Ethics by Design in AI: theoretical and practical perspectives
Abstracts

The Symposium will take place at Université Paris 1 Panthéon-Sorbonne.
Centre Sorbonne.17 rue de la Sorbonne.
Stair C, 2nd floor, right corridor, Lalande Room.

Thursday, June 8, (10 am-5.00 pm): Theoretical Perspective on Ethics by Design in AI

Ethics by design as a crossroad- Vanessa Nurock (CRHI, Université Côte d'Azur, France & UNESCO EVA Chair)

The expression « Ethics by design » is now commonly used in the field of Artificial Intelligence. However, what is meant by this expression is far from obvious. This talk aims at sheding light on the concept of Ethics by Design in AI understood as a crossorad. More precisely, I want to show that EBD is a crossroad interconnected with Ethics and politics of care. As such, Ethics by design can be analyzed through its verticality (or relationship with « time ») as well as its horizontality (or relationship with « space »). These two planes enable us to build the concept of ethics by design both from a methodological/practical and theoretical point of view.

Ethics by design : a 2.0 provisional morality ?- Damien Lacroux (CRHI, Université Côte d'Azur)

Identifying the key principles for thinking about the emergence of ethics by design has enabled the researchers to reveal the theoretical links with classical philosophies. With the ethical principles to be implemented in machines, it is easy to recognise the Kantian origin of the approach; the reflections on the intention of engineers or other players involved in the technology design process suggest a morality of intentionality close to Aristotle's theory of virtues; and the priority given to evaluating the consequences of a technology in order to make a moral judgement about its design undoubtedly bears witness to the influence of utilitarian thought. But what about Cartesianism? Descartes' name is nowhere to be found, perhaps primarily because his moral thinking has not left the same mark on posterity as Kant or Aristotle. Yet it was in the midst of doubt and uncertainty that the French philosopher proposed a provisional morality to help us find our way. This morality by provision, like the ethics by design, must be established upstream of the process of conceiving new knowledge, the effects of which are fundamentally unknown. There is every chance that ethics by design will take on important theoretical elements from Cartesian morality. The challenge, then, is to make these explicit and to question this unacknowledged filiation.

Damien Lacroux is a postdoctoral fellow on the MIRACLE project of the UNESCO EVA Chair. His research in ethics of care and algorithmic health attempts to identify the changes in the status of the doctor, the status of the patient and the responsibilities linked to the integration of algorithms in cancerology. Agrégé of philosophy and former ATER at the Université Paris Cité and the Université Paris 1 Panthéon-Sorbonne, he has also taught philosophy of science, history of philosophy and medical ethics. His doctoral thesis "Descartes à l'épreuve des théories neuroscientifiques des émotions" (Descartes to the test of neuroscientific theories of emotions) also makes him a specialist in Cartesian philosophy and the philosophy of neuroscience.

***Discipline and prescribe: ethics by design in the AI era*-Anthony Masure (HEAD, Genève, HES-SO)**

Released at the turn of the 2010s, deep learning technologies (a subfamily of machine learning) are based on opaque architectures. As they automate more and more tasks, these black boxes, that are not understandable to humans, overcome the usual ethical and moral categories. Faced with these challenges, some people want to "take back control." However, if machines (through a mirror effect) reveal the "inhuman" part of humanity, does the human/machine separation still make sense? On the other hand, embedding ethics in a computer program ("by design") requires making them explicit and fixed—which is the opposite of a dialogical dynamic. To overcome AI challenges, we must therefore "normative ethics" (prescription) and "capability ethics" (potentiation).

Anthony Masure is an Associate Professor and Dean of Research at Geneva University of Art and Design (HEAD – Genève, HES-SO). His research is currently focused upon the impact that artificial intelligence and blockchain technologies have upon design. He is also the author of the essays "Design and the Digital Humanities" (2017) and "Artificial Design: Creation Versus Machine Learning" (2023). He is a co-founder of the research journals *Back Office* and *Réel-Virtuel* and of Hint3rland, a creative studio for the decentralized world. Website: https://www.anthonymasure.com

[**Anthony Masure**](link) Anthony Masure is an Associate Professor and Dean of Research at Geneva University of Art and Design (HEAD – Genève, HES-SO).

"AI and education: the difficulty of aligning"- **Colin de la Higuera (Université de Nantes, UNESCO Chair RELIA) :**

AI in/and education introduce a huge range of challenges. Some of these are technical and could be sorted out with effort and money. Other are more profound and it is uncertain that there are good answers. Time issues such as the moving differences between the timelines of education, politics, technology and paradigm issues like the question of matching deterministic answers to stochastic problems are examples of such problems. We will illustrate these questions through examples from current AI and education problems.

Colin de la Higuera is Professor at Nantes University (France). He has been involved in a number of research themes, including algorithmics, formal language theory, pattern recognition and artificial intelligence.

He has been chairman of the International Community in Grammatical Inference (2002-2007), the founding president of the SIF, the French Informatics Society, and is currently a trustee of the Knowledge for All foundation where he is working towards the usage of technology for an open dissemination of knowledge and education.

In 2021 and 2022 he was co-chair of the Open Education Global Congress.

In 2017, based on the project Class'Code he contributed to launch in France, UNESCO has created a Chair in Technologies for the Training of Teachers by Open Educational Resources at University of Nantes. Since 2020 he is Chief Equality Advocate at IRCAI. In 2021 UNESCO renewed the Chair in "Open Educational Resources and Artificial Intelligence".

