An interdisciplinary conceptual study of Artificial Intelligence (AI) for helping benefit-risk assessment practices

Gauthier Chassang, Mogens Thomsen, Pierre Rumeau, Florence Sèdes, Alejandra Delfin

Title: An interdisciplinary conceptual study of Artificial Intelligence (AI) for helping benefit-risk assessment practices

Sub-Title: Towards a comprehensive qualification matrix of AI programs and devices

List of authors:

Gauthier Chassang 1,2,3,4

Mogens Thomsen 1,2,3,4

Pierre Rumeau 5,4

Florence Sèdes 6,4

Alejandra Delfin 1,2,4

Authors' affiliations:

1: Inserm, UMR1295, CERPOP, Team BIOETHICS, F-31000, Toulouse, France.

2: Université Paul Sabatier Toulouse 3, UMR1295, CERPOP, F-31000, Toulouse, France.

3: Ethics and Biosciences Platform (Genotoul Societal), GIS Genotoul, Toulouse, France.

4: Unesco Chair: Ethics, Science et Society, Working group: Digital and Robotics Ethics. Université Fédérale de Toulouse (UFT), France.

5: Groupement d'intérêt public e-santé Occitanie, Toulouse, France.

6: IRIT-CNRS, Toulouse, France.

Abstract:

We propose a comprehensive analysis of existing concepts of AI coming from different disciplines: Psychology and engineering tackle the notion of intelligence, while ethics and law intend to regulate AI innovations. The aim is to identify shared notions or discrepancies to consider for qualifying AI systems. Relevant concepts are integrated into a matrix intended to help defining more precisely when and how computing tools (programs or devices) may be qualified as AI while highlighting critical features to serve a specific technical, ethical and legal assessment of challenges in AI development. Some adaptations of existing notions of AI characteristics are proposed. The matrix is a risk-based conceptual model designed to allow an empirical, flexible and scalable qualification of AI technologies in the perspective of benefit-risk assessment practices, technological monitoring and regulatory compliance: it offers a structured reflection tool for stakeholders in AI development that are engaged in responsible research and innovation.

Text:

Several definitions of Artificial Intelligence (AI) have been proposed since the 1950s and the works of the mathematician Alan Turing [[109]] who became famous after having contributed to cracking the German Enigma code during the Second World War. He intended to reproduce human intelligence through automated technological processes and attached testing methods (Turing test [[45]], [[89]]). The computer scientist John McCarthy developed important programming tools (e.g. LISP) and he is currently recognised as the father of modern AI [78]. According to the Merriam-Webster modern dictionary, AI is a twofold notion covering, first, "a branch of computer science dealing with the simulation of intelligent behaviour in computers", and second, "the capability of a machine to imitate intelligent human behaviour". Recently, European experts proposed the following definition of AI: "systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals" [[54]], [[71]].

While these definitions might evolve, they reveal the final objective of AI, namely mimicking human intelligence. The rapid development of AI-based devices reaching professional and consumer markets (together with a hype about AI capacities) obliges governments to develop adapted ethical and regulatory frameworks intended to ensure the sole development of AI for the service of humankind. As a consequence, industrial policies go hand in hand with regulations to protect the fundamental human rights. The ethics frameworks try to find the best cut-off between the requirements of industrial development and rights protection within the conditions of a given ecosystem.

While AI research and innovation rapidly progress AI conceptualisations are evolving in disciplinary silos and are not structured to enlighten policymakers or regulators, nor to serve the public debate. Today, there is no universal understanding of AI, nor a common legal classification of AI programs and devices. Ethical and regulatory debates start to be organised at international, European, national and local levels with initiatives such as the recent Global Partnership on Artificial Intelligence (GPAI) proposed by Canada and France together with OECD in 2019. Other initiatives such as the Artificial Intelligence High Level Expert Group (AIHLEG) created under the auspices of the EU Commission in 2017, and the International Ad Hoc Expert Group (AHEG) of the UNESCO [[111]], are also in a quest for appropriate standardisation. But national AI strategies and visions for the field vary and make it difficult to agree on a single harmonised international regulatory framework [[75]]. More than 90 organisations around the world have proposed ethical principles for AI in the past few years and "leaders of the world's biggest economies – in the G7 and G20 organisations – signed a set of extremely vague ethical guidelines" [[76]] in the early 2019 [[66]]. One of the regrettable drawbacks of these initiatives is to directly focus on defining ethical principles without achieving essential upstream steps of qualifications of objects to fall within their scope. As a result, while paramount AI ethical principles are quite clearly established, translation of the principles into engineering practices and laws necessitates tools for defining AI programs and broader AI-based devices or systems in a meaningful way. Envisaging applied ethics and sounded regulations in AI necessitates first an appropriate methodology for conceptualising AI programs and devices in a standardised way. They should be based on essential characteristics and functions questioning the ethics and lawfulness of AI programs and devices. Due to the complexity of the field, this conceptualisation effort should be interdisciplinary and pragmatic. It should lead to comprehensive notions and practical tools that could be used by engineers, designers, programmers, policy-makers, lawyers, ethicists, end users and other stakeholders in AI for designing and implementing an ethics-by-design approach. Ideally, AI conceptualisation efforts will not only allow a community to talk the same language, but also ease implementation of specific benefit-risk assessments and help monitoring the technology. For example, regulators will need to legally qualify these technologies for identifying values at stake, existing rules to apply, and for performing impact assessments of AI technologies. This could then lead to clearing-house practices and eventually to the creation of new legal provisions. Also, the conceptualisation should facilitate honest and understandable communications towards the public, an ambitious and difficult task that needs to reach consensus. The way could be long, but it is a necessary step for avoiding the commercialisation of harmful technologies, as well as the adoption of conceptually unsounded or practically inapplicable regulations.

Through this paper, we aim at initiating the development of interdisciplinary and practical concepts for qualification of AI programs and devices that could be used by a variety of stakeholders looking for a

methodological framework; the identification of essential features to be considered will be eased when envisaging benefit-risk analysis and assessment of an AI-based product. Therefore, we provide a comprehensive picture of existing theoretical concepts emerging from three disciplines, namely Psychology, Cognitive Engineering and Ethics and Law, in order to draw an interdisciplinary common concept that can lead us to a proposal of a qualification matrix for AI programs and devices. This conceptual study will be the landmark for future papers in which we would like to trial the matrix with examples of qualifications of existing technologies in the health sector, a sensitive sector where AI-based innovations and related ethical tensions are numerous and where human rights are particularly developed and enshrined into practices. In addition, the health sector is stressing the interaction of AI with human beings in the achievement of a designed purpose. It is a good ground for case-studies. Nevertheless, the reasoning and proposals formulated in this paper could easily be applied to other sectors.

1. Considering accepted works in cognitive psychology for identifying "intelligence" in AI

While machines are not humans and do not have natural intelligence, AI was likened to human intelligence early on by its developers. Therefore, human intelligence is an essential conceptual benchmark for AI innovation.

Psychologists have developed numerous theories on intelligence, part of them taking advantage of researches and methodologies used in other disciplines such as anthropology, neurology and even genetics. One of these theories has been developed by Cattell and Horn [[74]] in the 1960s and further completed, notably by Carroll [[18]], in order to conceptualise human intelligence and related cognitive abilities, factors, processes and propose relevant tests. General Intelligence (G) refers to the human cognitive abilities and the mechanisms of learning, solving problems and meeting certain objectives. The domain is complex and relies on the study of conscious and multitask biological entities, mainly in humans. An interesting general theory defining two types of intelligence: the fluid (Gf) and crystallised (Gc) intelligence, provide useful elements in the context of AI. Any human being has both types of intelligence. Both types can be used together or separately, depending on the situation and functional objectives to accomplish.

Crystallised intelligence (Gc) is defined by Cattell as "the ability to use learned knowledge and experience". For example, this type of intelligence allows answering a problem encountered and appropriately solved in the past, by reproducing equivalent action. Learning abilities are generally supervised and include encoding capacity, middle-long term memory and ability to access memorised data in a logical way. But learning functions are limited. It is possible to search and select relevant information through the reference database for coping with a known situation or to make correlations between different datasets in order to find appropriate solutions, but without possibility to interpret the datasets in a way that would create new solutions, new strategies. Usually, acquisition or capture of reference data is guided. Gc does not include ability to generalise knowledge. In brief, the approach is based on an identified pool of skills, strategies and knowledge representing the level of cognitive development of an individual through his learning history. Gc is also referred to as "verbal intelligence". Concrete examples of Gc: being able to use semantic database and vocabulary to formulate correct sentences; being able to read learned language; being able to find similarities between different familiar objects.

Fluid intelligence (Gf) is defined as "reasoning ability, and the ability to generate, transform, and manipulate different types of novel information in real-time" [[122]]. This includes "the ability to solve new problems, use logics in new situations, and identify patterns" without having necessarily the prior experience of similar problems or prior strategic learning. Fluid intelligence is thus independent from learning, from experiences, culture and education (in AI we would refer to programming instead of education). Fluid intelligence is shaped by primary abilities such as induction, deduction, linkage, classification. It depends on factors such as processing speed, operative memory and, from a biological point of view, neuronal network activity. This type of intelligence has often been compared to the hardware for intelligence as it covers aspects supporting the development of the future abilities of an individual. Fluid intelligence includes adaptive skills and abstraction abilities. Extensive

learning abilities including broad or specific capture, integration (organisation, categorisation), interpretation, comparison, understanding and retranscription of novel external, environmental data in a comprehensive form are present. Gf can function on the basis of limited working memory [[9], [34]] (attentional system used both for data processing and for storage based on short-term memory), depending on the quality of the learning ability (in AI we would say learning algorithm) and of data retrieval capacity for problem solving (working memory can appeal to long-term memory data to improve data processing and can be diminished by parasite stimuli creating attentional troubles, informational or functional overloads). Acquired experiences are usually integrating Gc for feeding the reference database. However, in AI, it is possible that a program functions only on the basis of characteristics related to Gf, or Gc concepts. Gf is also referred to as "logico-mathematical intelligence". Concrete examples of Gf: ability to fully understand natural language; ability to apprehend implicit meaning or ambiguity for designing appropriate action; ability to interpret novel environmental data, from a simple term to a complex phenomenon, and to make deduction as to their meaning for finding relevant solutions. Gf does not need to appeal to acquired knowledge in order to analyse a new situation and to figure out logical relationships or underlying meaning of facts, attitudes or events, which are significant and useful to design an appropriate action.

