Global synthesis of NHNAI societal discussions (2023-2025)

Global-Transversal analysis

In 2023 and 2024, discussions on what it means to be human in the time of neuroscience (NS) and AI have been facilitated by NHNAI partners in 9 different countries. In each country, 3 lines of discussions have been opened to explore this question in the **3 thematic fields of education, health, and democracy**. Each partner then produced 3 **local syntheses** reporting on the content of discussions in these 3 fields in the corresponding countries.[1] On this ground, the coordination team proposed 3 **global thematic syntheses** (one per field explored, education, health and democracy). Finally, ideas of these 3 global thematic syntheses have been grouped to generate one **global-transversal synthesis**, gathering ideas that were more general and have been expressed in different thematic field.

This document presents **ideas of the global-transversal synthesis**, together with nexuses in which some ideas emerging from discussions enter in conflict and tension, manifesting possible complexities and delicate points of transversal questions.

[1] For an exact total of 8*3 + 2 local syntheses. In Canada (Québec), Cégep Sainte-Foy organized discussions focused on Democracy and Education, but not on Health.

Table of content

Part 1: Global-transversal ideas

Being human in the time of NS and AI means …

Relying on technology to improve ourselves and our lives

Participants to the collective discussions largely acknowledge that AI and NS developments in the last decades opened the way for various technological processes that (have strong potential to) improve human life.

- Automation of tedious tasks can improve our lives, notably by permitting to save time for more essential activities such as relationships or anything that fosters human flourishing.
- AI technologies can help us organizing the vast amount of information we must deal with (especially on social networks and the internet) and contribute to enhancing the quality of this information (fact checking, fighting against (deep) fake news, …).
- AI technologies can support humans in decision making (even perform better in some tasks). They may help us preventing or managing various problems and crises (ensuring better security in the public space with more efficient surveillance, detecting fraud or corruption, anticipating epidemics or the vagaries of the weather and climate change, …).
- AI, and notably generative AI, can be useful to stimulate creativity, find inspiration and new ideas, etc. Even finding the right prompts to write can be considered as a creative task.
- AI and NS outcomes may allow us enhancing our physical and mental abilities, improving our performance and efficiency. They may also support the most vulnerable and excluded persons (ranging from providing facilitated access to services and information, to empowering disabled persons and coping with aging issues).
- Finally, AI and NS may improve our lives by enriching and refining our understanding of ourselves as human beings.

Involvement in nexuses of complexity (see below Part 2: Global-transversal nexuses of complexities):

- Becoming more efficient without threatening the core of what makes us human

Corresponding ideas (to be consulted in the thematic global syntheses downloadable here):

- (Global – Health) Acknowledging the positive contribution of health technologies to healthcare 8 countries (BE, CH, FR, IT, KE, PT, TW, US) 28 ideas
- (Global – Health) Exploring the potential contributions of health technologies to humans' self-improvement 2 countries (FR, PT) 2 claims / ideas
- (Global – Health) Using health technologies to better the conditions of life of the most vulnerable persons 4 countries (FR, KE, PT, TW) 12 ideas
- (Global – Education) Using AI to improve performance and innovation 4 countries 6 countries (BE, CA, CH, PT, TW, US), 10 claims / ideas
- (Global – Democracy) Acknowledging the positive (potential) impact of AI on human life while asking the right questions 6 countries (BE, CA, FR, KE, PT, US) 14 ideas
- (Global – Democracy) Using AI to ensure Safety / Security 4 countries (CA, FR, KE, PT) 8 ideas

- (Global – Education) Using AI to free time for human flourishing 5 countries (CA, FR, PT, KE, US), 13 claims / ideas
- (Global – Education) Improving self and other-understanding with AI and NS 2 countries (FR, PT), 2 claims / ideas
- (Global – Education) Fostering creativity with AI 5 countries (BE, FR, PT, TW, US), 5 claims / ideas

Preserving human autonomy and agency

(notably in decision-making and (collective) cognitive abilities)

As many participants to NHNAI discussions worry, relying too much on AI technologies may lead to deskilling and cognitive impoverishment, overdependence and loss of resilience in case of technologies unavailability. In addition, AI may induce a kind of uniformization across individuals in more and more domains (as it becomes able to imitate more and more human traits and capacities), threatening people's uniqueness.

While participants recognize that AI can in many situations improve and support human decision-making, they also fear it may become difficult to preserve human independent decision-making, with the possibility to sometimes diverge from the machine recommendations (for instance based on human-reflection with trained intuition). This may become particularly problematic for professionals to whom we delegate and grant authority, with the risk of shifting authority delegation from professionals to machines (this worry has been expressed about the doctor-patient relationship but could probably also apply in the context of education about the learner-teacher relationship).

Participants also point a risk of reducing persons to their data, connected with an excessive focus on what can be measured and quantified. Then would arise the danger of prescriptive and coercive automated systems, notably with surveillance and algorithmic governance (but also in any field where decision making deeply impacting persons could be automated, such as in medicine).

Finally, some participants evoke the topic of automated editorialization of information. Although we need powerful algorithm to organize information for us (search engine, recommendation algorithm on social networks and other platforms), this automation may lead to information or cognitive bubbles isolating individuals in uniform informational landscapes (a problem reinforced with generative AI facilitating the production of (deep) fake news). Thereby, AI technologies can deeply threaten our (collective) intelligence. AI technologies can even be used to exert a form of control over citizens, undermining their freedom of choice, of expression and of thought.

Corresponding ideas (to be consulted in the thematic global syntheses downloadable here):

- (Global – Health) Preserving human agency and autonomy (in healthcare) 6 countries (BE, FR, CH, IT, TW, US) 11 ideas
- (Global – Education) Preserving human autonomy 8 countries (BE, CA, CH, FR, PT, KE, TW, US), 16 claims / ideas
- (Global – Education) Preventing the risk of cognitive impoverishment 9 countries (BE, CA, CH, FR, IT, KE, PT, TW, US), 30 claims / ideas
- (Global – Democracy) Preventing AI from undermining humans' critical thinking, decision-making abilities, and collective intelligence 9 countries (BE, CA, CH, FR, IT, KE, PT, TW, US) 39 ideas
- (Global – Democracy) Preserving the specificity of human beings (compared to machines) 3 countries (FR, PT, US) 15 ideas

Setting limits and regulations, even if it could prove challenging

A large consensus emerges from NHNAI discussions upon the strong need for regulation and norms to ensure AI and NS technologies deliver positive outcomes. Norms and regulations are key to allowing for trust building and for persons protection when deploying new technologies. AI should comply with human values (fairness, non-bias, ...) and should be human-centric (aiming at human flourishing). AI and NS technologies should benefit all (it is crucial to fight against the exclusion of poor and vulnerable persons).

