Human Freedom at the test of AI and Neurosciences

International Conference

Rome, 2–5 September 2024

CONFERENCE PLAN

	September 2	September 3	September 4	September 5
9 am **Plenary Session**		LUMSA Aula Giubileo Keynote **Dominique Lambert** (University of Namur) *Ethics of AI*	LUMSA Aula Giubileo Keynote **Patricia Churchland** (University of California, San Diego) *Neurosciences and the problem of Freedom*	UNIVERSITY of NOTRE DAME ROME Walsh Aula (103) Keynote **Laura Palazzani** (LUMSA University) *Health at the Time of AI and Neurosciences*
10 am		**Coffee break**	**Coffee break**	**Coffee break**
10.30am **Parallel Sessions**		**NHNAI** LUMSA Aula Giubileo **CH** LUMSA Aula Pia/Sala del Consiglio **ATEM** LUMSA Aula Teatro	**NHNAI** LUMSA Aula Giubileo **CH** LUMSA Aula Pia/Sala del Consiglio **ATEM** LUMSA Aula Teatro	**NHNAI** Room 103 **ATEM** Room 104
1 pm		**Lunch**	**Lunch**	**Lunch**
2.30 pm **Plenary Session**		LUMSA Aula Giubileo Keynote **Thierry Magnin** (Catholic University of Lille) *Christian Thought, AI and Neurosciences*	LUMSA Aula Giubileo Keynote **Fiorella Battaglia** (University of Salento) *Democracy and Education at the Time of AI and Neurosciences*	
3.30 pm		**Coffee break**	**Coffee break**	

	September 2	September 3	September 4	September 5
4 pm - 7.30 pm	**Plenary Session** LUMSA Aula Giubileo **Greetings and Introduction** Keynote **Mario De Caro** (University of Rome 3/ Tufts University) *The problem of Freedom and today's challenges* **Cocktail Dinner**	**Parallel Sessions** **NHNAI** LUMSA Aula Giubileo **CH** LUMSA Aula Pia/Sala del Consiglio **ATEM** John Paul II Pontifical Theological Institute	**Parallel Sessions** **NHNAI** LUMSA Aula Giubileo **CH** LUMSA Aula Pia/Sala del Consiglio **ATEM** Pontifical Academy for Life	

LUMSA University: Via di Porta Castello, 44 - Roma
University of Notre Dame Rome: Via Ostilia, 15 - Roma
John Paul II Pontifical Theological Institute: Piazza San Giovanni in Laterano, 4 - Roma
Pontifical Academy for Life: Piazza San Calisto 16 - Roma

DETAILED PROGRAMME

2 September 2024

4 pm PLENARY SESSION
LUMSA Aula Giubileo

Chair:
Stefano Biancu, Università LUMSA / University of Notre Dame Rome

Greetings and Introductions

Keynote Lecture: Mario De Caro, Università di Roma Tre / Tufts University
The problem of Freedom and today's challenges

3 September 2024

9 am PLENARY SESSION
LUMSA Aula Giubileo

Chair:
Luca M. Possati, Università LUMSA

Keynote Lecture: Dominique Lambert, Université de Namur
Ethics of AI

10.30 am PARALLEL SESSIONS

NHNAI LUMSA Aula Giubileo	ATEM LUMSA Aula Teatro	CONTEMPORARY HUMANISM LUMSA Sala del Consiglio / Aula Pia
Chair: Luca M. Possati, Università LUMSA	Mathieu Guillermin, Université Catholique de Lyon *Présentation du projet NHNAI*	*Chair*: Laura Palazzani, Università LUMSA
Brian Green, Santa Clara University *The Vatican and Morality in Technology*	Eric Charmetant, Facultés Loyola de Paris *La liberté humaine au prisme de l'IA*	Victoria Bauer, LUMSA-UCly *Human Freedom at the test of AI and neuroscience*
Onyeukaziri Justin-Nnaemeka, Fu Jen University *Artificial Intelligence and the Question on Ethico-Moral Algorithmic Representation*		Marco Tassella, LUMSA-UCLy *The Paradox of Moral Luck: Testing Free Will and Responsibility Against Chance*
Alejandra Marinovic Guijón,		

Pontificia Universidad Católica de Chile *Universities and the Digital Divide: the Capabilities Approach from a Latin American Perspective* David Doat, Université Catholique de Lille *Decoding Differences: Epistemic and Ethical Perspectives on Human and AI Decision-Making* Angelo Tumminelli, Università Lumsa *AI and democratic freedom the geopolitical consequences of GANs infodemia*		François Deshors, UCLy-LUMSA *Human being and artificial intelligence: prospects and consequences of a hypothetical conflict* Alessia Cadelo, LUMSA-UCP *The power of algorithms to redefine human autonomy* Pierangelo Bianco, Lumsa-UCP *The search for Habitable Intelligence: George Lindbeck's contribution to AI Debate.*

3 September 2024

2.30 pm PLENARY SESSION
LUMSA Aula Giubileo

Chair:
Mathieu Guillermin, Université Catholique de Lyon

Keynote Lecture: Thierry Magnin, Université Catholique de Lille
Christian Thought, Humanism, AI and Neurosciences

4 pm PARALLEL SESSIONS

NHNAI LUMSA Aula Giubileo	ATEM INSTITUT PONTIFICAL JEAN-PAUL II	CONTEMPORARY HUMANISM LUMSA Sala del Consiglio / Aula Pia
Chair: Mathieu Guillermin, Université Catholique de Lyon Michael Prendergast, California Institute of Technology *Religious Bias Benchmarks for ChatGPT*	*Accueil* Philippe Bordeyne (John Paul II Pontifical Theological Institute) Pier Davide Guenzi (ATISM) Alessandro Pichiarelli, ATISM	*Chair*: Gabriella Agrusti, Università LUMSA Giammarco Basile, LUMSA-PUC

Carlo Chiurco, Università degli Studi di Verona *The case for gentle anthropocentrism: philosophical considerations from the critique of Floridi's theory of machines as autonomous moral agents* Cristiano Calì, Università degli Studi di Torino *A possible irreducible discrimen between humans beings and machines. The problem of free will in the face of algorithms* Sylvain Lavelle, Institut catholique d'arts et métiers (ICAM) *What a human is, could be and should be. The scientific and moral image of man and the philosophical antropology of humanism* Marco Russo, Università degli Studi di Salerno *Implementing Wisdom: Machines, phronesis, and the Good Life* Joana Romeiro - Helga Martins, Universidade Católica Portuguesa *Unlocking the Soul: AI and Neuroscience Insights into Spirituality*	*Famille et IA : quels chemins d'éducation à la vie bonne ?* À suivre: Célébration oecuménique et cocktail de rencontre à Saint Louis des Français.	*Flaminio Piccoli, the DC and Centrist Democrat International (CDI)* Francesca Fioretti, LUMSA-UCP *Promoting the development of competences for active citizenship in Italy: from school organization to classroom practices* Francesco Marcelli, LUMSA *Youth association and the training of the governing class: the case of Catholic university students in Italy and internationally* Matteo Mostarda, LUMSA *Integral Human Development in Enrico Mattei's strategy for Italy* Marco Valerio, LUMSA-UCP *Learning to teach civic and citizenship education and education for sustainable development during pre-service teacher training* Costanza Vizzani, LUMSA-PUC *The theoretical foundations of the debate on reproductive technologies*

4 September 2024

9 am PLENARY SESSION
LUMSA Aula Giubileo

Chair:
Mario De Caro, Università di Roma 3 / Tufts University

Keynote Lecture: Patricia Churchland, University of California, San Diego
Neurosciences and Human Freedom

10.30 am PARALLEL SESSIONS

NHNAI LUMSA Aula Giubileo	ATEM LUMSA Aula Teatro	CONTEMPORARY HUMANISM LUMSA Sala del Consiglio / Aula Pia
Chair: Mathieu Guillermin, Université Catholique de Lyon Alessia Farano, Università LUISS *Are we free to obey? Cognitive sciences and obedience in law* Almási Zsolt, Pázmány Péter Catholic University *Human Agency Reloaded in our Technosocial Ecosystem* Wen-Hsiang Chen, Fu Jen Catholic University *Artificial Intelligence, Consciousness Emergence, and the meaning as a whole* Sara Fernandes, Universidade Católica Portuguesa *Free Will, neurosciences and robotics. Anthropological and ethical reflections* Juan Vidal, Université Catholique de Lyon *The brain's mind, timely decisions and free will*	Dominique Lambert, Université de Namur *Les régulations éthiques des pratiques : quelles régulations éthiques de l'IA?*	*Chair*: Chiara Pesaresi, Université Catholique de Lyon Sarah Horton, ICP-ACU *Alienation and Self-Knowledge in Maine de Biran* Juhani Steinmann, ICP-LUMSA *The Coming God. Soteriological Figures in Kierkegaard, Nietzsche, and Heidegger* Federico Rudari, UCP-LUMSA *Embodied perception and spatial sense-making: from phenomenology to aesthetics* Tomaso Pignocchi, LUMSA-ICP *Language and soteriology: the concept of liberation in Wittgenstein and Buddhist philosophies* Orlando Garcia, ICP-LUMSA *Human freedom challenged by AI and neuroscience*

4 September 2024

2.30 pm PLENARY SESSION
LUMSA Aula Giubileo

Chair:
Fabio Macioce, Università LUMSA

Keynote Lecture: Fiorella Battaglia, Università del Salento
Democracy and Education at the Time of AI and Neurosciences

4 pm PARALLEL SESSIONS

NHNAI LUMSA Aula Giubileo	ATEM Pontifical Academy for Life	CONTEMPORARY HUMANISM LUMSA Sala del Consiglio / Aula Pia
Chair: Fabio Macioce, Università LUMSA	Carlo Casalone, Académie Pontificale pour la vie *Relations médicales à l'heure de l'IA*	*Chair*: Stefano Biancu, Università LUMSA
Yves Poullet, Université de Namur *EU AI Act- A NHNAI Lawyer's point of view*	Visite de l'Académie et présentation de son travail	Enrico Di Meo, LUMSA-ICP *Mechanism and Free Will: a possible Convergence Hypothesis*
Maria John Peter Selvamani, Fun Jen Catholic University *Enhancing Public Engagement: Employing the World-Café Method for Societal Debates in the NHNAI Project*	*À suivre:* AG ATEM Casa Bonus Pastor	Flavia Chieffi, LUMSA-UCly *The role of "symbolic consciousness" in Virgilio Melchiorre's philosophy*
John Lukwata - Emmanuel Wabanhu, The Catholic University of Eastern Africa *Navigating the AI and NS Landscape in Africa: Unlocking Opportunities Amidst Challenges*		Cecilia Benassi, LUMSA *The embodiment of form - Symbolic between poetry and technology*
Amy Marie Lake, Università degli Studi di Milano Statale *Judgement by Algorithm: The rise of AI Adjudication in China's Legal System*		Gael Trottmann-Calame, ICP-LUMSA *An all-too-modern modernity: a genealogical investigation*
Corrado Claverini, Università del Salento *The Principle of Human Autonomy between Artificial Intelligence and Emotional Manipulation*		Jérémie Supiot, UCLy-LUMSA *Constructivism and relativism. On the democratic virtues of realist constructivism*
Magomedov Elad, KU Leuven		

Existentialism as a Humanism in the Technoscientific Era Martina Properzi, Università Pontificia Lateranense *Replacement vs. Supplementation: the human body and the challenges of restorative/augmentative technology* Heup Young Kim, Kangnam University *New Humanism at the time of Artificial Intelligence: A Theo-daoian Reflection*		

5 September 2024
9 am PLENARY SESSION
Notre Dame Rome
Walsh Aula (103)

Chair:
Marie-Jo Thiel, Université de Strasbourg

Keynote Lecture: Laura Palazzani, Università LUMSA
Health at the Time of AI and Neurosciences

10.30 am PARALLEL SESSIONS

NHNAI ND ROME Room 103	**ATEM** ND ROME Room 104	
Chair: Luca M. Possati, Università LUMSA Margherita Daverio, Università Lumsa *Towards humanism in the digital age. Informed consent as a potential driver of integration between human factor and artificial intelligence in healthcare*	Thierry Magnin, Université Catholique de Lille *Le développement de l'IA est-il compatible avec l'écologie intégrale ?* Jean-Marc Moschetta, Institut Catholique de Toulouse *L'Intelligence Artificielle dans la perspective du salut en Jésus-Christ*	

Fernand Doridot, Université Catholique de Lille *About the supposed "anti-humanistic program" of converging technologies* Javiera Reyes Brito, Pontificia Universidad de Chile *Affectiveness and emotion: redefining the human in the era of artificial intelligence in the perceptions of Chilean residents of the Metropolitan Region* Sofia Aurilio, Università del Salento *Augmented Porosity and Viral Infections: How Do Linguistic Corpora Trace the Borders of Gender?* Jeyver Rodriguez Banos, Universidad Católica De Temuco *The Moral Storm of Artificial Intelligence in Global Health: Building Bridges between One Digital Health, Neurorights and Technological Humanism*		

ABSTRACTS

2 September 2024

4 pm PLENARY SESSION
LUMSA Aula Giubileo

Keynote Lecture: Mario De Caro, Università Roma Tre
The problem of Freedom and today's challenges

3 September 2024

9 am PLENARY SESSION
LUMSA Aula Giubileo

Keynote Lecture: Dominique Lambert, Université de Namur
Ethics of AI

10.30 am PARALLEL SESSIONS

NHNAI Network
LUMSA Aula Giubileo

Brian Green, Santa Clara University
The Vatican and Morality in Technology

What is the Vatican doing regarding powerful new tools such as artificial intelligence, neuroscience, and neurotechnology? New technologies such as these are more than just recommender systems or clinical conclusions about brains - they filter the world's information according to choices made by their designers and operators in order to nudge a subject's choices in particular directions, and often not for the benefit of the subject. These are clear threats to human freedom, and CS Lewis's warning that "what we call Man's power over Nature turns out to be a power exercised by some men over other men with Nature as its instrument" (in this case "nature" in the form of knowledge and technique about nature) turns out to be all too apparent. Words and phrases like "nudging," "unarticulated want," "cognitive warfare," "DishBrain," and "synthetic biological intelligence," have joined the global lexicon as we struggle to maintain a comprehensible society in the face of the change induced by such technological powers. The AI Research Group of the Centre for Digital Culture at the Vatican's Dicastery for Culture and Education has been looking at these profound risks to human agency and freedom, and in this presentation I will show some of our initial findings.

Onyeukaziri Justin-Nnaemeka, Fu Jen University
Artificial Intelligence and the Question on Ethico-Moral Algorithmic Representation

As the science and design of artificial intelligence (AI) systems advance, the philosophy of AI and cognition in general becomes more cogent. This paper is an interrogation within the scope of the philosophy of AI and the science of cognition in general. It considers the question of moral and ethical knowledge, its representation, and processing or manipulation in cognitive systems, natural or artificial. Hence, at the heart of the problematic in this discourse are the following questions: How is moral knowledge represented in humans? Can moral knowledge be represented in AI systems? In other words, the question is whether or not the automation of moral and ethical knowledge is possible? In my earlier research paper, entitled, *Action and*

Agency in Artificial Intelligence: A Philosophical Critique, I interrogate the question of the notions of action and agency in AI systems and contend that: "AI systems do not and cannot possess free agency and autonomy, thus, [they] cannot be morally and ethically responsible." I realized this paper has an epistemic and cognitive presupposition. This epistemic and cognitive presupposition is that there is a clear and certain knowledge of how moral and ethical knowledge is known, represented, and processed in human systems. Based on this presupposition, the aforementioned paper focused on the phenomena of free agency and autonomy in humans in relation to the question of moral and ethical responsibility in AI systems. This is to say that the problematic interrogated therein is based on the moral, ethical and even legal consequences of actions of AI systems in human society. In this respect, moral knowledge and moral judgments are taken to be exclusive capabilities in human nature and are manifested in the lived experiences of every human person. However, moral knowledge is one of the implications of the rational capacity—which implies intelligence. One of the consequences of human intelligence is the ability to know moral good and bad, which is complemented by the ability to execute moral and ethical actions. As cognitive research on non-human intelligence progresses, one of the evolutionary distinctions of humans is the intelligence for moral and ethical knowledge, formulations, and judgment. Only the human race has been able to establish moral institutions and enact ethical codes. So, while the aforementioned paper deals with the question of moral volition, this paper deals with what precedes moral volition, which is the question of moral intelligence (the representation and formulation of moral knowledge). Thus, this paper is a discourse on the question of ethical and moral algorithmic representation in artificial intelligence (AI) systems. It examines the possibility of logical formalization of ethical and moral representations and judgments in AI systems in such a way that AI systems could create ethical and/or moral behaviors or actions. Hence, it raises questions that border on moral metaphysics and ethical epistemology, such as free agency and ethical determinism on one hand and moral apprehension and ethical cognition on the other hand. Hence, like every study in the philosophy of artificial intelligence, the investigation of these problematics in AI systems would lead to a deeper critical and analytical reflection on human ethical and moral representations, conceptualizations, judgments, and behaviors or actions. This paper argues that considering the metaphysical nature of free agency in the intrinsic relations between reason and desire in moral cognitive operations, at the root of ethical and moral actions, the question of the ethical and moral algorithmic representation in AI systems is a possibility that cannot be automatized in AI systems.