In sum, Gf is more "instinctive" than Gc that comes from "know-how". If fluid intelligence is compared to the hardware of intelligence, crystallised intelligence would be the software [[84]]. The two types of intelligence are complementary, some tasks will only need Gf abilities, others only Gc ones, and others will only be possible by combining Gf and Gc abilities. From an AI perspective, instinctive traits related to Gf can be approximated to "autonomous" automated processing leading to new data acquisition and new learning based on data initially unknown by the program. Traits related to the know-how are more related to those AI systems dependent on human programming and benchmark modelling. Gf and Gc include both encoding and data processing abilities.

Carroll (1993) developed further classification by identifying 3 intelligence strata [[18]]. According to many scholars, Carroll's "review of more than 400 data sets provides the best currently existing single source for organising various concepts related to intelligence" [[92]]. Figure 1 represents his concept, each stratum being cumulative. Stratum 1 represents specific or narrow cognitive abilities, stratum 2 represents derived broad cognitive abilities, and stratum 3 represents general intelligence, known as the "G factor" according to Charles Spearman [[22]], and defined as "the fluid ability to integrate multiple cognitive abilities in the service of solving a novel problem and thereby accumulating crystallised knowledge that, in turn, facilitates further higher-level reasoning" [[119]].
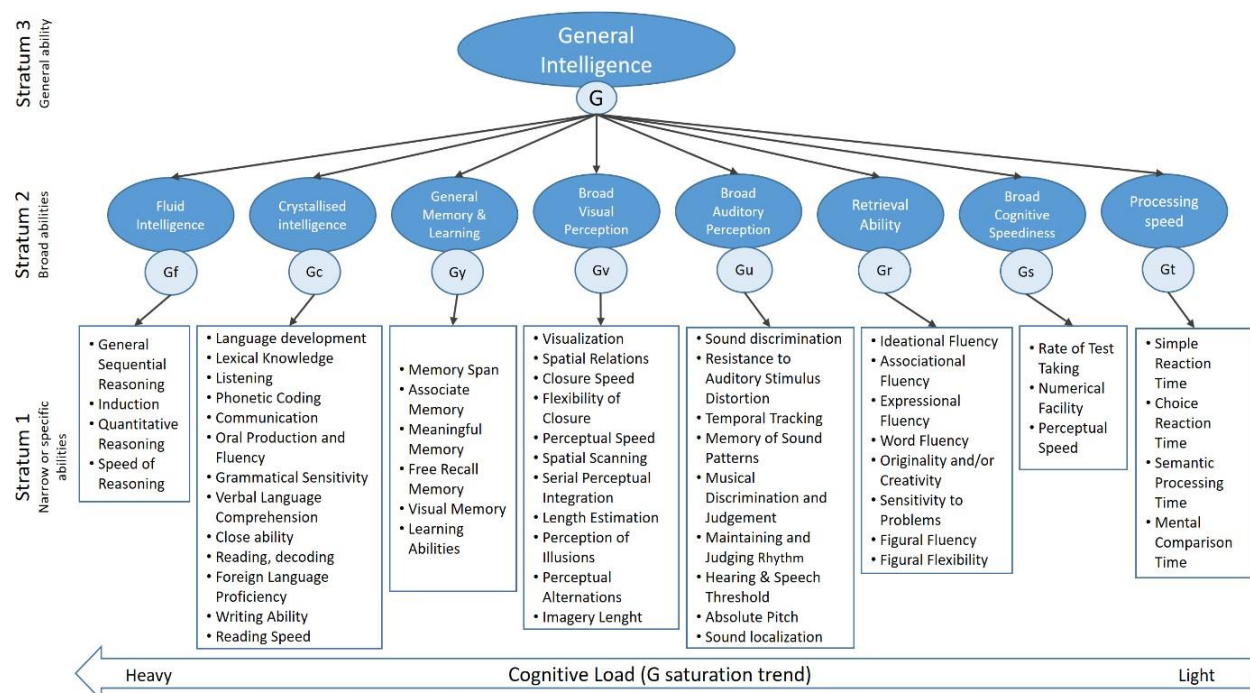

Fig.1: Carroll's Model of Intelligence. Adapted from G.L. Canivez, & E.A.Youngstrom [17].

Nevertheless, in AI, only the distinction between Gf and Gc (Stratum 2) seems crucial and conceptually useful [[36]]. The other features identified by Carroll in Strata 1 and 2 such as processing speed, visual perception, spatial scanning, sound localisation, are technical features translated into hardware and software algorithms, and are finally accessory and not so determinant for identifying the essence of AI. Similarly, regarding other explored psychological theories, such as the Gardner's theory of "multiple intelligence"[1] [[62],[21]], the study of G factors being an open and evolving research field in psychology [[78]], we considered only basic concepts applicable to AI. Nevertheless, such aspects are considered through various psychometric tests as intelligence factors. In AI, they could reveal specific risks leading to question proportionality of the accessory functionalities of a program and device, for adapting technological capacities to what is strictly necessary to meet an identified ethical AI purpose.

For simplifying without contradicting the heart of the psychological theories on human intelligence presented above, we will retain the Cattell-Horn's initial distinction between Gc and Gf essentially focusing on learning, memory, search and use abilities. Their approach allows fine-tuning the global understanding of intelligence factors and could inspire the naming of two broad categories of so-called "intelligent agents" composing AI programs, namely "Crystallised" and "Fluid" AI programs, and by extension, linked devices and software. The added value of such an approach is notably related to the representation that such a naming evokes to lay people. Without having performed an extensive study on this, we tested it with people in our direct environment with no specific skills in AI. In short, this naming essentially evoked reflections on the human control of AI programs and devices. Crystallised AI tended to be understood as a guided version of AI, more controllable by humans, more dependent on humans. On the contrary, Fluid AI tended to be understood as a more autonomous AI, based on advanced abilities, *a priori* necessitating less human intervention. Fluid AI shows creativity, in finding solutions to a problem, adaptability to various environments, evoking more caution towards such programs and devices regarding the perceived potential risks of applications.

The psychologist Bloom's taxonomy classifying human cognitive processes is also informative for understanding intelligent abilities in the meaning of thinking behaviours. Figure 2 summarises them as they have been lastly completed by the doctrine [[58]], and analysed in the field of AI [**Erreur ! Source du renvoi introuvable.**]. We also adapted it in order to integrate the fundamental concepts of Gc and Gf, and to better characterise where a system can be considered mature enough to be able to calculate or compute reasoned decisions. Levels presented in Figure 2 are cumulative, each building on acquired characteristics of previous levels.

---

[1] Gardner, in the 1980s, distinguished eight types of human intelligence namely visual-spatial intelligence, linguistic-verbal intelligence, intrapersonal intelligence, logical-mathematical intelligence, musical intelligence, bodily-kinesthetic intelligence and naturalistic intelligence. Nevertheless, Gardner's theory has been criticised due to its broadness and lack of empirical studies, but it remains popular and illustrative of human talents, personality traits, and abilities.

Lower order thinking skills ➡ Higher order thinking skills					
Crystallised Intelligence (Gc)			**Fluid Intelligence (Gf)**		
Level 1	**Level 2**	**Level 3**	**Level 4**	**Level 5**	**Level 6**
Remember	**Understand**	**Apply**	**Analyse**	**Evaluate**	**Create**
Information storage • Encoding • Knowledge base keeping **Recognizing** • Identifying **Recalling** • Retrieving	**Interpreting** • Clarifying • Paraphrasing • Representing • Translating **Exemplifying** • Illustrating • Instantiating **Classifying** • Categorising • Subsuming **Summarising** • Abstracting • Generalizing **Inferring** • Concluding • Extrapolating • Interpolating • Predicting **Comparing** • Contrasting • Mapping • Matching **Explaining** • Constructing models	**Executing** • Carrying out **Implementing** • Using **Decision-making** • Guided response	**Differentiating** • Discriminating • Distinguishing • Focusing • Selecting **Organising** • Finding coherence • Integrating • Outlining • Parsing • Structuring **Attributing** • Deconstructing **Decision-making** • Guided response	**Checking** • Coordinating • Detecting • Monitoring • Testing **Critiquing** • Judging **Decision-making** • Guided response	**Generating** • Hypothesizing **Planning** • Designing **Producing** • Constructing **Decision-making** • Innovative, unguided response
Evaluation criterion: Given answer is identical to the one that shall be memorised.	**Evaluation criterion:** Given answer has the same meaning than the information to reformulate.	**Evaluation criteria:** The imposed rule has been respected and the answer is correct (based on a unique answer possibility).	**Evaluation criteria:** Given answer is correct and complete	**Evaluation criteria:** Given answer is relevant, scientifically viable and original	**Evaluation criteria:** Given answer is relevant, scientifically viable and original

Fig.2: The cognitive process dimension. Revised Bloom's taxonomy of learning, adapted from Anderson and Krathwohl (2001), [[1], p. 67-68], and J. Parker and S. Jaeger (2016) [**Erreur ! Source du renvoi introuvable.**].