Friday, June 9, (10 am-5.30 pm): Ethical issues in case studies from Education and Healthcare

"Platformization, public values and stakeholder agency in the adoption of AI in education"- Tel Amiel (Universidade de Brasília, UNESCO Chair in Distance Education)

Normative documents and discussions on public values and ethics of AI emphasize stakeholder diversity, inclusion and participatory methods. But in light of the massive platformization of public education these imperatives might be more difficult to realize than expected. Here, I would like discuss how teacher and institutional agency might become secondary to corporate pressures, as diverse AI tools and services are deployed as services in educational platforms.

Tel Amiel is adjunct professor at the School of Education at the University of Brasília (Brazil) where he coordinates the UNESCO Chair in Distance Education, and adjunct professor at University of Nova Gorica in the Leadership in Open Education program. Co-leader of the Open Education Initiative, and activist research group.

"Ethical design issues arising from large language models in education settings"-James Katz (Boston Univeristy):

Important ethical design issues arise when AI technology is developed for educational settings. These typically include: 1. Privacy, Informed Consent and Data Protection; 2. Equity, Accessibility and Inclusion; 3. Transparency of Practices for Users; 4. User Autonomy and Consent; 5. Ethical Content and Fair Data Usage Practices; 6. Emotional Well-being; 7. Accountability and Oversight; 8. Sustainable Design. At least in terms of rationalizing frameworks, many of these issues are generally well-understood by ethicists, philosophers, but much less broadly by policy-makers and even less so by designers themselves (Cambridge-CRAASH, 2023).

As with all systems, inherent values are embedded in AI's designs as reflected in their physical operation, procedural rules and supporting structures. These values often remain unexamined or even invisible and consequently may lead to inadvertent trade-offs or the application of unexamined assumptions that conflict with other ethical desiderata. Because of their remarkable flexibility, the powers of which we are only beginning to perceive, large language models (LLMs) bring to the surface these previously submerged countervailing values. Unlike the issues mentioned above, these arguably receive much less weight in the popular and professional discussions of LLM educational regimes. Here these countervailing issues may

include: 1.The Right to Tinker; 2. Fair End-User Licensing; 3. Revisions to Authorship Rules; 4. Responsibility for Downstream LLM Actions; 5. Better Balancing of Equity with Individualism; 6. Allocating Dispersed Team Rewards and Responsibilities; 7. Maintaining and Enhancing Motivations to Create New Technology; 8. Supporting Instructors' Roles; and 9. Creating Authenticity in Communication.

What may have seemed like a fixed domain of issues in terms of their applicability to AI, with the arrival of LLMs, are now thrown into doubt. I provide a few brief cases to illustrate how this is happening and suggest future trajectories. As ethical questions play themselves out through deployment and development of LLMs, these seemingly timeless principles may even themselves become superseded. By addressing the full array of ethical aspects of AI in the construction and application of educational technology, and in particular those based on LLMs, designers can ensure that their creations respect learner and instructor rights, promote fairness and inclusivity of all groups, and contribute to the overall educational experience.

James E. Katz, Ph.D., Dr.h.c., is the Feld Professor of Emerging Media at Boston University. Katz's core interests revolve around societal and interpersonal aspects of communication technology. His pioneering publications on artificial intelligence (AI) and society, social media, mobile communication, and robot-human interaction have been internationally recognized and translated into a dozen languages.

His two most recent books are *Perceiving the Future Through New Communication Technologies: Robots, AI and Everyday life*, co-edited with Juliet Floyd and Katie Schiepers (Palgrave Macmillan) and *Journalism and the Search for Truth in an Age of Social Media*, co-edited with Kate Mays (Oxford University Press). His forthcoming book, also co-edited with Juliet Floyd and Katie Schiepers is *Nudging Choices Through Media: Ethical and Philosophical Implications for Humanity* (Palgrave Macmillan).

***MIRACLE project: Artificial Intelligence for lung cancer and application of an Ethic by Design approach*- Paola Ulivi, (Translational Oncology Unit, Biosciences Laboratory, IRCCS, Istituto Romagnolo per lo Studio de Tumori "Dino Amadori", IRST, Meldola, Italy):**

The MIRACLE (A machine learning approach to identify patients with resected non-small-cell lung cancer with high risk of relapse) project is a project funded by the ERAPerMed 2021 call, with the main aim to create and validate a prognostic algorithm for early-stage non-small-cell lung cancer (ES-NSCLC). In that project, a case series of ES-NSCLC patients underwent surgery will be considered. Several biological, clinical and radiological parameters will be analysed and taken in consideration all together by using a machine learning approach, in relation to patient prognosis and risk of relapse. The identified algorithm will be then validated on an independent cohort of patients. For the entire project an Ethic by Design approach will be applied, taking in consideration several aspects related to the role of medical doctors, patients, patient's association during the project and in relation to the artificial intelligence application in the clinical practice.

Paola Ulivi is a researcher/biologist at the Biosciences Laboratory at IRST-IRCCS in Meldola (Italy). She is the coordinator of the Translational Oncology Unit. Her main topics of research are related to translational research on solid tumors, in particular focalized on precision medicine. She is the Principal Investigator of many projects related to the identification of prognostic and predictive biomarkers in tumor tissue and liquid biopsy, mainly on lung cancer and gastrointestinal tumors.