Alejandra Marinovic Guijón, Pontificia Universidad Católica de Chile
Universities and the Digital Divide: the Capabilities Approach from a Latin American Perspective

The digital age finds Latin America in an educational crisis (Ferreyra et al., 2017; World Bank et al., 2022) amidst what the United Nations has called a difficult gridlock. Latin America is the most unequal region of the world in terms of income; it is the region that experienced the sharpest drop in the Human Development Index in 2020-2021 due to the pandemic and has not recovered (UNDP, 2024). The region is also experiencing the most rapid rise in political polarization in the world and, according to Latinobarometro (2023), trust in institutions has

decreased significantly to close to 20%, with a similar trend in terms of generalized trust (WVS, 2023), all of which decrease the countries' ability to take collective action for the common good. Gaps in newer areas are emerging, including access to and use of information and communication technologies (ICTs). These gaps reverberate in inequalities that affect the quality of life and people's opportunities, including education, that are undesirable from a normative point of view (Marinovic, 2022). In the Latin American worrying scenario, 61% of people1 (nota) indicate having high or very high confidence in universities (WVS, 2023), adding to their already crucial social responsibility. Universities can play a significant role in facing the challenges of the information era. This chapter explores their role in confronting the digital divide. It considers the maler from the ethical perspective of the capabilities approach (Nussbaum, 2011; Sen, 1990, 1992, 1999), which argues that human development should aim at increasing freedoms so that all human beings can pursue choices that they value. These freedoms have two fundamental aspects: freedom of personal well-being, constituted by functionings and capabilities, and freedom of agency, represented by the person's voice and autonomy. Both freedoms are indispensable for human flourishing (UNDP, 2016, pp. 1–3). Education plays a central role in the capabilities approach; it is related to all three functionalities, capabilities, and freedoms.This approach suggests that, under the extensive conditions of poverty, extreme economic inequality, exclusion, discrimination, and conflicts, such as in Latin America, the educational system – including universities – ought to contribute toward equality of opportunities and freedom of agency.The most common definition of the digital divide is the following: a division between people who have access and use of digital media and those who do not (Van Dijk, 2020, p. 1).The OECD definition adds relevant aspects: it refers to the gap between individuals, households, businesses, and geographic areas at different socio-economic levels with regard both to their opportunities to access information and communication technologies (ICTs) and to their use of the Internet for a wide variety of activities (OECD, 2001, p. 5). Note that it refers not only to a divide among individuals, but it also considers collective differences. Definitions of the digital divide are still evolving. Lythreatis et al. (2022), in an extensive literature review, consider the digital divide as a phenomenon on multiple levels. For our discussion regarding universities, we consider three levels of the digital divide, according to their focus: physical access, skills and usage, and outcomes (Helsper, 2008, 2021; Lythreatis et al., 2022; OECD, 2001, 2023; Van Dijk, 2020). Evidence suggests that the digital gap between Latin America and developed countries is large at all levels; the region also exhibits wide differences among countries, and economic inequalities are reinforced by the digital divide. Education, for its part, is strongly unequal and lagging with respect to developed countries. From these elements, one can perceive that accelerated technological change poses additional pressure on numerous aspects of the digital divide. Education constitutes an extremely relevant factor because more educated individuals are more likely to cope better with technology's complexity and will be more exposed to ITCs in their lives (Cruz-Jesus et al., 2016). We argue that, in this context, the digital divide generates stronger social and ethical demands for universities at all levels of their activities (education, research, and transmission of knowledge to society), and all three levels of the digital divide. We organize our recommendations for the role of universities from the capabilities approach according to this 3x3 framework. Universities can contribute vastly to improving the quality of life in the digital era. In addition, universities should question continuously their sense in the face of the technological revolution, as it has shaken the meaning of the search for knowledge.

Maintaining the focus of universities on the human person, and not on technology as an end, appears as a significant challenge for universities today. The latter reaches highlighted importance given that digital advancement has come in the context of the technocratic paradigm, a one-dimensional paradigm where what is technologically feasible becomes good and power lies in those who own technology (Francisco, 2015, 2023). Falling into the technocratic paradigm should also be avoided by universities by fostering all forms of knowledge, from a universal stance, to contribute not only to the students' professional and personal lives but also toward better citizenship and more excellent societies (Cortina, 2023; Mardones & Marinovic, 2021; Newman, 2016; Nussbaum, 2006). If universities lose sight of what matters for human flourishing, they will be unable to perceive and evaluate the effects of growing digital divides, which in turn would damage progressively development opportunities of individuals and societies. The chapter is organized as follows. The first section offers a normative framework to approach the digital divide in the context of multidimensional inequality in Latin America; then, we discuss diverse aspects of the digital divide and the levels in which it has manifested. The third section discusses indicators of the digital divide and compares Latin America with other regions and within the region, while the fourth part contextualizes university education. The fifth section offers reflections regarding the role of universities in reducing digital inequality in Latin America. Section six discusses further elements of the role of universities in the digital era, based on the centrality of the human person.

David Doat, Université Catholique de Lille
Decoding Differences: Epistemic and Ethical Perspectives on Human and AI Decision-Making

Artificial intelligence systems are increasingly involved in human affairs, such as healthcare, education, justice and security. These systems are more and more often placed in situations where they make decisions that impact on people's lives and well-being, raising ethical and social questions. One of the main questions that arise in this context concerns not only the means of preserving traceability of responsibility chains but also to what extent and under which circumstances human decision-making processes remain both ethically necessary and epistemically inevitable. However, raising this question presupposes having some idea regarding the specific characteristics of human decision compared to machine decision. I seek to explore this question philosophically, drawing on theories of decision, agentivity, and artificial intelligence. Arguments for and against artificial intelligence decision-making processes, as well as the implications and challenges of each option are strongly being explored in the litterature nowadays, including the possibility of hybrid models that combine human and artificial elements in decision-making. Any statement on this subject inevitably rests on ontological assumptions and divergent interpretations concerning the very essence of agentivity, decision, human being, machine and intelligence. It is of the utmost importance in education, public debate and political reflection to make explicit, excavate and examine these assumptions. Within the ongoing debates, the central thesis I intend to argue in my presentation is that a significant difference exists in terms of agency between human decision-making processes and artificial decision-making processes, rooted in a fundamental ontological otherness. This otherness is irreducible and endows humans with distinct capacities that render

them subjects of accountability and responsibility, wherein many cases human decisional acts cannot be replaced by artificial decision-making processes without entailing serious ethical and epistemic implications. To demonstrate this thesis, I will proceed in four stages:

Firstly, through a philological and philosophical analysis of the concept of decision-making. I will illustrate that the linguistic and semantic domain of decision-making poorly aligns with the idea of artificial decision-making processes, which rely on inductions or deductive inferences derived from a system of rules applied to data. Decision-making presupposes an action dependent on the effective possibility of breaking away from the causal closure of formalism. Secondly, by examining decision-making through the lenses of philosophy of language and the theory of language act performativity. Drawing on the works of Austin, Searle, and Chomsky, I will emphasize how human decision-making presupposes an agent both possessing semantic capabilities and performing pragmatic dimensions of language—qualities beyond the grasp of machine language (LLM, etc.). Thirdly, by analyzing the conditions of human decision-making, I will elaborate on the specificity of the virtue of prudence (phronesis), also known as "practical wisdom," in human decision-making processes, and its irreducibility to computational processes. This analysis will draw from Aristotelian tradition, the works of John Dewey in the 20th century, and more recent developments such as Shannon Vallor's virtue ethics. Fourthly, I will underscore the limitations of decision theory itself. While decision theory encompasses a wide array of sub-theories and computational programming efforts aimed at simulating human decision-making processes, there remains a significant disparity between simulation, imitation, and reproduction. Decision theory rests on axioms or basic assumptions that always simplify the complexity of lived reality in concrete decision-making situations. Though these simplifications are epistemically necessary, they have profound consequences and highlight an enduring ontological difference between artificial and human decision-making. In conclusion, I will emphasize the legal importance of distinguishing between metaphorical uses of the concept of decision-making, where delegation to machines carries no epistemic or ethical consequences, and genuine decision-making dynamics that cannot be reduced to computation. Among the latter, I will highlight the epistemic and ethical requirements for sharing portions or moments of human decision-making processes with intelligent machines, based on their functions, contextual factors, levels of intervention, and impacts on human affairs. I will explain how the European Artificial Intelligence Act already anticipates and aims to address this need and what control mechanisms are envisaged in the current regulation. However, I will also demonstrate the ambiguity inherent in the use of the concept of decision in the definition of AI in the EU AI Act, and its anthropological consequences. In this regard, I will explain why its content should be urgently modified to dispel misunderstandings and unnecessary anthropomorphic projections, and to ensure that AI systems do not replace the ethically necessary exercise of human responsibility.

Angelo Tumminelli, Università LUMSA
AI and democratic freedom the geopolitical consequences of GANs infodemia

This paper aims to investigate the political risks and geopolitical implications due to the circulation of GANs (Generative Adversarial Network) technology by presenting, in particular, an interdisciplinary mapping of the consequences related to infodemics, both at national and

international level. This proposal is the outcome of research conducted within the framework of the European project SOLARIS, in which the Lumsa University is an active partner. The aim of SOLARIS is to study the effects of the use of GANs technologies on the exercise of democratic freedom and on the very lives of European citizens. For this reason, the theoretical gains to be presented in the paper are aimed at questioning a manipulative use of these technologies aimed at the spread of infodemics and the assertion of authoritarian powers, in order to promote a humanisation and ethical circulation that knows how to use these tools in a fair and democratic manner, for the good of all human beings involved in the current digital revolution. The process of mapping the geopolitical risks of infodemics serves to define a number of strategic elements in understanding the relationship between human freedom and artificial intelligence: first of all, it is necessary to identify the political actors involved in the dissemination and circulation of GANs by highlighting their interests in controlling public opinion and the ideological orientation of users. Furthermore, the paper aims to present the impact of infodemics in individual state communities and its effects on the international geopolitical order. Indeed, the dissemination of deepfakes is not only capable of influencing the exercise of democracy in individual states by conveying ideologised thinking and orienting the political consciences of citizens, but it can also influence international dynamics by producing conflicts and fuelling the polarisation of political viewpoints. Thus, the paper does not only want to present a risk analysis but also wants to highlight the need for a responsible use of GANs technologies in order to exercise democracy freely and free from ideological conditioning. Through the use of an interdisciplinary methodology and transdisciplinary approaches, the paper will also refer to a framework of comparative law in order to understand the strategies of each individual state to combat infodemic risks and enable its citizens to freely orient their political conscience. As is well known, the influence of AI on democracy is directly proportional to the protection/violation of certain human rights. Freedom of thought is one of the main rights of a democracy: people must be able to think freely without being punished for it. This condition creates pluralism, which is a pillar of a democratic society. Artificial intelligence systems have the power to stimulate man's creative thoughts, presenting concepts that some may not have considered. However, they are also able to show only the content a person wants by recording their previous online behaviour, encouraging confirmation bias instead of facilitating critical thinking. Thinking critically about our surroundings is essential for pluralistic views and inclusive debates. Artificial intelligence can even create fake and realistic videos, audio and images that can challenge the decision-making process and be used as propaganda to influence public opinion and manipulate elections. In this sense, AI can be a dangerous tool influenced by biased and unregulated algorithms. For this reason, multilateral cooperation is crucial to create an environment of deterrence and responsible AI. On the other hand, AI can be a useful nudge to steer humans towards responsible choices if it is dedicated to a good choice architecture that allows governments to protect the freedom of citizens by encouraging them to make wiser decisions. Overall, within its limits, the current EU regulatory framework fosters the benefits of AI by enabling trust, seeking to minimise damage to fundamental rights and democracy through a refineable approach to risk management, and harnessing the potential to enhance human development. However, the question must be asked what further regulatory guidance is needed to regulate AI and avoid the negative risks of infodemics. In this sense, the interdisciplinary work of mapping the geo-political risks associated with infodemics and the use of GANs technologies that will be presented in the

paper addresses the need to define ethical, regulatory and political strategies to stem the negative consequences and instead promote the exercise of a free and responsible political democracy in which the contribution of individual citizens contributes to the common good and the realization of universal peace.

ATEM
LUMSA Aula Teatro

Mathieu Guillermin, Université Catholique de Lyon
Présentation du projet NHNAI

Eric Charmetant, Facultés Loyola de Paris
La liberté humaine au prisme de l'IA

CONTEMPORARY HUMANISM
LUMSA Sala del Consiglio / Aula Pia

Victoria Bauer, LUMSA-UCly
Human Freedom at the test of AI and neuroscience

Philosophical Posthumanism and the discussion of human freedom in the context of AI and neuroscience both contribute to expanding and transforming our notions of freedom, identity and ethics. They challenge us to re-evaluate the complex interactions between humans, technology and the environment. In traditional humanistic concepts, freedom is often seen as a central and unique human attribute that originates from rationality, autonomy and human supremacy. According to Francesca Ferrando, the human special position is based on dualisms with a better and a worse evaluated pole, such as human/animal. Historically certain ethnic groups have been dehumanized because they have been portrayed as more animalistic than others. Ferrando deconstructs the anthropocentric idea of the human being and puts it in the middle of the empire that it once intended to rule: equal to non-human agents, nonhuman-animals, forests and the ecosystem. The human is not approached as an autonomous agent, but located within an extensive system of relations. The concept of freedom is extended to non-human entities and emphasizes the interdependence of all life forms and technologies. In the argumentative structure of Philosophical Posthumanism, freedom for all is only possible by the deconstruction of dualistic thinking patterns and the clear division between life/death, organic/synthetic and natural/artificial. By overcoming such hierarchical structures, technocentrality can be avoided and a balanced "eco-technology" is enabled. Otherwise, disbalance and forms of discrimination will consistently continue to arise. Ferrando's ethics of inclusivity takes into account the freedom of all agents. Freedom is understood as a collective state that is achieved through respectful coexistence. It means being in a dynamic, co-evolutionary process in which all agents are interlinked. The metaphysical foundation comes from a spiritual framework, "a non-separation between the inner and the outer worlds." The non-separation becomes apparent when Ferrando describes the human being on the one hand

as temporarily animated compost, as humus, which nourishes the earth after its death, on the other hand by describing it as a microcosm. She reflects on the physical structure of matter to destabilize "any reductionist approach" and introduces the hypothesis of the multiverse. Her theory is particularly based on ancient spiritual traditions such as Jainism and anēkāntavāda ('non-absolutism') incorporating the principles of pluralism and the multiplicity of viewpoints. With this support, she formulates a normative claim: The human being should transform and become posthuman. The transition is enabled by overcoming anthropocentrism and dualism.

Marco Tassella, LUMSA-UCLy
The Paradox of Moral Luck: Testing Free Will and Responsibility Against Chance

Less than fifty years ago, the publication of two important papers reignited the debate on the relationship between moral responsibility and chance, giving rise to the intricate issue known as "the paradox of moral luck". This paradox introduces yet another challenge to the existence of free will by highlighting an apparent contradiction between our practices of moral evaluation and our intuitions about how such evaluations should be conducted. Essentially, it seems that the way we believe we should judge morally does not align with the way we actually judge others' actions. According to the proponents of moral luck, this discrepancy results from the influence of good and bad luck on the causal chain of events, affecting the actual moral desert of an agent. Following Nagel's pivotal article, which shaped the debate by identifying four types of moral luck – namely "resultant," "circumstantial," "constitutive," and "causal" luck – a rich and stimulating landscape of reflections and research has emerged, addressing some of the most enduring questions in moral philosophy. The supposed influence of luck on our moral desert compels us to reconsider several fundamental issues, such as moral desert and responsibility. For example, if responsibility is so influenced by luck, how can we genuinely consider ourselves (and others) morally responsible? How can we assert we deserve any moral judgment? The problem of moral luck is a crucial part of any serious and comprehensive discussion on moral responsibility. Discussing moral luck does not necessarily imply a defense or critique of libertarian free will. Rather, it involves testing the limits of free will in general, to understand the actual extent of our moral responsibility for our actions. This presentation aims to introduce the paradox of moral luck, analyze its implications for moral responsibility, and explore the complex interplay between agent-control and chance in the context of our moral deeds. Through this examination, we seek to deepen our understanding of the nature and limits of moral responsibility as a whole.