Bloom's taxonomy of cognitive processes could be completed by emotional skills (Bloom's Affective Taxonomy) and psychomotor skills (Simpson's Psychomotor Taxonomy) serving the simulation of human intelligence behaviours, notably in robotics. Parker and Jaeger explore these aspects together and concluded that "machines were able to attain about half the learning abilities that humans have" and "none of them possessed all of them" [**Erreur ! Source du renvoi introuvable.**, p.11-12]. Interestingly, the authors note that they have "seen that in some areas AI outperforms their human counterparts in the lowest rungs of the taxonomy. However, machines still have a long ladder to climb before reaching the top". Bloom's Taxonomy classification adds another interesting level of specification for achieving a comprehensive qualification of AI.

Psychology researchers also studied the **characteristics of human autonomy**, as a part of human self-determination. This concept deserves great attention due to the regular claims from industrials having developed "autonomous" AI devices and the impact that self-determination of AI devices would have regarding ethics and law if it was proved true. In common language, autonomy is defined as the "ability of someone to be independent, to not be dependent on others; the character of something that works or evolves independently of something else" [[83]]. Autonomy supposes to be able to act on its own and to

appeal to intelligence to find a way to achieve goals or tasks usually self-determined by the individual (by AI in our context). Autonomy is an essential component of self-determination that is a broader concept implying the expression of personality traits through autonomy's features of self-regulation, of empowerment and of self-implementation abilities. An individual is thus allowed to express his preferences through choices, to adapt and eventually refrain from behaviours contradicting personal values. Therefore, the autonomy concept seems to be twofold. First, it includes a capacity of something to self-regulate, this eventually leading to action going against the established rule, or against reason (reckless action). Second, it includes a capacity of something to react, resist or adapt to an external environment.

While differences exist between psychological and legal theories[2] of autonomy it is interesting to refer to the **three categories of human autonomy** distinguished in psychology [[97]8], namely the "basic autonomy", the "functional autonomy" and the "general autonomy". Each of them is being **expressed in two spheres**, an "executive" and a "decisional" autonomy sphere and **through different modes**, namely "direct autonomy", "assisted autonomy" and "delegated autonomy". In AI, the notion of "autonomous devices" is highly controversial and we are not pretending here to solve the issue. But with some adaptation and simplification, by taking AI programs or devices as the subject of analysis, we could consider relevant to distinguish between a machine with "basic autonomy", "functional autonomy" and "general autonomy" to question the ethics of AI devices. Calked on human psychology theory, a machine with **"basic autonomy"** is limited to actions or decisions strictly necessary to maintain its functions. In other words, such a machine could self-maintain or self-repair. A machine with **"functional autonomy"** can perform tasks necessary to community life. In other words, the machine is able to have social relationships with its pairs, between machines, or with humans. Finally, **"general autonomy"** corresponds to other activities necessitating autonomy without being determinant for the autonomous status of the machine. For example, the ability to self-entertain, what a machine cannot reasonably do. From a legal point of view, we could consider adapting relevant distinctions initially made in psychology between **"direct autonomy"** expressed without any intermediary, whether human or material, the highest degree of autonomy yet unachieved in AI, **"assisted autonomy"** expressed with the assistance of an environmental management device intended to increase, amplify, expand, regulate or distribute the effort made by the AI program or device and **"delegated autonomy"** where actions and decisions are totally entrusted to the AI by an external agent, here humans, on a case- by-case or task-by-task basis. Interestingly, questioning the autonomy leads to appraise the **dependence level**. Applied to AI devices, the dependence level should be envisaged first as regard to human intervention on machine functioning and decision-making, then, as regard to the natural/technological environment in which the device operates. Psychologists involved in autism [[15]5] identified three levels of dependence based on the assessment of defined behavioural items characterising human autonomy (known as ESDA[3], [[14]]) which could inspire AI stakeholders: 1- partial and discontinuous dependence; 2- partial and continuous dependence; 3- full dependence. This scale could be used regarding general functioning of a device or regarding specific characteristics such as cognitive abilities' implementation. It allows notably identifying a system's incapacities or disadvantages in a particular activity and to envisage human interventions, oversight, or further technological tricks for palliating these impediments to autonomous behaviour.

But in psychology, as well as in law, autonomy supposes willingness from the human subject, an essential ability for being free and responsible. Willingness of a machine is not acceptable, the machine functioning on the basis of advanced directives and methods designed by humans and integrated within their program (in software) or command laws (in robots). The machine does not express its own will but reproduces existing and limited schemes in several contexts more or less known in advance. Therefore, an AI system is not really able to govern its own behaviour. But appearances could easily mislead human observers and make them wrongly believe that the machine acts intentionally. Another limit of the concept of autonomy applied to AI devices is the notion of **independent action.** This particularly "refers to the properties of an entity that is capable of operating independently, without being externally controlled or without inputs (material, energy, etc.) from outside" [[116]]. In practice, any machine, even the more advanced, is dependent on external energy supply, on maintenance operations or on other needs such as an internet connection, a human-based support service. Therefore, so far, no AI-based machine can be considered independent of humans nor autonomous. But innovation goes fast and it is not excluded that real machine

---

[2] Cf. Section 3 regarding the legal notion of "autonomy".
[3] Echelle Stéphanoise de Dépendance dans l'Autisme – ESDA.

autonomy could be achieved one day, at least regarding energetic autonomy. For example, theoretically, a machine could be made independent either through renewable energy sources or perpetual embedded sources, or through dedicated resource-finding, extraction or crafting functions, achieving thus basic autonomy. An important question to address in the perspective of an ethical management of learning machines relates to **machine temporality**, in particular to machine sustainability. Indeed, if the rate of consumption of a finite energy source is so small that the machine would function longer than the average human life we could legitimately argue that such a machine reaches a particular level of autonomy in this regard, triggering special ethical and legal reflections to consider in relationship with other autonomies such as decisional autonomy. Another issue related to AI program or device temporality question their potential to self-improve their autonomy along the time, and to self-direct towards the famous and controversial singularity[4] point where AI systems would be self-sufficient and likely more difficult to control. Kraikivski notes "that machine intelligence can start growing with an initial intelligence capacity that is substantially lower than human intelligence if certain rules for the intelligence capacity growth are implemented in the machine algorithms. These rules include the ability of an algorithm to learn patterns from data, acquire new learning skills and replicate itself. The current computing technologies support the implementation of all of these rules in the form of autonomous algorithms." In such a perspective, even though the author notes that testing could be implemented in order to trial the validity of the singularity theory, he states that "human role should be limited by seeding initial rules for the intelligence growth and providing initial computational resources". The rest should be achieved by the machine. Finally, he concludes stating that "current machine-learning algorithms are already strong in retaining information and recognising patterns in Big Data but weak in deducing new skills and knowledge from it. If artificial intelligence will gain a high capacity in logic, planning and problem-solving, it will still lag behind the human model without the capability of understanding, critical thinking and self-awareness, emotional knowledge, creativity, and consciousness that humans possess. Therefore, building algorithms that are capable of conscious interpretation of perceived information (…) will be crucial for an artificial intelligence to be able to surpass the human intelligence capacity" [[82]]. Ethical reflections around the appropriateness of such developments should consider the observed decrease in human performances (not skills), based on IQ tests, known as the "negative Flynn effect" theory advanced by Dutton et al. [[43]] in 2016, even if this effect does not make scientific consensus and is still discussed.

Nevertheless, it is important to try disentangling complexities around this notion of autonomy in order to serve classification of future AI innovations and advances towards a potential consensus on the (in)existence of "autonomous machine".

This anthropocentric approach to AI based on theories and knowledge of human cognitive psychology could be criticised as it can lead to misleading messages and false expectations about (current) AI [[57]7]. But this is only true if we do not consider the limits of anthropomorphic approaches and the granularity of human intelligence, if we only focus on the General intelligence which, indeed, does not yet exist in a machine. We are aware of these limits and conceptually it makes sense to us to consider these theories as AI pioneers inspired from studies of the human brain and neurobiology to create the field, to innovate, by striving to translate and adapt our knowledge on human intelligence to technological contingencies. In such an approach, AI programs can compose the "artificial brain" of a machine. Furthermore, Gc and Gf are not the monopoly of humans and can be found in other living species. And today AI is not only conceived for devices mimicking human behaviour, but also animal intelligence or abilities, with the "animats", artificial animals [[16]6]. Therefore, it appears that these core notions are pretty useful to conceptualise AI basics in terms of abilities related to intelligence even if, taken alone, they are insufficient to shade lights on ethical and legal issues associated with this technology.

Further elements are necessary to consider in order to envisage tailored qualifications based on a finest typology of AI programs and devices that could lead, then, to legal categories based on legally relevant technical functionalities and effects. For doing so, an exploration of engineering theories is necessary.

2. Considering technical features of AI for determining legally relevant functionalities

---

[4] The singularity refers to an idea that once a machine having an artificial intelligence surpassing the human intelligence capacity is created, it will trigger explosive technological and intelligence growth.

As seen through Gc and Gf, the learning, the memory and the data capture and processing/reasoning abilities are essential to envisage qualifying an "intelligent" action and to start appreciating the potential of an AI technology. But a specificity in AI is rooted in the human control and intervention in these processes. Cognitive Engineering allows to better appreciate AI specificities in these regards and to question the importance of human action in the AI program and device design.