However, many participants also emphasize that regulation raises many acute issues making it a very difficult challenge. Among such issues, one can evoke the pace of technological development, the obfuscation of patterns of responsibility (with digital technologies in general and more specifically with machine learning), the often "easy" access to powerful tools (in the hand of badly intentioned actors, technology such as image/facial recognition can become extremely harmful), the global scale of research and development (with diversity of value systems around the world as well as constellations of conflicts of interest), the difficulty to enforce regulations (in such a diverse and international context). One should also take into consideration the economic or business model associated with digital technologies (cost-free models based on users' engagement and data collection might make it difficult to align with human-flourishing objectives).

Broadly speaking, regulation should foster reasoned and sound uses of AI and NS technologies. Nevertheless, identifying what is reasoned and sound and what is not can prove extremely difficult (take the case of social media moderation for instance: who are the legitimate actors? Or the case of health technologies with grey areas between curative and enhancement uses: who can decide whether a pathology requires/justifies the use of a given health technology?). Stakeholders, professionals, citizens and economic/industrial actors should be involved in regulation processes. In this respect, it is key to fight against the feeling of powerlessness citizens may experience when confronted with such regulation challenges.

Involvement in nexuses of complexity (see below Part 2: Global-transversal nexuses of complexities): this idea constitutes a nexus of complexity on its own.

Corresponding ideas (to be consulted in the thematic global syntheses downloadable here):

- (Global – Health) Regulating AI and health technologies in healthcare 6 countries (CH, IT, PT, US) 11 ideas
- (Global – Democracy) Being aware of challenges regulation raises 5 countries (BE, CH, FR, PT, US) 8 ideas
- (Global – Health) Being aware of challenges regulation raises 3 countries (PT, TW) 4 ideas
- (Global – Health) Limiting the use of health-enhancement technologies 4 countries (CH, FR, IT, PT) 11 claims / ideas
- (Global – Democracy) Acknowledging the positive (potential) impact of AI on human life while asking the right questions 6 countries (BE, CA, FR, KE, PT, US) 14 ideas
- (Global – Education) Reinforcing regulatory measures 5 countries (CH, KE, PT, TW, US), 9 claims / ideas
- (Global – Democracy) Setting limits, control and regulation of AI to preserve democracy 9 countries (BE, CA, CH, FR, IT, KE, PT, TW, US) 29 ideas
- (Global – Democracy) Assessing the economic model behind AI and its societal impacts – 1 country (FR) 2 claims / ideas
- (Global – Democracy) Acknowledging human free-will and the citizen power of influencing regulation and political choices – 1 country (FR) 1 claim / idea

Preserving human responsibility (only humans can be morally responsible)

Participants to collective discussions converge on the idea that only human beings, thanks to their awareness and critical thinking, are able to make ethical choices and responsible decision-making. Humans are therefore the only ones responsible for technological orientations and the consequences of AI uses. Except in certain specific legal senses (corporate responsibility, legal personhood allowing for instance for monetary compensation), moral, ethical, legal and political responsibility (and criminal responsibility) can never be attributed to machines. Dilution and obfuscation of chains of responsibility is highly problematic.

Corresponding ideas (to be consulted in the thematic global syntheses downloadable here):

- (Global – Health) Never believing we can delegate (moral) responsibility to machines 6 countries (BE, KE, IT, PT, TW, US) 10 ideas
- (Global – Democracy) Preserving human responsibility on ethical choices/decision-making 4 countries (BE, CA, FR, IT) 7 claims / ideas

Respecting the singularity and (cultural) diversity of persons

The NHNAI collective discussions largely converge on the idea that one must acknowledge persons as singular beings, and treat accordingly, in a comprehensive way, doing justice to their diversity. Any reduction of persons to measurable and quantifiable aspects (or to what can be accounted for and addressed through technological means) at the cost of not acknowledging persons experiences and feelings should be resisted. Even when machines can imitate, predict and reproduce very convincingly what humans do, imitating does not mean reproducing in all important dimension. These aspects do not exhaust what human persons are and what can be meaningfully said about them. Humans are all different with different spiritualities, and this difference is a richness for humanity that we should preserve from the threat of uniformization AI and NS can bring.

Participants thus insist that we should thus resist the kind of standardization across individuals AI tend to induce in more and more domains (as it becomes able to imitate more and more human traits and capacities). In fact, such a uniformization threatens people's possibility to be unique singular beings.

Corresponding ideas (to be consulted in the thematic global syntheses downloadable here):

- (Global – Health) Recognizing patients in their singularity and diversity (within a comprehensive approach) 4 countries (BE, CH, FR, KE) 6 claims / ideas
- (Global – Democracy) Recognizing that human persons exceed the sole measurable dimensions 3 countries (CA, PT, US) 3 claims / ideas
- (Global – Education) Preserving (cultural) diversity and human singularity 4 countries (FR, PT, KE, TW) 11 claims / ideas
- (Global – Health) Withstanding the overvaluation of performance, efficiency or productivity 4 countries (CH, FR, PT, USA) 5 claims / ideas
- (Global – Democracy) Preserving the specificity of human beings (compared to machines) 3 countries (FR, PT, US) 15 ideas
- (Global – Democracy) The (difficult) future challenge of distinguishing between AI and humans 3 countries (BE, FR, PT) 5 ideas

Preserving empathy, human contact and human relationships

Participants to the collective discussions make it clear that humans are social beings who can only flourish (and learn, teach, cure, care, or heal) in relationship with their fellow human beings. Unlike machines, humans have the indispensable social ability to put themselves in other people's shoes and form strong emotional bonds (importance of feeling and dialogue to do so). Trust and representativeness are built through human dialogue. AI is not able to replace human interaction. In this respect, the tendency to try to fulfill social needs with digital technologies and artificial companions may cause severe (mental) health issues.

For many participants, one should pay attention to the surrounding context that may in some cases reinforce the risk of degrading the quality of human contact (for instance in times of crisis or because of the exhaustion of healthcare or educational systems).

Corresponding ideas (to be consulted in the thematic global syntheses downloadable here):

- (Global – Democracy) Preserving empathy, human contact and relationships 2 countries 2 countries (CH, PT) 4 claims / ideas
- (Global – Health) Maintaining empathy and human relationship at the core of healthcare 8 countries (BE, CH, FR, IT, KE, PT, TW, US) 24 ideas
- (Global – Education) Preserving human relationships and in-person interactions 8 countries (BE, CA, FR, IT, KE, PT, TW, US), 32 claims / ideas

Seeking for self-improvement

Some participants point out that it is natural for humans to seek self-improvement and progress, in order to maximize their efficiency. Those are strong objectives for most of humans (which can lead to use cognitive enhancers or other enhancement technologies). AI and NS may be used to compensate human limits and could maybe lead to develop new kinds of cognitive skills.

