NHNAI Output 1.2

Transdisciplinary mapping of expert knowledge

Table of Contents

Authorship and contributors

Authors:

- Baptiste Bedessem (thematic coordinator for democracy)
- Stefano Biancu (thematic coordinator for humanism)
- Fernand Doridot (thematic coordinator for health)
- Brian P. Green (thematic coordinator for AI ethics)
- Corinne Mellul (thematic coordinator for education)
- Juan Vidal (thematic coordinator for neuroscience's ethics)
- Members of Carisma Project
    - Pascal Marin
    - Riccardo Rezzesi
- Fabio Macioce (Activity 1 leader)
- Laura Di Rollo (NHNAI research engineer)
- Mathieu Guillermin (NHNAI coordinator)

Contributors:

The content of this document draws on contributions presented at NHNAI 1st academic workshop (11-12 March 2022) by academic experts of NHNAI network:

- Sylvain AUCLAIR (International Observatory on the Societal Impacts of AI and Digital Technology, Québec): Thinking about privacy: from the individual to the human person
- Pr. Céline BORG (Sciences & Humanities Confluence Research Center, Lyon Catholic University): The role of the body in cognition (the case of Alzheimer disease)
- Dr. Chien-Wei CHUANG (Fu Jen Catholic University) with Pr. Ben-Chang SHIA: A Worldwide Bibliometric Analysis of Publications on Artificial Intelligence and Ethics in the Past Seven Decades
- Pr. Francisco DE LARA LOPEZ (Pontifical Catholic University of Chile): Explaining what it means to be human through AI and Neurosciences: limits and blind spots
- Dr. David DOAT (Catholic University of Lille): What theory of mind in the age of AI, neuroscience and transhumanism?
- Dr. Eric FOURNERET (Catholic University of Lille): Natural and artificial: do neuroprostheses question human nature?
- Dr. Augusta GASPAR (Catholic University of Portugal): Neuromarketing and the ethical commitment to human wellbeing
- Pr. Marco GRAZZI (Università Cattolica del Sacro Cuore): AI and its impact on employment
- Dr. Brian P. GREEN (Santa Clara University): Ways of Thinking about AI Ethics: Theory and Practice
- Dr. Nathanaël LAURENT (ESPHIN, University of Namur): Problematizing neurosciences to return to the human
- Dr. Everlyn M'MBONE ANDUVARE (Catholic University of Eastern Africa): Indigenous Knowledge and AI
- Pr. Fabio MACIOCE (Libera Università Maria Ss. Assunta di Roma): Safety and security: two challenges for the Ethics of AI
- Pr. Alejandra MARINOVIC GUIJÓN (Pontifical Catholic University of Chile): Citizenship education at schools in a context of multidimensional inequality
- Dr. Corinne MELLUL (International Federation of Catholic Universities): Technology in higher education: the ethical dimensions
- Dr. Carolina PIRES MARUTA (Catholic University of Portugal): Human-Machine error: Neurosciences, AI and clinical diagnosis

- Pr. Yves POULLET (ESPHIN, University of Namur): How International Public Organisations are envisaging AI Ethics? Convergences and divergences.
- Pr. Ben-Chang SHIA (Fu Jen Catholic University) with Dr. Chien-Wei CHUANG: A Worldwide Bibliometric Analysis of Publications on Artificial Intelligence and Ethics in the Past Seven Decades
- Dr. Warren VON ESCHENBACH (Notre Dame University): An Integral Human Development Approach to Ethical AI
- Rev. Dr. Emmanuel Charles WABANHU (Catholic University of Eastern Africa): HealthCare in African Context: An Ethical Consideration
- Dr. Tim WENINGER (Notre Dame University): Drowning in Data: AI Triage Systems to Aid Human Rights Workers
- Pr. Ping-Keung YIP (Fu Jen Catholic University): "Patient Right To Autonomy Act" In Taiwan, A Neurologist's Viewpoints

Introduction

This document constitutes an upgrade of NHNAI's output 1.1 (Interdisciplinary mapping of acadmeic knowledge) that builds upon questions co-constructed in each partnering countries with societal actors. This output 1.2 aims at proposing insights from academic experts that is as much as possible aligned with these co-constructed questions. These questions are recalled in grey boxes.

The document is divided in 4 main sections, one per field of applications of the whole reflection upon humanism (namely the fields of education, health and democracy) and one section devoted to topics that are transversal with respect to these three fields. In each of these for sections, the last subsection gathers co-constructed questions that, rather or in addition to raise issues for societal debates, seems to appeal for direct inputs from academic insights.

This is the first finalized version of the document and it is meant to evolve through successive iterations.

Humanism, AI and Neurosciences.

I. Expert insights on "Humans and humanism: definitions, questions, challenges."

Synthesis of co-constructed questions

What is a human being?

- What are the core traits of what it means to be a genuine human being?
- Is the human being only matter?
- Are transcendence and spiritual aspect fundamental to be human?
- Is interpersonal relationship fundamental character of humans?
- Does belonging to the human species is sufficient to count (and to be recognized as) a genuine human?
- Is there a need for a consensus on what is human?

What does it mean to be "intelligent"? How to use knowledge?

- Do physics, biology and neurosciences exhaust what can be validly said about the mind and brain of human beings?
- What relationship humans should have with knowledge about themselves? Can such knowledge sometime become a danger?
- Should we keep a part of ignorance upon the knowledge of human?
- How knowledgeable are humans of their biases in information processing and how can we improve such knowledge?

What is the meaning and purpose of human evolution? What goal's humans should pursue?

- Why do we develop technological tools? Are there some limits we should impose on our technological powers?
- Should the present human being be considered as the end point of evolution?
- How far can we go in our movement to integrate technology while respecting human life?
- Do technologically induced characteristics of humans belong to human nature?
- Can /should humans intentionally guide their evolution?

Humanism does not say what a human being is

The humanism of modernity itself does not constitute a definition of what it means to be human. **The general notion of humanism is extremely polysemic** (Biancu, 2019).

From an historical/historiographic sense, the term "humanism" refers to precise moments in the intellectual history of Europe and the West: Italian humanism of the 15th century, German new humanism of the 18th century (the *Goethezeit*), and the various humanisms of the 20th century – the pedagogical (Jaeger), the Christian (Maritain), the Marxist (Fromm, Merleau-Ponty, …), the existential (Sartre, Jaspers), the many humanisms of the Anglo-American humanist movements. Then, there are the reactions to these humanisms: the anti-humanisms of the 20th century (Foucault, Lévi-Strauss, Lacan, …) and the post- and trans-humanisms of the 21st century.

Humanism is also a broad term in culture. In this second sense, humanism is not only a term representative of a given historical period or intellectual current; it is rather a term that serves as catalyst for a *Weltanschauung*, an *ethos* that implies social and political institutions of a certain shape. Humanism is at this point nothing less than an "eponym" for European – and, consequently, Western – civilisation (Tognon, 2019). In this way, it could be identified as the fulfilment of Western metaphysics dating back to Plato (Heidegger, 1947).

Finally, in its axiological meaning the term becomes a regulative ideal that aims to establish a space of reciprocal recognition and a just order of relationships. Along this axiological meaning notably come *human* or *fundamental rights* putting the focus on any human individual "as a free-standing self-determining person with an identity and a name that is not simply a marker of family, birthplace or occupation but is 'proper' – belonging to [her or him] alone" (Davies 1997, p. 16). Freedom of speech and freedom of thought are emblematic examples of such fundamental rights. Rooted in his intellectual and moral freedom the human individual shall not blindly defer to external authorities (be it religious, aristocratic or of another form). On the contrary, human individuals can and must rely on their own cognitive resources.

Nevertheless, such a regulative ideal remains a partially effectuated (at best) human institution. This point was well appreciated by Michel Foucault, who underlines that humanism is, to all effects and purposes, a myth (Foucault, 1969). The idea of humanity and the category of the human are – within Foucault's perspective – a historical invention. Accordingly, the anti-humanisms of the 20th century represent a radical attempt to deconstruct humanism. Nevertheless, the statement according to which humanity is able to function without myths is itself a myth: the myth of a humanity devoid of myths.

Mythical does not mean false, i.e. the opposite of "true": if a mythical story is not true, it is so because it is not meant to be an account of the facts in a hypothetical past, a report providing empirically verifiable data. What myth lays out is rather the *truth* of history: a truth wherein everyone can recognise themselves and find a pattern on which to find fair relationships (relationships with themselves, with others, with the world and beyond). In this sense, terms like "humanism" and "humanity" are mythical categories that aim to establish a symbolic space where recognition and a just order of relationships are made possible (Biancu, 2019).

Although at core of what it means to be human in Europe and the West, humanism cannot serve as a definition of what it means to be human. As a local historical and political invention, it demands further scrutiny and reflection. It raises many questions. For instance, can such a space for recognition and just order of relationships be universal? In the negative case, is it a problem? How to make effective or progress in direction of such an abstract and formal regulative ideal?

Natural and *hard* sciences inputs on what it means to be human

Insights from NS

Neurosciences (NS) and artificial intelligence (AI) are two scientific (and technological) domains that explore human mind, human thinking, cognition and intelligence. Intelligence, ability to reason and know has traditionally been uphold as a core trait of human beings, notably in the dualistic cartesian anthropological and philosophical stream. Intelligence is often considered as an exclusive characteristic of humans (by opposition to artifacts and animals) or, at least, as a dimension that humans master without equals in an unchallenged domination. However, insights from AI and NS seem to call for a reconsideration or an enrichment of traditional views.

NS is a field of study that uses the scientific approach and method to achieve an understanding of the nature of nervous system (brain, spinal chord, nerves, neurons, etc.), its role and its mechanisms in animal (human incl.) behavior and mind. It is the study of the biological properties and mechanisms of this system in its interactions with the body, the physical and social outer-bodily environment.

The philosophical stance of most neuroscientific community is that the mind is a product of the operation of the nervous system in its interactions with the rest of the body which is always embedded in its surrounding, and often changing environment (adapting to it continuously). By this **NS assumes a naturalistic and materialistic approach to the understanding of mind**, i.e. understanding the material/biological laws or properties of this interactive system will reveal a realistic understanding of the phenomenon of mind.

Although it is often understood as or suspected of radical reductionism, this naturalistic and materialistic approach above all reflects a methodological point of view, a framework for studying the behavior and mind of lively beings possessing a nervous system (that does not intrinsically or necessarily forbid approaching the same topics from other standpoints). According to this methodological stance, NS cannot admit traditional dualism with the mind belonging to an absolutely different realm from the material one. Mind so understood is, by nature, outside of the scope of NS. This being said, the notion of matter, to which NS restrict themselves, cannot be reduced to a Laplacian mechanistic and deterministic conception. Contemporary understanding of matters is far more subtle. Physics includes light in matter in the broad sense of the term. Einsteinian relativity and quantum physics leads to deeply rethink our views on space, time and basic constituents of matter. Matter biology and NS confront with comes with its solid share of emergent complexities.

NS study of the nervous system extends into a multi-scale approach, ranging across all levels of biological organismic organization, from subcellular components (genetics, molecular biology) to behavior, to social and non-social body-mediated interactions with the biological and non-biological environment. NS thus span a large spectrum of specialized fields of study ranging from genetics to neural networks, from psychology to social interactions.

NS investigate how aspects of human experience, be that mental and psychological contents, i.e. thoughts, are related to certain features of brain function, often now expressed as activity of distributed neural networks. It is thought that all mental and behavioral pathologies or mental disorders can be traced back to some changes (dysfunction or dysregulation) of neural interactions within the brain. This strong causal link is an indication that mental (and spiritual[1]) dimensions of human experience are deeply rooted in our body neurobiological constitution. The progressive evidence that the system maintains a flexible relation between function and network structure (connectivity) due to neuroplasticity at the local and global level of organization, shows that the relation between mental processes, behavior, and neural activity is more complex than initially thought.

Historically, NS are part of a broader field called cognitive science, which includes disciplines focused on the understanding of knowledge systems and mind: neuroscience, linguistics, psychology, philosophy, artificial intelligence, and anthropology. NS has diversified in a series of branches. The branch of NS that has focused on a functional analysis of brain/neuronal activity, and that has adopted the informational-computational metaphor of cognition (cognitive proxy for mind), by studying the neural correlates of cognitive processes, is called cognitive NS. Other branches that focus more on the biological properties of neurons as cells and within neuronal networks are cellular and molecular NS (also known as neurobiology). Cognitive NS use neuroimaging and electrophysiology recording methods of brain activity to study the processes of cognition, often through a correlational approach. Because correlation does not entail a causal relation between elements, nowadays it has become possible to go beyond correlational analyses by producing specific localized interferences in brain activity (in animals and humans) by means of electric/magnetic/optogenetic stimulation to test the level of causal relation between certain brain/neuronal activity patterns.

Insights from AI

Since its premises in the 1950s with Turing, computer science and AI constantly pushed the boundaries of what programs can do, with a new impetus in the last decades due to the impressive results of deep learning methods. A number of definitions of AI have surfaced over the last few decades. **A very clear definition of AI** is the following: "AI is the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable" (McCarthy 2004). As the notion of intelligence proves peculiarly difficult to characterize, for humans as well for other animals or machines, one can borrow from Marvin Minsky (1968) another interesting definition of AI: “the science of making machines do things that would require intelligence if done by men” (p. v). From these definitions, an important point to highlight is that AI is not a thing or a machine or a program. Rather, AI is a scientific (and technological) domain that aims at producing machines and programs able to perform some tasks that would demand intelligence to humans, By extension, we could include tasks that human intelligence cannot achieve (such as very long and complicated calculations. Another goal often attributed to AI is to produce models of (human) intelligence allowing to better understand cognitive functioning. In that AI overlaps with cognitive sciences. In sum, AI aims at producing powerful programs and machines able to perform some cognitively demanding tasks. Without being restricted to such an approach, AI can pursue this goal under the direct inspiration of the way lively beings process information (humans but also other animals), attempting to mimic the manner we think consciously, or the way brains operate. An important technique in AI is machine learning, which can be understood as an automated assistance to human programing (through an automated search for adequate parameters of a given computational architecture).