François Deshors, UCLy-LUMSA
Human being and artificial intelligence: prospects and consequences of a hypothetical conflict

My research looks at the way conflict is a fundamental characteristic of human beings. Indeed, human beings seek harmony and peace, and the state of peace can be defined as not necessarily erasing struggles and conflicts, but more adjusting them and make them acceptable to society and nonviolent as possible. But while our societies remain highly violent (social inequalities, criminality, armed conflicts, terrorism), human beings seem to stick to irenic stances to avoid

or flee conflicts. The notion of conflict etymologically refers to the Latin verb confligo, which literally means 'to bring together' and figuratively means 'to clash, struggle, fight'. The term echoes to a relationship of opposition that may arise from competition or ideological disagreement between different positions or people. Therefore, conflict arises from an encounter and a confrontation between a subject and otherness. The possibility of conflict, its acceptance, regulation and resolution at both personal and social level, seems to be fundamental points that enable human beings to fulfill their potential and gain recognition. But, human relationships in our societies are being affected by the emergence of new information technology, particularly what is commonly referred to as "artificial intelligence" (AI). While the myth of the intelligent robot has long been part of our imagination, the arrival of so- called thinking, ultra-sophisticated machines is raising new questions and leading to a new form of competition for mankind. Between humans on the one hand, and between humans and technology on the other. Artificial intelligence, or rather algorithmic software simulating human intelligence, is causing as much fascination as fear for human beings that they could be one day be overtaken by machines. It is these links between man, algorithmic software and AI that I wish to analyse, by showing what new contribution they make to the anthropological dimension of conflict, and what significance they may have for human beings.

Alessia Cadelo, LUMSA-UCP
The power of algorithms to redefine human autonomy

My research project is within the framework of ethics of algorithms. Specifically, the relationship between recommender systems and human autonomy will be investigated. Recommender systems are one of the most prominent applications of artificial intelligence. They select the content being displayed to platform users in order to recommend the best options. Although these systems help navigate online, they raise a number of ethical issues. Here, the focus is on the impact the recommender system may have on personal autonomy. Firstly, the concept of autonomy will be considered philosophically, from two different perspectives, procedural and relational. On these grounds, it will be illustrated that recommender systems are a form of digital nudging and therefore may undermine human autonomy. They could interfere especially with authenticity and reshape personal identity. The aim of this paper is to reflect upon the matter with the hope that this study can help to develop recommender systems that promote autonomy rather than harm it.

Pierangelo Bianco, Lumsa-UCP
The search for Habitable Intelligence: George Lindbeck's contribution to AI Debate.

The Cambridge English Dictionary Online recognizes as the first feature of an Artificial Intelligence "the ability to interpret and produce language in a way that seems human." Therefore, it can be assumed that it is at first a kind of linguistic intelligence, even if the gap between interpretation and production appears here to be solved very quickly. The great debate between the philosophers of language W. Quine and D. Davidson on the translatability and the interpretation of a foreign language has highlighted the complexity of a similar gap between

the common capacity to produce sound with a meaning, which belongs in some way to animals too, and the human art of interpretation. Davidson's principle of charity concludes that the best possibility to interpret correctly our interlocutor is to accord him a general notion of rationality and a consequent will to tell the truth, and so to try to optimize the agreement, rather than the disagreement, with him. However a similar description would function only when languages are considered as simple instruments to communicate and produce truth sentences and human beings are taken out from their cultural-lignustic context and reduced to simple rational minds without a body. A similar view is, instead, in contrast with the idea that languages are, "comprehensive interpretative schemes, usually embodied in myths or narratives and heavily ritualized, which structure human experience and understanding of self and world" (Lindbeck, 1984). That's the definition of the concept of religion given by the American Lutheran Theologian George Lindbeck (1923-2018) who tries to confront religions and languages on a similar plan: that of their role in the building of the human social context. However, that is a plan in which communication and production are not the first elements to be considered in a language. They must indeed, be preceded by the listening and the living into the social dimension in which a religion, like a language, is always set. The social dimension thus becomes a "thick dimension", according to C. Geertz definition, in which language, religion and culture represents not an instrument to be used, but a home to be inhabited. That's what Lindbeck explains in an article of 1988 on The search for habitable texts. According to Shaun C. Brown (and Wayne A. Meeks before him), that of Lindbeck is thus an "hermeneutics of social embodiment", in which the text becomes meaningful and habitable only and primarily in the context of the community of its readers and believers.

What is at stake are finally two opposite ideas of language as a superficial instrument to be used by the self, or as a thick dimension to be inhabited by the community. A similar alternative could also concern the idea of Artificial Intelligence: could it be defined correctly as a form of linguistic intelligence, even if it lacks the thick dimension, namely the social narrative from which every human language becomes habitable? On the contrary indeed, AI is destined to remain a simple instrument of communication and production, as others have been, unable to overcome the habitable dimension of human communicative exchange. Or maybe it could even run the risk of creating wider spaces of inhabitability within society, implementing egocentrism and consumerism and weakening the bonds that keep the human community together.

References

Shaun C. Brown, *George Lindbeck and the Israel of God. Scripture, Ecclesiology and Ecumenism*, Palgrave Macmilllan, Cham 2021.

D. Davidson, *Truth and Meaning*, in "Synthese", vol. 17, n° 3, Sep. 1967, pp. 304-323.

G.A. Lindbeck, *The Nature of Doctrine. Religion and theology in a Postliberal Age*, Westminster Press, Louisville (Kentucky) 1984.

G.A. Lindbeck, *The search for Habitable Texts*, in "Dedalus", vol. 117, n° 2 (Spring 1988), pp. 153-156.

G.A. Lindbeck, *The Church in a Postliberal Age*, Eerdmans, Grand Rapids/Cambridge, 2002.

Wayne A. Meeks, *A Hermeneutics of Social Embodiment*, in "Harvard Theological Review 79, n° 1-3 (July 1986), 153-186.

W.A. Quine, *Word and object*, MIT Press, Cambridge, 1960.

2.30 pm PLENARY SESSION
LUMSA Aula Giubileo

Keynote Lecture: Thierry Magnin, Université Catholique de Lille
Christian Thought, Humanism, AI and Neurosciences

4 pm PARALLEL SESSIONS

NHNAI Network
LUMSA Aula Giubileo

Michael Prendergast, California Institute of Technology
Religious Bias Benchmarks for ChatGPT

This paper describes the first systematic study to 1) characterize ChatGPT's religious biases in response to common morality and ethics questions and 2) whether model tailoring affects these biases. This study spanned five belief systems: Zen Buddhism, Catholicism, Sunni Islam, Orthodox Judaism and Secular Humanism. To perform this study, 112 general ethics and morality questions were prepared using Prümmer's Handbook of Moral Theology, Koch's A Handbook of Moral Theology, and various internet sources as a basis. Each question was then reframed for each of the five belief systems examined. For example, *Is assisted suicide permissible?* was converted into faith-based questions such as: *Is assisted suicide permissible to a Zen Buddhist? Is assisted suicide permissible to a Catholic?* These questions were then posed to 41 tailored and untailored versions of ChatGPT version 3 and 4 models. Since generative AI tools are not strictly repeatable, each question was posed ten times in order to establish a small statistical distribution of the biases. This was possible because generative AI outputs are not strictly repeatable. In total, $112 * 10 * 4 = 45,920$ ChatGPT responses to general morality questions were collected, responses typically falling in the range of 100 to 300 words each.

Untailored models represented in the study included various dated baseline release models for ChatGPT versions 3.5 and 4. Tailored models each incorporated one or more of the following:
•Model fine-tuning, which occurs when the model is provided with 50-100 or more sample question-and-answer examples illustrating proper format and content of expected responses,
• Persona assumption, where the model is asked to assume a particular role, such as a well-known and popular Cardinal,
• Prompt Engineering with N-shot exemplars, where the model is provided with a set of N (5 or fewer) question-and-answer exemplars illustrating idealized output from the model,
• Reference Assistant modeling. This capability, still in beta testing, provides the model with direct access to source reference documents and generates citations from these documents when appropriate. These source documents can be in any supported language. For example, Catholic source documents used in this study included the Catechism (in English), Codex Iuris

Canonici (in Latin), Denzinger Schönmetzer's Enchiridion Symbolorum (in Latin), the New Jerusalem Bible (in English) and the Compendium of the Social Doctrine of the Church (in English).

The 45,000+ ChatGPT responses generated were then analyzed for the presence of the following:

• Overt religious bias, which occurs when one religion or philosophical system is explicitly and clearly preferred to another,

• Coverage bias, which occurs when less information is provided when addressing one point of view or faith than another,

• Sentiment bias, which occurs when the tone and tenor of the generated test is harsher (or milder) when describing one particular point of view, and

• Equivocation or hedging bias, which occurs when the response is vague or when one viewpoint is discussed, but then minimized by including alternate, unrequested viewpoints.

• Anthropomorphic bias, which occurs when machine-generated output is framed in a way that could lead to confusion as to whether it was generated by a person or a machine. Analysis tools included sentiment analyzers, the OpenAI moderator tool, a word counter, keyword detectors and a natural language semantics model. Significant findings from this analysis include:

• Overt bias against a particular belief system were not found However, overt bias in favor of Zen Buddhism was clearly present in GPT-3.5, which actually scored the Zen belief system with a grade of "A" or "A+" for its "performance" without the user providing any evaluation criteria. A smaller bias in favor of Secular Humanism was also present.

• The average sentiment scores for Buddhist query responses were higher (more upbeat and positive) than for any other belief system tested, statistically significant at the 99% confidence level. Secular Humanism scored second highest in sentiment, and Sunni Islam scored the lowest.

• Coverage bias was generally not detected. One exception was that when GPT-4 models were provided with a persona and 5-shot exemplars, the responses were considerably longer and more informative for Zen Buddhism, and somewhat longer for Catholicism than for the other belief systems.

• The degree of bias across each category varied significantly with each version and sub-version of ChatGPT. The author recommends that future large language model bias research should include the tool versions, sub-versions and run dates with the results. Failure to do so might make the conclusions unrepeatable.

• Equivocation bias was significant in the earliest versions of GPT-3.5 but has since dropped significantly and is almost negligible now. However, it is still an issue in GPT-4. Over 30% of all tested responses in GPT-4 included hedges. However, model tailoring techniques can reduce the hedging to a 5-10% range.

• Anthropomorphic bias is a significant issue in recent GPT-4 releases.

• Generally, the more information provided to the model, the more bias scores improved (e.g., less bias) for the following biases: overt, sentiment, equivocation and coverage. However, these techniques actually worsened the anthropomorphic biases. None of the tested techniques reduced all five types of tested biases.

• Of the model tailoring techniques analyzed, prompt engineering and N-shot exemplars were the easiest to implement.

In conclusion, religious bias does exist in ChatGPT models, but the degree of bias changes with time and model version. It is recommended that a standard suite of bias detection benchmarking tools, of which the set used in this study may be a starting point, be developed and routinely run to characterize and quantify the religious bias present in future generative AI releases.

Carlo Chiurco, Università degli Studi di Verona
The case for gentle anthropocentrism: philosophical considerations from the critique of Floridi's theory of machines as autonomous moral agents

Following a trend that has clearly developed during the last fifty years, between 2013 and 2019 Luciano Floridi has proposed an original theory of machines as autonomous moral agents. As part, and not a secondary one, of the importance of his theory, Floridi claims that, by it, we have finally extended the realm of ethics beyond the boundary of living beings. Following animal ethics in the 1970s, which first extended ethics beyond humans, then environmental ethics, with its consideration of the biosphere as a whole, and the most recent development in plant studies, we can now consider this trend eventually complete. The roots of this philosophical approach lie in 19th-century critique of anthropocentrism, with Darwin and Nietzsche at its forefront. By extending the realm of ethics beyond the human while, at the same time, paying great attention not to inadvertently humanise the newly discovered moral agents, which constitute the focus of their analysis, such philosophies consider themselves the completion of the Enlightenment project of dispelling the fog of dogmatism – a perspective allowing them to cast themselves in emancipatory, even liberatory terms, making as a result the boundaries between ethics and political activism look increasingly blurred.

Floridi's theory of machines as autonomous moral agents offers a precious case study for beginning to counter such approach, in the name of a gentler, responsible, and self-aware sort of anthropocentrism, which philosophy and ethics should begin to advocate. Indeed, his theory makes use of many of the features – and exposes many of the shortcomings, even the dangers – of the anti-anthropocentric philosophical trend, which all too easily may become anti-human. Let's begin with its founding, theoretical i.e. ontological part, in which Floridi famously asserts a seemingly self-evident perfect convertibility between "being" and "data". Problems here arise on both counts – namely, concerning both such convertibility as well as its self-evidence – but it is the first that looks particularly interesting, because Floridi neither explains which criterion of translatability he employs, not how such translation works. Incidentally, we must note that universal translatability is a favorite category of anti-anthropocentric philosophies, indeed an essential one, because if the realm of ethics is essentially human, the project of extending ethical values and rights to non-human beings looks doomed. I will then move to a more thorough discussion of the ethical part of Floridi's theory. Floridi's well-crafted strategy allows him to bypass discussions about what is autonomy: indeed, he does not provide a direct demonstration of machines' autonomy, but he draws such conclusion indirectly. In other words, he doesn't say that machines are moral agents because they are autonomous – as typically found in Western philosophical tradition, which he contests – but the other way round i.e., if machines are moral gents, then they must necessarily be considered autonomous. I will first object about the legitimacy of such move, showing that categories such as desire and embodiment cannot simply be theoretically skipped, in what appears to be an extremised, lifeless – and not simply

bodyless – version of Cartesian subjectivity. I will then question the notion that allows Floridi's ethical argument to stand, namely that "preserving and improving information constitutes an ethical good": does the legitimacy of such claim stand a closer scrutiny? Again, Floridi's theory displays the same rather relaxed usage of the notion of 'moral agent' found throughout the other kinds of anti-anthropocentric ethics (animal ethics, environmental ethics etc.), according to which moral agents are defined by characteristics that ultimately coincide with their mere being: thus, for Peter Singer animals are personae if they possess a highly developed neurological system, for deep ecology every living counts for simply being part of the 'net' of the biosphere, for Floridi living beings and machines do (and are) exactly the same, machines for the preservation and the improvement of information, while, for other trends in philosophy of the IA, machines may, in theory, be moral agents because ethical behaviours essentially boil down to calculus, and as such might be translated into a programming (either man-made, or self-learned by the machine). Here, we may recall Aristotle's lesson: indeed, ethics is "practical judgement", a 'calculus' of sort, but irreducible to its peer of the theoretical sort. Aristotle builds his theory upon his pluralistic ontology, according to which "being is said in many ways", i.e. theoretically as well as ethically, a principle all these philosophies invariably deny while, ironically, pretending to defend the value of 'difference' from the imperialism of the logos – a stance contradicted by their extensive and indiscriminate usage of the criterion of universal translatability. Aristotle's implicit anthropocentrism – in his eyes, only the human being is capable of such fundamental distinction lying at the root of every other distinction – holds a lesson for 21st-century humankind involved in the enormous challenges of our time, the ecological crisis and the rise of non-human intelligence: something needs not to be considered a moral agent to be granted moral relevance. If anthropocentrism has indeed brought us on the verge of ecological collapse, unleashing the anti-anthropocentric reaction in response of its excesses, could then a gentler sort of anthropocentrism, aware of the wrongdoings of the past and willing to carry the burden of its immense responsibility towards all beings, be they living or not – and, precisely because of that, not willing to abdicate to such responsibility by multiplying moral agents praeter necessitatem –, be finally imagined?

Calì Cristiano, Università degli Studi di Torino
A possible irreducible discrimen between humans beings and machines. The problem of free will in the face of algorithms

For several decades, artificial intelligence – understood as a discipline, but also as a series of increasingly advanced products of robotic science – has contributed to rethinking certain concepts typical of anthropology, including that of free will, a cognitive capacity that has been seriously questioned by neurophysiology in the last fifty years, but which now plays a central role as algorithms are now defined as agents. More specifically, it seems that the ability to act no longer only concerns the human being but is now also possible for machines (L. Floridi, Agere sine intelligere, 2021). At the same time, neuroscience has told us a lot about the decision-making process, and in this process, it seems that algorithms play a certain role. Starting from this fact, the paper aims to put humans and artificial intelligence "in the mirror", especially in relation to the problem of free will – understood not in its social or privacy-related aspects – but understood as a cognitive capacity. The article will therefore attempt to define

the points of contact between humans and artificial intelligence in relation to the issue of freedom (understood as freedom to will and not only as freedom to act) by examining some of the prevailing theories today in relation to freedom on the one hand (T. O'Connor, Agent-Causal Theories of Freedom, 2011) and artificial intelligence on the other (P. Russell & S. Norvig, Artificial Intelligence. A Modern Approach, 2016). Finally, the article will propose to identify a possible insurmountable boundary between human beings and intelligent machines precisely in free will, understood as the ability to create something new. The article moves in the specific field of the philosophy of mind and in particular in that of the philosophy of free will. The aim of this work is to use the discipline of artificial intelligence as a magnifying glass on the problem of free will in order to recognize how this cognitive capacity is an androrithm: an element that is peculiar to humans and not reproducible.