Involvement in nexuses of complexity (see below Part 2: Global-transversal nexuses of complexities):

- Becoming more efficient without threatening the core of what makes us human

Corresponding ideas (to be consulted in the thematic global syntheses downloadable here):

- (Global – Health) Constantly seeking for self-improvement and progress 1 country (PT) 1 claim / idea
- (Global – Education) Using AI and NS to better teach and learn 8 countries (BE, CA, CH, FR, IT, KE, PT, TW), 19 claims / ideas
- (Global – Education) Using AI to improve performance and innovation 6 countries (BE, CA, CH, PT, TW, US), 10 claims / ideas

Preserving and intensifying what makes us human and fostering human flourishing

Many participants to the NHNAI discussions highlight that certain values and features are unique to human beings, as spirituality, wisdom, emotionality, creativity, autonomy, critical thinking, imagination, consciousness, and empathy. AI and NS technologies should not

threaten or marginalize or minimize such core components of what it means to be human. In this perspective, one should resist any overfocusing on efficiency, performance and financial profitability only. For instance, not performing some tasks may lead us to impoverish ourselves, our cognitive abilities, our creativity (such as with abuses of AI assistance to creation). Also, gains in productivity may be mobilized to save time for activities fostering human flourishing.

Participants also worry about the tendency to systematically reject limits, to attempt (notably by the means of AI and NS technologies) at overcoming and transgressing all limits by principles can deeply undermine our humanity. Some limits and vulnerabilities (such as being "affectable" and thus susceptible to experience suffering, or being mortal) also are core to what it means to be human. In the same vein, fatigue and weariness are sometimes the sign that something is wrong in one's life, rather than mere limits to overcome (e.g. by using some enhancement technologies). These types of limits deserves acknowledgement and great delicacy in the context of reflection upon adequate technological development.

Involvement in nexuses of complexity (see below Part 2: Global-transversal nexuses of complexities):

- Becoming more efficient without threatening the core of what makes us human

Corresponding ideas (to be consulted in the thematic global syntheses downloadable here):

- (Global – Education) Preserving human creativity 8 countries (CA, CH, FR, IT, PT, KE, US), 11 claims / ideas
- (Global – Education) Using AI to free time for human flourishing 5 countries (CA, FR, PT, KE, US), 13 claims / ideas
- (Global – Education) Preventing the risk of cognitive impoverishment 9 countries (BE, CA, CH, FR, IT, KE, PT, TW, US), 30 claims / ideas
- (Global – Health) Acknowledging some of our limitations and vulnerabilities as inherent to our human nature 2 countries (FR, PT) 3 claims / ideas
- (Global – Democracy) Preserving the specificity of human beings (compared to machines) 3 countries (FR, PT, US) 15 ideas
- (Global – Democracy) The efficiency of technology should not lead to increase pressure to produce – 2 countries (FR, PT) 2 claims / ideas
- (Global – Education) Making people's flourishing a top priority - 4 countries (CH, FR, IT, PT), 16 claims / ideas
- (Global – Education) Preserving human creativity 8 countries (CA, CH, FR, IT, PT, KE, US), 11 claims / ideas

Fostering scientific/technical as well as ethical literacy and critical thinking

It is largely admitted among NHNAI participants that NS and AI literacy is key for concerned actors (stakeholders, professionals, developers, policymakers, economic/industrial actors) to be able to conduct proper ethical reflection on associated technological development and on adequate uses of available technologies (what are the limits of proposed technologies, what are the strengths and risks?). It may notably be key to look at the economic or business model associated with digital technologies (cost-free models based on users' engagement and data collection might make it difficult to align with human-flourishing objectives). Participants also worry about the tendency to present AI as infallible, or as by principle or nature superior to humans.

Participants also make a more general point: to allow for proper (collective) ethical reflection on NS and AI, it is essential to preserve and develop critical thinking (in a time where

disinformation is growing and relationship to truth and knowledge is threatened) as well as capabilities for ethical thinking itself.

Corresponding ideas (to be consulted in the thematic global syntheses downloadable here):

- (Global – Education) Encouraging ethics in education 4 countries (CH, FR, IT, PT), 9 claims / ideas
- (Global – Education) Fostering critical thinking, AI & NS ethics and literacy 7 countries (BE, CH, FR, IT, PT, KE, TW), 24 claims / ideas
- (Global – Democracy) Fostering literacy and critical thinking to preserve and strengthen democracy 6 countries (CA, FR, IT, PT, TW, US) 14 ideas
- (Global – Health) Fostering literacy and critical thinking 6 countries (BE, CH, IT, KE, PT, TW) 8 ideas
- (Global – Democracy) Assessing the economic model behind AI and its societal impacts – 1 country (FR) 2 claims / ideas

Ensuring that technology reduces (rather than increases) inequalities

The topic of inequalities appears very often in NHNAI discussions. The rapid development of AI and NS technologies poses the risk of increasing already existing social and economic inequalities. It is necessary to ensure that benefits and difficulties raised by these transformations are fairly distributed. One must for instance consider the problem of fairness in access to non-dehumanized services and to positively contributing innovations, or in protection against dangers and unwanted effects, such as automated discrimination and biases (in particular in the domain of human resources management). Inequalities can be in terms of access (skills and literacy, financial means, material infrastructures) as well as in terms of power or benefit-sharing asymmetries. Inequalities and discrimination can also occur when systems are less efficient or reliable with respect to minorities (groups of people less represented in training datasets). At the level of nations, inequalities can also lie in the ability to develop sovereign AI systems.

However, participants also point out the potential for mitigating inequalities. If correctly employed, AI (digital) and NS technologies can enhance social justice and human rights defense. AI technologies can foster social inclusion, notably by facilitating access to various services to the most vulnerable (poor persons, refugees) or by empowering persons with disabilities to help them become more independent. In this respect, it is important to ensure that such technologies genuinely help organizations and citizens by meeting the objectives they are designed to address. Technologies should not serve imposing universalized "tech values" at the detriment of local values upheld by users and impacted communities.

Corresponding ideas (to be consulted in the thematic global syntheses downloadable here):

- (Global – Education) Not exacerbating social and economic inequalities with AI 8 countries (BE, CH, FR, IT, KE, PT, TW, US), 20 claims / ideas
- (Global – Democracy) Taking into account vulnerable people and contributing to human rights, social and political inclusion 6 countries (BE, FR, IT, KE, PT, US) 30 ideas
- (Global – Health) Ensuring fairness and equality in opportunities for living a good life 6 countries (BE, CH, FR, IT, KE, PT) 23 ideas
- (Global – Health) Using health technologies to better the conditions of life of the most vulnerable persons 4 countries (FR, KE, PT, TW) 12 ideas
- (Global – Education) Fostering social inclusion thanks to AI technologies 9 countries (BE, CA, CH, FR, IT, KE, PT, TW, US), 21 claims / ideas

Privileging human - AI cooperation instead of human replacement

For many participants to NHNAI collective discussions, AI and technology must find their right place to contribute to a more humanized society. AI can be a useful tool to help humans save time on certain tasks, but machines should not replace humans (especially not in tasks they like performing). In particular, AI and automation technologies are often implemented in fields where actors have a lack of time or are exhausted (such as in healthcare systems). However, technology may not always constitute the right or primary answer to such major issues.

In the same vein, the problem of work automation and the risk of mass unemployment should be considered seriously, especially when it comes to the most vulnerable persons. Such major economic shifts have the potential to deeply affect our societies.