In AI there are two big categories of methods used in programming and driving inputs processing through mathematics formula. First, the so-called **Symbolic AI,** which processes visual/auditory and other sensor information in the form of symbolic representations of factual events. Symbols, as encoded knowledge, can define things, persons, activities, actions or states. They serve controls of AI applications such as movement, orientation, image recognition and recognition of stimuli, memory and speech functions. The Symbolic AI program is rule-based and causal. Using heuristic approach, it is notably used in expert systems, for medical diagnosis purposes for example, where human knowledge of a problem or of a feature is sufficiently advanced to be precisely described. Second, **Statistical AI,** which is based on alphanumerical big data inputs and statistics. Such an AI program is not based on knowledge but on large-scale data from which certain data are extracted and outputs are calculated using statistics. It can serve for example applications performing advanced planning, prevision or recommendation using statistical correlations. Symbolic and Statistical AI are now standard approaches to AI [[13]]. While the approaches are not exclusive, choosing to rely on one or another can significantly question the robustness of the AI and the quality of its results in a given situation. Top level institutes such as the ANITI in France consider going towards integrative AI, using a hybrid AI pattern fed by both approaches for improving cross-understanding of pattern strengths and weaknesses and develop cross-solutions intended to multiply AI potentials and systems robustness [[7]]. To achieve these methods in AI, European experts mention "several approaches and techniques, such as **machine learning** (of which deep learning and reinforced learning are specific examples), **machine reasoning** (which includes planning, scheduling, and optimisation), and **robotics** (which includes, in addition, control, perception, sensors and actuators into cyber-physical systems)" [[72], Glossary].

Machine Learning (ML), like its **Deep Learning (DL)** sub-area, is an essential component of an AI program. It has specific technical characteristics basically focused on a simple question: how best to feed the machine's program, as an "artificial brain", to reach "intelligent" action?
In summary, **ML** "is a method where the *target* (goal) is defined and the *steps to reach* that target are learned by the machine itself by *training* (gaining experience)" [[23]]. The important feature here relates to the training or self-learning function, the latter being understood as the "creation of algorithms which can modify themselves without human intervention to produce a desired output by using structured data" [[80]]. Four ML characteristics have been identified by scholars, namely **supervised learning, semi-supervised learning, reinforced learning and unsupervised learning** [[24]]**.** These four methods are based in particular on the necessity of human supervision and on the possibility to use labelled and/or unlabelled data for training the AI program. Most ML methods rely on Natural Language Processing (NLP) algorithms allowing image, text or sound recognition and classification[5]. In NLP the grammatical and vocabulary databases of the program, together with adequate procedures, will determine "the breadth" of the application in terms of data processing. A recent category of algorithms emerged as Natural Language Understanding (NLU) [[96], [113]], going a step beyond NLP by providing post-processing interpretation functions to the AI program (leading potentially to "highly complex endeavours such as the full comprehension of newspaper articles or poetry passages" [[117]]). In NLU, deep semantic and symbolic skills are crucial, they will represent "the depth" or "the mind" of the application. Some testing of NLU algorithms is already performed in the public internet area, for instance with the last updates of the Google's search engine "BERT" algorithm intended to improve contextual understanding of search queries [[90]]. **DL** can be defined as "a subset of machine learning where algorithms are created and function similarly to those in machine learning, but there are numerous layers of these algorithms - each providing a different interpretation of the data it feeds on. Such networks of algorithms are called **artificial neural networks**" [[80]] because their functioning inspires from the function of the human neural networks present in the brain. While there is always a need for human programming, these technical subtleties have consequences regarding the explainability of the algorithmic logic underlying a result, in

---

[5] Watson AI that played TV game Jeopardy.

particular where deep learning is used. This challenges the transparency regarding users and some other important legal rules and ethical principles.

Therefore, it is both the learning method, and the use of complex artificial neural networks that make the difference in terms of human intervention and explainability of solutions found by an AI program. Also for engineers, as seen earlier, it is foreseeable that advanced ML and DL programs do not need human intervention after initial launching to allow the AI to function. Such AI-based programs and devices could therefore self-instruct, self-behave and eventually self-maintain, provided that they have sufficient decision- making abilities and continuous energy supply.

Based on these learning methods, engineers and scholars highlighted different types of AI-based programs/devices. First, according to their scope, with the notions of **"Weak or Partial"** and **"Strong or Full" AI**. Second, according to their functionalities, with four different **AI Types**.

A starting point for conceiving general categories of AI using such a terminology seems driven by the initial objectives of the AI developers or investors. As Marr schematically notes [[86]], "generally, people invest in AI development for one of these three objectives: 1. Build systems that think exactly like humans do ("strong AI"); 2. Just get systems to work without figuring out how human reasoning works ("weak AI"); 3. Use human reasoning as a model but not necessarily the end goal" (what we can consider "moderate AI"). This approach based on the intention of AI promoters/developers does not really inform about the technological processes at stake which allow inscribing an item within the AI realm and within a special category. Therefore, we are not retaining the first intention of promoters/developers as a critical criterion for classification, but it is an element to consider and eventually readjust throughout the benefit- risk assessment. This could also play a role in judicial trials, notably against misleading commercial claims.

Weak / Partial AI programs are "rooted in the difference between supervised and unsupervised programming. Voice-activated assistance and chess programs often have a programmed response" [[81]]. In other words, Weak AI depends on programming (benchmarks, targets or objectives, functions) and classification. "Here each and every possible scenario needs to be entered beforehand manually. Each and every weak AI will contribute to the building of strong AI" [[23]]. It presents a "human-like experience" but is just a simulation as it is usually focused on narrow tasks. Therefore, for example, Siri (Apple) and Alexa (Amazon) bots are classified as Weak AI programs. The same is the case for the chess program from IBM that beat Garry Kasparov in the 1990s. All rules and moves were programmed beforehand in the machine. Weak AI also includes a notion of narrow AI programs as monotask program. **Strong / Full AI** programs are those tending to fully mimic human brain functioning by using "clustering and association to process data", coming closer to machine reasoning and to the highest G notion, as general intelligence, in human psychology. A strong AI program can go beyond what has been programmed for finding new strategies or solutions to reach an objective. Therefore, once initially programmed and launched by humans, these programs could work with a high degree of autonomy. This category tends to cover semi-supervised or unsupervised programs for which the agent has an extended leeway to calculate and find solutions. Here, the simulation of human brain is questioned as it is foreseen that such AI could be multitask and therefore come closer to human intelligence notions [[106]]. Nevertheless, as noted by AI experts, "there are still many open ethical, scientific and technological challenges to build the capabilities that would be needed to achieve general AI, such as common sense reasoning, self-awareness, and the ability of the machine to define its own purpose" [[71], p.5]. Today, there are no proper existing examples of a strong / full AI program but "some industry leaders are very keen on getting close to build a strong AI which has resulted in rapid progress" [[23]]. Thus we cannot ignore such a category.

AI programs can be classified based on functionalities or technical features allowing us to better appreciate the wide variety of AI-based programs/devices. Four types of AI have been distinguished [[73]]. **Type 1 AI** is a "reactive machine"; **Type 2 AI** is defined by "limited memory"; **Type 3 AI** is defined by social cognition, known as "theory of mind" [[94], [60]]; and **Type 4 AI** by "self-awareness" or "consciousness", a concept still difficult to apprehend [**Erreur ! Source du renvoi introuvable.**, [115], [102]]. These categories are relevant because they inform about potential inherent risks of these technologies. The following table presents and complements the main technical features of the AI program types identified above, each line being relevant for ethico-legal scrutiny. The criteria presented in the table for each type are not exhaustive.

AI Program Functionalities	Type 1: Reactive Machine	Type 2: Machine with Limited Memory	Type 3: Machine able to appreciate the mental status of others (Theory of Mind)	Type 4: Self-Conscious Machine
Use external data to inform action	✓	✓	✓	✓
Short-term memory (data retention <1 day)	(✓)	✓	✓	✓
Medium-term memory (data retention <45 days)	(✓)	✓	✓	✓
Long-term memory (data retention >45 days)	-	(✓)	✓	✓
Use past experience to inform action	-	✓	✓	✓
Understand and anticipate human expectations, emotions, beliefs, thoughts	-	-	✓	✓
Interact socially with humans	-	-	✓	✓
Be self-aware and sentient	-	-	-	✓
Interact with other technological devices (AI-based or not)	(✓)	(✓)	(✓)	✓
Legend of symbols used: −: Not applicable. (✓): Possibly applicable. : ✓Applicable.				

Table 1: A technical typology of AI programs based on machine functionalities

Going a level below in the table, a number of different specific technical functions and mechanical accessories can be enshrined within the device. The technical functions may for example be related to machine vision, NLP, data interpretation, speech translation from text to speech and/or conversely, data structuration, classification, planning etc. All will need at least one related "intelligent agent" to be designed and integrated for reaching an intelligent behaviour and a satisfactory result regarding initial program/device goals. While they are all contributing to cognitive computing, such as shown in human psychology with Carroll's theory, they are not critical for classification. They are only means to collect data, as inputs, therefore implying by default privacy and data quality issues. Nevertheless, their appropriateness and necessity regarding the purpose of the program/device shall be considered in order to appreciate the proportionality of a device functionality and avoid undue invasiveness[6].