[1] Before being approached by neurobiological terms, as any human behavioral and psychological phenomena, spirituality can be discussed in psychological terms, as the product of the mind and thus dependent on mechanisms of the physical mind. Spirituality can be considered a way to know and interpret experiences based on the belief in the existence of divine non-material and especially non-observable entities who can govern the material (observable) world and to whom causality in the material world is attributed. If spiritual entities were to be observable they would be subject to scientific validation and would not be spiritual anymore. Spirituality is closely associated to a belief (knowledge) in the absence of evidence, and thus somehow a process of arbitrary attribution. Very early in human evolution there was a presence of spirituality in human cultures, such as panvitalistic perception of the world. Extra-corporeal spiritual entities were attributed to all non-explained phenomena. Spirituality remains very important throughout all human cultures around the world. Importantly, spirituality is not exactly the same as religiosity. To be religious is to be spiritual and to adhere to a socially defined group that has normalized spiritual belief and organized into a social institution. People can be spiritual without per se being religious.

Combining NS and AI insights with other perspectives on what it means to be human?

Successes of AI and NS are indisputable. AI programs made striking progresses during the last decades thanks to machine learning (especially for problems involving the identification of patterns in vast databases). One can know consider delegating many tasks and decision-making to AI systems in field as various as layering, healthcare, recruitment, translation, defense, transportation … Such new possibilities raise deep interrogations about the place humans should occupy in an era of increasing automation.

In the same vein, medicine and NS have generated ongoing debates on the nature and essential understanding of fundamental concepts associated with the notion of humanism, such as agency, autonomy, self-determination and free will. Like astronomy at the times of Galileo, **NS is changing the way humans view themselves in the world and universe**. Some authors consider that, in the near future, NS will affect even more our views on what it means to be human, in comparison for example to genetics, because of the eminent proximity of the objects of study of NS at the level of the behavior and mental life of human beings.

In order to empower and enrich the collective societal questioning about *what it means to be human* at the time of NS and AI, it is important to have an accurate and clear understanding of what NS and AI as scientific and technological fields bring to the table in our understanding of who we are as humans, our behaving, our identity, our minds and everything that is associated to our *self* (such as affects, incl. emotions and feelings), and in our reflection on who we should be, on the place we should occupy in more and more technologically mediated societies. Some of these insights may affect and perturbate us. Nevertheless, a blunt and blind rejection of the inputs of natural sciences and their technological applications would be a great loss, as would be enclosure in these sole fields and exclusion of other potential sources of valuable insights, *e.g.* from human and social sciences, or from philosophical, spiritual or religious traditions.

Indeed, it is important to create or maintain open and lively a space of interaction and dialogue between the various sources of insights on what it means to be human. In particular, it is key to resist any ambient cultural discourse that would take advantage of discoveries in science and technology, and in particular in AI and NS, to forbid thinking any specificity of the human being and to announce its inevitable disappearance in the rest of the living world and in the machine. But the human being is still there and not so different from what he has always been, a historical being, caught in a fertile tension between nature and technology. In this context, it is a question of renewing today with the movement of philosophical anthropology (M. Scheler, H. Plessner, A. Gehlen), which was attached to the articulation of human ontology and scientific knowledge. Such a research can put in evidence that a philosophy of the human is not in competition with the scientific knowledge and its technical applications, but that it is quite differently in competition with a surreptitious philosophy, which does not say its name and which accompanies like its shadow the discourse emanating from the scientific discovery. This implicit philosophy identifies the being to the regime of objectivity of the sciences of the nature and supports so the reduction of the spirit to the living and the promotion of the ideals of the man-machine. This hidden philosophy, often ambient to the culture, the philosophical movement of the phenomenology gave him since E. Husserl the name "naturalism". It shapes a powerful and acting imaginary. To counter this imaginary and to reopen in this context an access to a thought of the human, it is crucial to make room to social and human sciences, to philosophical and spiritual resources.[2] Any knowledge that could help better understanding ourselves at the time of NS and AI is *prima facie* interesting and should be considered.

What goals humans should pursue?

Enriching and refining our understanding of what it means to be human, and specifically the understanding provided with the humanism of modernity, is of primary importance as we saw that these understandings play an axiological role, defining a kind of regulative ideal. Among the axiological material inherited from humanism, some elements may deserve criticism and amendment. For instance, humanism as a regulative ideal served the fundamental and questionable historical function of legitimizing anthropocentric claims. Now, some of the moral claims traditionally put forward by humanity, first at the European, then the Western, and finally the global level, are being contested from many standpoints, and often with good reason. Among them are: the

[2] In a parallel but complementary way to NHNAI, CArisMA project (Christian anthropology in the era of transhumanism and artificial intelligence) explores the contributions to such questions the resources of Christian anthropology could make, with its fundamental categories (the word, the body, the soul, the person, the community, the love, the divine etc.). The goal is not to artificially reactivate doctrines of the past but to reopen a thought of the simple of the human, which is always there active at the root of the knowledge, at the origin of the sciences and the techniques, even if these in their practices have no idea of it.

claim to dominance over nature and that of moral superiority over other animals. We are more conscious than ever of how the ecological crisis is bound up with the manifold threats to human dignity (Pope Francis, 2015). After the three blows to our narcissism (Copernicus, Darwin, Freud) and the teaching of the master of suspicion (Marx, Nietzsche, Freud), the human being is well aware of not being the master even of their own "house". The awareness of this irrevocability was already shared by one of the champions of humanism, Maritain: "Le malheur de l'humanisme classique est d'avoir été anthropocentrique, et non pas d'avoir été humanisme" (Maritain, 1936, p. 322).

Nevertheless, acknowledging such limitations of humanism does not mean rejecting the entire (axiological) heritage. Many other elements can prove extremely valuable. For example, humanism also performed a second historical function: **instituting a human measure that would allow for the defence of the weakest among human beings**, *i. e.* that of securing a boundary from below. This is to say that humanism has traditionally operated at the centre of a constellation of notions such as, amongst others, *freedom of thought* and *of speech*, *human dignity* and *human* or *fundamental rights*. In principle, these are terms universally shared, but as their content is not (any longer) self-evident, they need to be continuously renegotiated.

From this standpoint, important questions arise: is such an instance traditionally represented by the category of humanism still relevant today? Or are we dealing here with a legacy of mere historical and cultural value, no longer practicable after the end of eurocentrism and the political, scientific, and technological supremacy of the West?

Such questions are delicate. One can begin facing these challenges by highlighting that **there are two distinct but inseparable claims related to humanism in the axiological sense**: **the recognition of the shared humanity of all human beings** (the universal claim) **and the nature of this humanity revealed as a task** (the particular claim). This duality calls for an alliance between the humanist claim and particular cultures, which perform an irrefutable anthropogenic and humanising role. Humanism's universal claim – the recognition of the shared humanity, the realisation of which is a task falling to every human being – cannot develop in opposition to particular cultures. Without culture there is no humanity. The humanist claim of which we speak should not be mistaken for the dissolution of cultures into a conforming universalism. It acts, rather, as every culture's internal spur: as an axiological criterion and a stimulus to openness to the other. Humanism should, then, be understood not as a universal claim in which particular cultures are overcome for the sake of a common, abstract humanity, but as a recognition of the task of humanity, a task to be carried out in its multiplicity with creativity.

Accordingly, "humanism" is a term conveying mythical and axiological references that we cannot simply do without: not because we should make our self-esteem "great again" (as the populist rhetoric of our time would like us to do), but because of the claim of incommensurability and unavailability of the human measure and thus of the inclination to championing the cause of those who are unable to fend for themselves. This category is also crucial as a reference point for a just order of relationships. In this sense, we are certainly at liberty to discard the term "humanism", but then we will be required to find another, not yet available, notion. In any case, it is not the terminological precision of a Cartesian clear and distinct idea that we need to concern ourselves with. Rather, our concern lies with recovering a symbolic and mythical space that would allow for the unfolding and affirmation of a human measure as incommensurable, regulative ideal.

Conceiving humanism as a call for exploration, a task of self-understanding to be carried out, also provides and enriched framework for integrating new insights upon what it means to be human, especially those coming from the sciences. If a deep understanding of ourselves as humans is always to be explored and refined, notably to reflect upon who we should be and who we want to become, then scientific knowledge and its technological implications becomes highly relevant. It is notably the case with NS and medicine. They can for instance help human society and individuals becoming of some of their vulnerabilities. They have revealed the undeniable reactivity of the nervous system to any physical and psychological (informational) influence (pathological and non-pathological) from its environment and from within (from the non-neural parts of our body). Individuals are not all equal regarding their susceptibility to certain types of substance consumption or addictions (pharmacological/chemical or behavioral) because of their brain-body constitution. This type of information is highly valuable in an effort of humans to better understand what it means to be human and to determine who they want do become.

In the same vein, the natural and biological standpoint, which suggest to take distance from traditional dualistic views about human nature, highlight the role of the nervous system, but also of the body and the environment, in (human) cognition and intelligence. Such insights are key to understand what it may mean to be intelligent and to point the role of elements other than logical and formal reasoning, such as emotions and feelings, in rational decision-making (Damasio 2005). They are also very interesting to consider the claims about strong AI and an allegedly foreseeable future in which our AI creations would replace us (as with Hans Moravec's notion of "Mind Children"), who often relies on the idea that a computer program can exhaust the essence of what it means to be intelligent or to be human (computational theory of the mind), an idea tightly connected with functionalism and dualism. In addition, this type of reflection allows for a more balanced assessment of what AI programs are and a more fruitful reception of what they may teach us about who we are, as well as, consequently who we should be at the time of NS and AI.

II. Expert insights on "The human difference, at the time of NS and AI."

Synthesis of co-constructed questions

What makes humans different from AI?

- What makes us similar or comparable to it?
- In what sense or dimension could we think about comparing humans and machines/AI? In what sense would it be legitimate to talk about a competition between them?
- Is there not some characteristics that humans possess and not machines/AI? Could machines/AI somehow exhibit such human traits (learning, understanding, interpersonal relationships, feelings, creativity, morality, intuition, consciousness)? In what sense? What would it mean?
- Is there, or should there be - or not be - a definition of "intelligence" broad enough to include both organic ("human") and artificial intelligence?

How to preserve the human difference?

- How can humans become more human (and less like robots), to set them apart from machines?
- What are the core traits of what it means to be human that should be reinforced / preserved in the age of AI and NS? What roles should be exclusive to humans?
- Should technology (such as AI system) become so entangled with our lives that we can't clearly distinguish them anymore?

What relationships are consistent with the difference between humans and machines?

- What type of relationships should / could humans entertain with AI systems and robots?
- Can AI be more than a tool? How could a machine intelligence become more interesting than a human being?
- Do we trust AI too much?

The spiritual / religious dimensions

- Can AI interact, or even maybe compete, with human spiritual dimension?
- Some aspects of human spirituality can be approached in neurobiological terms. What does this mean for spirituality?
- What is the place of theological resources when exploring what it means to be human?

Understanding similarities and differences between humans, other animals and AI machines is crucial in an exploration of what it means to be human. It is clear that there is no all-purposes and ready-made set of criteria to fulfill to count as human being. There are many ways to be human. We can wonder whether Australopithecus was human, if he behaved as a human, based on biological/behavioral/cultural (etc.) aspects. Or do humans only emerge when individual from the Homo genus began resembling more or less to nowadays Homo sapiens. We can say that belonging to Homo Sapiens, and bearing its minimal traits, is indeed a minimally necessary to be considered human. Which are these minimal traits is however subject to debate, and is probably highly dependent on biological origin (evolution) and behavioral domain. It is also interesting to note that invoking traditional dualism as the ground for a demarcation between human and non-human (with, for instance, only humans possessing immaterial mind and soul) could prove question-begging or idle. How can we assess the presence or absence of an immaterial mind if not by comparing behaviors, by looking at features of organisms. As we evoked just above, appealing to dualism could also open the door to *strong AI* positions arguing in favor of the dilution of any demarcation between humans and AI based on the idea that only the immaterial and abstract structure of computation is what matters for having a mind.

This being said, lacking a clear and a firm consensus on a set of criteria allowing for an all-purposes demarcation between human and non-human may not be a genuine problem. It could even be a good thing. In any case, it does not prevent from highlighting certain specificities of humans, in particular traits humans possess and machines don't. **AI and Human Intelligence are not the same thing and should not be confused or equated**. We cannot use the same models for all things. For instance, programs or algorithms cannot reflect upon (their) meaning, they cannot *understand*, abilities or faculties that are typical for humans. The meaning (including the ones of algorithms and programs) comes from humans. In particular, machines do not *understand* or reflect upon the *meaning* of interpersonal relationships although this dimension is key for humans who are social animals, and mostly interdependent for their survival. In the same vein, although AI can manipulate data about persons' psychological and spiritual life, it cannot understand such dimensions as a person would.

Although the natural and the artificial, human intelligence and AI, can be clearly distinguished on very important aspects, it is also true that they often intersect in important and complex ways. More broadly, technology can never be totally separated from human nature. Some effects of technology on humans may be considered as modifying human nature (one can for instance wonder about the influence of the extensive use of screens and smartphones upon intellectual capacities or, less negatively, about the meaning of connecting a large fraction of humanity through digital networks). Moreover, humans can willingly use technology on themselves to induce changes at very fundamental levels (with NS and technological intervention on brains and nervous systems, but also for instance with genetics and nanoscience). One can legitimately wonder to which extent such *anthropotechnè* implies modification of human condition or nature (Jonas 1974, chapters 1 and 3).

Technology and human nature also interact with "technification," i. e. the delegation of human functions into machines. AI can be thought as the culmination of technification. The extent of hybridization with and delegation to AI systems thus become a central question. It is crucial to determine what can or should be delegated to machines and what necessitates human faculties or specificities. The ability to *understand* and reflect upon *meaning* is a key factor (especially when it comes to interpersonal relationships or psychological and spiritual live).[3] Artificial intelligence without reflection is "automated stupidity" (Nicholas Carr). When some say that all can be automated, should not we consider refusing automating certain tasks or dimensions requesting *understanding*, such as in philosophy, ethics or politics?[4] This is particularly relevant at a time of repeated demands for *ethical* or *trustworthy* AI (Moor 2006). One cannot expect reaching such goals without thick human inputs of what it means to be *ethical* or *trustworthy*. AI can only work with rules and algorithms – not relationships. So, to the popular question if AI can embody virtue, it should be answered by considering that it can only, at best, *simulate* virtue, as it simulates intelligence, through the use of rules. It can at best *reproduce* the right behavior in the right context.

