TALOS ERA CHAIR IN ARTIFICIAL INTELLIGENCE FOR HUMANITIES AND SOCIAL SCIENCES

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ
UNIVERSITY OF CRETE

Social/ educational issues and ethics

Eleni Katsarou, Professor,
Department of Primary Education

Funded by the European Union

Horizon ERA Chair TALOS AI4SSH Project funded by the European Commission
Grant Agreement n° 101087269, https://cordis.europa.eu/project/id/101087269

Introduction: Ethical challenges of AI

- The rapid integration of AI technologies into education and research brings **unprecedented possibilities**, but also **serious risks**.
- These include issues of **bias**, **transparency**, **privacy**, **security**, and the **changing roles** of educators and learners.
- International organizations and the EU have established **strong frameworks and guidelines** to ensure responsible use of AI.

Contents

In this section of the MOOC, we will present:

the social risks raised by AI technologies,

a vision of **ethical, inclusive, and human-centered AI** defined by EU and other global organisations.

Finally, we will highlight the **critical role of humanists and social scientists** in shaping ethical uses of AI.

Risks of AI in daily life, in professional life and in scientific research

Biased Outcomes

- AI systems may **reproduce** or even amplify existing social, cultural, or economic biases found in training data.
- Results can lead to **discrimination**, **inequality**, and **unfair treatment** in sectors like hiring, finance, healthcare, and law enforcement.

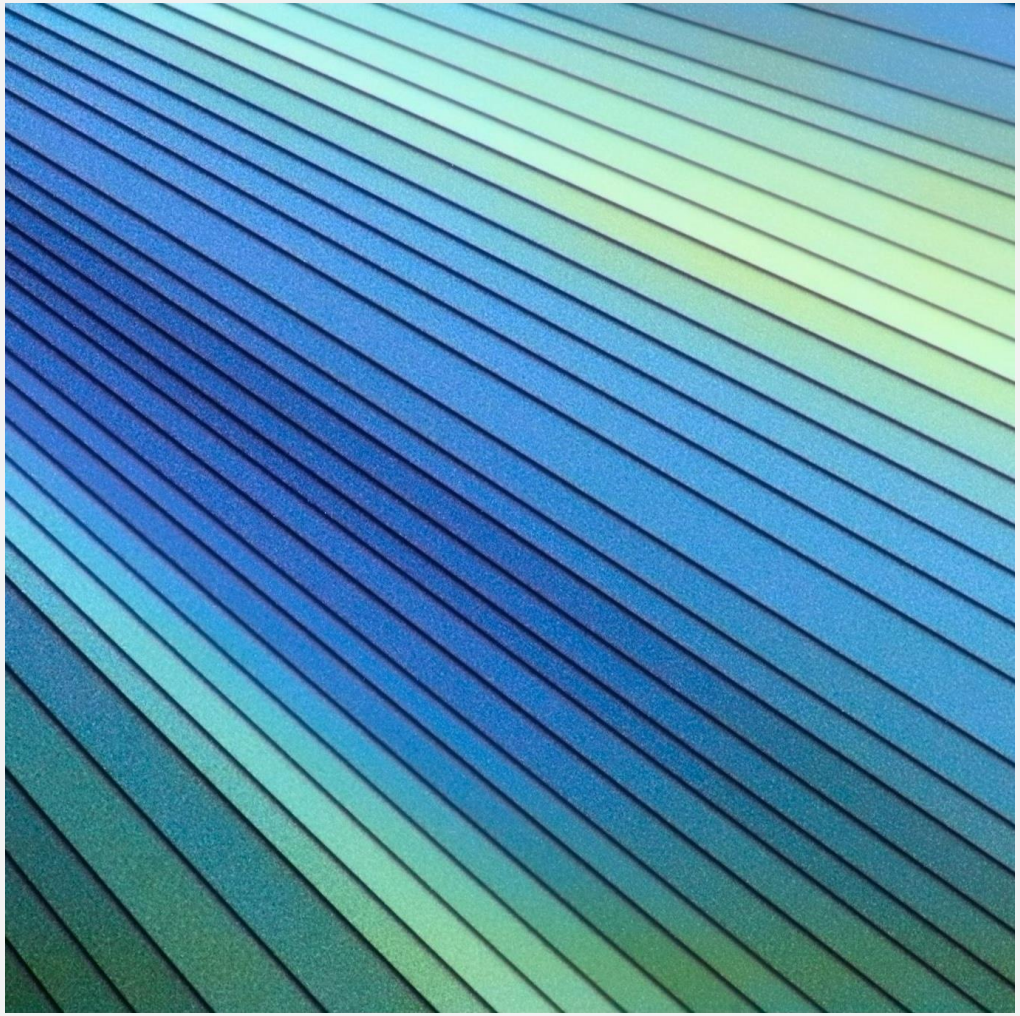
(Source: Cheatham et al., 2019; UNESCO, 2021)

Lack of Transparency

Many AI algorithms function as "black boxes" — their decisions are difficult to interpret.