Experts in AI disagree on what should be called an "intelligent agent" or "intelligent machine". Different definitions [[100]] have been designed by scholars, with more or less broad scopes, all being essentially based on AI technical features. Taking the human intelligence as a benchmark without some adaptations is not necessarily facilitating the answer due to unknown mechanisms involved in intelligent behaviour as an "agent's ability to achieve goals in a wide range of environments" [[83]]. And AI can contribute to improve understanding mechanisms of human intelligence. Five classes of intelligent agents can be considered, all being composed of sub-agents performing lower level functions and all being potentially combined in a "multi-agent system" for contributing to solve difficult tasks with behaviours and responses displaying a form of intelligence. These five classes are based on perception and actuation capacities [[118], [121]]. **Simple reflex agents** are based on current perception of the environment ("how the world really is now") and on an "if-then" rule. Because they are functioning on a condition-action rule such

---

[6] Cf. Section 3 regarding related ethical and legal challenges.

agents are quite limited in their action because they are unable to act without the appropriate condition being perceived. **Model-based reflex agents** are based on knowledge-model (data representing how the real world functions) and on a "find the rule" principle. Current perception of the environment is used to find a rule that matches with the observed environment in order to trigger the corresponding action. These two first forms of intelligent agents can handle partially observable environments. Options for actions are programmed upstream. **Goal-based agents** are focused on reaching a predefined goal set or a solution that is the close to the goal or desired state. There, the program has more flexibility to assess the situation, to assess the effect of an action and its distance from the initial goal of the device. The program's decision-making ability will need to choose the most suitable solution among the options available (e.g. GPS systems) to reach the goal. **Utility-based agents** are similar to goal-based agents, but include an "extra component of utility measurements" which makes the system able to measure the success rate of an action as regard to the goal. The program can identify the best way to achieve the goal, what is useful where there are multiple alternatives. The agent will have to choose based on a utility map serving to calculate the best way to act. These two forms of intelligent agents require fully observable or at least calculable environments for functioning properly. Engineers face challenges related to uncertainties and refer to the fuzzy logic mechanisms, modelling and programming [[105], [104]] and to Partial Observed Markov Decision Process (POMDP) [[42]] allowing a better mapping of complex and dynamic environments for reaching greater granularity in automated informed decision-making. **Learning agents** can use past experience to learn or has a learning component serving to translate observable environment in actable variables, for example, to consider user feedback for analysing and improving system's performance. Such agents can be integrated as companion agent of other classes of agents. Depending on the purposes of the AI program and final device, agents can be more or less complex and appeal to abilities and factors associated with crystallised or fluid intelligence to meet their aims. Also, except for simple reflex agents which operate in controlled environment, the other agents are questioning the computer's ability to decide about appropriateness of an action or inaction. This could be driven by logic, by statistics, but also by ethical values considered by the human programmer for ethical decision-making.

Any computing device, whether AI-based or not, will take **inputs (data), process** them, and calculate results as **outputs**. There are different issues according to the level of linkage between an AI program and the device actuation in terms of processes used to provide the expected service. The study of the causal link between the AI program and the device output will be particularly interesting from a legal point of view. Technically, challenges for providing or allowing the collection of significant inputs for the system are important, notably for avoiding statistical bias[7]. Similarly, important challenges exist regarding the building of efficient inputs' processing methods and outputs' formulation included in the AI program. Risks of mistakes or unfair actions through AI use are even greater where the AI device has an important actuator in terms of direct and indirect impacts on the environment in which it operates, including impact on human rights, calling stakeholders in engineering to pay attention to the guarantees brought through the technique to ensure AI program robustness. Outputs' variety shall also be designed in consideration of their effects regarding environment and in the perspective of always providing human control over the program or device. Testing the device throughout its development and post-marketing to ensure the sustainability of the system's performance, the quality of the results, and for assessing its predictability should help to meet a satisfactory level of control and confidence. This is of particular importance where data inputs are submitted to environmental evolutions and where the device is expected to act autonomously. On this basis, testing procedures and monitoring strategies for preventing functional mistakes or obsolescence of existing safeguards should be further developed for keeping a high quality of the AI system.

By considering these elements, it is possible to link psychological theories on Gc and Gf characteristics to technological classifications of AI programs based on learning methods. This may be achieved by inserting further criteria to assess the human role and machine functional autonomy regarding cognitive reasoning and technological abilities to interact with the environment, to self-maintain and eventually to self-represent and self-determine[8]. Keeping this in mind, it is now necessary to address the main legal and ethical issues related to AI programs and devices in order to envisage a precise qualification system

---

[7] In particular, the bias of selection and endogeneity. Other kind of bias exist and deserve considerations, cf. Section 3.
[8] Cf. Section 4.

allowing to identify relevant regulatory benchmarks for ensuring the protection of human interests and values.

3. Considering AI legal status for determining adapted legal qualifications and clarifying attached responsibilities

An essential complement to the previous qualification approach of AI tools is considerations of ethical issues and applicable laws representing the limits of the AI-based functionalities and the necessary options to plan for users in order to respect their rights. Such tools having multiple forms and functions should all be considered from an ethical and legal point of view at several levels, from the program set up to the final legal qualification of the device, in a context-specific approach. As mentioned by the EU Commission and EU AIHLEG, "AI-based systems can be purely software-based, acting in the virtual world (e.g. voice assistants, image analysis software, search engines, speech and face recognition systems) or AI can be embedded in hardware devices (e.g. advanced robots, autonomous cars, drones or Internet of Things applications)" [[54], [71]]. Let's go through these layers and underline some important elements to consider, in particular within the EU legal and ethical context, with some examples in the health field.

An AI program, whatever its characteristics and functions**,** should be legally qualified as "computer program" (or software), at least in the meaning of EU law [[48]]. Directive 2009/24/EC, recital 10, states that "the function of a computer program is to communicate and work together with other components of a computer system and with users and, for this purpose, a logical and, where appropriate, physical interconnection and interaction is required to permit all elements of software and hardware to work with other software and hardware and with users in all the ways in which they are intended to function. The parts of the program which provide for such interconnection and interaction between elements of software and hardware are generally known as 'interfaces'." The author of the program and editor of related software can claim copyrights on their creation, as literary works. Conditions attached to such protection claims apply to "the expression of a computer program in any form, but not ideas and principles which underlie a computer or any elements of it" [[52]]. The program must be original and represent the author's own intellectual creation. The holder of the rights may control the reproduction of the program or of a part thereof, the translation, adaptation, arrangement and any other alteration of the program as well as its distribution. As such, ethical and legal responsibilities related to AI program development and production rely on the manufacturer's research and development teams including engineers, developers, programmers and designers [[85]]. It should be performed with due consideration of other disciplines, notably those specifically concerned by the AI program, such as future users' representatives, and with other professionals such as lawyers. AI ethics is a recent field elaborating theories, principles and rules for the AI sector. Considering the ethics of AI algorithms is of paramount importance for reducing risks of misuses or side-effects of the technology before being placed on the market. Performing this ethical assessment is mainly the responsibility of the engineers, developers and designers. The end-users and general public should also be consulted as far as possible and their opinion should be duly considered. Several statements about AI ethics have been issued by different non- governmental and governmental public entities (e.g. Declaration of Montreal [[112]]; EU [[72]]) or private organisations (e.g. IEEE [[77]], ACM [[8]]). In the EU, the AIHLEG guidelines fix the ethical principles to scrutinise and provide an assessment tool to the stakeholders' community. The high level expert group identifies a global "human-centric" approach to AI developments and seven key requirements that should be met for achieving a "trustworthy AI". "Human-centric" approach "strives to ensure that human values are central to the way in which AI systems are developed, deployed, used and monitored, by ensuring respect for fundamental rights [...]" this including "consideration of the natural environment and of other living beings that are part of the human ecosystem, as well as a sustainable approach enabling the flourishing of future generations to come" [[72], p. 37]. The Human being is conceived as a referee deciding whether AI action is ethically sounded (desirable), lawful and finally reliable in a specific context (evidence-based approach) this leading to reinforce trust within AI. Therefore, **"trustworthy AI"** is defined by three components. First, it should be **lawful**, ensuring compliance with all applicable laws and regulations. Second it should be **ethical**, demonstrating respect for, and ensure adherence to, ethical principles and values. Third, it should be **robust**, both from a technical and social perspective, since, even with good intentions, AI uses can cause unintentional harm. AI ethics includes horizontal (cross-sectorial requirements) but also vertical aspects (sectorial requirements) to be considered through the design of the program and related devices.

Several important technical challenges to reach trustworthy AI with embedded ethical algorithms exist [[108], [33]]. Nevertheless, if human values could not be encoded directly within the algorithms, the regular human monitoring of ethical AI inputs, processes and outputs must be foreseen as fundamental blocks of safeguards pursuing trustworthy AI.

A common challenge to the trustworthy AI achievement relies on the intrinsic relationships between humans and AI. Three main categories of biases affect AI: cognitive bias from the programmer, statistical bias, and economical bias [[107]]. Because of human programming, at least for the initial algorithm, the same **bias** as for human thinking tends to be conveyed through AI programs which will automate them. As BBC mentions, "in short, AI is much more human than we ever realised. Which is perhaps the scariest notion of all" [[11]]. As an illustration, we can remember the bad outputs from the algorithms of Google [[29]] and Twitter [[65]] which showed to be biased and, at use, sexist and racist. Human stereotypes, empirical inequalities and other cognitive biases can easily be translated within the digital AI environment, through the training data used by AI and specific algorithmic instructions of data processing. This is particularly the case where the AI program is used in an open environment such as the internet or is used for politics, justice, police and other investigation areas focused on humans, health included [[93]]. Hence the need to develop upstream human intelligence and understanding of ethico-legal challenges in AI through education. This should ease responsible and ethical AI developments and ethical downstream human checks of AI processing, avoiding automating human bias at large scale [[55], p. 11]. In this perspective, the AIHLEG suggests to use "red teaming" to regularly challenge and finally improve the AI programs during Algorithmic Impact Assessments to be implemented throughout the AI system lifecycle [[72], p. 20, 37] [see also [31]]. The paramount role of ethics in the responsible development of AI programs and devices must be taken seriously. The results of these assessments about the acceptability of a specific AI program could be surprising, in particular if it dramatically improves the respect of human rights while replacing humans. At a certain point, we could also envisage AI to self-maintain ethical behaviour and maybe find new ethical ways of resolving persistent issues. Depending on the ethics referential mobilised, European AI could differ from US, Japanese or Chinese AI. In the field of health and biomedical research, many internationally recognised ethical standards could be used as the common ethico-legal benchmarks for AI developments embracing a global ambition (e.g. UNESCO, WHO, CIOMS, WMA declarations on ethical principles, Council of Europe International Convention and Additional Protocols on Human Rights and Biomedicine and Recommendations). Of note, the UN recently engaged a much needed action and is taking steps with regard to the ethics and regulation of use of personal health data. A first set of Guidelines [[110]] has recently been published, with a high potential for streamlining AI innovations and protecting international end users in this essential and sensitive field.