In general, one should pay attention to properly distinguish between to different meanings commonly attributed to the term *AI*. Any amalgam between each generates a confusion in the debate. The meaning that comes first is that AI is the sign of a technological revolution in progress, that of the digital. The other meaning is a philosophical, naturalistic meaning, which is announced in the very name of artificial intelligence, declaring an artifact that would reproduce, possibly in a more efficient way, all the faculties of the human mind. Regarding the technical meaning of AI, being human requires as we evoked more than ever the human thought and its ethical resources for the promotion of the best of the human. As for the philosophical sense of AI, which announces our obsolescence, it reaches in us the very sense of the human, it paralyses the thought and its ethical resources. There is thus the risk of an implacable and dangerous contradiction with the word AI. In its technical reality, AI requires the human and its ethical genius, but the antihumanist sense that AI conveys with the first one, comes to prohibit the deployment of the ethical resources, that the implementation of AI requires.

[3] What neuroscience, psychology, and other disciplines have shown is that it lies in the biological constitution of humans to behave socially and depend on social interaction for the normal development of cognitive emotional affective functions and behavior. It has been shown that isolation (in the sense of absence of human-human interaction) has terrible effects on human behavior and mind (and many other mammals). At a high degree of occurrence, human-machine interactions can be viewed as interfering with this natural social process of human-human (this also includes human interactions with the natural environment) which is beneficial for human overall health.

[4] Shannon Vallor, "The Thoughts the Civilized Keep," Noema, 2 February 2021. https://www.noemamag.com/the-thoughts-the-civilized-keep/

III. Expert insights on "Impact of AI/NS on human beings and human lives."

Synthesis of co-constructed questions

What is the relationship between humanity and technology?

- What should we do among the possibilities of hybridization between humans and technology?
  - Is artificial intelligence more than a tool or should we better call it augmented intelligence?
  - Does such hybridization leave some room open for other dimensions beyond the biophysical one?
  - What are the acceptable tradeoffs and the limits?
- If we determine "what it means to be human" and apply it to other forms of intelligence, e.g., biological, artificial, engineered, even extraterrestrial, how might these categories stress test our definition? Does it change in the short term vs. the very long term?

What should be the goals of neuroscience and AI?

- What is an advance in neuroscience, AI and neurotechnology?
- Could some development in these fields lead humans to lose a port of their identity or of their nature? (Work automation? Technological integration?...)

What should be the ethical criteria for the proper use of neuroscience and AI?

- What is the place of AI systems within the network of human relationships? How humans should relate to machines? What should they delegate to them?
- How should we arbitrate between the efficiency of high-tech products (such as AI systems) and environmental worries?

What ethics for the proper use of NS and AI?

A first general principle that deserves highlighting is the idea of distributing regulatory efforts according to the level of risk: high risk systems require more attention.

Concerning issues raised by NS, there is a branch of bioethics called neuroethics that has been specifically focused on NS, and that has developed due to a critical level of self-awareness by members of society on the impact of NS on different dimensions of human life and society. Current mayor research initiatives world-wide include neuroethic commitees that question, evaluate and stir the ongoing and future research in the field (Amadio et al, 2018). Neuroethics is not the same as humanism, neither one can be reduced to the other. However, neuroethics questions use in their background a reference to (explicit and implicit) human values which allow to draw a line between what actions (or thoughts?) are acceptable and which are not for human society and individual human life. In that respect, many questions and issues that are debated in the context of neuroethics are of concern in the process of drawing a map of modern-day humanism.

Interestingly, neuroethics is characterized by a direct involvement of scientists themselves in ethical reflection. This is a more and more recognized aspect of technology ethics in general. Ethical labor cannot be confined to a *a posteriori* reflection on the uses that could be made of ready-made technological innovations. Ethical also needs to be conducted at the research and development stage. And that limitation cannot be circumvented by asking to scientists and engineers to do the job alone. Ethics of new technology raises most of the time questions that have entangled societal and scientific or technical dimensions, and thereby call for more inclusive effort of exploration with scientific and engineers in addition to societal stakeholders. The same applies to AI ethics as, for instance, developed in Virginia Dignum's book *Responsible AI* (20??). Dignum insists on the need for bringing ethical reflection (with stakeholders) at the level of AI development with *Ethics for design* (concerning the level of ethics in organizations developping AI systems, with questions of research intergity, inclusiveness and gender equality ...), *Ethics in design* (focused upon the ethical consequences of tools developed) and *Ethics by design* (interested in providing AI systems with abilities of ethical reasoning ... which can be required in some cases to satisfy ethics in design objectives). At all these levels, it is important to enlarge the participation to the effort of ethcial reflection. Although division of labor is a key characteristics of our societies, a too pronounced division of ethical labor could prove damaging.

One can in particular wonder about the public / private division of ethical labor. AI is often developed in corporate settings, in which ethical concerns are not necessarily ranked as top priorities. Ethics can nonetheless become an important requirements for corporations that care about their reputation or that develop a culture more oriented towards thinking about their impact on society. Big tech companies suc as IBM or Microsoft

developed sets of ethical principles for AI development: Fairness, Reliability & Safety, Privacy & Security, Inclusiveness , Transparency, Accountability (Microsoft); Explainability, Fairness, Robustness, Transparency, Privacy (IBM). Both companies then took their principles and figured out ways to implement them in their corporate contexts through procedural tools and technical toolkits. It is interesting to remark that these sets of principles coming from corporate AI ethics are similar to those put forward by public institutions. UNESCO highlights the four main sets of values focused on Human rights and Dignity, Environmental flourishung, Diversity and Inclusiveness, and Peace and Social Justice.[5] In the same vein, the EU High Level Expert Group on AI put forward the need for an AI that should be lawful, ethical and robust and comply with seven requirements: Human agency and oversight, Technical Robustness and safety, Privacy and data governance, Transparency, Diversity, non-discrimination and fairness, Societal and environmental well-being, Accountability.[6] Despite the variations in formulation all these sets of ethical criteria largely overlaps. They remain at a rather commonsensical level where consensus is easily achieved. Implementation and enforcement of these principles and guidelines however prove far more complex and challening.

It seems that a reflection on humanism, on who we are and who we should be or become as humans being at the time of AI and NS can fruitfully complement these efforts and contribute to elucidate a context, to open a space where abstract principles and guidelines can find a more effective implementation.

IV. Expert insights on "NS and AI: risks and opportunities for human beings."

Synthesis of co-constructed questions

Risks

- Does the development of AI and NS risk depriving humanity of its agency and autonomy, of its ability to slow down or even stop what we have initiated?
- How to escape AI main limitations and dangers (inequalities, exclusion, discriminations, …)?
- Don't these technological developments lead to the possibility of the disappearance of "natural" humanity?
- In the digital age, personal identification can easily be concealed or falsely constructed. Will this inadvertently change self-identity of a person?
- In the future, in case AI systems become fully autonomous, should we change our manner to consider them? Would it be morally acceptable to use such systems as tools?

Opportunities

- How could AI and NS help us deepening our understanding of ourselves and our humanity? What are the important learnings from NS or AI upon what humans are? Can we learn from our interactions or confrontations with AI systems?
- To what end do we manufacture AI that can compensate for certain human "defects", or even exceed certain human performances (which all kinds of machines already do)? What values do we associate with such technical developments?

Challenges

- How can we make a reasonable assessment of the risks associated with the development of NS and AI when the developments in these disciplines are so recent, and given that consequences are unknown or uncertain and can emerge after long time periods?
- "Will AI be only for an elite population or for everyone?"
- Should data for training AI programs be a private good or a common good? Is there not something of a common legacy?
- Rather than miniaturizing ever more to make AI discreet, practically invisible, should we not on the contrary make it as visible as possible and highlight it as much as possible? This is to promote the educability of people who would live with these technologies.

Neuroethics addresses different issues, referable to two large groups of issues:

I. **Issues concerning consequences and motivations of direct and indirect interaction with the brain and its impact on global integrative aspects of the human mind and behavior** (neuroimaging, brain interventions, BCI, personal mind-behavior enhancement through

[5] UNESCO (2021). *Recommendation on the Ethics of Artificial Intelligence*, https://unesdoc.unesco.org/ark:/48223/pf0000381137
[6] High-Level Expert Group on AI (2019). *Ethics Guidelines for Trustworthy Artificial Intelligence*, https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

neurotechnological induced neuromodulation, memory alteration, placebo therapy, behavioral addiction mechanisms, neuromarketing). We can call it ethics about certain praxis related to NS.

II. **Issues concerning the knowledge and understanding about ourselves and the world as a result of NS research** (such as mind being a result of properties of living matter, non-dualistic mind-body relation, disorders of consciousness, concept of free will, neurodeterminism and personal responsibility, personal control mechanisms, etc.), **and how it will impact our decisions and actions, and society's evolution in the future**. These ethical issues address how certain new action possibilities, often mediated by technological developments, fostered by new knowledge on how brain-behavior relation is conceived, might affect individuals and society as a whole well beyond the field of research of NS. For example, because NS fundamental research has shown a causal link between certain brain activity patterns and memory performance, some private companies decide to develop technological devices or chemical substances that can modify brain function to increase the performance in memory utilization (cognitive enhancement). Then, what will be the impact on the quality and value of human life, the way humans conceive their behavior and how will it condition social relations?

<u>On the side of AI ethics, several risks can be highlighted:</u>

- AI is a risky form of power because it automates thinking and action, thereby increasing efficiency, efficacy, and speed in ways that can be large scale, long term, and irreversible. Humans should maintain control over AI. This should involve both ethical and political control.
- AI can be used as a form of mental, economic, and political control, perhaps analogous to colonialism, through control of information (or misinformation) and education, control of economic activities (jobs, capital, labor), and influence over democracy (voting, information flow, etc.).
- AI will have economic impacts that are very significant and are not well-understood. The dynamic between capital and labor will shift significantly in favor of capital and away from labor.
    - AI can be construed as a natural consequence of capitalism's logic of calculation, commodification, and cost saving.
    - AI represents a powerful trend towards increasing inequality not only in the Western world, but globally.

More broadly, there are a lot of issues in AI ethics, here are 16: 1) Technical Safety (failure, hacking, etc.), 2) Transparency and Privacy, 3) Beneficial Use, 4) Malicious Use, 5) Bias, 6) Unemployment, 7) Growing Socio-Economic Inequality, 8) Environmental Effects, 9) Automating Ethics, 10) Moral De-Skilling & Debility, 11) Robot Consciousness and Rights, 12) AGI and Superintelligence, 13) Dependency, 14) AI addiction, 15) Isolation and Loneliness, 16) Effects on the Human Spirit.

This last point is of particular interest in the framework of NHNAI. AI successes, combined with the confusion in meanings evoked above, put us at risk of distorting our human self-perception into a mere mathematical and reductionist perspective, thus reducing human dignity, purpose, complexity and wholeness. In a sense, this concern leads to warn against the manner we consider the place of AI and what we should delegate to it or not. We first need to pay attention to what is core to our human nature. If we tend to view ourselves as computers, then we risk to settle the debate about automation and human obsolescence in a distorted framework. Of course computers can be better computers that humans.

Additional challenges linked to AI development can be mentioned:

- Information quality: data science and AI algorithms could help us preserve or enhance the quality of information persons are provided with (moderation on social networks, image analysis to detect manipulation, ...). However, such techniques could also be diverted to serve less acceptable purposes (manipulation, propaganda, populations control, ...). Ensuring proper use of such AI techniques may prove extremely challenging.
- Privacy protection at a time of data "massification": there is a danger in the "massification" of data. What may seem insignificant when isolated is not insignificant when in aggregate. Harnessing the power of data science and AI would push toward more and more data massification, which can conflict with the objective of privacy protection.

- Environmental concerns: in the same vein, massifying more and more data collection and digitalization, despite serving the goals of AI efficiency or robustness, may conflict with ecological objectives (resources consumption, contribution to global warming ...).
- Diversity and inclusiveness: with AI we have at our disposal very efficient means to do our work. But an important dilemma is how to incorporate technology and yet retain our pluralism and diversity, our customs and cultures, and the huge differences among them. Natural language processing focuses on popular language and may thereby contribute to the loss of native ones.

V. Expert insights on "The human responsibility."

Synthesis of co-constructed questions

What becomes human responsibility at the time of NS and AI?

- Could machines bear moral / ethical / legal responsibility, have rights and duties? Doesn't the developer and/or the company and/or the user that "employs" it always share some responsibility?
- Where does human responsibility stand when more and more tasks are delegated to AI systems and when distances between humans constantly increase because of technology mediation?
- What is the responsibility of the scientist in the misuse of what he produces technologically?
- When AI-related technologies become a commercial product, should the possible impacts on the consumers be considered and incorporated into the product design?

VI. Additional questions for possible expert insights on Humanism, AI and Neurosciences.

Synthesis of co-constructed questions

What are the important learnings from NS or AI upon what humans are? Are there some limitations?

- Do thought and mind precede the appearance of the brain or do they follow the development of the latter? Should we at all costs trust the current dominating materialist monism?
- Do you think that it is possible to know the entire human mind (philosophical question)?
- Aren't neurosciences influenced by current psychological theories?
- Is the long-term effect of interventions on the human brain sufficiently considered, knowing that it is over several generations that it should be evaluated?
- Is intelligence, as AI envisages it, not always a logical-mathematical intelligence reduced to its usefulness, and which must therefore only function?
- What does it mean to be "intelligent"? Is there, or should there be - or not be - a definition of "intelligence" broad enough to include both organic ("human") and artificial intelligence? Or, where do respective definitions overlap and not overlap?

What are the limitations and dangers of AI and digital technologies (inequalities, discriminations, ...)?