Corresponding ideas (to be consulted in the thematic global syntheses downloadable here):

- (Global – Democracy) Privileging AI cooperation and support instead of human replacement 7 countries (CH, FR, IT, KE, PT, TW, USA) 7 ideas
- (Global – Democracy) Finding the right balance between human labor and AI task automation 7 countries (BE, FR, IT, KE, PT, TW, US) 14 ideas
- (Global – Health) Privileging AI cooperation and support instead of human replacement 8 countries (BE, CH, FR, IT, KE, PT, TW, USA) 22 claims / ideas

Protecting privacy

The rise of AI raises among participants to collective discussions concerns about privacy. For example, private and public entities have massive access to all types of personal data (about health, opinions, choices, habits and customs...) putting a strain on privacy (one should add to the top of that emerging problems concerning neurotechnology and brain privacy). To protect democracy and ensure individual freedom, it is imperative to strengthen privacy protection laws and clearly distinguish between private and public life, not only online (public opinions and online anonymity) but also on public space (the use of data obtained from video surveillance as facial recognition must be restricted to certain places, and their use should be justified).

Corresponding ideas (to be consulted in the thematic global syntheses downloadable here):

- (Global – Health) Ensuring privacy protection 6 countries (BE, CH, FR, IT, KE, US) 12 ideas
- (Global – Democracy) Ensuring Privacy protection 7 countries (BE, CA, FR, IT, KE, PT, TW) 24 ideas

The (difficult) future challenge of distinguishing between AI and humans

As AI systems progress, their ablity to mimic and simulate human behavior will develop. Some participants point out that it will become more and more difficult to distinguish between machines and humans, as well as between something real and unreal (like a picture generated by AI). Regulation should emphasize the need to inform citizens whether they are interacting with humans or AI systems (and whether products or services they receive are human made or not).

Extracted from the global synthese on Democracy downloadable here)

Part 2: Global-transversal nexuses of complexities

Being human in the time of NS and AI implies carefully exploring nexuses of complexities where valid ideas are nonetheless in tension, manifesting subtleties and challenges one should not overlook. Here are below some examples of **transversal nexuses of complexities**, identified in NHNAI discussions based on **local and global syntheses**.

What link, what interactions with machines?

Some participants point out that, with the progress of AI, we will tend to develop machines (robots, conversational automatons) capable of imitating or simulating behaviors and capacities specific to humans and living beings, such as empathy, assertiveness, emotional and affective life. As a result, it will become increasingly tempting to become emotionally attached to this type of machine capable of simulating relational capacities (such as companions or artificial assistants, or robots for personal care).

These discussions also raise the question of the rights to be granted to advanced robots or intelligent systems.

At the same time, many contributions to the discussions emphasize the importance of not losing sight of the specificity of the living and the human by comparison to machines. Machines are not conscious, do not feel emotions, cannot be wise, creative, critical or autonomous, are not capable of spirituality in the usual sense of these terms, which implies rootedness in lived experience, in a biological body. At best, they can simulate convincing behaviors in these registers (notably through conversation), behaviors that human beings or living beings would have in given circumstances. But imitating does not mean reproducing in all important dimensions.

From this point of view, many participants agree that AI cannot be a subject of law. The question is widely described as speculative or science-fictional, without being uninteresting (as success in generating strong AI would lead to deep interrogation about many philosophical and religious foundations of our societies).

Thus, it is quite widely expressed in the discussions that it is necessary to resist the (increasingly real and powerful) temptation to perceive certain robots or AI systems as genuine people and to try to connect with them affectively (as one would with a human, or even with another living being). We must resist the temptation to substitute interactions with machines for genuine human relationships. The tendency to try to fulfill social needs with digital technologies and artificial companions may cause severe (mental) health issues.

This being said, discussions progressively emerge about the manner we should relate with machines that, although just machines, could be more intelligent than us (at least in some large domains).

Corresponding ideas (to be consulted in the thematic global and local syntheses downloadable here):

- AI systems and machines cannot be confused with humans and therefore cannot be endowed with rights similar to those of humans.
    - (Global – Transversal) Respecting the singularity and (cultural) diversity of persons
    - (Global – Democracy) Preserving the specificity of human beings (compared to machines)
    - (France – Democracy) Undesirable: The recognition of a legal personality for AIs is not desirable
    - (France – Democracy) Desirable: Algorithms remain tools
    - (USA – Democracy) Machines are to serve humanity, therefore humanity must maintain appropriate control of AI
    - (France – Democracy) The complex question of the legal status of artificial intelligence is widely debated
- AI cannot be subject of laws + worries about the emergence of superintelligence and strong / conscious AI
    - (France – Democracy) Artificial intelligence: between ethical limits, rights and human control
    - (France – Democracy) Ethical challenge for AIs: risks, inequality and human nature
    - (France – Democracy) Undesirable: Ethical and Social Challenges of Artificial Intelligence: Fears, Inequalities and Questioning Fundamental Values
- AI systems should not replace human relationships
    - (Global – Transversal) Preserving empathy, human contact and human relationships
- AI systems will increasingly have behaviors that enable / encourage the tendency of humans to want to connect with and attach to them.
    - (Portugal – Democracy) Humans and machines may bond
    - (Global – Democracy and Transversal) The (difficult) future challenge of distinguishing between AI and humans

Expertise input:

Based on insights from Brian P. Green,[2] Mathieu Guillermin,[3] Nathanaël Laurent,[4] Federico Giorgi[5]

It's more than legitimate to marvel at recent developments in AI technologies, which have enabled programs such as ChatGPT and other large language models to sustain a convincing conversation with humans and to answer more and more correctly to advanced questions on human knowledge. These performances may deeply impact human relationships and interactions humans have with machines.

As noted in many thematic areas of the NHNAI project, relationships are of great importance in human life and their protection and enhancement should be a serious concern of all those working with AI systems and their effects. In general, AI systems should assist and not replace humans – but especially in relationships. As social creatures, theologically we are made in the image of a relational Triune God who is love itself, but this is also a philosophical and empirical point, and logically necessary. Humanity cannot live alone, and anything that erodes our relationships is a risky and dangerous thing. AI must be used to strengthen human relationships, whether familial, friendship, economic, political, or otherwise. AI which damages relationship attacks a core part of what it means to be human.

[2] Professor in AI Ethics, Director of technology ethics at the Markkula Center for Applied Ethics (Santa Clara University, USA)
[3] Associate professor in ethics of new technologies (UCLy (Lyon Catholic University), UR CONFLUENCE: Sciences et Humanités (EA 1598), Lyon, France)
[4] Associate professor in philosophy of biology (Université de Namur, ESPHIN, Belgium)
[5] Post-doctoral researcher in philosophy (Université de Namur, ESPHIN, Belgium)

A. With AI, we do not create radically a new kind of entity

The sense of wonder we may feel with AI technology must be for the right reasons. After all, its successes have nothing to do with the creation of new forms of life, new intelligent beings, we would call the AIs. It is just as dizzying, if not more so, to realize that mankind has been able to build machines, artifacts capable of simulating or reproducing intelligent behavior (convincing behavior that could have come from humans), with absolutely no life, no lived experience, no consciousness, but with pure mechanisms (inert mechanisms, but dazzlingly complex and miniaturized).