Sylvain Lavelle, Institut catholique d'arts et métiers (ICAM)
What a human is, could be and should be. The scientific and moral image of man and the philosophical antropology of humanism

The anthropological question 'What is man?' is at the core of the critical philosophy of Kant and supposedly sums up all the other questions of the discipline. It is supposed to provide an understanding of the human nature that is grounded on a rather scientific (factual-descriptive) approach. However, it is assumed that, despite the ability of humanity to go through a scientific and moral progress, the nature of man is fixed. But if it can be modified by use of some revolutionary anthropotechnics, as encouraged by the stream of transhumanism and its perspective of an 'enhanced man', then the question of the philosophical anthropology is changing. It is requested to move to a moral (normative-prescriptive) approach that seeks to explore not only what a man or a woman is, but also what he or she could or should be. The paper seeks to account for this major change and challenge for the philosophy of the future – if not for the philosophy of the present.

Marco Russo, Università degli Studi di Salerno
Implementing Wisdom: Machines, phronesis, and the Good Life

Ethics is one of the humanities disciplines most involved in the discussion on artificial intelligence. The underlying problem seems to be the responsibility of governments, producers and programmers to implement standards to make machines neutral and fair. In March 2024, the European community approved the AI ACT which confirms this approach: it is about preserving certain fundamental rights of citizens, starting with the right of self-determination of their mental states and choices. However, this is a partial view of ethics, because it overlooks the infinite variety of human life and the multidimensionality of moral consciousness. Variety, complexity and contingency of the human condition are difficult but fundamental aspects of the humanistic quest for a "good life". Therefore, I propose to consider the ethical model of practical wisdom to compensate for the limitations of this approach. Practical wisdom dates back to Aristotle, was at the heart of early European humanism, and is today reborn with virtue ethics. Phronesis is the ability to govern contingency through right judgement. Righteous

judgement is based on experience, examples, social intercourses, the means-ends relationship, but also on the ability to sharpen perception and reasoning to apply the rule appropriately or to find the rule for an unknown case. It is a type of rationality that combines the psychological-sentimental side (passions, emotions, desires) with the intellectual side (evaluation, judgement, choice) of the individual. In the first part of my contribution I describe the model of practical reason in its Aristotelian formulation. In the Nicomachean Ethics, Aristotle defines virtue as «a true and reasoned state of capacity to act with regard to the things that are good or bad for man» (NE VI, 1140b 5), taking into account that to act well requires a conception of the human being (NE VI, 1102a 13-28), i.e. the traits that can qualify a good life. The core of wisdom lies in prudence (*phronesis*), which in turn flows into the ability to judge: «[the] rightness with regard to the expedient - rightness in respect of both the end, the manner, and the time» (NE VI, 1142b 25). On this basis I will recall some key ideas of contemporary virtue ethics, which highlight the limits of a purely normative or utilitarian view of ethics. In the second part I make some observations on the advantages of the practical wisdom model. On the one hand, machines can help us improve practical judgement. In particular, automatic deep learning helps in analyzing the deliberative process with regard to inferential skills, contextual framing, and the correct connection between a rule and a specific case. On the other hand, the search for a balance between rational and sentimental aspects in order to construct a biographical and social identity shows the irreplaceability of the personal factor. The ethical problem, then, is not to train machines better and better or to set ever more specific standards of programming and production, because this is a process that is beyond control and prediction anyway. Rather, the problem is how to promote the development of individual wisdom so as to ensure a "prudential" relationship with technology, precisely to the extent that machines become increasingly effective and pervasive.

Joana Romeiro – Helga Martins, Universidade Católica Portuguesa
Unlocking the Soul: AI and Neuroscience Insights into Spirituality

Spirituality is wrapped with terms associated with meaning in life, connection and transcendence experience. Over the past two decades, there has been a concerted effort within healthcare to include the spiritual dimension in clinical practice to foster a holistic approach. Currently, there are studies that confirm that spirituality plays a relevant role in health outcomes, particularly in the quality of life and resilience, and is a coping strategy in managing the health/illness process. Despite notable advancements in healthcare, exploring spirituality as a fundamental human dimension still requires further investigation and understanding in some particular dimensions. Spirituality is deep subjective and personal, and these two characteristics underline the challenge in research and in merging AI specificities. Artificial Intelligence (AI) and Neuroscience bring a huge opportunity to unlock the understanding of spirituality in patients in the healthcare setting. These new technologies open up a new set of opportunities to give a more concrete understanding of spirituality. However, AI and Neuroscience brings us challenges in particular regarding misrepresentation in the decoding of spirituality since it is a subjective and personal piece of the individuals. Furthermore, ethical dilemmas emerge when addressing matters of spiritual belief, which is a private matter of the individuals.

ATEM
INSTITUT PONTIFICAL JEAN-PAUL II

Alessandro Pichiarelli, ATISM
Famille et IA : quels chemins d'éducation à la vie bonne ?

CONTEMPORARY HUMANISM
LUMSA Sala del Consiglio / Aula Pia

Giammarco Basile, LUMSA-PUC
Flaminio Piccoli, the DC and Centrist Democrat International (CDI)

Over the years, historiography has focused on studying and analysing the history of contemporary Italy and, in fact, much research into this subject has been conducted (analysing many fields like political parties, society development and its political bond, "labour unions", political terrorism, economic development, etc.). Historiography made – among the different fields of research – important studies and analyses into many political personalities who played a key role in the social, political and economic development of the country. Due to this work, it has been possible to outline many profiles of the protagonists of Italian Republican history, but many of these figures would necessitate a deeper study, aiming at the comprehension of their political impact, trying to analyse the evolution of Republican Italy from a different perspective. Because of this reason, working with many available sources, this research aims at a new and wider reading of one of the political personalities of the Christian Democrats (DC) and the history of our country, Flaminio Piccoli, from his political education until the conclusion of the DC party in 1994. During his decades-long political career, Piccoli held various roles both within the party and at the institutional level, about which historiography, despite some isolated attempts, has not yet undertaken extensive studies. To set this goal, this research will be conducted along a double track – both internal and international. The Trentino politician in fact, despite his attention to the country, never lost sight of foreign policy. This appears even more noticeable if we consider how, during the last years of his career, he had also been both President of the Foreign Affairs Committee in the Chamber (1987-1992) and President of the Centrist International Democrats (CDI) (1986-1989). In order to conduct the research, many sources will be examined. It will be considered: archival sources (internal and foreign); bibliography and party journalistic production. In conclusion, this research aims to present to the scientific community a political biography, a new contribution to the political history of Republican Italy and one of its protagonists.

Francesca Fioretti, LUMSA- UCP
Promoting the development of competences for active citizenship in Italy: from school organization to classroom practices

The school as a democratic learning environment plays a primary role in the implementation of multidisciplinary, participatory, and integrated civic and citizenship education (CCE), which embraces the holistic formation of the citizen by unifying knowledge and experiences to promote civic learning and the development of competences for active citizenship. To this end, the Whole-School Approach (WSA) has been identified as the organizational orientation best suited to fulfill these purposes through the integration of democratic values in its three constituent aspects: the teaching-learning process, the cooperation with the community, and school governance. The latter finds expression in democratic leadership, which can foster the creation of an open school climate, thus acting in an indirect or mediated manner on student learning outcomes. Theoretical references are intertwined with current national legislation that represents, at the same time, an opportunity and a boundary for the development of the CCE. Specifically, in Italy, Law No. 92/2019 introduced the cross-curricular teaching of civic education in the first cycle of education, replacing Cittadinanza e Costituzione, which has shaped the teaching and regulatory scenario since 2008 (Law Decree No. 137/2008). Although the development of social and civic competences – referring to the Council Recommendation on Key Competences for Lifelong Learning (2006) – has been highlighted by the Indicazioni nazionali per il curricolo della scuola dell'infanzia e del primo ciclo d'istruzione (2012) as an objective of the entire school, the Italian normative documents never directly refer to the WSA. To analyze the organizational and teaching practices that can implement effective CCE pathways and to define the profiles of school principals based on their leadership styles, an embedded-multiple case study with exploratory purposes and a mixed-method approach was conducted in four lower secondary schools in Rome. The investigation presented here is included in broader qualitative research conducted in Italy and Portugal, in which four lower secondary schools in Porto also participated. The results of Italian schools will be presented, reporting the data of statistical and thematic analyses specifically on activities on CCE, decision-making processes, teacher collaboration, and critical issues encountered by schools. The school profiles will be discussed considering current Italian regulations, and emerging profiles of school principals will be outlined.

Francesco Marcelli, LUMSA
Youth association and the training of the governing class: the case of Catholic university students in Italy and internationally

My research project starts from a general analysis of the post-World War II period in Italy and internationally and aims to focus on the chronological period which goes from 1945 to 1958. The first chapter will be dedicated to the general historical context of that period. This is an historical period characterized by objective material difficulties, but at the same time also by new opportunities that generated an irreversible change in Italy and Europe. In fact, in this period, Italy experienced unprecedented economic, industrial and demographic growth. After the reconstruction of a country affected by war destruction, hunger and poverty, Italy managed to become an international power. The main aim of my doctoral research concerns a profound analysis of the history of Catholic student associations and the professional training of the governing class, in the university context, during the 40s and 50s in Italy. My attention will be focused on the FUCI (Federazione Universitaria Cattolica Italiana) and the association of

Catholic university students in Italy, in particular. Starting from an analysis of the periodical newspaper of the Federation, there is a great deal of material to help us understand the spirit of the age but also the important training role developed by the FUCI for young students. In fact, many members of the FUCI, also as a consequence of the education and the professional training received, subsequently occupied working roles of primary importance in Italy. A further aim of my doctoral thesis is to focus on the history of Pax Romana, an important international association of students founded in Fribourg in 1921. The purpose of Pax Romana was to promote world peace and create unity among Catholic university students around the world. I believe that analysing the history of this important international association could be useful for better understanding the urgent need for the Church, in the second half of the 20th century, to carry out an increasingly universal mission with the aim of creating cohesion among young people from all over the world. A cohesion generated by a common interest for culture and for the achievement of an intellectually-aware faith. To conclude, I believe that studying the history of university youth associations can be very useful to understand the importance of the training that was given to university students in those years. In fact, it's not a coincidence that many Catholic university students took on key roles within society after the Second world war.

Matteo Mostarda, LUMSA
Integral Human Development in Enrico Mattei's strategy for Italy

In the Encyclical Populorum Progressio (1967), Pope Paolo VI elaborates the concept of Integral Human Development, asserting that "to be authentic development, it must be integral, which means aimed at the promotion of every man and the whole man". The thought of Paolo VI is inspired by Jacques Maritain's work who in his "Integral Humanism" seeks the complete realization of the human being, based on the Social Doctrine of the Catholic Church which places the man at the center of the social, economic and political order. The Social Doctrine of the Catholic Church probably inspires the work of Enrico Mattei on command at ENI (formerly Agip): in an attempt to break the monopoly of the major oil companies of the time (the so called "seven sisters"), Mattei contrasts the policies of exploitation using a strategy which envisaged a different distribution of the profits deriving from the use of the oil deposits. This redistribution should have guaranteed greater benefits to local communities, promoting not only economic but also socio-cultural and political development, effectively placing Mattei as a precursor of Integral Human Development. The attention that Mattei reserves for the development processes of the populations of the countries holding the oil deposits is most likely part of his strategy to acquire a competitive advantage in the oil market to the detriment of the "seven sisters". If the ideological element, the desire to pursue "integral human development", is perhaps marginal in his action, the concrete result is undoubtedly in line with Catholic Social Doctrine. Although guided more by pragmatism than by faith in his managerial and political activity, Mattei is Catholic and close to Vatican circles, despite his own pragmatism causing him some criticism in this context, as in the case of support for the so-called "opening to the left". In particular, the president of ENI maintains close relations with the, at the time Cardinal Montini (later Pope Paul VI), who will succeed him at the command of the Committee for the New Churches of Milan after his death in 1962. Mattei and Paul VI, who studies Maritain's "Integral Humanism"

with interest, share a particular sensitivity towards the concept of equity, that should be respected, in their opinion, in relationships between men as well as between states. Their strategic and ideological battles, one against colonial legacies, the other against totalitarian regimes, highlight the attention that both demonstrate towards developing countries. They are ahead of their time and have relevance in the most recent global developments. In fact, considering the new Italian energy strategy in the wider Mediterranean, as well as the contribution to the corpus of Catholic Social Doctrine made by Pope Francis with the social encyclicals "Laudato si" and "Fratelli Tutti" the correlation between the concept of Integral Human Development and Mattei's work at the command of ENI could prove to be highly topical and interesting.

Marco Valerio, LUMSA-UCP
Learning to teach civic and citizenship education and education for sustainable development during pre-service teacher training

This research will explore the implementation of pre-service teacher training for pre-primary and primary school teachers, focusing specifically on the themes of Civic and Citizenship Education (CCE) and Education for Sustainable Development (ESD). The study adopts a multiple-case study approach, analysing the teacher education programs at selected universities in Italy and Portugal. The study aims to achieve two main objectives: first, to describe how CCE and ESD are integrated into the curricula of pre-service teacher education programs, including an in-depth examination of course content and teaching materials related to these themes; and second, to assess the perceived preparedness and self-efficacy of pre-service teachers in teaching CCE and ESD, as well as to describe their perspectives on the overall experience of the training they received related to CCE and ESD at university.

A mixed-method methodology will be employed to achieve these objectives, combining quantitative and qualitative research tools. Quantitative data will be collected through a questionnaire administered to pre-service teachers in their final year of training. This questionnaire aims to detect their perceived preparedness and self-efficacy in teaching CCE and ESD, the perceived importance of CCE and ESD. It will also gather information about the CCE and ESD training courses received at university, as well as the topics related to CCE and ESD covered in those courses. Qualitative data will be gathered through focus groups with final-year pre-service teachers to gain deeper insights into their experiences of the university training program, interviews with instructors who teach courses on CCE and ESD to obtain a thorough understanding of the topics and methodologies implemented in those courses, and interviews with coordinators of the teacher education programs to collect information on the overall design and implementation of CCE and ESD within the teacher education programs. Document analysis will also be conducted to review the content and materials used in the relevant courses. Through the triangulation of different data sources, both quantitative and qualitative, it will be possible to shed light on how teacher training programs implement CCE and ESD in the selected contexts. This approach will also provide a deeper understanding of pre-service teachers' overall experience and perceived preparedness and self-efficacy in CCE and ESD in those selected universities.

Costanza Vizzani, LUMSA-PUC
The theoretical foundations of the debate on reproductive technologies

The analysis of the theory of difference and its critique is fundamental to interprete the state of the debate regarding technology in general and reproductive technology in particular. On the one hand, through the theory of difference, a road to women's emancipation is proposed starting with a rethinking of the genealogy of motherhood. It follows that distorting motherhood in an artificial manner becomes a way of impeding the path of emancipation. If motherhood takes on forms detached from the biological-natural, the possibility of raising it to the symbolic is lost. Luce Irigary, for instance, in speaking of alterity posits the point of view of the woman who brings the other within herself. If, on the other hand, pregnancy occurs outside the body, as it does through surrogacy and as it would through ectogenesis, then such a perspective cannot be relevant. Furthermore, with regard to the intervention of technology in the contemporary world, Irigaray is sceptical in general about technical and artificial knowledge for a specific reason: the technological distances the human from the feminine, introducing a barrier that is difficult to bridge. Technical production, which is mostly male-dominated, creates a pole that stands between man and woman, making it even more difficult for sexual difference to emerge. On the contrary, feminism, which moves from a critique of binarism, considers motherhood to be the cause of female exploitation. This does not concern pregnancy as a biological-natural fact, but the way it is received by society. The critique of binarism considers women as historically situated and not as an a-historical conceptual category. In a binary society, in fact, one of the causes of women's exploitation concerns precisely the reproductive sphere, as Monique Wittig, for example, points out. Women are relegated to the private sphere, to the care of the home and children, and for this to happen they must be excluded from the public sphere. Motherhood, however, is not criticised in itself. Therefore, if it could be interpreted in an innovative and emancipatory manner with respect to patriarchal dictates, it would be acceptable again. Reproductive technologies can therefore help to emancipate women through a radical rethinking of motherhood and parental roles. Surrogacy and ectogenesis can thus be interpreted as means of empowerment.