- How AI can reinforce poverty, exclusion and dehumanization?
- How should / will AI affect the relationship between humans and the place they live (their spatiotemporal, contextual and cultural location)?
- Don't some studies show that exposure to screens decreases empathy?
- Is AI not lacking in intuition, creativity, and power of abstraction only to be able to claim to compete with humans ("artificial intelligence")?
- Is it sufficiently recognized that consciousness and language are irreproducible? Is not language insofar as it induces a perception of the world impossible to implement in a machine?
- Can a machine that learns on its own already be associated with humans? Isn't the latter much more complex than that?
- Are we exaggerating when we announce the advent of a strong AI at short notice?

Education.

I. Expert insights on "Why do we educate?"

Synthesis of co-constructed questions

Objectives and expectations

- What is this 'human' education should help flourish?
- How does one become human? How to foster comprehensive / integral development?
- Can happiness be taught?
- What are the basic developmental needs of the child?
- What is the place of cognition, knowledge acquisition and information processing in human nature, in what it means to be human, especially in the educational context?
- How can we prevent humans to be reduced to their cognitive functioning (e.g., avoid labelling students as low or high achievers)?
- What expectations are reasonable regarding human cognitive functioning (e.g., for students)?
- If AI becomes able to do everything that humans can do, and then proceeds to take care of all of our worries for us (a very theoretical concern indeed) should humans bother to educate ourselves anymore, and if so why? If not, why not? And if we are not being educated, what should we do instead?

Strategies and paths to educate

- On what condition(s) can an education (i.e., the development of a free, critical mind) take place?
- What does a child need to develop as human being?
- What is the role, if any, of the reward system in knowledge acquisition and processing?
- Besides cognitive functioning, what human dimensions, values, and skills (such as critical thinking) should be valued in educational contexts? Is transcendence and spiritual aspect fundamental to be human? How to foster a comprehensive development of children and students? Can NS help? Could they threaten or impede it?
- How can humans enhance cognitive functioning without losing their humanity? Do our innate characteristics define us more as humans than the characteristics we purposefully create?
- Which are the differences between AI and human beings and how can we teach them?

Objectives and expectations.

Education, insofar as it accompanies growth and permits the promotion of the human person, is necessarily part of a humanist perspective.

- In our 21rst century, there is an urgent need to reconsider the nature and role of education to face major societal challenges (such as climate change, technological revolution, poverty…). Without undermining the importance of the core components of education (learning to read and write, to count, mastering basic scientific and technical knowledge) or negating the necessity to teach persons the knowledge and skills they will need in their future career, more and more voices call for a new paradigm with the integration of other dimensions (such as relational and emotional skills, artistic aspects, …).
    - UNESCO's report on The futures of Education (2021) invites us to think about education as a new social contract that leads toward more just and sustainable worlds, notably by strengthen human relations, solidarity, cooperation, interculturality and interdisciplinarity and by seeking to develop human well-being.
    - OCDE's report "Trends shaping education" (2022) highlights: "Examining the future of education in the context of major economic, political social and technological trends is necessary for education to support individuals to develop as persons, citizens and professionals."
    - In his encyclical "Laudato Si'" (2015), Pope Francis argues that education will be ineffective and its efforts will be in vain, if it does not also seek to disseminate a new paradigm regarding the human being, life, society, and the relationship with nature.
- Accordingly, it seems that education is tightly associated with humanism, with what it means to be a genuine human being and how do humans become a genuine moral person. One of the goals of education is to grant the opportunity to each individual to flourish as a genuine human person. In this perspective, it is very important to consider the topic of social inequalities in education as well as examine the manner the economic systems shapes educational systems.

Strategies and paths to educate.

Education implies knowledge acquisition (theoretical (incl. moral) and practical), but not as a passive process. Knowledge acquisition should come with the develoment of critical thinking. The individual should be taught to be an actor of his state-of-knowledge. Education should help flourish a process of knowledge acquisition and know-how (practical knowledge) against a state of ignorance of human beings, ignorance includes not knowing but also possessing false knowledge, e.g. that is contradicted by scientific knowledge (such as the belief that Earth is flat.

It is crucial to conduct a careful reflection upon the various components that should constitute knowledge corpuses transmitted in education, as well as upon the relative importance granted to these components (notably as reflected in the coefficients attributed to them in graduation cursus). For instance, one could mention:

- The relative importance of scientific disciplines over humanities and arts.
- The place of relational, emotional or even spiritual dimensions

Contemporary NS highlights the importance of these multiple dimensions. Though the brain is a key seat of memory and learning processes, the "information processing" heuristic view of the brain, inherited from the cognitive revolution, is somewhat limited nowadays, as the nervous system is considered an integral part of the body, and depends on its self-organized life properties at all times. For example, nutrition and sleep are two factors that affect the body and brain development and functions. Emotional experience through human interaction (family, friends, others) also affects brain functions, and deeply affects learning. These defining features of brains go well beyond the computer-analogy of the brain. In line with these new views on brain functioning we can mention the lack of proper social and emotional interaction and experience seriously affecting brain plasticity by altering negatively neuronal and whole brain functioning.

Therfore it is key to determine how to articulate the various dimensions that may be legitimate in education according to a more humanist paradigm (traditional skills and corpuses of knowledge, but also critical thinking, creative mindset, personal development abilities, human and spiritual dimensions ...).

II. Expert insights on "NS and AI in education? Roles, opportunities, challenges?"

Synthesis of co-constructed questions

What is the place of technology in the training and development of humans?

- As we increasingly delegate typically human tasks to IA machines, how can we prevent it from constraining our growth as humans?
- How is technology also linked to the atrophy of certain cognitive abilities?
- How does technology dispossess man of technical knowledge?
- What is the influence of technology on the manner with allocate our attention?
- Are virtual communities (in social networks) more important than physical communities (in real life)?

How AI and NS impact educational processes?

- How can AI help with education? Has AI made some aspects of learning easier and authentic?
- How NS knowledge and tools can help improve education?
- AI relies on formal and mathematical computation: what place could digital teaching leave to other skills and processes that are essential to humans (spontaneity and relationship to the unexpected, imagination, emotional investment, etc.)?
- Does the development of AI and NS not risk leading to a dehumanization of teaching as well as reducing the diversity of sources?
- Isn't social control of teachers and students inevitable with digital education? How to preserve autonomy and privacy in education at the age of AI and NS?

AI and the place of human educators and human relationship?

- What in education could be delegated to automata?
    - What educational functions are inherently human?
    - Could AI be better teacher than humans in some respects?
- What is the place of genuine human relationship in education, of personal experience sharing (to illustrate or facilitate understanding for instance)?

- What kind of relation between human and machine would be desirable at school?
- Will not a replacement of human contact by the machine lead to relational impoverishment?

Influence of AI and NS on solidarity and care for the most vulnerable in education?

- How can NS and AI help to rethink "human frailty" in schools and universities?
- How to deal with the risk of educational and social inequalities linked to difficulties to access to AI and technological tools for excluded persons?
- How can AI technologies contribute to the assistance and the inclusion of students with disabilities in schools and universities?
- Can NS studies allow more accurate diagnoses and more efficient care for pupils with (cognitive disabilities (SLD (or DSA) and SEN pathologies)? Should pathologies such as ADHD and Dys syndromes be seen as deficits or particular cognitive functions?
- What is the place of classificatory systems such as the ICF-CY (International Classification of Functioning, Disability and Health for Children and Youth)?

What to do with technologies for cognitive improvement?

- What motivates students to "artificially" (e.g., by using neuropharmaceuticals) improve their cognitive functioning? What are the harms of overstimulating the brain? How can we prevent harmful uses?
- Will focusing on improving the human intelligence through neuro-interventions fail to recognize the holistic aspect of being human?

Can / should we use AI and NS to enhance human rationality, to support humans in their quest for truth or knowledge?

- Can AI, assuming it is sufficiently efficient and neutral, provide additional training toward critical thinking and enlightenment?
- Can AI assist in improving of the quality of data retrieved on internet?
- Could neuroscience not help us to better understand how the brain can develop without being too influenced by the external environment (to become sufficiently neutral and objective)?

Who should possess data and AI programs used in the educational context? Is it acceptable to re-use such data for other purposes?

- Who should possess data linked to digital education and AI programs that are designed based on these data? A human legacy? A common good?
- Will digital education not lead to the risk of exploitation of data by technological companies or governments for economic or control purposes?

AI and Education Technology (EdTech)

In the last decades, the impact of digital innovation on higher education has grown. Among the many, we can mention here two EdTech-reliant practices by way of illustration of the ethical questions that can be raised by the use of technology in higher education. These practices are, first, the metric-based approach to learning, and second, customized learning.

In line with the digital innovation shaping our societies, emerging technologies are spawning the rapid development of AI-aided and cloud-based technology. As a consequence, universities began to introduce data collection and processing about a decade ago. One of the main purposes of this growing practice is to put in place and expand systems of measurements for learning outcomes, with the aim of quantifying the benefits of education to students. In doing that, universities broadly mimic the methods used by corporate actors in consumer markets for the purpose of assessing the performance of products and services in terms of customer satisfaction, of employees in terms of productivity, of companies in terms of meeting sales objectives, etc. This practice enables universities to repackage the prospect of a degree program into a good whose benefits can be scientifically measured and assessed, with the double advantage of providing a quantitative account of student success and consequently of institutional success.

EdTech plays a role as a prime agent in accelerating the shift of universities and colleges from a focus on general education, the transmission of knowledge, the cultivation of a critical mind and of guided reflection on the broader dimensions of the human condition, to a concentration on employable skills that is threatening to turn universities into little more than vocational schools in the near future.

Most important in this endeavor is the imperative of climbing further up in the university ranking systems that have become ubiquitous in the sector over the past decade and increasingly guide the policy choices of universities across the world. In Australia, North America, the United Kingdom and other European countries, the use of learning metrics is now well entrenched. Yet many studies have already showed that the metric-

based approach, which is also increasingly used in secondary education, has a minimal net effect on student learning, and makes the career prospects and evolution of faculty significantly more dependent on student evaluations, just as the fate of businesses has become increasingly dependent on customer reviews. Overall, this practice raises the question of whether the holistic benefit of higher education does indeed lend itself to quantification, thus yielding metrics that can be used as a sales pitch. Additionally, critics have argued that the focus on a quantified learning experience prompts the entire ecosystem to shift the blame for student failure, when it occurs, unto the university, thus suppressing more adequate ways to remedy the difficulties that students with adverse socio-economic backgrounds may encounter when struggling with college-level work.

One product of the use of the metric-based approach to learning has been the rise of technology-based customized learning, prompting a growing number of institutions to use learning data to produce learning analytics. Learning analytics gives an institution the ability to assess a student's competency in the different areas of a curriculum, and subsequently to adapt learning methods and curricula to fit every student's ability to move through a learning program at a customized pace. Some systems go so far as to track student mobility on campus such as trips to cafeteria, the gym, the library, etc., in order to flesh out each student's profile as a learner, and the technology used sometimes also includes prediction (Predictive Analytics), which enables the institution to identify potential trouble spots and intervene proactively in order to boost the student's performance. A customized curriculum is created in the process for each student, as he or she moves along through the study program. These customizing methods mimic the ways in which tech giants collect and process data from internet users in order to customize the advertising that they are then exposed to.

As more and more data are collected and analyzed, learning analytics and other AI tools within the educational process raise the ethical question of privacy and autonomy. Some digital tools can even be able to identify the learner's psychological state of mind, by analyzing his time response, the way he moves his cursor…etc. According to Jocelyn Maclure (Champagne, 2018), more than talking about a right to privacy, we should talk about "the right of interiority". Indeed, are not thoughts and mental states the inner, private and intimate world of human? How can we preserve such intimacy in education at the age of AI from a social and economic control these collected data could lead?

The EdTech industry justifies the use of customized learning by arguing that automated ability and learning difficulties assessments lead to better outcomes for each student, and promotes learning analytics as a way to reduce or even close achievement gaps among students of increasingly diverse socio-economic and ethnic backgrounds. The overall claim is that more students can therefore perform well at university. This of course increases retention and graduation rates and enables universities to offer enrolled students a virtual guarantee of success, grounded in fact in the overall lowering of assessment standards (of which grade inflation is an example). Insofar as the penetration of EdTech in the higher-education sector is a byproduct of the commoditization of higher education, the commoditization process that has unfolded over the past four decades should itself be the first ethical issue to examine critically. Who do EdTech tools serve? Are they really the real needs of students or the needs created and defined by an economic market logic?

A more focused dimension of EdTech that should be debated is the industry's strategy (and, very often, selling argument) that EdTech makes learning increasingly easy and less constraining for the customer-student, with the growing introduction of game-like activities in the pedagogical process aimed at drawing in virtually anyone who desires to go to university regardless of whether that is the best option for them, or whether they do have the ability to thrive in higher education, though the latter was designed as only one of the possible paths toward adult life.

More to the point, the transformative role of an industry that generates enormous profits and uses markedly aggressive marketing techniques as it constantly strives to expand to new markets while unwaveringly purporting to provide a public good and to serve the goals of social justice should constitute the core of a broad ethical debate at society level. This is rendered all the more necessary by the fact that EdTech operates in a branch of activity among the most crucial to all societies: educating young people to become the self-accomplished, mature, responsible adults of tomorrow.

Ultimately, the impact of the EdTech industry on the higher-education sector raises the fundamental question of what education is for. Should higher education be akin to a consumer product whose design and promotion

are guided by the same rationale as any other in the market? Or should it retain at least part of its status as a service to society, most often provided by the public authority, for the broad higher purposes it has historically pursued? This also mean wondering about the vision of the human we deploy in education, to determine its goals. Shall education produce humans provided with the right employable skills (e.g., technical and scientifically) to ensure the economic market? Should it also help persons developing their own personality? Should it accompany them in the growth of a critical and open mind, in the development of their social skills, and their ability to contribute to the citizen life?

AI and the place of human educators and human relationship?