An AI-based system is composed of three main capabilities according to European experts: **perception, reasoning/decision making, and actuation** [[71]]. Perception as data acquisition or capture ability can be achieved through various interfaces (e.g. cameras, microphones, keyboards) or sensors (e.g. measuring temperature, pressure, distance, force/torque, rates of certain products, presence or tactile sensors). Reasoning and decision-making relates to the machine's ability[9] to interpret the data in a consistent way in order to envisage an action. It supposes an information processing module intended to make the data understandable for the machine. In any case, AI devices must be legally qualified as data processing[10] devices. Hence existing data protection laws applying to personal and eventually public data [[50]] must be respected. Actuation is the machine's ability[11] to execute the action it identified as being appropriate to reach a defined goal. Actuation can be performed through actuators, whether physical (e.g. articulated arms) or digital (e.g. activation or denial of a software function) and change the environment, including by producing legal effects regarding third persons or goods. The notion of "robot" [[19]] is at the heart of reflections about the AI devices. According to CERNA, the robot is a material entity including software.

---

[9] Cf. Section 1 regarding the concept of "intelligence".

[10] A data processing can be understood by reference to the definition provided by the EU General Data Protection Regulation 2016/679 (GDPR) Article 4(2) which, by extrapolation, can usefully apply to all data. The GDPR defines data processing as "any operation or set of operations which is performed on […] data or on sets of […] data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction".

[11] Cf. Section 1 regarding the concept of "intelligence".

This excludes purely software-based AI such as conversational bots. The capacities of the robot to move, interact and decide by itself are conferring to it a certain degree of autonomy[12]. But as we will see below, such autonomy should be relativized and nuanced when it comes to the legal autonomy notion.

Qualifying the AI-based device in law questions the **legal nature** of the AI device. Is the machine a good (Res) or a person (Personae)? In 2017, the EU Parliament [[41]] proposed the creation of a digital legal personhood for "autonomous robots", a notion of "an electronic person" consisting in a legal regime close to the corporate personhood applied to legal persons, "where robots make autonomous decisions or otherwise interact with third parties independently". This was including a proposal for a new mandatory insurance scheme for companies producing such devices in order to cover damages caused by the robots. However, this controversial proposal has been criticised [[40], [88]]. Some national advisory bodies close to regulators even departed from the EU report's recommendations (e.g. in France with the OPECST [[39]]), mainly because such a regulatory perspective would diminish the responsibility of the different persons acting upon the commercialisation of the program or device (programmers, manufacturers, operators). A majority of legal experts definitely tends to consider AI-based devices, including the so-called "autonomous" robots, as legal goods and only goods. Why? Because the robot's full autonomy is a fiction. Based on the difference between technical and philosophical autonomy notions [[60]], Daups states that, **in law, automated behaviour must not be confounded with autonomous behaviour** [[37], [33]], [[38], p. 84-87]. Automated behaviour is rooted in a panel of programmed operations allowing a machine, in defined and limited cases, to execute an action without the need of human intervention. But the action itself and the method is mastered by humans and does not appeal to creativity, but to responsiveness. Adaptive ability in responses to environmental evolution is usually not present or not so important due to the controlled environment in which automates are placed. Autonomous behaviour appeals to intelligence, reasoning and enhanced adaptability of the machine's actions outside the normative human control. Based on Daups' proposal in the context of robotic law, a machine's autonomy could be defined in substance as its capacity, in the meaning of the immaterial and material possibility[13], from its own will[14], to decide, in the frame of its competence, about a legal act that is enforceable to thirds and punishable, and whose responsibility can be related to a legal or natural person. The concept of machine autonomy relies on the hypothetic ability of a machine to be able to act appropriately in unplanned situations or unprogrammed scenarios, this being only, to date, a human prerogative. Indeed, machines cannot be considered as having a philosophical autonomy because they cannot create their own law, their own values, or algorithms, nor their own purposes. In fact, even if after initial programming the machine is able to develop its own behaviour based on new data acquisition, this will only be based on data recognition, not on plain understanding, not on self-acquired knowledge. It is not entailing willingness nor capacity to self-decide about new values, as matrix rules to perform unprogrammed choices, and this is not constitutive of legal personality or capacity. Without a highest degree of intelligence based on related intelligent agents determining wider potentials for machine's inputs and processing self-design, a machine cannot be presented as reaching full autonomy. Thus it cannot achieve philosophical emancipation regarding pre-existing states allowing to reach a plain self-governing state, a sovereignty. It might be oriented towards such (very hazardous) behaviour by humans, but so far none reached such a philosophical autonomy, remaining thus in a variable level of automating. According to this approach, a machine cannot (yet?) be legally autonomous, implying the impossibility of a machine to perform legal acts[15] by itself in the absence of intrinsic philosophical autonomy and willingness. For instance, a machine cannot provide valid consent on its own behalf. But an AI device can help humans to convey a human-sourced legal act, to formalise or to implement legal acts by being driven by programmed procedures and criteria, in an automated way (e.g. with smart contracts [[15]]). Therefore, considering a machine as personally accountable is a fiction, the machines being anyway programmed by a human. They should only be tools under the human responsibility. Calculated solutions, even if self-determined within a panel of options, are used as a support of human purposes, in human decision-making or into human-controlled

---

[12] Cf. Section 1 regarding the notion of "autonomy" and its granularity.

[13] Referring respectively to the machine ability of perception and of actuation according to Daups.

[14] Meaning from the machine capacities of appreciation and judgment leading it to select and execute a program deemed adapted to the situation.

[15] Defined as voluntary act that generates legal consequences sought by their author (in our context the machine). For example, contracting.

environments, falling thus under humans' laws such as contract law. While algorithms and AI programs leading to the adoption of a particular solution or pressing humans to adopt a behaviour have a normative effect [[10]], this effect should not be solely attributed to the machine. It appeals to responsibilities of humans that intervened in the process that led to adopt the problematic solution, this ranging from the manufacturer and programmers to the end-user by passing through persons in charge of the program or device maintenance. Therefore, the machine should not be considered as having a legal capacity, nor as being a person with legal personhood allowing it to be source of legal acts, even if this would simplify the legal reasoning in case of prejudice. Keeping this in mind, because an automated machine can execute legal acts commanded by humans, on their behalf, they can generate involuntary legal facts[16]. These could be related to defects in usual technological processes or be caused by external unplanned or unforeseeable events. In this regard the EU law on general product safety [[47]] and on defective products [[46]] establishes horizontal responsibility regimes of interest which could need some adaptation regarding AI, but which already plan relevant legal rules protecting end users. Theoretically, we can highlight a slightly different conceptual understanding of the notion of autonomy in law and in psychology. In law, the focus is made on willingness' autonomy and does not consider intelligence per se. This legal approach would better fit to the broader notion of self-determination defined in psychology, which includes autonomy among other criteria allowing to qualify something autonomous but not self-determined. In psychology, autonomy is focused on the cognitive ability to accomplish tasks, this including the notion of functional autonomy in the task achievement, a notion further explored through engineering regarding technological actuators and machine's material autonomy.

As a result, we consider that, at least in the EU legal order, AI devices are first of all legal "Res". They are things, as goods in the meaning of "products which can be valued in money and which are capable, as such, of forming the subject of commercial transactions*" [[26]]. Most of existing AI devices are **assisting human actions**, even when they replace a human factor through task automation, they remain cooperation tools. Robotic autonomy, as well as robotic consciousness, does not exist, and, from a legal point of view, could never exist. Therefore, any personal rights given to existing AI seem themselves legally artificial. The current realistic legal approach of AI programs and devices aims at clarifying and, only if needed, at creating the rights and obligations of the humans behind the machine in order to establish a consistent and practical responsibility regime. This latter will probably prove complex to apprehend at first look as it necessitates to approximate several existing and often complex legal regimes. In our opinion, the driving legal principle in AI regulatory approach should be the subsidiarity of any AI program or device to humans, as a legal good, leading thus to shared competences and responsibility regimes solely attached to humans.

This approach is not universally accepted. Indeed, certain States have accepted to give some rights to "autonomous" robots. The case of Sophia, a humanoid manufactured by Hanson Robotics and the first robot to officially get Saudi nationality and citizenship on 25 October 2017 marked the century [[70]]. Similarly, we can mention the case of a Chinese man who married a robot he created himself [[68]], a practice foreseen to increase by 2050 [[20]]. But does this mean that robots shall have rights similar to human beings? Again, this probably does not make sense. We could for example envisage providing certain rights such as robot integrity and dignity calqued on what already exists for animals. Indeed, in many jurisdictions, animals are qualified as "tangible property" on which the owner can have property rights and guardianship duties. In EU countries like in France, animals are recognised as sentient living organisms legally qualified as special tangible goods. Thus, animals have an opposable right to dignity protection [[59]] obliging humans to provide an ethical attention to animals and to respect them in activities involving animals, including scientific research [[49]]. Such a special category also protects animals from humans, from bad treatments and violence. If conscious AI devices, humanoids or animats, exist one day, creating such an intermediary legal qualification and related rights, based on animals' protection regime or on a new one like the status of an electronic person[17], could make sense.