This **lack of explainability** affects public trust, accountability, and the protection of democratic processes and human rights.

(Source: Cheatham et al., 2019; UNESCO, 2021)



Privacy and security concerns

Invasion of Privacy

- AI systems often rely on vast amounts of personal data, raising **serious privacy issues**.
- Misuse or over-collection of data can **threaten human dignity** and **individual autonomy**.

Safety and Security Risks

- Inadequate safeguards in AI applications can lead to **cyberattacks**, **data breaches**, and **unauthorized surveillance**.
- Potential misuse of AI for **malicious purposes**, including disinformation campaigns and automated hacking.

Risks in Educational Settings (like HE)

Changing Role of Educators

AI tools can shift focus from human-led teaching to automated instruction, reducing teacher agency. Risk of **de-skilling** educators.

Changing Role of Student

Everything is ready-made. GenAI answers any question. No effort, no agency. Risk of **disempowerment**

Pedagogical Concerns

- The methods embedded in AI systems may not align with critical thinking, creativity, or inclusive learning goals.
- Over-reliance on AI may prioritize standardized content over holistic education.

(Source: Holmes, 2023)

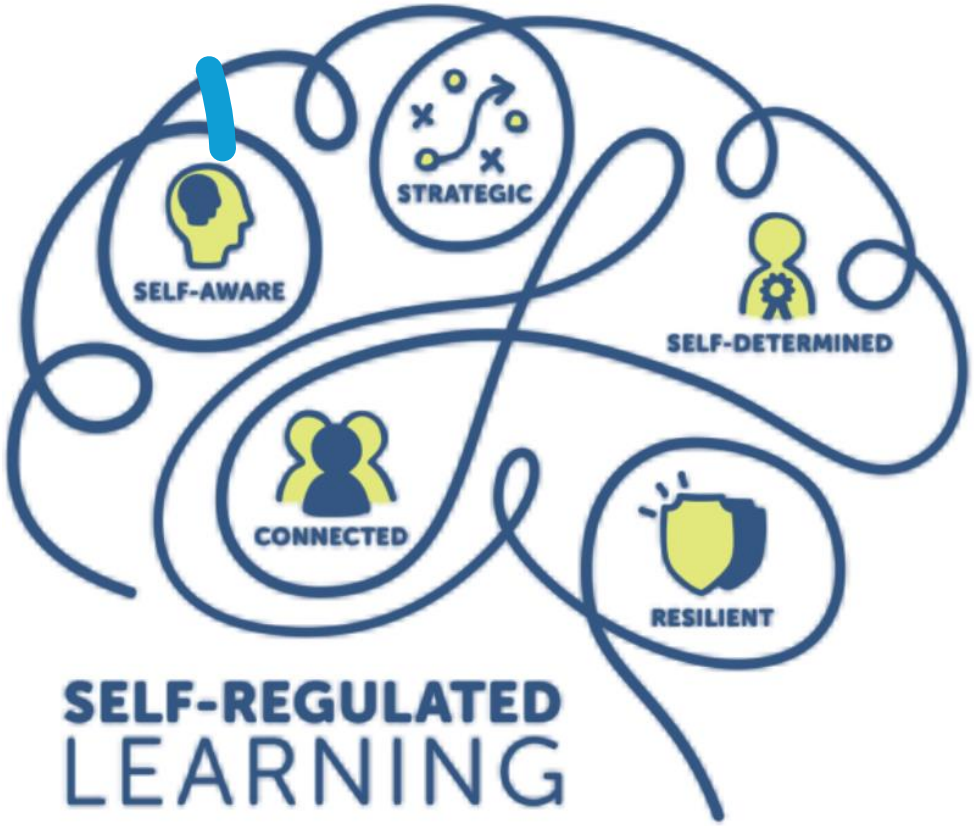
Equity and Accessibility Challenges

Digital Divide and Motivation for Self-Regulated Learning

- AI-enhanced learning may **widen the gap** between well-resourced and under-resourced institutions.
- Even when students gain access to AI tools, **equitable use is not guaranteed**.
- Effective learning with AI platforms often requires **high levels of digital literacy**, **access to expensive platforms**, **self-regulation**, and **intrinsic motivation**.
- Students from disadvantaged backgrounds may face difficulties in managing **all the above**, leading to **unequal outcomes**.