Social interactions are crucial for any human and mostly for young children. The automation of teaching in schools can present the risk of an impoverishment of relationships that can be harmful for pupils and students. The covid-19 period, during which pupils and students were on digital application at home, put an even more acute light on such dangers. Many users of digital remote learning systems reported issues lined to the lack of social interactions. It is the case of users of "Summit Learning," put in place by Brooklyn University and hosted by Facebook to customize learning.[7] It is also the case of Chinese students using the application "Squirrel AI" employed to customized learning with AI.[8] Humans are not just machines processing information, therefore education is not just about transmitting content, where AI might be useful. Education also deals with intentional capacities of learners to significate and understand the world, to develop their own view about the world, their own reflections and their critical mind. Is it not this educational aspect which ask to commit into a significative process that escape to AI functionalities?

It seems thus important to reflect upon the place of AI tools in education and in the learning process. What functions/aspects in education are desirable to automated and what are the functions/aspects in education that should not be automated? What educational functions are inherently human? Will not a learning experience entirely delivered by AI impoverish some cognitive mindsets in addition to social interactions? However, will not AI be a better teacher than humans in some respects? As AI will be more and more used in education and at school (through learning analytics, customized learning with digital tools…), we should ask ourselves about what kind of relation between human and machine would be desirable at school, said differently, what is the place of AI in education.

NS and education

Neuroscience or cognitive science knowledge are increasingly part of education as the comprehension of brain functioning and its causal relation to behavior enable to better understand the learning processes which are part of the educational plan, notably at school. In the 1990, there was an increased interest of educators and policy makers for understanding learning processes in education (OCDE, 2007). For instance, neuroscience studies cognitive and learning processes that are relevant for understanding literacy and numeracy which are fundamental capacities acquired through education.

Currently, the relation between neurosciences or cognitive science and education consists of growing interactions at a careful distance because both communities have understood that immediate use and application from neuroscience to education is rarely possible for now, because of the complexity of real-life context to learning. Importantly, this scientific knowledge contributes to demining false beliefs on the brain-behavior relation, also called neuromyths, which pervade the formation of educators. Neurosciences studies can contribute to pedagogy by opening it to scientific methodological observation and analysis as well as bringing awareness about the brain-body-behavior relation. That is why communication and dialogues between different fields, neuroscience incl, and actors approaching pedagogy and education are of paramount importance nowadays.

- Processes like learning, memory, attention, language, spatial representation and others, are being analyzed from a brain perspective. Yet, neuroscience findings, obtained in a simplified behavioral

[7] Tate, E. (2018, 15 novembre). 'Dear Mr. Zuckerberg': Students Take Summit Learning Protests Directly to Facebook Chief. EdSurge. https://www.edsurge.com/news/2018-11-15-dear-mr-zuckerbergstudents-take-summit-learning-protests-directly-to-facebook-chief
[8] Hao, K. (2019, 2 août). *China has started a grand experiment in AI education. It could reshape how the world learns*. MIT Technology Review. https://www.technologyreview.com/2019/08/02/131198/china-squirrel-has-started-a-grand-experiment-in-ai-education-it-could-reshape-how-the/

context in the laboratory (for heuristically observations), and often linked to neuroimaging data, have a hard time finding real application in the real-life complexity of classrooms, creating some skepticism in some sectors of the field of education. The data is also very complex to be easily accessible for educators.

- Though the brain is a key seat of memory and learning processes, the "information processing" heuristic view of the brain, inherited from the cognitive revolution, is somewhat limited nowadays, as the nervous system is considered an integral part of the body, and depends on its self-organized life properties at all times. For example, nutrition and sleep are two factors that affect the body and brain development and functions. Emotional experience through human interaction (family, friends, others) also affects brain functions, and deeply affects learning. These defining features of brains go well beyond the computer-analogy of the brain. In line with these new views on brain functioning we can mention the lack of proper social and emotional interaction and experience seriously affecting brain plasticity by altering negatively neuronal and whole brain functioning.
- The complexity of neuroscience research results sometimes makes them largely inaccessible for non-specialist audience. This can in some cases lead to the misapplication of neuroscience knowledge. For example, it can lead to exhaustive neuro-labeling for marketing and commercial purposes without thorough scientific guidance. This may be detrimental for education.
- An important contribution of neuroscience for education is that it has shown that our brain is not fixed and molds to experience throughout life, enabling and disabling the development of certain behavioral capacities. Early scientific findings showed how enriched environments modify brain connectivity and neuronal homeostasis, deeply correlating with the organisms' health and behavior. By revealing that neuroplasticity is essentially context-driven, neuroscience has shown that learning capacity is also context-dependent, and that learning capacity of individuals is not solely pre-determined by genetic nor social factors, but results from a complex interaction between them. The fact that the brain is shaped through experience and everyday life highlights the importance and the role of the environment in the human's and child's development.

Manipulating the reward system for educational purposes?

The tendancy toward gamification and the promotion of an effortless learning process coming with EdTech can be connected with knowledge we progressively acquire upon the human brain functioning, that comes from psychology, neurosciences and cognitive science.

Indeed, the reward system plays a crucial role in knowledge acquisition and processing. Learning is an active process that needs motivation. There is no learning if there is no reward associated to it, even implicitly. Anything that motivates someone and stimulates his reward system by triggering the dopaminergic system will have an effect upon his memory and his ability to learn. The notion of reward is not limited to physical stimuli, it can also be a change in understanding that is associated to a reward or pleasure in a theoretical construct, as is the case of grades in evaluations. If having good (bad) grades is viewed as improving (downgrading) social recognition and status, it will constitute a positive (negative) feedback that will motivate (or not) the same behavior. In another stream of thought, curiosity can also constitute intern motivation stimulating dopaminergic system and enhance memory (Gruber, Gelman & Ranganath, 2014).

Nevertheless, stimulation of the reward system can also lead to addictions. By being created according to what psychologists know about the human brain functioning, especially the reward system (e.g., by adding elements of games notably), technological applications are successful and continue to be more and more used, notably by children and adolescents, given the low development of the brain areas linked to rational abilities and, at contrary, the rapid development of the brain areas linked to emotional process. Brain areas associated to emotional process are the first developed in childhood. That is why children and adolescents are more likely to develop addictions.

This discussion can also be extended to the possibility of influence, by using AI and NS, on the human capacity to stop the things and say "no" to what one imposed to us. With AI and NS, one can play with the human free will and the human autonomy.

Thus, it seems important to raise the following questions:

- How can we prevent digital addiction? Has education a role to play?
- How to preserve autonomy in education at the age of NS and AI? How can education foster human free will?

DYS syndroms and educational failure

Another important contribution of neuroscience or education is that it can demystify certain academic failure of children who suffer from a DYS syndrome, by showing that alterations in brain development and brain functioning can affect normal cognitive development, and thus learning and school or academic performance. Thus, neuroscientific knowledge can give access to a scientific understanding of behavior, thought and emotional processes, that could be relevant to improve the understanding of certain academic or learning failures of pupils. That is why, classification of individual traits (such as, DSM or IFC-CY) in medicine lead to discover appropriate scientific laws about certain behaviors, and enable us to make sense of them. Moreover, diagnostics lead also to news ways of acting that can be more suitable for concerned persons (Cigman and Davis, 2009). However, these learner categories (such as Dys syndrome, ADHD…) come with criteria of diagnosis that are complex and not consensual yet. They could lead to the temptation for parents, teachers, and doctors to labelling children and put them into cases, to over-diagnosing and misusing or abusing of such diagnosis. The vice of a diagnosis is the essentialist and reductionist interpretation it might lead to, as putting a scientific word on human's difficulties may constrain his nature, somehow restraining possibilities of change, evolution, and progress.

The manner educational systems handle this type of learner can be highly ethically significant. It can pertain to an approch aimed at accepting differences between humans, sheding some light on particular cognitive functioning. But it can also participate to discrimination and stigmatization, putting people into cases, without acknowledge the complexity of learners' situations and conditions? This question underlies the manner we see the normal and the pathologic in society: should pathologies such as ADHD and Dys syndromes be seen as deficits or particular cognitive functioning? Indeed, the way we act with these persons will differ depending on how we see these pathologies: should we seek to correct these deficits or should we consider them as a way of embody our plurality as human? However, it might be really difficult to bring a clear answer to this question. Can we always consider a pathology as a particular cognitive functioning without seeking to correct or alleviate some symptoms that someone is suffering? Is it enough? In what cases should we see pathology as particular cognitive functioning and in what cases should we see it as something that must be corrected? Or, said in another way, should we see all pathologies as something to correct? Is it the person who should be modified or the environment? Or both?

In addition and according to the particular claim of humanism (defined above), learner categories do not consider the specificity of each person. It seems to be more a tool that help scientists and physicians to better understand symptoms, prescribe the right treatment and reassure patients by putting a word on their differences. But diagnosis tend to uniformize people by aggregating different people in common categories. There is a need to go beyond the diagnosis and being careful to the person in all his human dimensions, in health and also in education, since neurosciences can help teachers and parents to better identify and learn more about the different pathologies that can impact learning in children and students (Cigman and Davis, 2009).

From curing to enhancing?

Persons diagnosed with ADHD are prescribed and consume psychoactive substances (such as, ritalin, adderall …) that alleviate their symptoms. But these drugs also have effects on cognitive performance by enhancing some cognitive abilities (such as attention, concentration…). That is why a lot of students consume such drugs, even if they don't have ADHD, to enhance their cognitive performances during exams. This phenomenon is particularly developed in USA. It manifests the pressure exerted on students by schools and colleges where performance and success are a priority. Therefore, it seems important to ask ourselves: do we want to make education a path to success at all costs? By risking drug abuse and addiction?

III. Expert insights on "What is the place of knowledge and literacy about NS and AI in education?"

Synthesis of co-constructed questions

What information and knowledge about NS and AI should be better shared?

- What are the most important elements of AI and NS to explain to the public? Does these elements change depending upon the types of population (teachers, children, parents)?
    - Are we exaggerating when we announce the advent of a strong AI at short notice?
    - Shouldn't we make it so that every individual can read a code?

How to foster NS and AI literacy?

- How can digital literacy be fostered in a constantly changing world?
- Can the current curriculum of education provide the digital natives with an adequate formation and civic literary to tackle the challenges in the technological and Internet era?
- Should the government take on more responsibility to re-configure the educational programs in order to foster youngsters who can think independently in the AI world?

Though direct neuroscience applications in the field of education are treading carefully, educators should be aware of the most fundamental knowledge of the link between neuroscience, psychology and learning. By having access to grounded knowledge base on how the brain is involved in cognition, behavior and learning, educators will be correctly informed about neuromyths and will avoid propagating them in their pedagogy, teaching and interactions with the children in the classroom. This "neuroeducation" will also help developing an acute sense of behavioral analysis of child behavior, using more scientific reasoning when tackling a problem instead of using a "known myth" or mysterious theoretical constructs. But teachers are not scientists, neurosciences or cognitive science knowledge should thus be brought in addition with other types of knowledge from other fields approaching pedagogy and education. That is why, it seems important to reflect on the possibility to foster collaboration between these fields, the manner we are supposed to add these new types of knowledge into the training of teachers and pupils, and the place of neuroscientific knowledge among other kind of knowledge regarding education and pedagogy.

IV. Additional questions for possible expert insights on Humanism, AI and Neurosciences in the field of Education.

Synthesis of co-constructed questions

What is the most significant knowledge we can get from NS studies?

- About the human mind and its learning capabilities?
- Is happiness measurable, quantifiable?
- Are there any limits to neuroscientific diagnosis with young children?
- What are the differences between a human and a superhuman level of cognitive functioning?

What is the place of the human?

- What are the harms of replacing human and in-person relationships with technological activities and/or technology-mediated relationships?
- Since the human brain is permanently influenced by the environment, doesn't technology tend to put more pressure on our system 1 thinking? Living in a society with a highly automated organization mode, doesn't this put humans at risk of functioning more with their system 1, which also risks impoverishing their thinking and reducing their functioning to a machine?

How can AI and NS help with education?

- From a scientific point of view, which scientific knowledge about NS and AI should be necessarily taught to teachers and students?
- Which technological tools can AI provide for a deeper and more efficient learning process in schools?
- How can AI contribute to the child development and how can AI threaten it?

Health.

I. Expert insights on "What are the goals of medicine and healthcare?"

Synthesis of co-constructed questions

How AI and NS (and more broadly new technologies) impact goals and values in medicine and healthcare?

- What are the main ethical and moral values that should be preserved in the field of health? Do NS and AI change the order of priority given to these values? Do they impact patients autonomy and freedom?
- Is the human being only matter (only biophysical)? How to combine scientific and technologic approaches of health with other more humane dimensions (affects, relationships,…)? How to combine the understanding of medicine along the model of the machine to repair and approaches based on maintaining balance or recovering? What is the place of natural healing and alternative medicines?
- In the age of precision medicine (with extremely refined molecular analyses), how to distinguish between normal and pathological conditions? Won't these technological developments transform everyone in an unknowing sick person? To which extent medical prescriptions should enter into persons daily life and constrain their lifestyle?
- Should neuroscience seek to normalize, correct or cure neurodivergence conditions like autism, anxiety, or dyslexia. Etc.?
- Is immortality compatible with humanity? Is it desirable to prevent humans' natural aging process?

What are the challenges with brain implanted devices?

- What are the alterations of or conflicts with what it means to be human because of brain implanted devices or brain-computer interfaces?
- Could cthe implantation of artificial devices in the brain change human nature? Could it affect our freedom and our conception of freedom?

How to deal with the possibilities of human-enhancement? Are there some limits?

- Should there be limits to improving humans' normal level of functioning? What level of functioning transcends humanity?
- What is the point of considering non-strictly medical uses (therapeutic purpose) of technologies based on NS and AI? What would be the point of intervening on an already healthy body?
- How does cognitive enhancement impact our conception of ourselves as humans? Are there some acceptable uses? Do some of them threaten our humanity?

Can / should we try to overcome death through AI and NS?

- The AI and VR technologies can be used to clone or reconstruct a dead person, whose "resurrection" may offer consolation to the family and friends. Despite all the good intentions, can there be negative effects?
- When mind uploading is possible, can the uploaded mind be considered to be a person? How can I distinguish between the COMPUTERIZED me from the REAL me? Can I negate the computerized me, saying that it is NOT-ME? If the cognition of an individual can be preserved with tools of NS, can whatever be preserved represent that individual?
- Is immortality compatible with humanity? Is it desirable to prevent humans' natural aging process?