In addition to demystifying machine learning (including deep learning, based on artificial neural networks)[6], it's also crucial to remember that **all programs (from the most traditional and conventional to the most advanced AI program produced by machine learning) run on computers or similar machines that are not (or less) programmable**. What a machine like a computer does is to transform material configurations to which humans have associated precise meanings (a series of magnets on a hard drive disk symbolizes a sequence of 0s and 1s, itself associated, for example, with a sequence of words or a sequence of numbers coding the colors of pixels in an image) into new material configurations associated with other meanings (for example, a new series of words, a modified image or a description of the image). This type of machine, designed to transform material configurations into others according to what these configurations signify, is not new. The computer can be seen as the culmination of a long evolutionary history of information techniques and technologies, probably dating back to the very beginnings of writing. From this perspective, the abacus can be seen as an ancestor of the computer (mechanical transformation of configurations symbolizing, for example, numbers to be added, into configurations symbolizing the result of addition).

So, strictly speaking, there are no meanings, images, words or numbers in computers, let alone emotions or consciousness. They are, however, fantastic machines for mechanically manipulating (with incredible efficiency and precision) countless material configurations to which we humans attach meaning. A series of magnets on a computer hard drive disk will cause different pixels on the screen to emit different colors, which will be more than just tiny sources of colored light for us, which will become texts telling us about feelings, images of faces feeling such and such emotions... But the computer only processes information by mechanically and automatically manipulating magnets (or other hardware configurations). This makes it all the more breathtaking to see what we can get computers to do with programs derived from machine learning techniques.

Accordingly, and as evoked by many participants in NHNAI discussions, one can hardly assume that machines are capable of discernment and ethical decision-making in the strong sense. In this perspective, the famous three (or four) Laws of Robotics as portrayed by Isaac Asimov cannot constitute a firm ground for ensuring that machines behave ethically. In this perspective, Frank Pasquale (2020)[7] has recently formulated four "new laws of robotics," clearly intended as a continuation of Asimov's laws, with the fundamental difference that, unlike

[6] Learn more about machine learning in an expert's contribution to the reflection of a nexus of complexity in the field of democracy: https://nhnai.org/focus-on-nexuses-of-complexity-democracy/

[7] F. Pasquale, *New Laws of Robotics: Defending Human Expertise in the Age of AI*, Springer, 2020.

Asimov's, Pasquale's laws are directed at humans, not robots. Among other things, the four new laws of robotics state that robots and AI should not be conceived as substitutes for human professionals but rather as their complement. As Cabitza (2021)[8] also affirms, it is therefore desirable to overturn the classical view according to which humanity can and should rely on machines, and instead to rediscover the irreplaceability and centrality of human beings—especially when it comes to ethical decision-making.

B. But AI, like any technology, shapes what we are and how we live

According to the discussion of the previous section, acknowledging the powers of computers should never come without a clear understanding that computers and AI systems are not entities emerging *aside* from us. As we just saw, they are nothing like Science-Fiction AI that become conscious and autonomous in a strong sense. However, there is another crucial sense in which AI systems are not aside from us: they are not mere tools that we could mobilize only when we need them and that would otherwise remain quietly and neutrally on the shelf. **Technology deeply transforms us. It shapes and mediates our ways of being and of living together**.

Bruno Latour's sociological view can help us grasp this important point. For him, the 'social' is an associative composition[9]. A situation is seen as a 'hybrid collective' made up of human and non-human interactants. Neither objects nor subjects, these interactants are themselves envisaged as relational networks. A digital application, for example, cannot be envisaged without its designers, or the maintenance staff, or the user interface, or of course without its presumed users and intended uses. But users may well hijack these uses to adapt them to their own experiential context. An AI like ChatGPT is a composite formed by all the human authors who generated the texts that trained the model, plus all the designers of the model, plus all the agents who filter the AI's productions, plus all the users and the expected and unforeseeable contexts of use.

C. Imitation capabilities of AI systems are a deep gamechanger

Large language models like ChatGPT speak to us convincingly (with credible affective or emotional content). We can also try to automatically analyze emotions and feelings in what people say, or in videos capturing body or facial expressions. These new technologies open up the possibility of ever richer and more interesting interactions with machines, with modalities that reproduce or simulate a growing number of characteristics of interactions and relationships between living beings in general, and between humans in particular. To properly consider the consequences and challenges of these new possibilities for interaction with machines, several points need to be emphasized.

a. AI's extreme usefulness and uniformization issues

Before looking at the stakes with human (and life) imitation *per se*, it is important to point out that these imitation capabilities deeply transform the manner we interact with machines. This interaction can be rendered extremely fluid and easy, by comparison with digital skills that are normally required to use a computer. Now, more and more tasks can be launched and driven

[8] L. Floridi & F. Cabitza, *Intelligenza artificiale. L'uso delle nuove macchine*, Bompiani, 2021.

[9] See: https://www.erudit.org/fr/revues/cs/2022-n4-cs07915/1098602ar.pdf

by vocal control in natural language. This also means that digital systems will no doubt become even more ubiquitous than they already are.

In this perspective, a first issue we must circumvent to maximize the positive outcomes of AI technologies is not the problem of the human appearance of an object, or of the objectification/datafication of a human. Based on Latour's insights (humans and their technology form an intricate network of interactants, humans cannot be isolated from), what is important to avoid is that AI systems lead to a uniformization of human lives and become an impediment to their creativity. Standardized forms of mediation between AI systems and humans interacting with them could overwhelm and threaten the possibility to learn and innovate in concrete local situations. Local learning resulting from uncontrolled interactions with the environment is just as crucial as standardized data recording and processing systems. This is what Amitav Ghosh has formulated[10], for example, about the problem of climate change:

> *For those who carefully observe the environment in which they live, clues to long-term change sometimes come from unexpected sources. (...) The people who pay the most attention to ecological change are more often than not on the margins; the relationships they have with the soil, the forest or the water are barely mediated by technology.*

b. Never hiding who's who (or what's what)[11]

Returning to the question of the human appearance of machines per se, and contrary to what behaviorist approaches might suggest (in connection with the famous Turing test), it seems first important to maintain a distinction between on the one hand simulating a behavior resulting from a lived experience and on the other hand having this same behavior while experiencing this lived experience. What can we say, for example, about a machine that expresses words of compassion to an elderly person at the prospect of the end of life? This cannot be confused with the same words uttered by a person capable of experiencing his or her finitude, feeling and sympathizing in a shared lived experience. If AI technology is properly understood, what we have with a machine emitting words of sympathy must not be described as a machine having such feelings. Rather, it is interesting to look at what type of human will, feelings and intentions are really involved. Latour's analysis is deeply illuminating in this perspective as it leads to consider the AI systems as part of a network of human and non-human interactants, in this case organized to automatically utter words of sympathy. Human intentions exist here, but it looks extremely general, remote and abstract. They are those of developers and other persons involved in the decision to build this system. Such feelings, will and intentions are radically different from the one of a singular person expressing her sympathy to someone she's in direct contact with. The value of the uttered word cannot even be compared.

