://unesdoc.unesco.org/ark:/48223/pf0000381137

7. UNESCO. (2023). Guidance for generative AI in education and research. UNESCO Publishing. https://unesdoc.unesco.org/ark:/48223/pf0000386693?posInSet=1&queryId=0aa578a6-1511-404b-ba92-92442c7c46f6