1. Changes in the goals and values and advent of a bioeconomy

One of the most important evolutions of our time for which concerns our relationship to medicine and health in the context of the development of the new technologies is the increasing emergence of a representation of the body as a capital. As Lafontaine points out, in a world characterized by the cult of perfect health and the biomedicalization of identity, the value attributed to individual life appears to come from individuals' capacity to maintain their biological capital. The "bioeconomy" of which Gary Baker speaks – the ultimate stage of globalized capitalism – fosters a representation of body as capital. So, not only does the bioeconomy model dictate all neoliberal politics in innovation and research, but also redefines the very foundation of collective life. In this sense, maintaining and extending health becomes an investment that increases individual social value.

To better understand the meaning of body as capital and its theoretical and ethical consequences, we can look at two phenomena: the development of private cord blood banks and translational medicine. In the post-genomics era, the identity becomes more biological than ever, to such an extent that, according to Rose, the

conception of human life has totally changed. The molecularised bodies of contemporary biology present themselves as fully malleable, perfectible, and improvable precisely by the technosciences. In that "culture of life" pointed out by Karin Knorr Cetina, the perfectibility of human body through the biotechnologies becomes the ideal of contemporary world. There is a new "biopolitique moléculaire", which does not aim to change the world anymore, but to improve it in order to make it more efficient.

At the basis of the bioeconomy society we can find the regenerative medicine, which perfectly represents the ambition to extend life, to enhance the existence. This new type of medicine does not aim to heal, but to regenerate. Referring to this idea and to the body as capital, in the last years the female body – very fertile in stem cells – is becoming a gold mine for the biomedical industry. Collected and freezed, the cord blood has given rise to two opposing models of banking: one public and the other private. The cord stem cells could eventually allow the treatment of cancer and, in this sense, women who choose to invest to the biological future of their baby, therefore deprive unintentionally the community of this valuable lifesaving resource, that could potentially be beneficial to patients. From this perspective, the women become the preferred target of "biosurance" marketing.

Another interesting phenomenon which sheds light on the meaning of the body as capital is "medical tourism". The cult of perfect health and the desire to maintain, improve, and extend the biological potential of individuals have fostered the realization of a huge global health care market. According to Song, pilgrims of biotechnologies come to occupy an interesting double position within the bioeconomy: they are at the same time consumers and subjects for biomedical innovation. It is in this sense, then, that the regenerative medicine should be considered the core of the bioeconomy of human body: it represents the double hope of regenerating bodies and revitalizing economy.

One could also mention medically assisted procreation or surrogacy, which, in certain circumstances also undermine the principle excluding the body from the market.

Considering these developments in the field of health at the time of Neuroscience and Artificial Intelligence, at least one question becomes suddenly urgent: does the human body have value only as capital of the bioeconomics and so barely in its material sense?

2. Operations on the human brain and their ethical challenges

There exist currently many fields of scientific activity dealing directly with operations on the human brain, and all of them open some specific ethical issues.

Already developed for many years, neuroimaging is a scientific area which makes use of quantitative techniques to study the structure and function of the human brain in a non-invasive manner. Neuroimaging as such raises a set of ethical issues. According to Racine, they can be summarized under the following categories: 1) *human nature*: risks and concerns that relate to the transformation of human nature; manipulation of people conducive to a view of humans as mere means; misuses; 2) *confidentiality*: risks and concerns that relate to the potential harm that could result from sharing confidential information; 3) *troubling findings*: risks and concerns that relate to the potential harm to individuals or society of findings, of actions based on Magnetic Resonance Imaging findings or of incidental findings; 4) *recruitment*: risks and concerns that relate to the recruitment of research subjects and protection of subjects enrolled in protocols; 5) *knowledge transfer* from basic research to clinical research must be carefully operated (new neuroimaging techniques and methods need to be thoroughly investigated and validated before being incorporated into routine diagnostic procedures); 6) *the difficulty of access* to this expensive brain scanning material by rural populations might generate inequalities regarding access to this health treatment.

Another crucial field concerning NS and health is that of brain interventions, which consist in any interference or perturbation occurred physically upon the brain. Neurosurgery is the most radical form of intervention (legally and ethically allowed) which is often proposed when no other form of treatment is possible. It consists in the damage or removal of brain tissue with the aim of affecting radically a person's behavior. Patients undergoing this type of intervention receive information of the high risks associated to it and are asked to give their consent. Current brain surgery procedures include Deep Brain Stimulation, and ablative neurosurgical procedures (in case of tumor removal, local coagulation for epilepsy, etc.). Deep Brain Stimulation consists of introducing electrodes in the brain (often into basal ganglia structures) and applying rhythmic electric

stimulation, for activating or inhibiting stimulated regions. Deep Brain Stimulation is being used mostly for Parkinson's disease and other movements disorders. It is however also considered with promising results in various neurological and neuropsychiatric disorders, such as depression and Obsessive-Compulsive Disorders.

Pro's and Con's of neurosurgery can be evaluated on three levels: reversibility (adaptability), effectiveness, invasiveness. Some important ethical questions concerning neurosurgery can be mentioned:

- What is the risk/benefit ratio of the different methods?
- Who shall decide how the patient's mood and personality are to be changed and according to which criteria?
- Can the autonomy of patients be respected, in particular if their capacity for autonomous decision-making is affected by their psychiatric disorder or as a consequence of the intervention?
- For which kind of disorders is psychiatric neurosurgery justified?
- Can these procedures be justified in cases of self-inflicted disorders such as drug addiction or anorexia nervosa?
- How should researchers and physicians deal with conflicts of interest, for example, conflicts between research interests and benefits to patients?
- Should psychiatric neurosurgery be applied in forensic contexts, for example, to reduce the risk of re-offense of pedophilic child offenders or violent psychopaths?

Two areas in link with neurosurgery and with particularly important ethical challenges are neuromodulation and memory alteration/enhancement:

Neuromodulation is the goal-directed intervention in the nervous system by altering cellular interactions, which short-cuts this self-interacting system with the body. It can be induced by implantable devises and by pharmacological products. Decisions about how *best* to modulate brain function can be a source of ethical conflict between clinician and patient. Clinicians can feel they do not have to take into account the level of happiness of patients in order to recommend the use of neuromodulatory artefacts. What justifies a clinician abiding by a patient's neuromodulation request or refusing it? Should clinicians invoke notions of clinical integrity or the goals of medicine in refusing patient requests for neuromodulation?

Concerning memory alteration/enhancement, it can be performed in particular through Deep Brain Stimulation of the enthorinal-hippocampal system. Recent advances in NS consider the case of boosting memory capacity and slow memory decline in people with Alzheimer's or epilepsy. The possibility to edit human memory (decoding, enhancing, incepting, or deleting) has been studied for therapeutic reasons (neurodegeneration, Post Traumatic Stress Disorders), but raises considerable ethical concerns, especially agency and identity of self. Our memories, good or bad, are intrinsically part of our identity and make up who we are. Having the possibility to erase the bad ones, for example in the context of Post Traumatic Stress Disorders, could on the one hand improve the patients' life by enabling relief from psychological and emotional pain, but on the other could disturb normal identity and natural resilience processes which feed on memories of painful experiences. What are the main dangers for the individual to have the possibility to eventually manipulate some of their memories (alteration, selective removal)? In the case of trauma, what is the experiential value of bad/painful memories?

A final field concerning the brain is that of brain implants and brain-computer interfaces, where Artificial Intelligence and Neuroscience are converging. Artificial neuromimetic neural networks are currently developed to process neural signals fully automatically through unsupervised learning, in a way which simulates the natural capacities of biological neurons, in particular cerebral plasticity. These achievements will pave the way towards a new generation of autonomous brain implants processing and interpreting automatically brain signals, opening new perspectives in neural rehabilitation in case of severe paralysis or degenerative diseases. In particular, this will make a step further towards seamless brain interfacing, allowing ideal neural prosthesis where artificial devices controlled by a brain-computer interface become fully assimilated as an extension of the body, or as an extension of the mind. Achieving full success with this kind of technology, however, requires continually reducing the distinction between the biological and the electronic, the natural and the artificial, the human and the machine. The more seamless the brain implant is to the human individual, the more efficient the brain implant will be. The development of these technologies of course raises new questions about the singularity of human nature, and about the status, between humans and machines, of the individuals implanted.

3. Human enhancement and its ethical issues

At the time of Neuroscience and Artificial Intelligence, one of the main questions of humanism in the field of health arises considering the transhumanist topic of the endless enhancement of the human being. In the Anglo-American literature the expression "enhancement" refers to every intervention of body and mind alteration compared to the normal physic and psychic functioning. We can find this type of intervention especially in some specific fields: in the biological one, for example, as "gene-editing" and "genetic-enhancement", or even as "cognitive enhancement", or finally as "moral enhancement".

The debate around this topic can be broken down into two sides: the *technophiles* and the *technophobes* (or *bioconservatives*). To the first category belong authors like Julian Savulescu and John Harris, who think that health consists in full psycho, physic, and social wellness. In this perspective, the borders between therapy and enhancement result blurred and three main principles are outlined: 1) biotechnological shortcut (to come to the same result in faster times, accelerating the natural evolution), 2) evolutionary duty, and finally 3) respect for the free market. To the second position belong, instead, Michael Sandel, Jürgen Habermas, and Francis Fukuyama, among others. These authors state that it is incorrect to do a parallelism between drugs and technologies, on the one hand, and education and natural training on the other. In this regard, one question would immediately arise when the boundaries between drugs and natural training disappear: was it really me who reached a certain result? Then, the bioconservatives affirm a so-called "freedom argument", which contests the enhancement as a form of obligation: in a technocratic society, we are obliged to enhance, not free to choose to do it. Finally, there is the "justice argument": due to the possibility or not of enhancement, inequalities significantly increase. Both drugs and technologies are dual-use, therapeutic and enhancing, but there is always the possibility to use this knowledge in the wrong way. Approaches that protect at least the fundamental human rights must be found. For instance, it is fundamental to guarantee security and efficacy, starting from a right experimentation process.

II. Expert insights on "How can NS and AI improve medical and healthcare practices?"

Synthesis of co-constructed questions

- How can IA contribute to lessen the burden of medical professionals?
- What benefits may result from having robot caregivers in nursing homes?
- What AI and NS should bring to contribute centering / focusing healthcare systems upon the human person? How can we reconcile the use of IA in healthcare contexts with the humanization of healthcare?

Artificial Intelligence and Neuroscience have both the ability to improve medical and healthcare practices.

Artificial Intelligence can be used to provide clinical decision-support to the clinician. In the years to come, this use is expected to improve significantly clinical capabilities in terms of diagnosis, determination of therapy, delivery of care, but also drug discovery, epidemiology, personalised medicine, operational efficiency, etc.

Neuroscience as such is the multidisciplinary scientific study of the nervous system, and of the biological correlates of mental and cognitive activities. Its applications to medicine and health encompass disciplines such as neurology, psychiatry, neurosurgery, which address diseases of the nervous system, including their diagnosis and treatment. Neurotechnologies aim at establishing new connection pathways to the brain for the recording or alteration of the neuronal activity, through diverse techniques such as neuroimaging, intracranial electrophysiological monitoring, brain-computer interfaces, neurostimulation, etc. Medical neurotechnologies can help people with neurological or psychiatric conditions, but of course raise at the same time major ethical and legal challenges.

Neuroscience can help in the understanding of the functioning of the human brain and of the animal brain. Today, finding the minimal neuronal mechanism sufficient for any given cognitive ability remains the Holy Grail for many neuroscientists. Nevertheless, it is progressively understood that cerebral functions are not always localized, that there is systemic functioning, and that it is sometimes illusory to try to associate a particular cerebral function with a single particular neuronal mechanism. There is rather a trend to characterise metacognitive functions in a "hodotopic" approach, by considering that cognitive functions are subserved by

distributed, large-scale brain networks. The concept of "multiple realizability" has been also advanced, according to which a single function can be performed by very different cerebral architectures, which means that two brains with the same functions do not necessarily look alike. More generally, contemporary neurobiologists have developed both a "structure change principle", according to which a same function can be exhibited by a variety of different structures, and a "function change principle", according to which a same structure can manifest a variety of different functions.

At term, this kind of researches will probably find applications in the care of mental and nervous pathologies. But current knowledge and experiences on mental and nervous pathologies can also contribute to the understanding of the real functioning of the human brain, and the different paradigms proposed for that. In this matter, the contemporary research tends to state the insufficiency of both the cognitivist and the connectionist paradigms. It pleads rather for an "enactivist" paradigm, according to which in the activity of cognition intervenes a sensory-motor part linked to the activity of the body of the subject, to its interests and to its preservation. In particular, studies of patients with Alzheimer's or Parkinson's disease demonstrate that knowledge in memory is not stored and retrieved, but it is rather recreated, simulated from the sensory-motor representations. The embodiment of the cognition is also proven by the discovery of the "mirror neuron systems", which are activated both when an animal performs a task and when it observes the same action performed by another animal.

Advances in neuroscience and medicine also shed new light on the understanding of altered, diminished, modified states of consciousness, and on the legislations that are developed with regard to them. How are these states of consciousness defined and diagnosed? What defines a conscious person, and how can one know if another person is conscious? Of course from an ethical point of view, having clear diagnosis criteria is crucial in the management and care of patients. For example, we distinguish now between irreversible coma, persistent vegetative state, and severe dementia. A vegetative state as such can lead to two kinds of evolution. Either an improvement, which can lead to a better conscious state called "minimally conscious state", then in some cases to awakening; or a stagnation or deterioration, which can lead to a "persistent vegetative state". It is increasingly possible to appreciate the differences between a persistent vegetative state and a minimally conscious state with the help of functional neuroimaging. Some clinical investigations reveal that, in some medical contexts, up to 40% of patients are wrongly diagnosed as in a persistent vegetative state, and should be rather classified as in a minimally conscious state. The diagnosis of persistent vegetative state is made after a period of observation, the length of which differs between the different states (for example, in case of traumatic brain injury, it is currently twelve months in the United Kingdom and the United States, and six months in Taiwan). The autonomy and dignity of the patients facing these different states are increasingly respected through the worldwide legislations. For example, in Taiwan, the 2019 Patient Right to Autonomy Act allows a patient to decide to renounce to some life-sustaining treatments in case of terminal illness, irreversible coma, persistent vegetative state, severe dementia, or some other disease conditions.