The challenge of self-regulated learning

- As education shifts toward **AI-driven personalized learning platforms**,
- success will increasingly depend on a learner's ability to **set goals, manage time, and reflect on progress** — skills that are not equally developed across all learners.



Policy and Ethical Responses to AI Risks
Strong Global and European Regulatory Efforts

In response to the growing risks of AI, key international and European bodies have established **frameworks and ethical guidelines** to ensure safe, fair, and inclusive AI use — especially in education.

🔷 **European Commission:**

- **Ethical Guidelines on AI and Data in Education (2022):** Promotes transparency, accountability, privacy, inclusion, and human-centric AI in educational environments.
- **Digital Education Action Plan (2021–2027):** Aims to support digital transformation in education with emphasis on equity, quality, and digital competence.

🔷 **EU Competence Frameworks:**

- **DigCompEdu**: For educators' digital competences including responsible use of AI tools.
- **DigComp 2.2**: Updated in 2022 to include AI literacy and data-related skills for all citizens.

🔷 **International Initiatives:**

- **UNESCO (2021):** Recommendation on the Ethics of Artificial Intelligence – emphasizing human rights, sustainability, and inclusive development.
- **OECD (updated 2025):** AI Recommendation framework focused on transparency, safety, accountability, and democratic values.

The Talos AI4SSH Project – Aligned with EU and Global AI Ethics Guidelines

Alignment with European and international guidelines (EU, UNESCO, OECD) on the ethical and responsible use of AI in education.

All research and educational actions are guided by **10 core ethical principles**:

1. **Human Agency**
   AI use aims to empower both educators, researchers and learners, enhancing participation, autonomy, and critical engagement — not promoting passivity.
2. **Fairness, Inclusion, and Non-Discrimination**
   Systems are designed and tested to avoid bias and promote **equity**, with special focus on **students with disabilities and diverse backgrounds**.
3. **Human-Centric Approach**
   AI tools and uses respect human dignity and well-being, prioritizing human needs over algorithmic efficiency. Systems remain under **human supervision**.

Ethical principles (continued)

4. Transparency and Explainability
Algorithms used are understandable by researchers, educators and learners. Decision-making processes of AI systems must be **clear and accessible**.

5. Societal and Emotional Impact
TALOS promotes awareness of how AI affects **democracy**, **social cohesion**, and **emotional well-being**, stressing that AI lacks empathy and consciousness.

6. Privacy and Data Protection
Strict mechanisms for **data anonymization**, **informed consent**, and **secure data handling** are implemented. Clear communication on what data is collected, why, and by whom.

Ethical AI Principles (continued)

7. Accountability
TALOS cultivates a sense of **responsibility and awareness** among all stakeholders regarding the implications of using generative AI in education.

8. Security and Safety
AI tools are developed and deployed with robust protections against **cyber threats**, ensuring **safe use** for both learners and educators.

9. Democratic Participation
Policy planning and AI implementation in Talos involve **educators, learners, and stakeholders** in **inclusive, participatory processes**.

10. Combatting Disinformation
Educational content includes **critical AI literacy**, helping users recognize AI-generated content and **resist disinformation** or manipulative outputs.

Closing Reflections – Ethics as a Foundation for AI4SSH

Without ethics, AI is not just ineffective — it is dangerous.

- The use of AI must **always be guided by strong ethical and human-centric principles**.
- Technologies that affect learning, autonomy, identity, and democratic life require **vigilant oversight**.