---

[10] A. Ghosh, *La malédiction de la muscade. Une contre-histoire de la modernité*, Wildproject 2024, pp. 170-171 (our translation).

[11] In the following sub-sections, we draw on the work of the *AI Research Group of the Centre for Digital Culture* (Culture and Education), and its book "Encountering Artificial Intelligence: Ethical and Anthropological Investigations." *Journal of Moral Theology* 1 (Theological Investigations of AI) 2023; especially chapter 4. https://doi.org/10.55476/001c.91230

c. The problem of treating human-looking machines as machines

Secondly, it's also important to say that simply acknowledging that machines are just machines, and treating them as pure tools, is not necessarily the answer to every problem. Indeed, from this perspective and in all likelihood, artificial companions (as in Spike Jonze's 2013 film *Her*) will be built and programmed to find their place in a market and therefore behave in a way that satisfies the user (for example, who would want to pay for an artificial companion that might betray or leave its human?). We will therefore be faced with systems that are perceived as objects, as possessions, but which will derive all their specific appeal from their ability to resemble a genuine person, to manifest an appearance of humanity, personality or life. Gradually becoming accustomed to the combination of these two characteristics could prove extremely destructive for humanity. It could be tantamount to gradually developing a capacity to feel comfortable with slavery: "Where there is no "other," but only the *appearance* of an other at our disposal, concurrent with the absence of the demand that would be exercised upon one's own self-gift by confrontation with a true other, we risk being conditioned in a dangerous talent for exploitation."[12]

In the same vein, this combination of object or tool status and personal appearance can also lead us to become accustomed to a consumer attitude towards other people's behavior, gradually reducing our tolerance of other people's behavior that would disturb us. It's not impossible that the constant presence of artificial companions, whose disturbing behaviors will be perceived as defects (by virtue of their status as tools or objects), surreptitiously leads us to view genuine people who disturb us in the same way, "as *simply* faulty human beings, viewing them with the same sort of idle dissatisfaction that we would feel with a robot that did not deliver the set of behaviors and reactions that we wanted to consume."[13]

This may lead to reconsider the question of what rights should be granted to robots and AI systems. Admittedly, their status as machines means that we can legitimately refuse to consider them as subjects of law. This does not mean, however, that we should let everyone do as they please with them, just as we might with a table. A regulatory framework may be desirable in this area, if only to prevent the development of behavior or habits that are extremely toxic for human beings and other living beings.

All these factors encourage us to reflect deeply on why developing machines increasingly capable of presenting the appearance of humans or other living beings. We need to reflect upon what we can really gain from such technologies.

[12] Ibid., p. 119.

[13] Ibid., p. 121. The full sentence reads: "Is it possible that we will no longer see this as a glimpse of a wider array of humanity, that we will not struggle toward a charitable response? Perhaps instead, we may come to think of these others as *simply* faulty human beings, viewing them with the same sort of idle dissatisfaction that we would feel with a robot that did not deliver the set of behaviors and reactions that we wanted to consume."

Setting limits and regulations, even if it could prove challenging

(based on the eponym global-transversal idea: Setting limits and regulation, even if it could prove challenging)

A large consensus emerges from NHNAI discussions upon the strong need for regulation and norms to ensure AI and NS technologies deliver positive outcomes. Norms and regulations are key to allowing for trust building and for persons protection when deploying new technologies. AI should comply with human values (fairness, non-bias, ...) and should be human-centric (aiming at human flourishing). AI and NS technologies should benefit all (it is crucial to fight against the exclusion of poor and vulnerable persons).

However, many participants also emphasize that regulation raises many acute issues making it a very difficult challenge. Among such issues, one can evoke the pace of technological development, the obfuscation of patterns of responsibility (with digital technologies in general and more specifically with machine learning), the often "easy" access to powerful tools (in the hand of badly intentioned actors, technology such as image/facial recognition can become extremely harmful), the global scale of research and development (with diversity of value systems around the world as well as constellations of conflicts of interest), the difficulty to enforce regulations (in such a diverse and international context). One should also take into consideration the economic or business model associated with digital technologies (cost-free models based on users' engagement and data collection might make it difficult to align with human-flourishing objectives).

Broadly speaking, regulation should foster reasoned and sound uses of AI and NS technologies. Nevertheless, identifying what is reasoned and sound and what is not can prove extremely difficult (take the case of social media moderation for instance: who are the legitimate actors? Or the case of health technologies with grey areas between curative and enhancement uses: who can decide whether a pathology requires/justifies the use of a given health technology?). Stakeholders, professionals, citizens and economic/industrial actors should be involved in regulation processes. In this respect, it is key to fight against the feeling of powerlessness citizens may experience when confronted with such regulation challenges.

Expertise input:

Based on insights from Brian P. Green and Nathanaël Laurent

As participants make very clear, regulation will be vital to make sure that AI is directed towards its best uses and away from its worst. This is a serious concern from all of the global partners in the NHNAI project, and this concern is heightened by the knowledge that bad or neutral actors will attempt to exploit – or at least not have the common good in mind – as they deploy AI systems onto our world.

AI should not only respect human values and be exclusively centered on the human being, because how can we aim for the fulfilment of a living species whose existence depends on countless interdependencies with other living species and with its terrestrial environment?

Limits and regulation could then come from a decentered approach like the one introduced by Aldo Leopold one hundred years ago:[14]

> *By extending 'the boundaries of the community to include soil, water, plants and animals, or collectively, the earth', Leopold's Land Ethic not only goes beyond the boundaries of humanity (the ordinary boundaries of morality), it becomes that of a mixed community, including diverse populations of different species. This should enrich our understanding of the variety of duties within the biotic community.*

With these values in mind, important questions arise:

- What becomes humanism when it takes into account the whole relational network of existence on our planet?
- What become scientific projects and technological advances if we try to render them compatible with interdependences which render living experiences possible?
- More specifically, what are the potential benefits of AI for earth in its globality (globality of all interactions and complexity of apprehending them)?

Improving ourselves without threatening the core of what makes us human

Many participants to NHNAI discussions highlight the fact that AI and automation technologies could help us saving time for essential activities such as relationships or anything that fosters human flourishing by delegating tedious tasks to machines. It also pointed that AI and NS outcomes may allow us to enhance our physical and mental abilities, improving our performance and efficiency. In addition, AI, and notably generative AI, can be useful to stimulate creativity, find inspiration and new ideas, etc.

For some participants, this way of considering the possible contributions of technology corresponds to a core part of human nature, with a strong drive toward self-improvement and progress, toward the maximization of efficiency.