III. Expert insights on "What is and should be the role of human actors in healthcare?"

Synthesis of co-constructed questions

What are the limits of technological tools in healthcare? Can technology (especially AI systems) handle everything?

- What is the place and role of human beings in medicine as an art of healing?
- What is the place of NS knowledge and tools within patient care (diagnosis, cure, …)?
- According to NS studies, is it always possible to treat emotional and psychological disorders by means of drug treatments?
- What in healthcare is inherently human and should not be delegated to AI or robots?
- In the age of precision medicine (with extremely refined molecular analyses) and big data, medical analysis and prescription can enter more and more deeply into persons day-to-day life (tracking and recommendations through connected devices for instance)? Should we put some limits?

How to combine AI input and human autonomy in healthcare decision making?

- Do robot-assisted or AI interventions in medicine alter the autonomy of a human person?
- Shouldn't humans be able to keep their distance from what the machine is telling them?

- How can the conflict of authority between an AI clinical decision support system and a human medical doctor be navigated?
- If we consider that AI knows "what is good for a patient" better than the patient himself, what decision will be taken? What future for consent? For the individual control of our lives?

How NS and AI influence the doctor – patient relationship?

- How to ensure that the gain brought by AI in terms of diagnosis (via the analysis of big data, etc.) is not accompanied by a loss in the relationship between the doctor and the patient (loss empathy, risk of dehumanization of the care relationship, …)?
- Can health data analyzed by a computer replace face-to-face professional consultation with a physician?
- What human characteristics are essential in communicating medical diagnoses and in the healing process?

The evidence argues for a collaboration between humans and technology in healthcare, rather than for a complete replacement of humans by technology. Because the technology itself is not perfect, it has biases, and it also makes mistakes.

The use of Artificial Intelligence (AI) technologies for diagnosis purposes concerning brain pathologies offers a particularly meaningful example of this. The ability of artificial neural networks to identify patterns in large, complex datasets has proved very effective in term of clinical diagnosis, by helping with diseases classification and prediction, including those that arise from brain damage. Given the subtle presentation of symptoms typical of several neurological conditions, clinical neurosciences are particularly well positioned to benefit from the advances of AI and machine learning. Those are able to reply to the current need for early and more accurate detection of clinical patterns of specific brain disorders, with the aim to initiate potential disease-modifying interventions. Such applications already exist, in particular concerning the early detection and prediction of dementia, of Alzheimer's disease, and the classification of primary progressive aphasia. Nevertheless, as for humans, there exist risks of machine limitations and machine errors in these applications, giving rise to misclassifications and misdiagnosis. Machine learning cannot be considered as fully objective and unbiased. Based in particular on association of data, its potential inability to disentangle correlation from causation can be a source of error. Machine learning can also be mistaken by the subjectivity of data used for its training. The differences between machine error and human error should be analysed. The types, implications and consequences of those two genres of errors vary, despite the commonalities. Even if the margin of human error remains vaster than the margin of AI error, there is a recurrent need to enhance and to reflect upon errors detection, with the aim of mitigating their impacts.

The need for a collaboration between humans and technology is also clearly demonstrated by the example of applications of AI to healthcare in non-occidental contexts. In Africa for example, AI-enabled medical chatbots make possible an increased access and connectivity to healthcare providers and actors (including specialists, hospitals, pharmacies, etc.). AI-enabled diagnosis tools assist the doctors in rural areas. AI helps also by improving medical knowledge and emergency healthcare, and by reducing medical cost. These applications open a certain number of ethical issues. Some of them are very general, such as the risk of AI-powered machines misdiagnosis, the AI's bias leading to racist robots, the dehumanization of health care and public health policies, the disrespect for human autonomy, the risk of unemployment and economic inequality. Others are proper to traditional contexts and indigenous knowledge. Indigenous knowledge is defined by UNESCO as referring to the understandings, skills and philosophies developed by societies with long histories of interaction with their natural surroundings. For rural and indigenous people, decision-making about fundamental aspects of day-to-day life tends to be informed by local knowledge. In Africa for example, healthcare is traditionally based on spirituality, meditation, rituals, beliefs, and connected with nature. Additionally, Africa is the place of the notion of Ubuntu solidarity, according to which emphasis must be given on "being self through others". These conceptions are challenged by the development of contemporary AI-powered medicine, in its individualistic and commercial aspects. Rather than as a liberator, AI-powered healthcare can then be seen as a threat, or as an instrument of power.

IV. Expert insights on "What kind of regulations do we need for NS and AI in Healthcare?"

Synthesis of co-constructed questions

What rules do we need?

- What would regulations to protect data from medical brain scanning look like?
- Do we need new human rights in the context of AI/NS and healthcare? Why or why not?
- In AI applications which require massive data and resources, how can we limit its impact on the environment or on human health?
- Should public funds be invested in extending human life as much as possible?
- What should we do with the digital remains of the dead? Should our digital traces be truthful?

What economic model for AI and NS in healthcare?

- Regarding health technology, how can we work for profit without losing our humanity?
- How should we share the benefits health technology could bring? Especially, AI codes from health data shared by patients?

How to assess responsibilities?

- Does an AI-powered robot used in the healthcare have freedom and responsibility for their actions as do human beings?
- How to deal with responsibility in case of AI-assisted / driven health decision making? What happens in case AI systems outputs result in physicians' decisions which harm patients?
- Where does human responsibility stand? Haven't we opened the door to a legal disempowerment of the medical profession and society?

Under the effect of current and future technical developments in the field of Artificial Intelligence (AI) for healthcare, and the upheavals they will inevitably produce, the adoption of a clear legislative and regulatory framework is a need expressed today by many actors. The increasing use of AI in healthcare raises many ethical issues, including data privacy and health equity issues. Of course the healthcare sector as such is already heavily regulated ; and in general, some already existing data privacy laws, as well as industry specific regulations, can be applied to the use of AI devices. Nevertheless, there is not yet an effective jurisdiction of AI in the field of health. Such a regulation is currently complicated by the difficulty of the exercise of anticipation that it presupposes. Faced with the profusion of technology and its uncertainties, some authors call for the promotion of simple principles, such as the reaffirmation of the human being at the center of the healthcare system, the strict control of machines, and the creation of new ethical and regulatory committees. Several institutions around the world have started making proposals for regulation. The proposals demonstrate both the desire to protect the patient and that to not stifle innovation. Proposals for the regulation of AI in the field of health are often inspired by the main ethical principles that regulators have already begun to implement in the field of AI in general: control measures related to human agency and oversight ; robustness ; privacy and data governance ; transparency ; traceability ; explainability ; diversity ; fairness ; societal and environmental well-being ; accountability, etc. Another issue is also to adapt the traditional principles of bioethics to the case of AI: dignity ; autonomy ; social justice ; benevolence. Particular attention must be paid to not going into an inflation or a redundancy of principles. Such a profusion of additional ethical values could be certainly a source of confusion. Among the concrete proposals made by certain authors, we can also mention the labeling by the State of companies authorized to manage medical data, the guarantee of the right to human intervention as a last resort for any medical act, the establishment of ongoing training in technological developments for healthcare personnel, the systematic appointment of a referring doctor who takes on legal responsibility for the medical actions produced by technological artefacts, or even the systematic definition for any medical act involving digital or mechanical prostheses of the border between restoration and improvement.

The case of neurotechnologies shows the same type of regulatory uncertainties. Cerebral implants for medical use are developing rapidly, particularly for the treatment of depression or certain neurological disorders. They are not subject to specific regulation, but in many countries they are subject to the regulations governing medical procedures. Some authors are nevertheless concerned about the inadequacies of this regulation. In particular, they criticise a lack of requirements in the authorisation procedures, the absence of long-term public monitoring of the side effects of implants, and an evaluation that is too dependent on manufacturers' statements. There are also calls for more regulation of brain computer interfaces, where particular concerns arise, such as the risk of external piracy. There are wider concerns and a expressed need for regulation about

the future possibility of non-medical uses of brain implants, for example to enhance cognitive abilities. But regulators are reluctant to anticipate, and are more easily focused on short-term management. It should be noted that, faced with the future applications of neurotechnologies and the challenges of regulating them, some people are calling for the recognition of new human rights described as "neuro-rights", such as the right to freedom of thought and conscience, the right to mental privacy, the right to mental integrity, the right to personal identity.

V. Additional questions for possible expert insights on Humanism, AI and Neurosciences in the field of Health.

Synthesis of co-constructed questions

Are there some studies about the value of the use of AI in healthcare?

- Do recent (psychological and sociological) studies done on patients about the effects of AI in healthcare encourage us to continue on the path of increasing such use, or do they make us cautious about it?

How NS can inform diagnosis and cure of mental diseases?

- Are there better tools or ratings scales developed to determine the degree of dementia and to determine whether the dementia patients are able to assess the value and meaning of their life?
- According to NS studies, is it always possible to treat emotional and psychological disorders by means of drug treatments?
- Should neuroscience seek to normalize, correct or cure neurodivergence conditions like autism, anxiety, or dyslexia. Etc.?

From the NS perspective, no. Autism, anxiety or dyslexia affect humans but are not from the same category of human affections. They all affect human behavior differently but are not to be considered as pathologies. Except maybe in the case of anxiety, which is an emotional state resulting from a stress-response mechanism. Prolonged, it can negatively emotionally, psychologically and healthy-wise affect the individual and thereby its life quality. Finding "natural" ways to reduce chronic stress may help improving life quality and health.

What are the possible uses of AI (and limitations) in health policy and health care?

- What lessons do the pandemic years leave us with about the use of AI in health policy?
- How can technology be used to foster health? How can we foster individual informed decisions regarding health (e.g., regarding the use of cognitive enhancement)?
- Isn't Ivan Illich's thesis according to which some technologies in the field of health sometimes create other pathologies still very topical?

What can we learn from institutions producing ethical reflection / regulation on AI and NS in health care?

- Are National bioethics committees proving capable of responding to the ethical challenges that are arising in health care due to the extensive use of AI and NS?
- Are national and international bodies charged with overseeing ethical issues in the field of health care efficiently operating in the area of NS and AI? What could or should be improved? And at what level (national or international)?

What do we know about the impact of robots in health care and care giving?

- What benefits may result from having robot caregivers in nursing homes? As increasingly more robot caregivers are used, what are the harms of the lack of human contact for people that are ill?

What are the possibilities and limitations or risks with cognitive enhancement?

- What are the risks of prolonged use of cognitive enhancement?

Democracy.

I. Expert insights on "What will be democracy at the times of NS and AI?"

Synthesis of co-constructed questions

How new technologies modify on our idea of democracy?

- What world do we want to live in? With what place for technology? Are our technological choices committing us (or even locking us) in some pathways?
- Do AI and NS change significantly what it means to be an individual in a democracy? Do they influence how individuals in a democracy are viewed and treated by democratic institutions and processes?
- Can NS and AI allow to better assess persons' ability to participate democratic life? (by contrast with age as sole criteria).
- Could we, with the help of technology, transpose at the collective level the model of individual enlightened consent?
- Shouldn't the fact that the machine can replace man in political decision-making be excluded from progress, whatever the limits of the human politician?
- By instrumentalizing what comes under assistance (via devices such as Alexa, Siri, chatbots, etc.), isn't AI inducing an impoverishment of mutual aid and listening relationships between humans? In short, isn't AI invading areas where human relationships are important/crucial?
- Can AI compete with the collective intelligence of all of humanity? Can we imagine that it will one day itself become a collective intelligence? According to what logic? With what ethical implications?

What can be the role of technology in fostering democracy?

- Which uses of AI and NS could help use becoming more human in society, enriching the meaning of citizenship?
- How can neuroscience and AI improve peace between human beings?
- Can new technologies using AI promote exchanges between humans?
- Can AI promote the advent of a more inclusive society, taking into account the most vulnerable?
- How can IA be used to prevent the online dissemination of false information?

Democracy may be conceived both as:

- **a form of political organization**, grounded by political rights. These political rights are based on a notion of citizen participation in the making of laws and public decisions. This participation might be indirect, through elected representatives, or more direct, following models of participative democracy. This notion of political participation also protects citizens' right to manifest or express their political preferences, and to gather into associations and political parties (in the limits raised by the protection of human dignity and individuals' fundamental rights). These rights go with the technical possibility for all citizens to engage in public debates, and to access high-quality and diversified information.
- **a form of social organization**, grounded by fundamental rights (freedom of speech, freedom of association, rights to independent and impartial justice systems), and a notion of human dignity. These ones are dependent on an idea of what it is to be human.

Developments of democracy in both senses are historically fueled by an ideal of social progress in the wide sense. Then, it is crucial for the future of our democracies to interrogate how current developments of AI&NS impact both our conceptions of humanism and/or their practical applications in the social world. Are these applications compatible with both political and fundamental rights associated with democracies ? In the following, we will explore this question by considering the influence of NS and AI on the pillar of democracy, and the possibilities of regulation of NS&AI.

II. Expert insights on "What is the influence of NS and AI on the pillars of democracy?"

Synthesis of co-constructed questions

Influence on citizens' autonomy?

- Does the development of AI and NS risk depriving humanity of its autonomy and freedom? Can persons decide to live outside such technological developments when their deployment is decided at the collective level?

- Can some aspects of democracy be automated or delegated to machines without undermining human persons' centrality?
- How to prevent AI and NS uses tending to manipulate the will of persons (formatting of needs and thinking, especially during elections)? Can some use reinforce persons autonomy?
- What happens to human responsibility in a society where everything is regulated by AI or other artificial devices?
- Do AI technology trigger the risk of impeding, disturbing or inhibiting the expression of the will of the people?
- May AI and NS raise the risk of excessive social control (especially by the state), conforming individuals to the role that has been socially assigned to them and rendering political dissent more difficult?
- Can AI and NS lead to other criteria for assessing decision-making capacity of people in a democracy than the sole criterium of age? Should the level of (mis)information of citizens be taken into account?