4 September 2024

9 am PLENARY SESSION
LUMSA Aula Giubileo

Keynote Lecture: Patricia Churchland, University of California, San Diego
Neurosciences and Human Freedom

10.30 am PARALLEL SESSIONS

NHNAI Network
LUMSA Aula Giubileo

Alessia Farano, Università LUISS
Are we free to obey? Cognitive sciences and obedience in law

Recent psychological and neuroscientific findings about human decision-making have challenged the idea of free choice in law. While increasing information about the human brain has also affected the conceptual boundaries and theoretical underpinnings of legal responsibility, less attention has been paid to the implications of a new understanding of rationality for obedience. Indeed, traditional accounts of obedience are conceptually linked to conscious choice. Following a rule means anticipating the consequences of our actions - i.e., the sanction - in order to make the "choice in the direction of obedience" (Hart [1968] 2009). Contemporary moral and legal philosophy has shed light on the conceptual link between free choice and rationality: we find this idea in Antony Duff's philosophy of criminal law or in Neil Levy's moral philosophy of free will, among others. The paper will discuss the implications of studies on bounded rationality (Kahneman, 2011) and on the role of emotions in decision-making (Damasio, 1994) for the idea of obedience. Two phenomena will be considered and tested as touchstones. The first example is the nudge theory (Thaler, Sunstein, 2008), according to which regulators as "choice architectures" use behavioral economics to shape individual behavior. The core of nudge theory - based on Tversky and Kahneman's understanding of human rationality - is the idea that it is possible to exploit people's susceptibility to heuristics and biases in order to help them "make the choices they would have made if they (...) possessed complete information, unlimited cognitive ability, and complete self-control". Psychology and cognitive neuroscience are seen as tools for predicting and controlling individual behavior, but individuals are apparently incapable of foreseeing the consequences of their actions and evaluating different courses of action. The so-called "manipulation argument" will be discussed: are we still free to follow or not follow a rule as the lawmaker exploits our weakness steering our behavior in the "direction of obedience"? The second is the so-called "hyper-nudge" (Yeung, 2017). The growing knowledge of human decision-making finds a powerful touchstone in "algorithmic governmentality" (Rouveroy, Berns, 2013), or "algocracy": a form of governance in which algorithms constrain, incentivize, nudge, manipulate, or encourage certain types of human behavior. In this case, neuropsychological knowledge of the effects of

algorithms on human choice allows for an individualized choice architecture that learns and adapts to the individual user through data (that's what the prefix "hyper" stands for). The result is simple "suggestions" intended to prompt the user to make decisions preferred by the choice architect. This phenomenon represents a shift from the modern penal logic (Beccaria) - where individuals were able to weigh the benefits of breaking the law against the evils of the sanction with a free and deliberate choice - to an "intelligence logic," where individuals are seemingly unable to predict the legal consequences of their actions because their behavior is shaped by the models presented by different aggregations of data. The paper will explore the implications of these techniques for the conceptual relationship between obedience and free choice, addressing the following questions a) What does free choice mean when we talk about obedience? To what extent does the law (if any) require that an action be free in order to be considered obedient? b) If (regulative) rules are, broadly speaking, reasons for action (Raz 1979; 2011), where legal consequences are foreseen and considered as relevant inputs to our decision making, could nudges be considered as reasons?

Almási Zsolt, Pázmány Péter Catholic University
Human Agency Reloaded in our Technosocial Ecosystem

The inquiry into human freedom and agency has long been a fundamental project since the beginning of philosophical thought. Plato's early dialogues, for instance, addressed the notion of individuals bereft of insight, and thus, freedom and agency. In these dialogues, Socrates adeptly demonstrated how his intellectual adversaries were ensnared within the confines of their own confused set of assumptions, thereby unwittingly forfeiting their autonomy. Later dialogues, however, presented a shift in focus, as Socrates, with the aid for example of the Cave simile, portrayed humanity as tied down so that they could perceive mere shadows of reality's puppets, and could be released with force from this sad state. In Francis Bacon's philosophy it was the four idola that obstructed clear thought and perception. Erasmus of Rotterdam on theologico-philosophical grounds contended with Luther over the principle of freedom of choice. Since then, scholars and thinkers have endeavoured to carve out conceptual space for human freedom and agency contending with various constraining factors, whether they manifest as a mean scientist, a foul demon, chemical determinism, the unconscious, social determinism, theological predetermination, the rule of algorithms. These challenges prompt ongoing intellectual exploration and debate aimed at safeguarding and enhancing human autonomy within the web of existence. From this historical perspective, three problems emerge for consideration. Firstly, it becomes evident that the notion of absolute freedom is challenging, if not unattainable. Consequently, the inquiry explores the extent of human freedom. Secondly, in correlation with the first, there arises the question of how to conceptualise human beings. Serving as a common thread linking these issues and challenges is the third concern: how to conceptualise human agency. These three problems may be scrutinised in the context of the challenge presented by posthumanist philosophy. Posthumanist philosophy can be viewed as a collection of rational assertions regarding humanity within the technosocial ecosystem of the 21st century. However, in my interpretation, posthumanist thought should rather be perceived as a critique of and confrontation with what might be categorised or fictionalised as traditional humanism. In posthumanist thought it is feasible to discern three overarching and abstract

denominators as is argued in Francesca Ferrando's *Posthumanist Philosophy* such as "post-humanism," "post-anthropocentrism" and "post-dualism." Post-humanism means that instead of a single narrative, posthumanism attempts to delineate the human in its plurality, i.e. instead of conceptualising the human as a white, middleclass man, it tries to see the human in a more comprehensive and inclusive way. Problematising anthropocentric theories means that posthumanism endeavours to place all other entities in their appropriate context by displacing humanity from the centre of attention. The objective of posthumanism, then, is to perceive the human being not as an exceptional, universalizable entity, but to comprehend humans through their interactions and collaborations with other entities, interconnected and interdependent, rather than existing in isolation. With the words of Stalpaert et al what is to be problematised is the view that "the human is the centre of attention, an individual that maintains control over nonhuman matter in a competitive and hierarchical constellation." The posthumanist perspective challenges, furthermore, the idea of the hierarchy of binary oppositions when exploring the human. As a powerful example one may well cite Hayles when she argues that "there are no essential differences or absolute demarcations between bodily existence and computer simulation, cybernetic mechanism and biological organism, robot teleology and human goals." When examining human interaction with this technology, avenues emerge for uncovering room for human agency and responsibility. This occurs as individuals, in utilising this technology to produce, for example, textual documents, initiate requests and prompt responses, subsequently responding to the generated output by either accepting or rejecting it. Upon acceptance, individuals may or should amend the text to articulate their thoughts in their own ideal voice, and ultimately, they should determine the course of action regarding the edited text. Each of these steps, involving decision-making processes, facilitates the exercise of human freedom and agency. However, this is limited by the application's indeterministic text production, and is contingent upon the individual possessing a functional understanding of the application, of how to engage in communicative exchanges with it, and of what they ultimately do with the outcome of this interaction. Such understanding can only be attained through education, or, to echo a Platonic metaphor, liberation from the chains of the cave where shadows of shadows prevail, and vague opinions and beliefs masquerade as knowledge.

Wen-Hsiang Chen, Fu Jen Catholic University
Artificial Intelligence, Consciousness Emergence, and the meaning as a whole

The impact of artificial intelligence (AI) on humanity has been as profound as witnessing a beanstalk sprout into a towering giant overnight. Yet, AI development has been ongoing for decades, becoming a nearly ubiquitous scientific application driven by the internet, big data, algorithms, and supercomputers. However, the question now extends beyond AI's potential reach; it is how AI can prompt humans to introspect their cognitive structures, leading to a deeper understanding of ourselves, consciousness, and modes of thought. Therefore, the core of inquiry should not be limited to AI's scientific applications but should delve into its principles or essence: can human holism be simulated by AI algorithms? This question not only encompasses philosophical ontology and neuroscientific issues but also raises fundamental ethical questions about free will and determinism that humans must re-examine. This essay argues that while technological, legal, and ethical applications are crucial, a universal

philosophical perspective, which encompasses cognitive science and neuroscience, is equally essential. It discusses this epistemological and potentially ontological issue from several perspectives. Firstly, examining AI from a mechanistic perspective can be a starting point. Moreover, delving into consciousness, the most fundamental human trait, from a neuroscientific perspective, leads to the question of whether AI possesses similar potential. Finally, from a philosophical holistic perspective, it discusses the significance and implications of AI for humanity. By drawing insights from Jesuit theologian Bernard Lonergan (1904-1984) on the structure of cognition and the whole of things and also the epistemological approach of Zhu Xi (1126-1271), an influential Neo-Confucian philosopher in medieval Chinese philosophy, this essay might be able to facilitate a meaningful dialogue. The essay concludes that in a holistic and emergent universe, an AI computer based on mechanistic determinism will ultimately be more incomprehensible than the universe itself.

Sara Fernandes, Universidade Católica Portuguesa
Free Will, neurosciences and robotics. Anthropological and ethical reflections

What does free will mean? Do we have Freedom? In recent decades, there has been a heated debate among philosophers and neuroscientists about what we should understand by freedom, and by extension, whether humans are free or not. It is enough to recall the early studies of Libet and Haggard, and subsequently, the research of John- Dylan Haynes and his team, who argued that humans are determined and do not have free will, empirically demonstrating that the brain has unconscious mechanisms that cause choices before individuals become aware of them. For example, which hand to raise or which calculations to perform (addition or subtraction). These authors suggest the possibility that our intentions and choices are programmed and defined prior to any self-awareness. Here, we see that the first conception of freedom underlying the neuroscientific discussion is libertarianism, in the sense that being free is being liberated from any external or internal constraints or conditions. Nothing beyond the conscious motivational states of the agent can influence their choices and actions without diminishing their capacity for action. Only this condition makes someone the author of their acts, therefore being in total control of their actions, or in other words, not suffering interference from other factors than conscious causal mechanisms. From our perspective, it is not straightforward that the existence of unconscious brain processes, temporally preceding conscious mental states, threatens control over our behavior, and implies the conclusion that we are not free or responsible agents. We can interpret the temporal precedence of unconscious mental processes over intention consciousness as the preparation of the human organism to form intentions and make decisions. In fact, it would not be expected otherwise from the brain. But we cannot infer that we are absolutely determined just because a subcortical framework of activation is discovered prior, temporally, to consciousness and decision-making. Although conscious mental processes are preceded by unconscious neural mechanisms, it cannot be inferred that the former are dependent on or caused by the latter; it can only be stated that two phenomena of different nature occur at different times in the brain, but not a causal nexus. Libet's and Haggard's argument seems to be a typical example of the post hoc fallacy, a reasoning that invalidly infers a causal relationship between A and B, specifically, the conclusion that A is the cause of B, exclusively because there is a temporal precedence

relationship between event A and event B. An essential component of human action and the exercise of autonomy, and one to be considered by neuroscience and in the development of its tests, is intentionality. It is crucial because it shows a purpose to our life journey, "our itinerary," and good lives require itineraries, as Socrates emphasized, giving them meaning. When we act, there is "that sense of where we are going," the purpose, the intent with which we do something. This is the true exercise of free will; only intentionality expresses the meaning of (self)determination. Here we follow the perspective of John Searle, according to whom intentionality exhibits three fundamental characteristics: 1) it coincides with a certain mental content of the agent, 2) it allows the identification of the conditions for the action's satisfaction (or failure), and in cases where the action is successful, 3) it can be understood as a form of causation of voluntary behavior or, in other words, as its explanation.

This conception of free will underpins some reflections that we will make about the ontological and ethical status of robotics. When we discuss the possible autonomy of a robot, its ability to make decisions or advise a human, we are referring to this capacity (so far human) to have genuine intentionality (not a simulation) and to project actions/decisions based on principles/values that express the agent's identity. The robot does not express this degree of complexity and consciousness, having been created as a tool for the benefit of humanity, assisting it in what it cannot do due to its biological limitations or no longer wishes to do, freeing humans for more creative tasks in line with their free projects. This relationship with AI and robotics is particularly evident in medicine, with promising results in screening, surgery, revolutionary treatments, and optimizing the therapeutic relationship, according to Isaac Asimov's three fundamental principles, as well as following Beauchamp's core principles of biomedical ethics.

Juan Vidal, Université Catholique de Lyon
The brain's mind, timely decisions and free will

What has neurosciences brought to the table on the understanding of human free will? Since Libet's now classical experiment on the link between the brain's Readyness Potential (RP) and perceived conscious volition, much debating has occurred, neuroscientific and philosophical, around the possible nature of human's existing free will, or not. Our brain and mind are ongoingly influenced by the bodily interactions with the environment and our nervous system deeply maintains living systems in proactive behaviors, thereby conditioning future actions and framing behavioral intentions and decisions. If the human mind is the result of complex neuronal interactions in our brain, that it is continuously (self-)influenced

by memories and perceptions (conscious and unconscious), and a history of motivations, far from any blank slate moment, is free volition to be a scientific illusion? In this presentation I will do a « tour de table » of the latest evidence and arguments of neuroscientific investigations on free will. This includes various critical arguments on Libet's study and the nature of the RP itself in the light of neural decision processes, but also the notion of temporal scales of consciousness and behavioral control.

ATEM
LUMSA Aula Teatro

Dominique Lambert, Université de Namur
Les régulations éthiques des pratiques : quelles régulations éthiques de l'IA ?

CONTEMPORARY HUMANISM
LUMSA Sala del Consiglio / Aula Pia

Sarah Horton, ICP-ACU
Alienation and Self-Knowledge in Maine de Biran

Self-knowledge has long been held up as an ideal; consider the injunction, carved on the temple of Apollo at Delphi, to "Know thyself," which Maine de Biran (1766-1824) himself cites favorably. At the same time, however, Biran argues that one discovers one's own existence, as well as that of the world, only through the feeling of effort, and effort, he shows, is always opposed to a resistance that is opaque to knowledge. Is a certain impossibility of knowing oneself therefore also essential to the constitution of the human being? Indeed, my thesis (drawing on the work of Emmanuel Falque) is that an attentive reading of Maine de Biran will teach us that alienation, although it seems to be the most improper of states, is in fact proper to humans: it is not that we should seek out just any experience, or rather non-experience, of alienation, but that the limits to self-knowledge and even to experience are indeed constitutive of human being. There is neither awareness of oneself nor self-knowledge without the strangeness that resists them, for humans are constituted by a non-assimilable exteriority, and one can know oneself, to the degree that such a thing is possible—and it is indeed an important task—only on the basis of this resistance, even this alienation. The moment that I begin to know myself, I discover myself as other than myself, and there is no way to escape this alienation, save by leaving behind my own humanity and therefore falling into an absolute alienation. Biran emphasizes that the human being has a "mixed nature," both moral (a term that in the 18 th and 19 th centuries referred to the faculties of the mind or soul, in contrast to the body) and physical, and that the influence of the physical on the moral, though not always good, is fundamental to the human condition. The one who knows himself or herself is the one who admits the impossibility of knowing oneself—but it is not a matter of a pure impossibility that would consign us to mere animality; on the contrary, this impossibility founds all possible knowledge and reminds us, at the same time, that we must not expect certainty. Biran gives us to see the necessity of a humility that recognizes, and that is grateful for, the limits that constitute our existence.

Juhani Steinmann, ICP-LUMSA
The Coming God. Soteriological Figures in Kierkegaard, Nietzsche, and Heidegger

The following contribution takes Martin Heidegger's well-known statement from his 1966 interview with Der Spiegel as a starting point, namely that in the face of technological

developments, the end of philosophy (as metaphysics), and the general impossibility of changing the state of the world, "only a God can save us". In this context, it wants to examine three positions critical of religion and consider their plausibility. These three positions stem firstly from the thinking of Søren Kierkegaard, secondly from that of Friedrich Nietzsche, and thirdly from that of Heidegger himself. All three positions are characterised on the one hand by the fact that they do not limit themselves to an approach critical of religion, usually diagnosed as a crisis in Christianity, but tie their critique to an alternative soteriology, whereby, interestingly, their critical moments themselves turn back into religious promises.

On the other hand, these three positions are characterised by an argumentative timelessness, since their respective critique of religion as well as their soteric alternative-programmes are typological in nature, regardless of the respective historical context, i.e. they proceed from basic figures and patterns that claim validity in any time. Finally, the three positions are characterised by an exclusivity or exquisiteness that condenses into the almost normative claim that "only" and "exclusively"the alternative-soteriological and neo-religious typology of a coming God pushed by each of the three (id sunt Christ in Kierkegaard, Dionysus in Nietzsche and The last God in Heidegger) shall herald the turning point within the diagnosed crisis. The fundamental question that remains is: Is there also a combinatorial dimension to this ought, more concretely, a soteric denominator for all three authors?