Nevertheless, participants also worry about an uncritical and systematic quest for augmentation and improvement of efficiency and performance. It could lead to sacrifice aspects that are essential for humans, such as autonomy, creativity, relationships or to negate some limits and vulnerabilities that are at the heart of what it means to be human (mortality, affectability for instance). Overreliance on AI may also lead to uniformization and threaten people's possibility to be unique singular beings.

Corresponding ideas:

- AI and automation technologies can enhance our abilities and allow us to free time for essential activities: (Global – Transversal) Relying on technology to improve our lives
- It is a core part of human nature to seek for self-improvement: (Global – Transversal) Seeking for self-improvement

[14] Larrère, C. (2010). Les éthiques environnementales. *Natures Sciences Sociétés*, Vol. 18(4), 405-413. https://shs.cairn.info/revue-natures-sciences-societes-2010-4-page-405?lang=fr (our translation).

- Preserving the core of what makes us human:
  - (Global – Transversal) Preserving and intensifying what makes us human and fostering human flourishing
  - (Global – Transversal) Respecting the singularity and (cultural) diversity of persons

Expertise input: Subtleties with the notion of (self-)improvement

The notion of self-improvement might prove tricky, and we should mobilize it with caution.

A. From the perspective of cognitive sciences

Juan R. Vidal[15]

From the cognitive science point of view, seeking self-improvement is something that does not exist as such in human behavior, if it is not attached to a goal-oriented action and in a broad temporal context (ex: we want to assure access to food and water, shelter, …). This goal carries a value for the human that motivates (or not) to further learning and development of certain capacities and behaviors. Humans think they maximize their efficiency, but as Herbert Simon has mentioned, humans have a bounded rationality, and thus limited capacities to really maximize thought processes and thus behavior. Human rather "satisfice" their behavior in order to become as satisfied as quickly as possible, which is not the same than maximizing their capacities. This bias also applies regarding the use of technology, and with AI it is strongly potentiated. Yet, as has been shown, it also reduces dramatically the learning possibilities of the person and *in fine*, its freedom for action in the world. So, seeking self-improvement should resonate with the possibility to increase learning (embodied) and the possibilities for future learning (keeping doors open…) instead of accelerating certain performances that further ahead deprive the human of learning and thus adapting to changing conditions (if we consider that its adaptability greatly depends on its capacity to learn new behaviors/thoughts to face new problems). Thus, self-improvement cannot be equated with faster output generation, as quality learning takes time. Equating learning quality exclusively with a quantitative production process could be detrimental to quality, even if it can be usefull.

B. From the philosophical, anthropological and theological perspective

Based on insights from Brian P. Green and Nathanaël Laurent

In a general manner, the tension between the will to improve ourselves and the need to preserve what makes us human could be discussed on the ground of a collective book published in March 2024 by Editions du Cerf (Paris) entitled "The human being at the center of the world: For a humanism of the present and future times. Against the new obscurantisms." Daniel Salvatore Schiffer sums up one of the key messages[16]:

> *In short : the insidious and gradual erosion, if not evaporation, of the human being, in all his anthropological complexity (to use a key concept in Edgar*

[15] Associate professor in cognitive neuroscience (UCLy (Lyon Catholic University), UR CONFLUENCE : Sciences et Humanités (EA 1598), Lyon, France)

[16] Salvatore Schiffer, D. (ed.) *L'humain au centre du monde : Pour un humanisme des temps présents et à venir. Contre les nouveaux obscurantismes*, Les éditions du Cerf, 2024, ISBN : 9782204162661 (our translation). https://www.opinion-internationale.com/2024/03/09/lhumain-au-centre-du-monde-un-livre-a-lire-sous-la-direction-de-daniel-salvatore-schiffer_119419.html

> *Morin's philosophico-sociological edifice), to the benefit of a world that is all too often alienated, directive and reductive, it is a totalitarianism that ignores itself or does not speak its name, and so, in the face of increasingly Manichean thinking, it advances masked, sly and silent, but all the more dangerous for the freedom of the mind, of speech and thought, if not of conscience!*

This evaporation of what it means to be human is highly threatening. In fact, we cannot know what is of value in us if we do not know what and who we are.

Biblically speaking, the core of our human nature can be interpreted as love because we are made in the image of a God who is love (1 John 4:8) who commands us to love (Lev. 19:18, Deut. 6:4-5, Matthew 22:35–40, Mark 12:29–33, Luke 10:27) – even our enemies (Matt. 5:43-44) – and by that love become more fully human and divine. However, from the first chapter of John's Gospel we also know that God is Logos, word and reason, and that therefore the universe is rational, meaningful, and grounded in the most profound wisdom.

If, then, we have a dual nature (at least dual, if not much more) as loving and logical creatures then AI presents an opportunity and threat to us in these two key areas. We can use AI to help us learn new truths and gain new wisdom about the universe, to better care for each other and build peace around the world. Or we can abuse AI to replace our thinking abilities, thus leaving us mindless, and stunt our ability to love, or even worse turn our love into hate. We are already seeing these evil uses of AI move into society, in the form of using generative AI to cheat in school, and AI algorithms driving social media and app engagement through content that appeals to addiction, vice, and disdain for others.

This opportunity and threat of AI goes right to the core of our being, and thus demonstrates the validity of the existential angst that AI instinctively raises in some people. Indeed, it should raise this angst – or at least concern – in all of us.

Insofar as AI can help us become more logical and loving beings, then it is a blessing to humanity. Insofar as it makes us less logical and less loving it will be a curse. While these two assumptions about humanity are grounded theologically, there are good reasons to assume that it is not merely a theological grounding: it is also psychological, anthropological, sociological, philosophical, ethical, and more. There is an intuitive sense – and rational case to be made – that these features of humanity are legitimately near the core of human identity, and are therefore concerns regarding our engagement with AI.

Lastly, an empirical case can be made regarding the importance of autonomy and agency. From the data collected in this project itself. With three major thematic syntheses covering education, democracy, and health, coming from every country involved in the project, with dozens of claims/ideas made, this is clearly a topic of preeminent importance.

Regarding autonomy and agency, AI threatens both. Because AI automates agency, it effectively delegates that agency from some humans to other humans using AI as the implement (recalling CS Lewis, who said the same of technology in general, as a distilled form of nature, in chapter 3 of *The Abolition of Man*). Whomever control these agential AIs therefore has the power to disempower other people through automated systems.

This is only one way that AI might remove our autonomy and agency. Another is that we might be deskilled – both technically and morally – and through that lose our own ability to be full moral agents. Whether we are being actively disempowered by others or are instead disempowering ourselves passively or through inaction, AI presents a genuine threat that must be met with great care and urgency.

Remembering that autonomy and agency are at the core of what it means to be human also reminds us that responsibility is ours as well. We have responsibility for our actions, whether small or large, whether we are choosing to empower or disempower ourselves, whether we are acting through commission or omission, or acting directly or through intermediaries – human or AI. Responsibility rests with those humans making decisions, even if AI ultimately executes those decisions, once or a billion times.