Influence on democratic plurality? The risk of impeding pluralism?

- How to deal with the risk that the development of AI and NS leads to impoverishment of social interactions, to standardization of points of view, and to reduction of opportunities for dialogue? Can AI be sensitive to diversity in its various dimensions (political, cultural, …)?
- Does AI and NS contribute to fragmenting and diving society (as for instance with AI use during the Brexit or Trump political campaigns)?
- Can new technologies using AI promote exchanges between humans rather that locking them into closed circles and making it more difficult to have constructive contradictory debates?

Influence on policy making?

- How can we prevent individual data to be harmfully used through technology (e.g., to manipulate elections)?
- Can technology and AI dispossess political leaders of their leadership?

Influence on justice and Law?

- What will be the limits, in terms of privacy, in the construction of judicial truth? Should we use AI profiling based on (personal) data or NS imaging tools (e.g. to assess someone's truthfulness)?
- How could AI assist lawyers and magistrates? What part of their tasks could be fruitfully alleviated by AI systems?
- Could an AI serve on a jury? (assuming it was legal) If "Yes" - would it serve as a jury of one, judging only by data presented in court? And if "No" - why not?

Influence on defense and military matters?

- What type of use could be done of AI in defense domains?
- In military matters, should we use AI not only for tactical matters, but also for human politics, international relations, and war in all its dimensions?

1. Democracy as a political organization

At the European level, the "numerical democracy" is often presented as an opportunity for public deliberation and co-decision. However, various issues can be raised:

a) Algorithmic governance has many faces. For many authors, algorithmisation and datafication makes governance more intrusive and powerful, and at the same time it (paradoxically) dilutes administrations' accountability and responsibility regarding decisions and public action by putting the accent on the 'objectivity' of the algorithmic processes. However, an important point here is to recognize that human agency remains central behind massive data and algorithms. A second point is that of the competencies of public actors: by delegating the epistemic aspects of governance (collecting and interpreting data), the administrations may lose their capacity to analyze social realities.

b) Globalization of exchanges raises a sovereignty issue regarding the collection and use of individuals' data collected in a given territory by governments. For instance, identification and/or profiling systems are developed only by a few number of companies, which de facto will gain an unjustified control over data from foreign countries.

c) Social media and internet platforms constitute political spheres, in the sense that they allow people to design, consume, and share political news, seek political information and discuss, make decisions, donate money, or engage with political parties and other organizations. They may also foster political engagement and democratic participation. On the other hand, the formation of opinion is heavily dependent on the circulation of (false) information, which is now biased by individual targeting allowed by profiling systems. Social platforms

and new media are increasingly perceived as conducive to the creation of ideological "echo-chambers", eroding the space for public dialogue. They are seen as fostering polarization, radicalization, politicization, spreading misinformation and subject to manipulation. At the same time, they have been used in attempts to covertly influence the political choices of citizens, thus sapping their democratic credentials.

d) These numerical political spheres may exclude some citizens who do not have the technical capacities to enter online deliberations. Besides, they make democracies dependent on private companies owning or developing these plateformes. They also make democracies at risk of hacking or manipulations.

2. Democracy as a set of fundamental rights and values

a) AI & NS allow the collection of data which are individual-specific (for instance, NS tools aiming at modifying/improving individuals' neuronal circuits lead to infer individual identity from cerebral activity and structure). Yet, even protected by adapted systems, these data are all at risks of being cracked. The uncontrolled collection and circulation of these personal data interrogate the notion of the right to privacy: how is it possible to regulate the diffusion and use (for instance through predictive analytics tools) of personal data at the age of AI & NS? This issue is traditionally considered under an individualistic approach, which stresses out the right for individuals to control their personal data. Another perspective would adopt a more collective view, which takes as a central point the protection of the human person against the risks arising from the use of their personal information.

b) Similarly, the developments of AI & NS provide government with many new tools to control citizens' actions and opinions. AI tools (such as facial recognition) may be used to generate new biometrics systems, and to create databases of citizens' identities without authorization. These data might be used for individuals' profiling, that is, the processing of personal data to analyze, categorize, and predict factors specific to an individual. This makes possible different levels of manipulation of citizens' opinions and behaviors. This impact is reinforced by the use of neuromarketing as a way to better predict individuals' reactions to a product or a discourse. The collection of data on individuals makes possible the development of a form of algorithmic governance, that is the use of AI to produce social ordering at different levels: predictive policy and public policy evaluation, automated content moderation, rating and scoring of citizens, criminal justice.

c) The impact of AI on labor markets and wage inequality is not easy to assess and predict. In some domains, it is easy to predict the consequences of AI-related innovations: for instance, the development of automated cars will threaten bus and taxi drivers. However, generally speaking, the type of jobs affected is much more diffused and difficult to identify, as well as the type of firms and sectors which are and will be impacted. AI impact strongly depends on local economic context (the rigidity of the job market, for instance), on the specific technology which is considered, and on the internal organization of the firms.. Out of this question of wage inequalities, the growth-based economy which justify the development of AI, as well as the kind of jobs which are available in AI-driven job markets, deserve to be discussed. Under this perspective, one should consider the fact that the energetic needs linked to AI development are in contradiction with the current exigency of environmental transition and energetic negative growth.

III. Expert insights on "What kind of regulation do we need?"

Synthesis of co-constructed questions

How to regulate AI and NS? What can be the role of states and public institutions?

- What values and goals do we seek with the developments of AI and NS?
- Shouldn't humanity ensure that the basic needs of every individual are met before embarking on further technological developments? Especially with respect to environmental constraints? Should we favor the development of low-tech (easy-to-use technologies with low environmental impact; source)?
- How to organize and steer technological development in society?
- Beyond the mobilization of technicians, should participations of citizens be increased in ethical deliberation about technological developments? How to involve society in technological orientations?
- How reversible are our choices to implement AI? Are they irreversible? If so, they require much more thought, and more of a "precautionary principle" approach.
- Should private research be regulated differently from public research?
- Should countries seek for technological sovereignty?

- What is the correct level to regulate the use of AI and NS, especially in democratic life (e.g., in election campaigns)? National, International or European? Can ethical and regulatory framework be universal, or should several cohabitate? How to ensure such a cohabitation (what form of international coordination)?
- Which traditions of political and philosophical thought seem best able to meet the challenges to democracy in times of NS and AI?
- Can a model of "ethical" use of AI and NS in politics be developed? Would it be valid for Western democracies in general, or is it necessary to develop separate models for, say, Europe and the United States?

How to protect fundamental rights?

- Should AI and NS developments lead to deep and fundamental changes in legal regulations and fundamental rights?
- What about the right to physical and mental integrity, the right to privacy, the right to data protection, the right of defense? For example, what level of interference in private life is (legally) acceptable?
- What about the question of rights and legal personhood for AI systems? Can certain forms of AI be considered as the next subjects of (international) law? On a par with humans?

How NS and AI affect individual right to privacy?

- AI services are often proposed at the cost of privacy. How important is privacy to humans? When using technology, what do humans value more than privacy? Which threats to humans result from private companies accessing data of individual citizens through technology? What are acceptable bargains in these matters?
- What democratic use for the enormous amount of data generated by citizens?

How can misinformation be reduced?

- AI and digital technologies sometimes foster misinformation. How should we deal with information and truth in the age of AI?
- How can humans access the truth? How important is knowledge to humans?
- How can IA be used to identify manipulated data and to fact-check information?

AI may be regulated at two levels: that of the way these technical tools (platforms, applications) are used; and that (more radical, but also more democratic) of their very existence and opportunity for development.

1. Regulation of the technical aspects of AI

a) The use of AI in our democracies may be first regulated regarding their ecological impacts. An economic regulation of AI would apply if their real effects on the environment (notably through the emission of greenhouse gas) were taken into account.

b) The need to include all citizens (including those without access to online tools) in public debate is a strong instrument of regulation of the use of AI.

c) The notion of political responsibility (of elected representatives and institutions) should also be redefined in order to better delineate the role of AI in public decision making.

d) The exigency for protection of personal data also puts under pressure the development and use of AI. For instance, it should be possible for citizens to refuse the sharing of their personal data. Regulation of information (manipulation, propaganda, censorship) is another starting point to regulate the spread of AI.

e) More generally, the constitution of a form of digital citizenship is a central challenge for democracy at the time of AI. This digital citizenship applies both to information diffusion and treatment: it implies that individuals are responsible regarding the news and data they diffuse, and critical regarding the information they receive. Another regulation strategy would consist in opening the 'black boxes' of social media algorithms in order to counter the echo-chamber effect.

2. Regulation of the developments of AI

a) In democracies, rules and laws could constraint the use of innovations. This implies to give citizens access to transparent information on technology assessment, for them to build insightful opinions on the actual and possible consequences of using new technologies.

b) A regulation could also be done by approving ethical rules constraining the kind of scientific research which are acceptable – that is, by forbidding the use of some objects or technologies in research. Regarding these

two first points, it deserves to be noted that the formulation of rules and laws is at risk of being under pressure from lobbies and conflicts of interests — as for many other issues.

c) A third direction would consider the fact that scientific research itself can be regulated through funding decisions. Currently, research funding is largely left aside from democratic decisions (citizens never vote to set-up the public research agenda). Yet, AI & NS progresses directly depend on how private and public funds are allocated to scientific research. Taking the democratic ideal seriously implies that people should have their say on the way (at least) public resources feed in scientific programs. The challenge here is to find procedures to democratize the governance of science, mainly the funding decisions: which research should we (as a society) support, and how is it possible to make this choice in a democratic manner? Obviously, public funding is often intertwined with private ones. A reflection on the democratization of the funding of science should take this issue into account.

IV. Additional questions for possible expert insights on Humanism, AI and Neurosciences in the field of Democracy.

Synthesis of co-constructed questions

Do AI and NS change significantly what it means to be an individual in a democracy?

- Has the relationship between citizens and democratic institutions been altered due to developments in AI?
- Don't some studies show that exposure to screens decreases empathy?
- Italy has been a forerunner for the extensive use of technology and social networks in politics, especially with the experience of the Movimento 5 Stelle. Almost 15 years after these developments, what teachings can be drawn?
- Is AI undermining democracy by accelerating the spread of misinformation and disinformation, especially the spread of misinformation related to elections and important public policy issues? If so, can we use AI to help limit the spread of misinformation and disinformation (and if so, how)? Are there other ways in which we can use AI to enhance democratic governance?
- During the pandemic years, the use of AI and social networks in politics experienced a significant increase. What lessons do the last years leave us with?

What about the question of rights and legal personhood for AI systems?

- Can certain forms of AI be considered as the next subjects of international law?
- Should an AI be given legal personality? If so, how do we assume the absence of a limit between an AI and non-IA personality (who are we dealing with?)?

How to regulate NS and AI?

- At what level is it right to regulate the use of AI in democratic life (e.g., in election campaigns)? National or European?
- Which traditions of political and philosophical thought seem best able to meet the challenges to democracy in times of NS and AI?
- Can a model of "ethical" use of AI and NS in politics be developed? Would it be valid for Western democracies in general, or is it necessary to develop separate models for, say, Europe and the United States?

Can NS be employed in justice?

- How reliable is functional magnetic resonance imaging for lie detection?

Can AI be at the service of societies and cultures?

- Can culturally sensitive AI be produced? Not an AI biased by cultural factors, but an AI capable of interacting considering cultural specificities (AI adapted to the socio-cultural profile of those for whom it is designed).
- How can the online dissemination of false information be prevented?

Bibliography

Biancu, S. (2019). "The human measure and the (impossible?) legacy of humanism. Guest editor's preface," in *Etica & Politica / Ethics & Politics*, vol. XXI, no. 2, pp. 9-23, doi: 10.13137/1825-5167/28131.

Damasio, Antonio (2005). *Descartes' Error: Emotion, Reason, and the Human Brain*, New Jersey: Penguin.

Davies, T. (1997). *Humanism*. Routledge.

Gruber, M. J., Gelman, B. D., & Ranganath, C. (2014). "States of curiosity modulate hippocampus-dependent learning via the dopaminergic circuit," in *Neuron*, vol. 84, no. 2), pp. 486-496.

Cigman, Ruth and Davis, Adrew (2009). Editors: *New Philosophies of learning (vol. 2)*, Wiley-Blackwell.

Foucault, Michel (1969). "Qui êtes-vous, professeur Foucault?", in M. Foucault, *Dits et écrits (1954-1988)*, tome I (1954-1975), Texte n. 50, pp. 601-620 (or. ed. in P. Caruso, Conversazioni con Claude Lévi-Strauss, Michel Foucault, Jacques Lacan, Mursia, Milano 1969, pp. 91-131).

Heidegger, M. (1947) *Platons Lehre von der Wahrheit mit einem Brief über den "Humanismus"*, Francke, Bern.

Jonas, H. (1974). *Philosophical Essays: From Ancient Creed to Technological Man*. University of Chicago Press.

Maritain, J. (1936). « Humanisme intégral. Problèmes temporels et spirituels d'une nouvelle chrétienté », in Maritain J. et R., *Œuvres complètes*, vol VI (1935-1938), Éditions Universitaires Fribourg Suisse - Éditions Saint Paul, Fribourg-Paris 1984.

Minsky, Marvin (1968). Editor: *Semantic Information Processing*, Cambridge: MIT Press.

Moor, James H. (2006). "The Nature, Importance, and Difficulty of Machine Ethics," in *IEEE Intelligent Systems*, vol. 21, no. 4, pp. 18-21, doi: 10.1109/MIS.2006.80.

Pope Francis (2015). *Laudato si'* (open access here: https://www.vatican.va/content/francesco/en/encyclicals/documents/papa-francesco_20150524_enciclica-laudato-si.html)

Tognon, G. (2019). "Humanism. Reflections on an Eponymous Idea," in *Contemporary Humanism – Questioning an Idea: A Time of Fragility, a Time of Opportunity?*, «Munera», special issue (open access here: https://www.cittadellaeditrice.com/munera/tutti-gli-articoli/contemporary-humanism-giuseppe-tognon-humanism-reflections-on-an-eponymous-idea/)