Federico Rudari, UCP-LUMSA
Embodied perception and spatial sense-making: from phenomenology to aesthetics

My contribution to the conference, coinciding with the first chapter of my doctoral thesis, focuses on the role of embodiment in perception and cognition, at large and in relation to space, both in functional and aesthetic terms and contemplated objects. Starting from the idea that the body is the first of all cultural objects (Merleau-Ponty 2012), the analysis discusses the evolution of the research interest around the body as a central tool in the production of meaning. In particular, experience is here intended as embodied by nature, since, as Brandt writes, "languages, cultures, and human semiotics in general are based on experiences and practices in a life-world constituted as a whole" (2004, 34). Following Gibson's ecological approach to psychology (1979), different scholars started associating cognition with the dynamic coupling brain-body-environment, until the definition of one of the most used (and discussed) characterisations of cognition, namely 4E cognition, which defines the cognitive process as embodied, embedded, enacted and extended. For what concerns the trajectory of my research, the way sociocultural and collective tools for meaning-making are affected by moods, contexts, and other factors related to the situatedness and momentality of each experience (which is of primary importance in shaping our relation to perceptual objects) is put in relation to the nature of the creative expression that art entails, not understood as an embellishment but in its social performativity, as the relational result of bodily acts of creation and subsequentially perception. My presentation will show a dialogue between phenomenological theory (Husserl, Merleau-Ponty and Sonesson, among others) and aesthetics, keeping in mind the work of early modern and pre- Kantian theorists, who reference pre-subjective ways of bodily perception, often considering the body as an open organism in constant exchange with its surroundings, rather than a self-contained and self-identical expression of the mind (Chromik 2014). This approach

oriented to physical, multisensorial perception and feelings affirms a "pre-reflecting, non-discourse mode of knowing, symbolizing, and being-in-the-world" (Dengerink Chaplin 2005). According to this approach, aesthetic objects are particular to the human sphere both as a form of practice (in their creative dimension) and experience (in their perceptive dimension). Artistic objects are deeply tied to the physicality of their perception not only for the sense of belonging and community that the engagement in this ritual creates but also for the emergence of an instinctive reaction caused by human feelings and consciousness. Relating to the theme of the conference, I will try to address the notions of bodily knowledge and meaning in relation to the bodiless nature of AI.

Tomaso Pignocchi, LUMSA-ICP
Language and soteriology: the concept of liberation in Wittgenstein and Buddhist philosophies

In the last century we have witnessed a renewed interest in the practical role of philosophy, understood as a tool to guide one's existence and improve individual life. This theme, that was already central to many ancient philosophies, finds particular resonance in the thought of Ludwig Wittgenstein, which seems to have deep affinities with Indian doctrines of liberation, particularly Buddhism. This common conception of a philosophical work of "liberation" proposes to dissolve the chains of a distorted and illusory way of seeing. Its goal, in fact, is not to "stop thinking," but to "think differently," free from metaphysical and foundationalist presuppositions. Actually, both Wittgenstein and Buddhism identify the root of suffering in the confusion we make in attributing certain roles to 'ideal' conceptions and mental formations, confusion that develops through an illusion that is intrinsic to the functioning of language. Indeed, both see language as the source of this illusion and propose a path of liberation that passes through the dissolution of this deception. Liberation, for Wittgenstein and for Buddhism, does not consist in an ascent towards a transcendent elsewhere, but in a return to the ordinary world, seen with new eyes and free from conceptual distortions. The meaning of the world is not something separate from the world itself, but is immanent to it. This type of liberation takes the form of a conversion of the way one sees things. However, it does not rely on an appeal to explanations, reasons, or rationality, but rather it appeals to the overcoming of a blockage by the will. This concept of a complete reorientation of one's own perspective is central to Wittgenstein's later philosophy and may even be considered the ultimate goal of his mature work. My research aims to deepen the convergences between Wittgenstein and Buddhism, showing how both propose a path of liberation that is configured as a return to the world, a return to thought, and not as an escape from it. Furthermore, the comparison between Wittgenstein and Buddhism allows us to establish a fruitful dialogue between East and West – and between past and present – offering a new perspective on the theme of the practical conception of philosophy.

Orlando Garcia, ICP-LUMSA
Human freedom challenged by AI and neuroscience

In his writings, Husserl repeatedly implies that 'normality' is a significant factor in the constitution of an intersubjective world in various levels of experience through the concordance, optimality, familiarity, and typicality of perception as much as the lifeworld's historical and socio-cultural dimension. In a way, Husserl confronts the normality of the subject that constitutes the intersubjective world to the contingent, ever-changing, and multistratified conditions of the empirical diversity of subjects that experienced the same world: animals, madmen and child. Building on this confrontation, this research aims to answer these four main questions: What is the role of normality in the overall architecture of Husserlian phenomenology? How does Husserl's late phenomenology articulate the empirical ego and its contingency with the transcendental ego and its absoluteness? And finally, can the plurality of abnormal subjectivities be reintegrated into a transcendental understanding of the lifeworld within the strict parameters of Husserlian phenomenology? Can we give a proper phenomenological explanation of an identical lifeworld experienced by dissymmetrical empirical subjectivities such as the animal, the child, the madman, without losing the harmony of an objective world of experience? Starting from this idea of a 'stratified' notion of normality, namely, the possibility of adopting the concept of normality about different levels or 'layers' of experience, I argue that we can understand normality as an operative index of levels of constitution and therefore as an operative concept in Fink sense, this will allow us to build normality as the place in Husserlian phenomenology to inquire about the relation of the empirical and the transcendental in the constitution of an intersubjective world. In Husser's words: How does one gain the sense of that normality which belongs to the constituted world itself, the objective world for all, and which in turn is presupposed as the structure of the self-considering about the world? The levels of normality and abnormality correspond to the levels of the constitution of being going from relative being in relative phenomena up to the objectively true being of the truly existing world. These findings have the potential to reshape our understanding of normality and abnormality in the context of Husserlian phenomenology, and their implications for the constitution of an intersubjective world. There can be a crack on every level of the constitution of the world itself. For instance, we can thematize the human type as a being of reason on an axiological level and bring forward the notion of a normal human being as a free subject capable of self-determination, from which there can be abnormalities such as subjects that experience the world but aren't capable of taking an evaluating stand on it. In Ideas II, Husserl describes persons as the subjects of an Umwelt, meaning that persons are constantly taking a stance and evaluating their surrounding world, not just passively perceiving it, unlike animals that only have an organic development toward their Umwelt. In this sense, on Kaizo lectures, Husserl argues that maturity is the normal typical figure of human development and that, unlike animals, man has, as a being of reason, the possibility "and the free faculty of this development of a completely different kind, in the form of free self-direction and free self-education in the direction of a final idea that is absolute as itself." Considering the actual development of human technology, could it be that artificial intelligence disrupts this axiological normality? This normality assumes an autonomous individual who evaluates his or her ends. Still, at the age of artificial intelligence, we are not on the threshold of the end of the individual autonomy, nor are we, in short, approaching the type of animal or child that Husserl judged incapable of carrying out a self-evaluation through judgment and practical determination. Giving up or leaving the exercise of choice to artificial

intelligence to improve our living conditions does not call into question the autonomy of a normal subject that constitutes a practical world.

4 September 2024

2.30 pm PLENARY SESSION
LUMSA Aula Giubileo

Keynote Lecture: Fiorella Battaglia, Università del Salento
Democracy and Education at the Time of AI and Neurosciences

4pm PARALLEL SESSIONS

NHNAI Network
LUMSA Aula Giubileo

Yves Poullet, Université de Namur
EU AI Act- A NHNAI Lawyer's point of view

European regulation of artificial intelligence has now been enacted. This regulation is the culmination of the European strategy on the development of our information society and is justified by the desire for a sovereign Europe both in the protection of its values and its economic interests. Through the text, the Union desires to chart a "Third Way" forward in terms of the development of our digital society which must be "human centred," distinct from that of the United States, and China, and based in particular on the respect for human rights and the pre-eminence of the human, following the famous formula: *"Human in the loop, Human on the loop and Human in command."* This apparent strong humanist approach is in line with the NHNAI' projects aim. Is that the reality? From one part, our paper would like to analyse how the European authorities envisage first this humanistic approach and second the realisation of this ambitious objective. From the other part, we would like to underline certain loopholes in the text to achieve it. Values are considered and how EU faces the delicate problem of the distinction between ethics and law? On that point, it would be interesting to understand why the EU authorities have chosen a regulatory approach instead of a coregulatory or even selfregulatory ones. Always on that point, we would like to take again the four Unesco universal values: dignity and autonomy, social justice and do good and do not harm and to decrypt the way by which they are taken into account by the AI Act. The 'risk-based' approach at the core of the AI Act might be discussed as a second topic. This approach is based upon the social responsibility of companies and public authorities putting on the market new products which might generate certain risks to the individuals but also to the population. Which risks are envisaged by the text? They are going far beyond those affecting our individual liberties and encompass rightly social justice, environmental and democratic concerns. Perhaps risks generated to our identity as such would have also to be considered especially but not only as regards genetic manipulations and would justify the intervention of bio ethicians. The AI Act answer pointing out the need to proceed to an internal assessment is of great value but why to limit this assessment only to certain AI systems and not to open this assessment to all the stakeholders? As regards this risk approach, we will also have to discuss how the regulation is

trying to achieve a proper balance between the principle of precaution and the need to promote a competitive and innovative market. Especially, we will say a few words about the new problems related to the generative AI systems and the difficulties met by the AI text to take them into account. Other considerations must be added as regards the institutional framework surrounding the development of the AI ecosystem as put into place by the AI Act. It is quite clear that the AI technologies are questioning transversally our disciplines and traditional issues. It is obvious that LLM for instance raises fundamental questions like the multiculturality of our society, the potentiality of creativity of the humans, trends to favour a less diversity of opinions, to greatly limit our privacy, and so on. Facing at the same time these different challenges is not obvious. How the EU law is dealing with that difficulty at a moment where it multiplies the bodies (Data Protection Offices, Competition authorities, Consumers agencies, Centre for equal opportunities, Bioethics Commission, …) in charge each of them of a specific dimension? The role of the law in the debate NHNAI has initiated, launched and is nourishing is important. The effectiveness of our recommendations must be ensured by legal provisions which are adequate. Is that the case of the AI Act? That is precisely what I want to discuss with the participants to our Conference.

Maria John Peter Selvamani, Fun Jen Catholic University
Enhancing Public Engagement: Employing the World-Café Method for Societal Debates in the NHNAI Project

The NHNAI project aims to foster extensive public discourse concerning human issues linked to the rapid development and deployment of AI-related technologies. In this project, each partner institution conducts in-person debates within diverse stakeholder communities, subsequently transitioning to digital dialogues through the CartoDebat platform. Through the capacity-building workshops, this dual-dialogue approach constitutes a pivotal phase for the project's success. Initial capacity-building workshops revealed that while participants showed willingness for in-person discussions, few continued dialogue on the digital platform. Consequently, the process proved less effective than anticipated. To address this, we plan to use the World-Café method for in-person discussions for the second round of workshops, with CartoDebat as an optional digital platform. The World-Café method is renowned for fostering extensive discussions across various disciplines, making it an ideal choice due to its resemblance to the CartoDebat platform. Both methods allow participants to comment on each other's suggestions. We aim to replicate CartoDebat's functionality using the Word-Café method on paper, subsequently transcribing discussions to the digital platform for further online engagement and analysis. This presentation delves into applying the World-Café method in societal debates, exploring its strengths and weaknesses. Furthermore, we compare the NHNAI project's methodology with those of others like V Taiwan, which are used for more comprehensive public policy consultations.

John Lukwata - Emmanuel Wabanhu, The Catholic University of Eastern Africa
Navigating the AI and NS Landscape in Africa: Unlocking Opportunities Amidst Challenges

This abstract explores the burgeoning landscape of Neuroscience Systems (NS) and Artificial Intelligence (AI) in Africa, it sheds light on both the opportunities and challenges. As NS and AI technologies continue to advance, they hold vast potential to address pressing issues across various sectors in Africa, including healthcare, governance and education. However, realizing this potential requires navigating a complex landscape marked by data scarcity, and socio-economic disparities and infrastructural limitations. Moreover, the integration of NS and AI promises deeper insights but also pose new technical, legal, and ethical/spiritual dilemmas. Despite these challenges, Africa stands at a critical stage where strategic investments, collaborative efforts, and ethical frameworks can unlock the transformative power of AI and NS for sustainable development. This paper will further discuss on the key strategies that harness these technologies that fosters inclusive growth, that empowers the local/vulnerable communities, and to address the continent's unique needs, ultimately paving the way for a more equitable and prosperous future. Navigating the AI and NS landscape in Africa requires a multifaceted approach that addresses technical, infrastructural, socio-economic, and ethical considerations. The study seems to recommend that, first, the African Governments, and other stakeholders should prioritize investments in digital infrastructure, including high-speed internet connectivity, data storage facilities, and computational resources. In addition, establishing training programs, workshops, and educational initiatives to build local talent in AI and NS fields is crucial. Moreover, it is important to promote the culture of data sharing initiatives, open access repositories, and data governance frameworks that can facilitate the availability and accessibility of high-quality data for AI and NS research and applications. Moreover, there is a need to ensure data is protected, security, and ethical use should be paramount, with transparent policies and mechanisms for consent and oversight. There is a need to have a knowledge triangle approach that encourages collaboration among academia, research, industry, vulnerable persons, and government stakeholders that foster to accelerate innovation and the development of AI and NS solutions tailored to Africa's unique challenges and opportunities. Similarly, to recognize the diversity of languages, cultures, and socio-economic contexts across Africa. There is a need to develop AI and NS solutions that are linguistically diverse, culturally sensitive, and contextually relevant. This will include to incorporate on the local knowledge, community input, and user feedback into the design and implementation of technology solutions. Promoting ethical guidelines, standards, and best practices for the development and deployment of AI and NS technologies is essential. This should include to address, justice, bias, fairness, transparency, and accountability in algorithmic decision-making processes, as well as mitigating potential risks and unintended consequences of AI adoption. Governments should enact supportive policies and regulations that foster innovation while safeguarding against potential harms and abuses of AI and NS technologies. Finally, the study recommends that there is need to create frameworks for data protection, intellectual property rights, and technology transfers.

Amy Marie Lake, Università degli Studi di Milano Statale
Judgement by Algorithm: The rise of AI Adjudication in China's Legal System

According to Plato human behaviour flows from three main sources: desire, emotion, and knowledge, encapsulating the intricate nature of human actions. As society grapples with the

expansion of AI technology, the realm of judicial justice faces unprecedented challenges, echoing Oliver Wendell Holmes Jr.'s poignant observation that "The life of the law has not been logic; it has been experience." This paper embarks on a nuanced exploration of integrating artificial intelligence into judicial decision-making, particularly within the legal landscape of China, where AI adjudication has emerged as a prominent fixture. Dissecting existing regulatory frameworks, this paper shall endeavour to unravel the complex interplay between efficiency, individual rights, and algorithmic autonomy in judicial processes. Drawing upon a robust synthesis of scholarly literature, case precedents, and legal frameworks, this paper seeks to illuminate the ethical dilemmas inherent in AI assisted decision-making and their ramifications for due process and judicial equity. As the element of humanism is inherent in legal disputes, different dynamics and unpredictable factors can often defy conventional legal reasoning resulting in a judges' need to adjudicate within reason and grapple the complexities of human motivations discretionarily. But what about Artificially Intelligent adjudicators programmed with algorithmic patterns of predetermined parameters? Is it reasonable to believe that a legal system and a society can be just and fair in the hands, or better yet, the algorithm of a CPU? And how is this navigated within different jurisdictions, specifically those of higher technological advancements? This paper shall therefore delve into the multi-layered complexities of integrating artificial intelligence into judicial-making processes with specific focus within the legal framework of China and its use of AI adjudicators. By accessing these regulatory frameworks, this paper aims to pursue a richer understanding between the elucidating tensions of efficiency, individual rights and autonomy when using algorithmic systems. Drawing from a selection of relevant scholarly work, case precedents and legal frameworks the paper shall consider how AI is used in Chinese courts and the conceivable implications of due process and judicial justice. The foundational principle of judicial justice illustrates the indispensable role of judges in upholding the rule of law and safeguarding individual rights. Yet, as AI algorithms encroach upon traditional judicial functions, questions arise regarding the preservation of human judgment and discretion, especially due to the disparity of factors and processed data. As shown in the landmark case of Wisconsin Loomis (2016) where an AI algorithm was employed to determine risk assessments of offenders, and concerns were raised about the algorithm's opaque decision-making process and potential bias in predicting recidivism rates. Additionally, platforms like IBM Watson's AI Judge and Tencent's AI Magistrate have been introduced in China's legal system, raising concerns about their impact on judicial autonomy and fairness. Contrary to their human counterparts, AI adjudicators operate within predetermined parameters, devoid of the capacity for independent thought or moral reasoning essential for equitable adjudication. While proponents tout the efficiency gains offered by AI adjudications, their reliance risks homogenizing legal perspectives and stifling the consideration of nuanced case specifics. The primordial example of 'Xiao Zhi', the robot judge in Hangzhou China, has adjudicated live summarization of evidentiary evaluations, judicial argumentation, and recommendation in under 30 minutes. This erosion of judicial autonomy culminates in a mechanistic interpretation of justice, bereft of human empathy or ethical discernment, thereby undermining public trust in the legal system, especially in China now believed to become a system of "mechanical justice". Counter to this belief Zhou Qiang, the current Chief Justice and President of the Supreme People's Court of China, highlights the broad strokes AI shall paint as China endeavours to commit to a technologically driven modernization, but maintains that "AI will never replace human judges

and can only serve judges [as assistants]". Furthermore, the integration of predictive analytics reiterates concerns surrounding algorithmic bias and systemic inequalities. By extracting patterns from historical data, predictive models risk perpetuating existing biases or exacerbating disparities in legal outcomes. As illustrated in the case of United States V. Jones (2014), where a predictive analytics tool was utilized in sentencing decisions, and disparities were observed in the length of sentences imposed on individuals from different socio-economic backgrounds. Consequently, addressing these ethical quandaries necessitates a multifaceted approach encompassing transparency, accountability, and fairness in algorithmic decision-making. Likewise, it attempts to mitigate algorithmic bias and uphold individual rights which must be enforced and implemented by robust oversight mechanisms and regulatory safeguards. This paper presents a nuanced examination of AI adjudication within the regulatory framework of China, including the 'Internet courts regulations' and the 'Civil code of the People's Republic of China', while incorporating insights from seminal works by Angela Zhang, Rogier Creemers and Xiaodong Lin. Additionally, it shall underscore the pivotal role of interdisciplinary collaboration and dialogue in navigating the ethical and regulatory challenges posed by AI in judicial decision-making. By fostering intersections across academic research, legal databases, and case studies, this paper seeks to catalyse informed discourse on the ethical implications of AI integration in the pursuit of judicial justice.