---

[16] Defined as a voluntary or involuntary event that generates legal consequences that have not been sought by their author. Examples of legal consequences: creation or modification of rights held by legal persons. For example, the birth and death are legal facts.

[17] As proposed at the EU Parliament level [[41], Point 59]: "creating a specific legal status for robots in the long run, so that at least the most sophisticated autonomous robots could be established as having the status of electronic persons

If AI devices are goods and only goods, the next issue that is important in most jurisdictions is to define whether they will be mobile or fixed by nature or by destination. In some contexts, the notion could be evolving and an AI device (such as prosthesis) could even integrate the body of an individual becoming then part of the legal person.

Qualification according to the **purpose of the device** brings interesting elements to the reflection. Several actors could have critical role in determining an AI program or device purpose. Defining the purpose of a program and of a broader device are challenges as such. In any case, AI is pursuing OUR objectives, making us responsible for the trajectories and tasks defined and delegated to the machines. The manufacturer of the device, as well as the AI-program developers, engineers and programmers, will be the initial actors in the definition of the task and purpose of their product. First, they need to identify the broad economic sector they are addressing with their creation in order then to determine the technical means to involve and the desired functionalities to include in the AI tool. A broad identification of the economic sector is necessary according to the three sectors identified in economy science [[27]], namely primary (agriculture), secondary (industry) and tertiary (services) economic sectors. By going a level down, the identification of the targeted precise branch of activity will be necessary. The device could be addressed to actors in different fields as finance, business, law, health, insurance, telecommunication, tourism or research. By taking health as an example, the device could be used in administration of health services, for healthcare provision, for health surveillance etc. Each wide sector and branch has legal and sometimes ethical benchmarks to consider for building the algorithm and related program, software and device [[33]]. Second, they need to inscribe their creation in a specific product group [[53]] to which specific quality rules are attached. In the EU, as in other parts of the world, the manufacturers must ensure that "products placed on the extended Single Market of the EEA are safe. They are responsible for checking that their products meet EU safety, health, and environmental protection requirements. It is the manufacturer's responsibility to carry out the conformity assessment, set up the technical file, issue the EU declaration of conformity, and affix the CE marking to a product. Only then can this product be traded on the EEA market" [[53]]. Despite the lack of AI-specific technical quality norms that would include AI ethics check, international quality standards have an important role to play in the global governance of AI as notes P. Cihon [[25]]. Existing ISO norms, although unspecific to AI, are of interest to ensure the quality of materials, products, processes and services. For example, AI programs used in self-driving cars will be integrated in the broader legal category of vehicles. An AI program serving military purposes will be legally categorised as a military good, it could ultimately be qualified as weapons, surveillance, tracking or detection devices for example. AI programs used in puppets will be related to toy regulations. Connected devices will also have specific requirements to fulfil. Indeed, AI programs can evolve in various environments including more or less risks regarding users, persons concerned and society. Basically, it can either evolve into a **closed-to-the world environment**, without access to the public internet and using a limited number of identified and controlled databases, or into an **open-to-the-world environment** where the program is connected to the public internet, and uses uncontrolled databases, including through the use of observational sensors. No matter whether AI is the driver device or an accessory, quality norms and legal rules attached to the development, commercialisation and uses including limits fixed for protecting public interest will apply to the AI device as a unique good. Questioning the purpose of a program leads to identifying the **target users** of the final product. This is also relevant in the quest for applying a relevant ethico-legal framework. AI devices, including software, can be designed for professionals or consumers. Their leeway in the use of the program or device, including regarding functionalities and purpose modification, shall also be addressed by manufacturers in order to define safeguards limiting potential misuses, by default, such as through Application Program Interfaces[18]. This distributive approach in the qualification of AI products has been recently recognised by the EU and further detailed within the 2020 White paper on AI [[55]]. Nevertheless, some special features would require more research from a legal perspective.

The **modalities of use** of the device are to be considered as an important element. As we said, AI

---

responsible for making good any damage they may cause, and possibly applying electronic personality to cases where robots make autonomous decisions or otherwise interact with third parties independently".

[18] An application program interface (API) is a set of routines, protocols, and tools for building software applications. It is a specification of possible interactions with a software component.

technologies are intended to either assist or replace human action. **Assistance technologies** can be subdivided in two groups based on program/device functional behaviour: executive technologies or vector technologies. Executive technologies perform automated actions fully commanded by algorithms strictly defined by humans to achieve specific tasks (e.g. an algorithm delivering a precise quantity of a product like insulin at determined time intervals). Vector technologies are those that are dependent upon and adaptive to human action (usually physical action, e.g. driver assistance systems or surgical robotic arms). These technologies are assisting/corrective technologies prolonging a human action thanks to AI programs. They are activated on command, and limited to the human action they will act on. **Replacement technologies** can be considered as fully automated programs/devices which do not need humans to work and achieve their goal, to decide, entirely or partly, when to start, when to stop an action, when or how to learn and to implement a new way of acting. Decision-making ability is therefore central here. The machine's decision considered here is the one that produces legal effects for humans, users or not, regardless of the result. For being considered as a decisional device, the technology shall produce a judgment based on which action or inaction is implemented. In law, as we have seen, the autonomous character of a decision is an important criterion to distinguish between autonomous and automated decision-making and for attributing this latter to a particular actor allowing identification of corresponding responsibilities or, most probably, responsibility chain. Whatever the supposed degree of AI decisional freedom that could be admitted in certain States, AI decisional action should always engage human responsibilities and be conceived as the result of previous external legal decisions taken by humans regarding the AI device. These decisions concern AI device elements (sensors, algorithms…), purposes, terms and conditions of use, promotion, and monitoring of computational methods, of technical processes enabling the device. Daups suggests that, "in fine, the guardian of the autonomy would be the effective responsible of the damage provoked to thirds". The guardian of the autonomy is further defined, in substance, as the legal or natural persons that have the material and the immaterial control of the machine autonomy [[38], p.115-120]. Several persons could qualify, notably the end-users, the employers, or the device's rental service establishment. Other persons involved could be the manufacturer of the device, the designers, the programmers etc. Several levels of responsibility could then be determined according to the causality link between their acts/inactions, legal duties, ethical prescriptions, and the damage occurrence. The capacity of the human agents to constrain the level of autonomy of the machine will necessarily be considered to answer the question "who had the authority"? Modalities of use also rely on **subsidiarity games which can be evolving** and crucial. One AI technology could be at a certain time considered as executive and at another time as a replacement technology, this mainly depending on device autonomy. This interplay is normally guided by humans, but we could also imagine autonomous systems such as with airplane piloting AI programs. In certain circumstances a human pilot could be the master, in others, AI could take the lead and eventually keep it even in presence of human-sourced counter indications. It is precisely for regulating such delegations of powers in AI interfaces that AI ethics codes would be important, provided that it is feasible to program ethically. Indeed, such tools should ensure that the AI will take and let the control at the right time, for the right reasons, and establish "which decision-making procedures can and cannot be transferred to AI systems and when human intervention is desirable or mandatory" [[56]]. This remains an open research area.

AI legal effects should be envisaged through **testing, use-cases and specific benefit/risk assessments**. These assessments should be part of the ethics-by-design and cover, among others, the effects of the use of the final device on user behaviour. Which values do these systems effectively and demonstrably serve? Is the AI device misleading the user in the realisation of a legal act and which are the measures to reduce potential misleading effects (e.g. bias; authority effect…)? Is it creating a disruption on essential elements of human being protection (e.g. an AI used for deciding about granting a right to an individual)? Is the involvement of the AI-based device creating a new legal act as such (e.g. AI self- decision making)? Will the end-user be able to modify the program in order to personalise the use experience or to influence the functioning of the AI device? What is/will be the broader societal impact of the program and device use and spreading? How is privacy affected and respected? These examples of questions should be tackled collegially, interdisciplinarily and where possible with end-users' representatives. Identified risks should be accompanied with detailed measures intended to avoid, reduce or eliminate the risks. The efficiency of such measures should be assessed upstream and downstream as part of the quality management process. As far as possible, testing methods, data and assessment results should be made public.

For example, in health, the patient-physician relationship is crucial in many instances due to the fact that only humans can really understand the complexity of humans' behaviour, health status, potential influences on individual choices, and can interact with the patient properly, such as for communicating health information in the respect of biomedical ethics. Nevertheless, in specific instances, the human action is not perceived as essential to protect the rights and interests of the patient. This is the case for performing the medical data analyses, often complex and time-consuming. The literature about AI medical devices shows impressive results demonstrating the added value of using AI compared to human experts in certain tasks (e.g. identification of tumour cells). Even though using AI in clinical settings has the potential to decrease prognostic or diagnostic errors, by considering the complexity or scarcity of certain medical conditions, AI uses could also entail risks of mistakes and lead to wrong conclusions (e.g. due to bias in algorithms, lack of quality of reference data, or automated misinterpretations). Therefore, the appropriateness of the use of AI system must be assessed case-by- case. In certain countries, whatever the use case, sectorial assessments methodologies [[69]] start to be elaborated. Of note, the human action of competent health professionals is considered essential to review, confirm or eventually infirm solutions found by AI and to inscribe it in real life through concrete measures which will benefit the patient. Such a position affirming the necessity of human control, transparency and auditability of AI systems, in particular where they have a meaningful degree of autonomy, appears essential in a European ethical context [[43], p. 9-10] [[51], Article 22]. In France, the position of the National Council of the Physicians' Order (CNOM) [[28], p. 20-28 and p. 48-52] and the one of the State Council [[30], p. 207] on the matter have been echoed in the current debates on the bill of bioethics law.