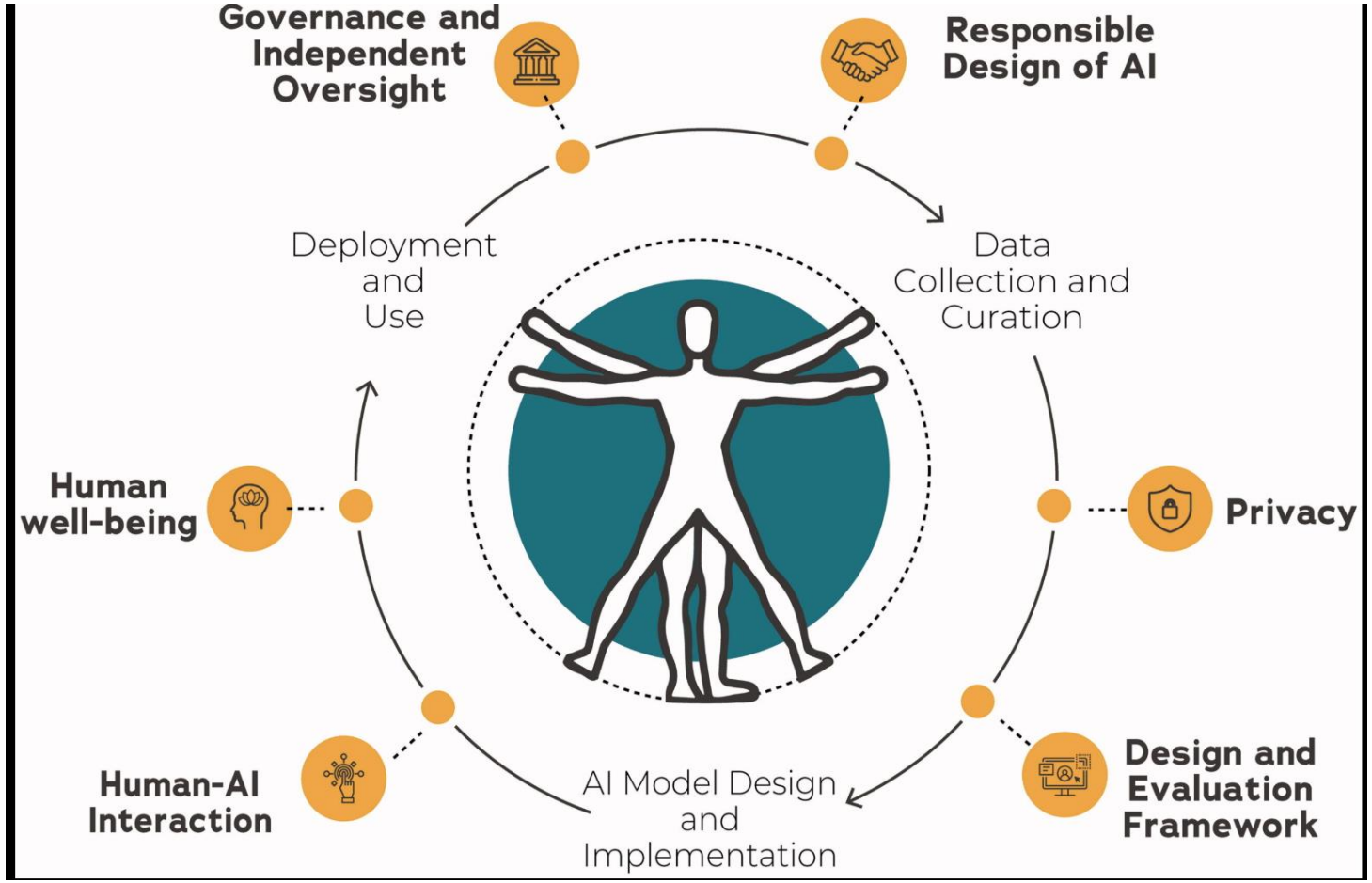
It is **not enough** for AI tools to be innovative — they must also be:

- **Fair**
- **Transparent**
- **Inclusive**
- **Respectful of human rights and dignity**

The crucial role

The role of **humanists and social scientists** is **critical**:
They must **monitor, assess, and shape AI use** in ways that protect fundamental values and public good.

They bear **significant responsibility.**



References

Cheatham, B., Javanmardian, K. and Samandari, H. (2019). Confronting the Risks of Artificial Intelligence. *McKinsey Quarterly*, 2 (38). https://www.mckinsey.com/capabilities/quantumblack/our-insights/confronting-the-risks-of-artificial-intelligence

European Commission. (2017). *DigCompEdu: The Digital Competence Framework for Educators*. https://joint-research-centre.ec.europa.eu/digcompedu_en

European Commission: Directorate-General for Communications Networks, Content and Technology & Grupa ekspertów wysokiego szczebla ds. sztucznej inteligencji. (2019). *Ethics guidelines for trustworthy AI*. Publications Office. https://data.europa.eu/doi/10.2759/346720.

European Commission: Directorate-General for Education, Youth, Sport and Culture. (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://data.europa.eu/doi/10.2766/153756.

Holmes, W. (2023). *The Unintended Consequences of Artificial Intelligence and Education*. Education International. https://www.ei-ie.org/en/item/28115:the-unintended-consequences-of-artificial-intelligence-and-education

OECD. (2019). *Recommendation on AI* (Amended on 2024, March 5). OECD Legal Instruments. https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449

UNESCO. Global Education Monitoring Report Team (2021/2). *Global education monitoring report, 2021/2: non-state actors in education: who chooses? who loses?* UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000379875