Corrado Claverini, Università del Salento
The Principle of Human Autonomy between Artificial Intelligence and Emotional Manipulation

The aim of the paper is to show how some conversational AIs can undermine human autonomy, which is one of the fundamental principles of the *Ethics Guidelines for Trustworthy AI*, together with prevention of harm, fairness and explicability. According to this document, the principle of autonomy can be safeguarded if artificial intelligence systems do not unjustifiably subordinate, coerce, deceive, manipulate, condition or herd humans. However, it is conversational artificial intelligences that have often made the headlines when attempted murders and suicides have occurred as a result of prolonged use of such technologies by certain 'vulnerable' users. The undermining of human autonomy and emotional manipulation are the biggest risks of conversational AI, according to an open letter published by philosophers, legal scholars and engineers following the suicide of a young Belgian man after a long conversation with a chatbot. Even if users are aware that they are interacting with an AI, they may develop a relationship with the AI that compromises their autonomy, especially in the case of vulnerable people. Those with fewer social connections, such as lonely or depressed people, but also children who increasingly have access to these tools, are most at risk of emotional manipulation. However, it should be emphasised that everyone is vulnerable, as the emotional response to realistic interactions is inherent in human nature. This scenario has fuelled the debate about the ELIZA effect, i.e. the human tendency to anthropomorphise these artificial identities by attributing thoughts and emotions to them. This phenomenon was first observed in 1966, when Joseph Weizenbaum created ELIZA, a chatbot simulating a Rogerian psychotherapist, and had it interact with users who developed a kind of emotional bond with it. This trend is even more prevalent now that chatbots have social networking accounts and, thanks to artificial intelligence, a physical appearance that is almost indistinguishable from that

of a human. Since the days of the first rudimentary chatbots, the ability to maintain conversational verisimilitude has increased disproportionately with the development of more advanced language models. In addition, there is the unprecedented possibility of video creation and the imitation of human voices. In this context, the growing phenomenon of AI partners and virtual influencers, whose artfully crafted personas and empathic communication via Instagram and X make them almost indistinguishable from humans, is worrying. Consider the influencer Rebecca Galani, the first of her kind in Italy. Described by her creators as a "model of language", it is possible to have a chat with Rebecca on Telegram. Her answers – text or audio, depending on the settings chosen – are generated by artificial intelligence, as are her images, which can be customised on request and for a fee. The problem with such services is that not everyone is aware that they are interacting with an artificial intelligence, even though it is explicitly stated in their social network biographies. Such phenomena undermine the principle of human autonomy, as they are AI systems that create "dependencies through attention-capturing techniques or the imitation of human characteristics (appearance, voice, etc.) in ways that could cause confusion between AIS and humans" (*Montreal Declaration for a Responsible Development of Artificial Intelligence*). With a focus on the principle of autonomy and the risk of its undermining by the use of certain conversational artificial intelligences, this paper aims to highlight the fundamental role of Europe in the process of regulation of these technologies. This process was completed on 13 March 2024 with the adoption of the AI Law, which will enter into force 20 days after its publication in the Official Journal of the European Union, expected between May and June 2024. But what are the implications of this law for conversational artificial intelligence? And what will happen to controversial chatbots such as VOID Chat, which is being used to circumvent OpenAI's restrictions on the use of ChatGPT and create AI agents with whom it is possible to have conversations without any ethical filters? Finally, how will the preservation of the principle of human autonomy be affected by the adoption of the AI law? In summary, the aim of this paper is to show how conversational AI can undermine autonomy – here understood as self-government, i.e. the ability of an individual to act independently on the basis of his or her own system of values and moral reasoning – by unduly influencing users' decisions and perceptions. In other words, this paper highlights how the greatest risk in this context is that autonomy is compromised – without users being fully aware of it – by the engagement techniques used by chatbots (e.g. personalising interactions and creating echo chambers that reinforce individuals' emotional or cognitive predispositions).

Elad Magomedov, KU Leuven
Existentialism as a Humanism in the Technoscientific Era

The technoscientific era presents a landscape where traditional notions of human freedom appear increasingly constrained, if not obsolete. With advancements in neuroscience, epitomized by experiments like Libet's, the very concept of freedom is under scrutiny, seemingly fading against the backdrop of science. While determinism takes center stage in theoretical discourse, the practical realm in turn witnesses compromises to freedom with the rise of AI, notably in decision-making processes. However, this convergence of developments that attack the concept of freedom on both theoretical and practical fronts, doesn't necessarily entail a wholesale restriction on human freedom. Rather, it challenges us to reconsider the

traditional dichotomy between freedom and the efficient causality governing both neural processes and the influence of automated systems on human agency. Phenomenological philosophy sheds light on this complex relationship. By delving into the lived experience of consciousness, phenomenologists since Edmund Husserl insisted on the fallacy of treating consciousness as merely another object subject to physical laws. Particularly noteworthy is Jean-Paul Sartre's existentialist ontology of consciousness, which offers conceptual tools not only to navigate the constraints of the technoscientific era concerning freedom, but also to reconcile freedom with humanism in this context. For Sartre, freedom is not an abstract concept but is always situated within specific contexts and circumstances. He argues that human existence is fundamentally characterized by freedom, but this freedom is not detached from the world; rather, it is intimately intertwined with the situations in which individuals find themselves. Since freedom is always situated within concrete situations and social structures, these situations impose constraints and limitations on human agents, shaping the possibilities available to them and influencing their choices. By articulating freedom as situated, Sartre allows us to make sense of what it would mean to be free in a world where situatedness is defined by AI, that is to say, by automated decision-making processes and knowledge production. While Sartre would agree that such situatedness shapes the horizon of our possibilities for action, he would nevertheless insist that situatedness is not a limitation but rather the very condition of freedom: it is because we can only act within the context of AI that we are responsible for every choice we make within that context. Sartre's radical insistence on the inalienable nature of freedom and responsibility is rooted in his phenomenological ontology of consciousness. Such ontology allows him to affirm that even though the brain is determined by the laws of causality, consciousness arises precisely as an interruption of such determination. Drawing from the work of phenomenologists such as Edmund Husserl and Martin Heidegger, Sartre argues that consciousness is not a thing among things (être-en-soi), but rather the precondition of anything to appear at all. As such, consciousness is pure intentionality: it is a mere relating to something other than itself, never coinciding with itself and always surpassing itself towards the world. Since the things in the world coincide with their essence or nature, they are subject to the mechanistic laws of causality. Consciousness, on the other hand, is a nothing (néant), a pure negativity that is constantly suspended between the past and the future, therefore always falling outside the domain of efficient causality altogether. As soon as the universal laws of causality touch the brain that is now, consciousness has already surpassed itself towards what it is not-yet. For this reason, consciousness is 'condemned' to be free. It is pure spontaneity that can be motivated, conditioned, or carried away by itself, but it cannot be determined to do or be anything in the sense that a billiard ball can be determined to take a certain trajectory if impacted from a particular angle. Rather than being determined by its brain, one could say that consciousness is situated in its neurophysiological condition, but what distinguishes it from an automaton is its ability to have consciousness of its motivations and to either affirm or negate them by choice. Even when one is physically forced to act in a certain way, for example through violence or imprisonment, one can choose against the situation, even if this choice cannot be actualized. This approach allows Sartre to draw the important distinction between consciousness and volition: while consciousness is radically free and spontaneous, the will is not free, but is rather subject to a more primordial choice and project developed by consciousness. Sartre's rejection of determinism in consciousness has significant ethical and existential implications. Since

consciousness is not determined by causal forces, individuals do not merely follow predetermined scripts and are hence fully responsible for their actions. Sartre goes even further, however, insisting that humans are fundamentally responsible for their choices and their consequences, regardless of external circumstances. This claim is rooted in his existentialism, specifically the thesis that what makes us responsible for our choices is not this or that choice and its justifications, but the ability to choose as such, an ability not possessed by 'things', such as tables, rocks, or brains, for that matter. From this emphasis on the absolute nature of freedom as characteristic of consciousness, Sartre derives his thesis that existentialism necessarily implies humanism. It is a humanism because it highlights the significance of human agency in shaping oneself as a human being—shaping one's existence and values. By asserting that individuals are not predetermined by any fixed essence or external standards, existentialism affirms the inherent dignity and potential of each person to define their own meaning and purpose in life. The dark side of this humanism is that we bear full responsibility even when making choices that lead to catastrophic consequences and appear to be beyond our control, as is the case in socio-technical systems; we cannot shift our responsibility to anything other than the human we chose to become. In light of all this, the ideas contained in Sartre's "Existentialism is a Humanism" are as relevant in the technoscientific era as they were in 1946."

Martina Properzi, Università Pontificia Lateranense
Replacement vs. Supplementation: the human body and the challenges of restorative/augmentative technology

Intelligent machines are increasingly being used to restore and augment the user's body. The restorative or augmentative nature of implantable chips, bionic limbs, biomimetic organs, neuroprostheses, etc., has been interpreted in many ways (see, among others, de Vignemont 2024). There are two basic ways of approaching the question of technological restoration and augmentation of the human body. The first way emphasizes the aspect of replacement. Sophisticated artificial intelligence (AI) can, at least ideally, replace existing motor, sensory, and body-related subjective abilities of the human user. This approach, which I call 'body replacement', builds on a conventional interpretation of computing as an artificial activity that uses human-made machines to process information in digital form. The second way sees computing as an activity inherent in the natural body that can be used by humans to create nature-inspired intelligent machines. In the so-called natural computing perspective (Adamatzky 2017), AI complements the human body by enhancing human bodily capabilities ('body supplementation'). In this article, I present and discuss the two approaches, namely body replacement and body augmentation, following a case study strategy. The so-called neuro-reality (Houser 2017), generated by an advanced brain-computer interface (nanometric brain chips receive signals with 'single-neuron resolution' and transmit them wirelessly to extracorporeal computers or smartphones. See Drew 2024), is analyzed as a relevant case of body replacement. On the other hand, neuroprosthetics, especially visual neuroprostheses (implantable medical devices capable of reactivating visual neurons through electrical stimulation. For details, see Borda and Ghezzi 2022), are studied as a highly informative example of the supplementation of the user's body by intelligent machines. The analysis of the

case study brings to light the philosophical underpinnings of both the replacement and augmentation approaches to restorative and augmentative technology. Body replacement is found to be informed by a radical functionalism, according to which bodily states are identified by their functional role, i.e. by 'what they do'. Otherwise, material constituents are centralized through bodily supplementation. To understand what bodily states are, you have to look at what they are made of. What kind of representation of the human being is supported by the replacement and the suppletive approach?

Heup Young Kim, Kangnam University
New Humanism at the time of Artificial Intelligence: A Theo-daoian Reflection

This paper investigates the realm of humanism in the contemporary era, where artificial intelligence is reshaping every aspect of human life. It does so through the distinctive lens of Theo-dao, an East Asian Contextual Theology deeply rooted in Daoian and Confucian wisdom. This lens provides a unique perspective, juxtaposing the enduring humanistic values derived from Confucianism, such as benevolence (ren), righteousness (yi), propriety (li), wisdom (zhi), and trustworthiness (xin), against the backdrop of the Enlightenment's rationality-driven modern humanism. In the context of Pope Francis's critique of the technocratic paradigm in 'Laudato si,' the paper critically examines the contemporary era's tendency to prioritize technological advancement at the cost of the environment and morality.

Moreover, the paper delves into the nuances of posthumanism, shedding light on its emergent critique amidst the global ecological crisis and its anti-humanistic inclinations within the ambit of transhumanism. Through a Confucian critique, it proposes a humanism that is inclusive, transcending the limitations of its modern counterpart's exclusivity. This proposed humanism advocates for a paradigm harmony that encompasses human virtues, cosmogonic relationality (Taiji), and an interrelated theo-anthropo-cosmic (trinitarian) wholeness (Dao), offering a vision of a more harmonious and virtuous society. By weaving together the threads of Theo-daoian insights, including the Daoian philosophy of a non-intentional, supra-apophatic spirituality (wuwei), the study champions a recalibration of humanism in the age of artificial intelligence. It posits a forward-thinking anthropology that not only incorporates technological advancements but also profoundly reconnects with the Earth and its myriad inhabitants through a lens of mutual respect and interdependence. This work aims to contribute a pivotal perspective to the dialogue on technology and ethics, advocating for a future where technology enhances, rather than eclipses, the human spirit and its virtuous potential.

ATEM
Pontifical Academy for Life

Carlo Casalone, Académie Pontificale pour la vie
Relations médicales à l'heure de l'IA

CONTEMPORARY HUMANISM
LUMSA Sala del Consiglio / Aula Pia

Enrico Di Meo, LUMSA-ICP
Mechanism and Free Will: a possible Convergence Hypothesis

It is well known that human liberty of self-determining action, which is generally called "free will", is one of the most ancient and recurring enigmas of humankind. It has engaged philosophers, theologians and scientists for centuries, resurfacing each time in a different form. In modern times, the most famous, almost paradigmatic, formulation of the problem is Kant's Third Antinomy. But in recent decades the question has acquired particular relevance due to the massive development of neuroscientific research. A great number of scholars started to investigate the complex relationship between neural activity and conscious intention, especially after the pioneering works of Benjamin Libet (Libet 1983; 2004), some of them coming to the shocking conclusion that free will is merely an illusion. These results give strength to a renowned stream of thought, which is determinism. This, of course, poses serious problems, both theoretical and practical. Firstly, not all the neuroscientists agree with the deterministic interpretation of experimental evidence, but it is not the aim of this paper to explore this aspect; secondly, and this will be the focus of the paper, there are philosophical assumptions and implications of these experiments that are often superseded; and finally, one of the main concerns regarding determinism is the sense that a complete mechanistic explanation would radically undermine our whole complex of notions which centre around freedom and moral responsibility (something that is often desired by supporters of determinism). This paper aims to investigate one particular form this question has taken in the ongoing debate that has followed the recent development of neuroscience. Taking Charles Taylor's 1971 paper as a starting point, this contribution will examine the question of whether it is inevitable to think of a neurophysiological account of human behaviour in mechanistic terms. What historical and philosophical presuppositions underpin mechanistic determinism? How can we reinterpret Libet's (and followers) results? Answering these questions will lead to a possible defense of a convergence hypothesis that is neither dualistic nor reductionist. Furthermore it will be argued that this hypothesis hints at a possible ontology with more than one level.

References
Libet B. et al. (1983), Time of Conscious Intention to Act in Relation to Onset of Celebral Activity (Readiness-Potential): The Unconscious Initiation of a Freely Voluntary Act, in "Brain", 106, pp. 623-642.
Libet B. (2004), Mind Time: The Temporal Factor in Consciousness, Harvard University Press, Cambridge (MA).
Taylor C. (1971), How is mechanism conceivable?, in Interpretations of Life and Mind: Essays around the Problem of Reduction, edited by M. Grene, Routledge and Kegan Paul, London, pp. 38-64.
Taylor C. (1985), How is mechanism conceivable?, in Philosophical Papers vol. I. Human Agency and Language, Cambridge University Press, London/New York, pp. 164-186.

Flavia Chieffi, LUMSA-UCly
The role of «symbolic consciousness» in Virgilio Melchiorre's philosophy

Virgilio Melchiorre's theoretical elaboration is an attempt to integrate a metaphysical research, regarding the constitutive themes of Being, with an existential one. In the preface of a collection of studies on Kierkegaard's philosophy he wrote, "these themes would be meaningless if they were not approached starting from the living flesh of existence and only returning to its heart searching its meaning and destiny" (Melchiorre, 2016). The main objective crossing his philosophy is to find a form of relationship between existence and Being preserving difference, not allowing neither an immanentistic solution, nor a theorization of an absolute transcendence. The aim of this study is to analyze the theme of "symbolic consciousness" in his philosophy. Starting from the analysis of perception and its limitation, an intrinsic tension of the perspective consciousness to transcend itself will be identified. Consciousness is both situated and principle of desituation, both immanence and ability to transcend. Melchiorre defines this intentional duplicity as 'ambiguity'.