Supporting without undermining human decision-making

Many participants to collective discussions acknowledge that AI technologies can support humans in decision making in various domains (even perform better in some tasks). They can help us organizing the vast amount of information we must deal with (especially on social networks and the internet) and contribute to enhancing the quality of this information (fact checking, fighting against (deep) fake news, …). They may allow preventing or managing various problems and crises (ensuring better security in the public space with more efficient surveillance, detecting fraud or corruption, anticipating epidemics or the vagaries of the weather and climate change, …).

However, it is also largely expressed that AI support to decision making can raise extremely acute difficulties. First, it may become difficult to preserve human independent decision-making, with the possibility to sometimes diverge from the machine recommendations (for instance based on human-reflection with trained intuition). This may become particularly problematic for professionals to whom we delegate and grant authority, with the risk of shifting authority delegation from professionals to machines (this worry has been expressed about the doctor-patient relationship but could probably also apply in the context of education about the learner-teacher relationship). In addition, chains and patterns of responsibility can suffer dilution and obfuscation. In this perspective, one should never loose sight of the fact that only human beings, thanks to their awareness and critical thinking, are able to make ethical choices and responsible decision-making. Humans are therefore the only ones responsible for technological orientations and the consequences of AI uses.

In addition, and as the discussions in the field of democracy focused on, the involvement of (generative) AI in the processing, management and editorialization of our informational landscape triggers worrisome issues, with serious risks of undermining and impeding collective intelligence. Biased and/or unfair algorithms may automatically and silently propagate discriminations, create information or cognitive bubbles isolating individuals in uniform informational landscapes. (Generative) AI can facilitate and foster the production and dissemination of (deep) fake news. AI can damage our ability to find accurate, trusted and

sourced **information**, introducing **mistrust** among uninformed citizens, compromising good democratic choices and pluralism.

Corresponding ideas (to be consulted in the thematic global and local syntheses downloadable here):

- AI technologies can support human decision-making: (Global – Transversal) Relying on technology to improve our lives
- AI technologies can undermine human decision-making:
  - (Global – Transversal) Preserving human autonomy and agency
  - (Global – Transversal) Preserving human responsibility (only humans can be morally responsible)
  - (Global – Democracy) Preventing AI from undermining humans' critical thinking, decision-making abilities, and collective intelligence

Expertise inputs:

Nathanaël Laurent & Federico Giorgi

The thesis that it might be possible to program an algorithm to make ethical decisions on our behalf is sometimes referred to as algorethics. In addition to the many critical issues that such a perspective understandably encounters—some of which are highlighted by the participants in the NHNAI debate—it is interesting to note how this kind of project tends to reduce the morality of an action to the intention of the agent to align their behavior with a set of ethical principles.

This deontological view of ethics, although not without supporting arguments, appears somewhat reductive, as it does not give sufficient consideration to the outcomes of the action taken (Cabitza, 2021). When reflecting on how a new technology ought to be used, it therefore seems more appropriate to adopt a consequentialist approach—one in which the moral character of an action is evaluated primarily based on the consequences it produces.

AI and ecology

NHNAI discussions manifest a deep point of tension concerning the relationships between AI and ecological transition or environmental questions.

On one hand, AI technologies can be used to adapt and combat global warming. For instance, AI is efficient to forecast future climate disasters, thus ensuring the security of inhabitants. In addition, AI can reduce pollution and greenhouse gas emissions, and monitor the health of ecosystems. It can also help fighting against illegal and dangerous activities undermining the preservation of environment and biodiversity.

However, on the other hand, AI technologies may induce an important impact on the environment and biodiversity. The rapid and excessive development of AI technologies do not seem compatible with the ecological transition as it uses a lot of resources (metals, water, energy…) and increases global warming by augmenting the greenhouse gas emissions. This is particularly the case for generative AI.

Given the considerable impact of technology on the environment because of its whole life cycle, should we use AI to preserve biodiversity and the environment? Is "AI for green" possible?

Ideas from local and global synthesis mobilized in this nexus of complexity:

- (France): Desirable: using AI to better adapt and fight global warming
- (Kenya -Democracy): Desirable: Climate change mitigation
- (USA): Environment and Sustainability
- (Belgium – Education): Desirable: an ecological employment of AI technologies
- (France): Managing the uncertainty with or without AI
- (France): The excessive development of digital technologies and AI is not compatible with the preservation of the environment and life on Earth

Expertise inputs:

Samuel Nyanchoga[17]

The contribution of AI to climate mitigation: the case of Kenya

Artificial Intelligence (AI) plays a significant role in climate change mitigation in Kenya and Africa at large, by enhancing data-driven decision-making, promoting sustainable resource use, and improving early warning and response systems. To begin with, one of its most impactful contributions lies in climate data analysis and forecasting. AI models can process large datasets, including satellite imagery and weather records, to predict extreme weather events such as droughts, floods, and heatwaves. As a result, this capability allows governments, farmers, and communities to anticipate and prepare for climate-related challenges, ultimately reducing their vulnerability. In the area of agriculture and food security, AI supports the development of climate-resilient practices. It recommends drought-resistant and fast-maturing crops, optimizes planting schedules, and guides irrigation needs using predictive modeling. Through the use of smart farming technologies that combine remote sensing and machine learning, AI improves crop yields while minimizing environmental impact. Furthermore, in the energy sector, AI enhances energy efficiency by predicting demand and improving the integration of renewable energy sources such as solar and wind into national grids. It also supports the deployment of microgrids and smart energy systems in rural areas, thus contributing to a cleaner and more sustainable energy future.

In addition, AI contributes to natural resource management by tracking deforestation, monitoring land degradation, and managing water resources in real time. This, in turn, enables better conservation efforts, such as reforestation and ecosystem restoration, and helps monitor biodiversity and wildlife migration patterns affected by climate change. Likewise, in urban planning, AI supports the development of climate-resilient infrastructure by modeling climate risks and optimizing transport and construction systems to reduce emissions and environmental impact. Another critical area of impact is disaster preparedness and early warning systems. AI technologies enhance the ability to detect and respond to natural disasters by simulating scenarios, identifying high-risk zones, and guiding emergency response plans.

[17] Professor of History and Director of Research at the Catholic University of Eastern Africa (CUEA).

Moreover, AI supports climate finance and policy by helping governments track carbon emissions, design carbon credit systems, and create data-informed environmental policies. Notably, AI is also transforming agriculture through precision farming, which helps farmers make informed decisions using real-time data collected from sensors, drones, and mobile technologies. This includes monitoring soil moisture, temperature, and other environmental conditions to optimize the use of water, fertilizers, and pesticides thereby increasing productivity while reducing environmental harm. For instance, in Ghana, AI is used to analyze satellite imagery and weather patterns to predict crop yields and manage resources more efficiently, enabling the government to formulate informed agricultural policies that support both productivity and sustainability.

In conclusion, AI is a powerful enabler of climate change mitigation across Africa. Its effectiveness is greatest when integrated with local knowledge systems, mobile technology, and community participation. When responsibly applied, AI offers transformative potential for building a more resilient and sustainable African continent in the face of climate change