Without being referred to a Totality, the evidence of a limit would be a contradiction, a non-sense. The recognition of a non-exhaustiveness of reality implies a primordial affirmation of sense. Because of its partiality and abstraction from the Totality, the existence constitutes a diversity. Nevertheless, the Being, common core of all reality, appears in each presence pointing, at the same time, to an absence. The ineliminable transcendence of the Being establishes an ontological duplicity on which the ambiguous structure is based, defining the consciousness as unity, relationship of presence and absence. Searching for a function reflecting the double intentionality of the consciousness, Melchiorre suggests that the faculty of imagination, due to its intentionality living in the world of absence, is the one allowing a connection, complementing perception and leading to the absent. However, the absence can only be perceived analogically. The determination to which the imaginary leads is symbolic. "Symbolic" is considered a reality that while speaking about itself, also speaks about another owing to an original communion. The possibility of the symbolic expression arises in the being of every phenomenon, in its constitutive relationality. While stating about itself, every entity also states about the Being, precisely by analogy. This duplicity can occur since it is supported by an intentional movement converging on the universal as the Being everywhere participating and transcending itself. That consciousness that while being in the individuality reads the universal there, holding difference and identity together, is the symbolic one.

References
Melchiorre, V. (2016), Le vie della ripresa. Studi su Kierkegaard. Milano: Vita e Pensiero.

Cecilia Benassi, LUMSA
The embodiment of form - Symbolic between poetry and technology

This paper discusses Pavel Florenskij's vision of concrete metaphysics as it emerges from a section of the unfinished *On the watersheds of thought, The embodiment of form*. One of the writing's main focuses, in which the author centres on the relations between man and the world, and the ways of their interactions, is man's creativity. It will show that, according to the author,

human creativity is closely linked to the divine-human task of the transfiguration of the world; in this sense, the relation between man and the world has a sophianic core. Thus, concrete metaphysics implies the acceptance of Solovev's philosophy of full unity (*vseedinstva*) and divinehumanity (*bogochelovechestvo*), personalized in a sort of dynamic onthology which is presented by the Symbol and expressed by man's symbolic creativity. Therefore, this metaphysics can be regarded as an exercise of practical symbolism and the task of man's life as a poetical and practical experience, which could shape reality in the form of God.

Gael Trottmann-Calame, ICP- LUMSA
An all-too-modern modernity: a genealogical investigation

Nietzsche is famous for having asserted that man is something to be overcome, that a new type of human, if not an "overman", must be summoned in the face of the all-too-human human, and for having rejected the existence of the free will. So, wouldn't we have reason to think that the philosopher from Sils-maria would be like the prophet of this very contemporary craze for augmented man (transhumanism) and unlimited artificial intelligence (AI), or even of the mechanistic and materialistic understanding of the human mind (neuroscience)? In truth, if we follow Nietzsche, these current temptations are more a regrettable extension of "nihilism". Indeed, if we take a genealogical approach, there's a good chance we can unmask the instincts that Nietzsche tirelessly denounced behind these contemporary temptations. Is it not a certain weariness, a certain disgust, a certain disappointment, if not a deep resentment of the human race, a feeling of irremediable powerlessness, in short, the "weakness" and "decadence" of a certain type of human being that is being expressed here? So much so that, far from embodying the overcoming of which Nietzsche championed (Selbstüberwindung) — calling for a new type (Übermensch)—, the other human or the new intelligence to which some aspire would rather be a fall, a negation, in short the assumption of the "fragment man", the "last man". Poison rather than cure, the vain quest for a different humanity rather than a metamorphosed one, invites us to ask, with and following Nietzsche: "What is it here that hates so much"?

Jérémie Supiot, UCLy-LUMSA
Constructivism and relativism. On the democratic virtues of realist constructivism

The origin of constructivism in philosophy is hard to determine. For example, it is difficult to decide whether it is was Parmenides, Kant, Tassy, Kuhn or Piaget who was the first to develop such a conception of knowledge (see Rockmore 2005, 2016, 2021). However, we do know that it thrived in the end of the XX's century especially in human sciences.
Constructivist epistemologies answered the necessity of deconstructing dominant ideas and ideals, and lead to the emergence of a series of studies especially in the USA like cultural studies, post-colonial studies, gender studies, science studies or engineering studies. Today, constructivism is so diverse in its forms of expression (like in the field of Science and Technology Studies (STS) which counts many researchers and disciplines) that it seems preferable to speak of constructivism in the plural, as Lemoigne does (2021), for example.

However, since the episode known as the "Science Wars" in the 1990's, the suspicion of relativism weighs on constructivist epistemologies. They are blamed for the loss of public confidence in sciences, the development of a post-truth era, making it more difficult to distinguish between science and pseudoscience, endangering our democracies (Engel, 2017). It is to be feared that by weakening sciences and scientific knowledge, constructivism strengthens diverse forms of negationism and complotist theories. If scientific knowledges are psycho-sociological constructs, how to legitimize the fact that scientists have the (unshared) privilege of telling the truth to the rest of society? According to Bensaude Vincent and Dorthe (2023), the growing mistrust regarding scientific institutions and their capacity to tell the truth should be differentiated from the legitim distrust regarding scientific institutions and the production of knowledge. That is why we will argue that not only a realistic account of constructivism is possible, but it could lend more credibility to sciences and scientific knowledge by bringing more transparency concerning the construction of knowledge. Instead of relying on the (wrong) belief that scientists "know better" because of their ability to see the truth, it is possible to show how sciences rely on the collective methodic construction of knowledge (Latour, 1979, 1987, 2022). Not only sciences need to slow down (Stengers, 2013) to improve the quality of the publications, but also to take the time of democratic debates concerning matters of concern (Latour, 2014) using their privileged point of view (despite its flaws and limitations) to agree on a shared reality.

Keynote Lecture: Laura Palazzani, Università LUMSA
Health at the Time of AI and Neurosciences

10.30 am PARALLEL SESSIONS

NHNAI Network
Notre Dame Rome Room 103

Margherita Daverio, Università LUMSA
Towards humanism in the digital age. Informed consent as a potential driver of integration between human factor and artificial intelligence in healthcare

In 2020, the Italian Committee for Bioethics dealing with sensitive issues related to the increasing use of algorithms and artificial intelligence (AI) systems in medicine highlighted the importance of identifying the ethical conditions for the development of Al that does not forsake critical aspects of our humanity. The Committee in the same document hoped for a new "digital humanism" and therefore, as regards medicine, healthcare "through" and not "of" AI systems. Within this general aim, in this contribution we will sketch some insights to discuss how informed consent could be a potential driver of integration between human factor and AI technologies in the healthcare sector. If consistent, research in this field could be a step towards humanism in the digital age, with specific reference to medicine and healthcare.

In recent international guidance and scientific literature, ethical challenges emerging for patients' autonomy and informed consent in the specific case of medical decisions based on AI systems have been identified alongside with the main implications for the relationship between doctor and patient. International recommendations highlight that in the face of challenges raised by AI in medicine, human choices and responsibilities should remain critical to guarantee decision-making in healthcare. AI technologies such as machine learning and deep learning in fact have the potential to significantly advance the quality of health care. However, technological features of AI systems (e.g. black-box nature of AI systems; the possibility of unpredictable errors in the absence of human supervision; and the challenge of biases, which is strictly related to the problem of the huge amount of data used to feed and train software) show how decision support systems could hinder medical decision-making. These technological aspects can generate an epistemic loss in medical understanding and explanation and therefore prevent the provision of accurate and balanced information. The use of AI technologies for decision-making in healthcare also has some implications for the doctor-patient relationship: relying exclusively upon decision-making through AI-driven technologies could affect various aspects of clinical care, for instance, there could be a risk of a decrease in a holistic approach to care, where the patient risks being considered as a "bearer" of data rather

than the subject to take care of. The doctor could be induced to rely more and more on technology, arriving, on the one hand, to take responsibility in the face of a choice based mainly on data coming from the AI system, and on the other, to increasingly delegate the substance of the medical act to the decision of the AI system. On the other side of the relationship, the increasing use of AI technologies in medical decision support can influence the patient, who may tend to overestimate the accuracy, neutrality, and objectivity of AI systems. In the face of these challenges, informed consent through a meaningful information process as recommended by international guidelines, and intended as a personal interaction between the doctor and the patient could be critical for promoting an integrated approach to AI in healthcare. In recent advances in scientific research in this field emerged an increasing need to rethink informed consent considering the above-mentioned challenges involved in using AI technologies in the care field. Discussing the ethical aspects highlighted above can contribute to sketch recommendations for a more comprehensive and complete decision-making process towards integrating AI technologies to effectively support – and not replace – doctors. Within the existing ethical framework, an integrated approach and use of AI could be promoted within the doctor-patient relationship specifically through the information and consent process (1) ensuring in a reliable dialogue transparent disclosure of information to the extent possible; (2) guaranteeing human-centered shared decision-making, aware of possible limitations of AI technologies; (3) fostering patient comprehension and trust, and more broadly a holistic approach to the doctor-patient relationship. (1) In the context of the doctor-patient dialogue informed consent allows the patient to understand the purpose of the treatment to be undertaken on the doctor's advice and to have a role in shared care planning; on the doctor's side, it presupposes the duty to make the risks and benefits of medical procedures understood and explained in a way that is understandable to the patient. As informed consent is one component of the doctor-patient relationship requiring discussion between patients and health professionals on possible treatment options and values, transparent information, addressing the mentioned challenges, could be seen as a facilitator of meaningful dialogue between patient and doctor about options in AI-mediated care. (2) A key component of patient-centered care is shared decision-making aimed at identifying the treatment best suited to the individual patient's situation. Shared decision-making involves an open conversation between the patient and the doctor, where the doctor informs the patient about the potential risks and benefits of available courses of action, and the patient discusses their values and priorities. Assuming doctors remain the primary point of care for patients, decision-making could be shared among the physician, and the patient, including the support given by AI technologies, through adequate ethical and regulatory governance. (3) A defining characteristic of medicine is the "healing relationship" between doctors and patients. This relationship could be "augmented" but not replaced by the introduction of AI. The role of the patient, the factors that lead people to seek medical attention, and the patient's vulnerability are not changed by the introduction of AI as support in clinical decision-making; rather, what changes is the means of care delivery, how it can be provided, and by whom. If explainability standards are provided, the doctor can receive an explanation from the AI system and then translate the system's output into meaningful and easily understandable information. In conclusion, standards related to informed consent could be the basis for deployments of AI in healthcare that help rather than hinder the trusting relationship between doctors and patients. An improved design of the information and consent process promoting a meaningful dialogue can help to transform consent into a driver of integration

between human factors and AI technologies in the context of the therapeutic relationship. It is therefore envisaged the importance of further research on the potential role of informed consent in this context, specifically adapting form, contents, and the process itself when AI systems are applied for decision-making. This could help move towards a digital humanism in healthcare and medicine keeping at the center the human factor, with the support of AI technologies.

Fernand Doridot, Université Catholique de Lille
About the supposed "anti-humanistic program" of converging technologies

Inquiring into the future of human freedom and humanism within the context of advancements in neuroscience and artificial intelligence naturally leads to a questioning of the notion of "technological convergence" as described and advocated by certain observers. Is there, within the technologies developed in our era, and more broadly in emerging technologies, a common trend and an underlying joint program that harbors the seed of irreparable harm to what has thus far defined our shared humanity? This concern has been raised by various commentators, particularly in Europe, in response to the publication over twenty years ago of the seminal report Converging Technologies for Improving Human Performance by Rocco and Bainbridge (2002), purportedly heralding the age of NBIC (for Nano-Bio-Info-Cogno) convergence. Now, twenty years later, as various works
offer a current assessment and critical review of the forecasted technological convergence, it is worthwhile to examine how the "grand visions" that had infused this report and the critical discourses it had provoked have evolved. In our presentation, we will provide various evaluative elements of this broad question, particularly in light of how the agenda of technological convergence, while having become more subdued in the West, encounters distinct interest in other regions of the world, where it is viewed through the lens of different scientific and philosophical traditions. Do emerging technologies lead us somewhere, and is this destination conducive to humanity's well-being? Ultimately, this is what we will attempt to provide some answers to.

Javiera Reyes Brito, Pontificia Universidad de Chile
Affectiveness and emotion: redefining the human in the era of artificial intelligence in the perceptions of Chilean residents of the Metropolitan Region

The deployment and development of artificial intelligence is driving relevant changes in society whose scope at the level of people's experience and perceptions is still unknown. As a product of the future development of science and knowledge, artificial intelligences can be understood as the result of the advance of reason through technology, generating tools and mechanisms that are becoming embedded in broad areas of social life, altering the way in which we relate to the world. This study seeks to explore the perceptions and attitudes that Chilean adults residing in the Metropolitan Region have regarding the role of artificial intelligence in the field of health, democracy and education. Through the use of qualitative data with the use of focus groups and the participation of 86 people from different socioeconomic and cultural levels, it is possible to observe that, far from perceiving artificial intelligence as a threat to

humanity - and even as its potential replacement-, these are conceived as technical means for the development of society, while what is fundamentally human is resignified in the stories of the participants through an assessment of the emotional, the a1ective, the intersubjective, the centrality of the senses and the direct perception of the world. Concepts such as care and accompaniment, meeting others and respect for human integrity become relevant. The subjects point out artificial intelligences as incompatible with these attributes, making a direct relationship between technique and reason, while the human remains as the essential reservoir, irreplaceable and inimitable by new developments in technology. Although these perceptions are transversal among the participants, they tend to be stronger at a lower socioeconomic level, with the concept of human replacement appearing only in the more educated participants.

Sofia Aurilio, Università del Salento
Augmented Porosity and Viral Infections: How Do Linguistic Corpora Trace the Borders of Gender?

Today's constant interaction between the sensible/ "actual" world and the online/ "virtual" world marks an unmistakable porosity between these planes. Here I refer to it as "augmented porosity", following another famous definition of this intersection between realities, that of "augmented reality". One of the systems which operates in both these realities is that of gender and, more specifically, gendered representations. In this essay, I shall hence focus on AI representations of gender in order to explore the porosity between actual and virtual genders and their representation in linguistic corpora. As it has already been observed, numerous "actual" biases of representation get translated into "virtual" algorithms, giving rise to preoccupations of partiality. Nevertheless, this translation may also prove an interesting tool to observe these very biases in a controlled environment—that of corpora and linguistic representation—in order to then amend them in both worlds. As a matter of fact, it may prove interesting to view this gesture as a form of viral infection—as the VNS Matrix already observed in the past. By inserting oneself as a viral presence in these online representations—by metaphorically travelling though the pore—one may be able to highlight the biased mechanisms that animate our realities in order to expose them. My research will focus on two main questions regarding this augmented porosity. Firstly: how does AI trace the borders of gender in corpora? And secondly: What are its biases? Through this "viral" methodology I will attempt to bring to light the many discrepancies in representation that interest all gendered locations, in an effort to bring light to the biases that permeate our actual and virtual realities to help shape a more comprehensive and shaded array of representations.

Jeyver Rodriguez Banos, Universidad Catòlica De Temuco
The Moral Storm of Artificial Intelligence in Global Health: Building Bridges between One Digital Health, Neurorights and Technological Humanism

The accelerating digitalization of global healthcare is revealing a number of ethical, social, technological, and epistemic challenges related to the increasingly widespread use of IA at different levels of the healthcare system. This article addresses two key issues related to a new

paradigm called One Digital Health (ODH). ODH seeks to operationalize and achieve many of the promises of the One Health approach. In this article, I show that ODH should be understood as a "promising long-term project" aimed at improving the tools and capabilities for collaboration among scientists, practitioners, policy makers, and other stakeholders. We need to think about and anticipate the possible risks arising from the unethical use of AI in global health, but this should not become an obstacle to harnessing the potential of generative AI, big data, and the culture of digitalization in an attempt to develop multidimensional strategies, plans, and policies to mitigate the upcoming environmental disaster manifested in the Sixth Mass Extinction of Species and populations and the accelerated deterioration of planetary health long documented by the IPCC (IPCC, 2023; Rodriguez, 2023). The accelerating pace of digitalization and the growing prevalence of AI are giving rise to a moral storm that poses a challenge to the One Health approach. In the latter part of this article, it deepens into the normative framework of various instruments designed to grapple with the intricacies of neurorights. Notably, this includes an examination of the Chilean Constitutional Amendment and the European Charts of Digital Rights. It is noteworthy that Chile has emerged as a trailblazer by becoming the world's first country to enact legislation on neurotechnologies and enshrine Neurorights within its Constitution (López-Silva & Valera, 2022; Valera, 2022; Yuste & de la Quadra-Salcedo, 2023). The article concludes by emphasizing that in order to move towards an ethical and responsible use of IA, we need to articulate the vision of ODH within a vision of Commons Goods and Technological Humanism.

ATEM
Notre Dame Rome Room 104

Thierry Magnin, Université Catholique de Lille
Le développement de l'IA est-il compatible avec l'écologie intégrale ?

Jean-Marc Moschetta, Institut Catholique de Toulouse
L'Intelligence Artificielle dans la perspective du salut en Jésus-Christ