Based on a clear classification of AI devices, it is important to consider **AI-based systems** which raise specific issues. According to experts, "trustworthy AI" must comprise "the trustworthiness of all processes and actors that are part of the system's life cycle" [[72], p. 5, p. 37]. Defining case-by-case the **actors involved** in an AI system will be important to design shared responsibilities regime. The EU Parliament, through the Delvaux Report [[41]], identifies device "manufacturer", "the operator" (that can be a distributor, exporter, importer), the owner, the user". This list should nevertheless be completed with some more granularity, in particular behind the term "manufacturer" that should include the AI developers, engineers, designers and programmers who can be different entities from the final manufacturer assembling the device. The notified bodies undertaking quality control over the device prior to the marketing and post-marketing should also be considered as important actors. Advertising companies that can sometimes create hype, false expectations or misleading messages for commercial purposes around AI should also be considered from an ethico-legal perspective. On this basis, existing liability regimes and liability distribution rules can provide an adequate framework for responsible AI development, marketing and use.

In clinical settings, the main actors whose responsibility regarding patients could be engaged are the physician in charge of the medical act, the hospital, the AI device producer/manufacturer and the AI programmer. Each of them has different responsibilities at different levels which could potentially be engaged in the court, case-by-case. In research settings, the participant could try to engage the responsibility of the researcher, of the research promoter or institute in charge of the management of the specific project action that generated the damage, the AI device producer and the AI programmer. Here again, each of them has different responsibilities at different levels. Interestingly, each actor could look for engaging the responsibility of the others depending on the circumstances. The main difficulty for the patient or research participant side will be to identify the situations where AI devices have been used and to demonstrate a causality link between this tool and the damage. Hence the necessity to enhance transparency about the use of such technology during the informed consent process. Further ethico-legal issues are today well identified such as risks of **discriminations**, AI **explainability and robustness**, including in the field of health [[12]].

More generally, it should be recommended that AI innovation's stakeholders reflect on the broader societal impact of their AI programs and devices, in a systemic or epistemological approach. This should notably cover ethical issues related to the ubiquitous development of AI, to the product's capacity to modify social relationships, personal identities, relationships with the environment and to related consequences in short and long terms. Engineers should be able to express their views about the ethics and desirability of a given development to which they contribute, and be protected if they alert the public

authorities of any doubtful practices. Policy-makers and regulators should continue paying attention, for example, to new human dependencies on technologies, including the problem of **excessive trust** in AI-based devices resulting from their uses. Attention should also be paid to what we call here the "AI **imperium effect",** designating the phenomenon of progressive divestment or abandonment of the human capacity to act to the benefit of complex automated AI systems, the related potential risks regarding AI paternalism on human dignity and psychology [[120]], as well as the new power games underlying the AI market [[104]]. Such reflections should serve awareness-raising of all AI actors and complete other targeted assessments. They could lead to the elaboration of further risk mitigation strategies, initiatives, collegial opinions, or policies at several organisational levels, from a company to the State.

Ethics by design is necessary but will not suffice alone. Legal safeguards should be clarified for ensuring appropriate protection of human rights as an essential part of the AI consumer new deal, in particular through **post-marketing monitoring**. This includes environmental and social responsibility in AI development and use. In this regard, we presently note the scarcity of relevant labels and related audit processes. However, efforts are ongoing in that way[19].

4. Proposal of an integrated matrix for qualification of AI programs and devices

Figures 3 to 5 summarise the main elements for qualifying an AI program/device. We provide instructions for use in this section.

First step consists in inscribing the AI program in one of the broad categories of the Figure 3 based on identified AI program's or device's general purpose, needs, or its already known features. A given program or device can be inscribed in both categories and the qualification can be segmented regarding the processes implied for one task and not another. Nevertheless, the main processes at stake in normal AI functioning should trigger general qualification of the device (e.g. an AI program only uses Gc mechanisms for a quarter of the tasks while more than half of the tasks operates on the basis of Gf. In such a case the qualification to retain is Gf and, thus, Fluid AI).

Crystallised Artificial Intelligence	Fluid Artificial Intelligence
- Basic Learning Abilities (encoding) - Memory abilities - Ability to use acquired knowledge/experience to find solutions and base actions (retrieval) - Described as "verbal intelligence" of human beings based on "know-how"	- Advanced learning abilities (encoding + full understanding) - Ability to use, transform, generate different types of novel information in real-time (creativity) - Abstraction and Adaptive skills - Ability to have independent process from acquired knowledge, experience, patterns - Described as "logico-mathematical" intelligence of human beings, a more instinctive intelligence
Main limit: Cannot serve to find solutions to new situations without reference materials related to these new events, such as patterns, operational database	Main limit: Needs crystallised intelligence basics in order to properly function

Fig. 3. Broad qualification of AI Core Elements according to Intelligence Mechanisms adapted from Cattell/Horn/Carroll's theories.

Then, the procedure is to go through each column of the Figure 4 and Figure 5. In some columns, qualification will be alternative (item must enter in one or the other class, exclusively). In other columns, the qualification could be cumulative (the second class covering automatically aspects of the first one). These modalities are specified at the end of each column by an "A" (alternative mode) or "C" (cumulative

---

[19] E.g. Algorithm Data Ethics Label (ADEL).

mode).

Of note, regarding the assessment of autonomy as a functional behaviour characteristic, we have seen that there is no universal concept of autonomy whether in psychology, in engineering or in law. But each discipline has an entire literature providing criteria and clues allowing to assess and document autonomous behaviour at general and specific levels. In sum, the assessors should first tackle separately aspects related to cognitive autonomy (main reference: psychology), to functional and material autonomy (main reference: engineering, robotics) and to legal autonomy (main reference: law) for assessing the program or device abilities and envisage more precise qualification. In terms of functional purpose qualification, by taking the example of a chatbot using an AI program for human-machine conversation, the relational purpose could qualify as informational only but could also cover aspects qualifying predictive, propositional of even self-decisional aspects, depending on case-by-case analysis.

Going through Figure 4, it can be assumed that the more an AI program or device inscribes in the lower categories of the table, the higher are risks related to their use.

	General Category	Cognitive Level	Technology Type	Learning Method	Functional Scope	Functional Previsibility	Functional Behaviour	Functional Purpose
AI Program or Device	Crystal-lised AI	L1	T1: Reactive	Supervised	Narrow / Monotask	High	Executive	Implemental
		L2						Informational
		L3	T2: Limited Memory	Semi-supervised		Medium	Vector	Predictive (estimation)
	Fluid AI	L4	T3: Theory of Mind		Broad / Multitask			Propositional
		L5	T4: Self-consciousness	Reinforced		Weak	Auto-nomous	Self-decisional
		L6		Un-supervised				Conscious
Modality	A	C	A	A	A	A	A	A or C

Fig. 4. Qualification regarding AI product's performances and technical features

The next step is to classify the ethics and legal aspects related to the product based on Figure 5. Each column must be addressed one by one and necessitates specific assessments and developments from the producers, before commercialisation. For a quick explanatory quotation of qualification only the central column (Legal qualification) of Figure 5 can be used. The left and the right columns are indicators of ethical benchmarks that should be adapted to the context in which the AI is developed or to the rules applying in

the territory where the program or device will be commercialised.

	Ethical purpose Based on EU "ethics by design" principles	Legal Qualification			Specific protective measures Based on EU "Human-centric approach" for "trustworthy AI"
		1. Technology		2. Sectoral regulations / quality norms	
Qualified AI Program or Device (Fig.4)	Beneficence, Sustainability: "Do good" (e.g. meet UN Sustainable Development Goals) Non Maleficence: "Do not harm" (e.g. side-effects assessment) Autonomy of Human Being: "Preserve Human Agency" Justice, Equity, Solidarity: "Be fair and accountable" Explicability: "Operate Transparently" **Evaluation criteria:** contextual and periodic risk assessment (professional, individual, populational levels)	Good ("Res") **Evaluation criteria:** private/public; related property and responsibility regimes	Software	Technical rules Responsibility rules Service provision regulation Deontological rules Technological transfers regulations	Accountability Data Governance Design For All Human oversight Non-discrimination / Non-stigmatisation Respect (or enhancement) of Human Autonomy Respect for privacy Robustness Safety Transparency
		Sentient Good **Evaluation criteria:** private/public; related property and responsibility regimes	Hardware (physical device or robot, e.g. humanoids and animats)	**Evaluation criteria:** depends on the target sectors/users (e.g. consumer or professional goods), claims from manufacturer, and on specific product classification (e.g. medical device, toy, manufacturing machine, profiling, surveillance device, military or dual-use device)	
		Person ("Persona") **Evaluation criteria:** legal or natural person; related responsibility, rights and freedoms			
Modality	C	A	C	A or C	A or C

Fig. 5. Qualification according to identified ethical and regulatory compliance achievements or requirements for the AI product.

As a rule, the ethical and legal benchmarks identified and used for the qualification and assessment should be recorded. Each ethical or legal principle and rule at stake and measures taken to comply with should be documented. Related compliance measures too. At the end of the assessment, stakeholders could decide to qualify their product as ethical and legal.

It is important to remind that the result of the risk assessment performed according to this framework should be available to authorities and made public for ensuring transparency. Also, it should be made clear that in case of important residual risks for users or other persons impacted by the AI program or device which could not be contained or reduced, the stakeholders should refrain from going to the market before finding acceptable solutions. This important rule should be extended to devices which successfully passed the risk assessment, but which reveal defective or disproportionately risky at use. Therefore, we remind here an underlying and unwritten principle of any risk assessment, the **precautionary principle**. This latter shall lead to continuous technological refinement for reducing or eliminating risks or to renunciation of the product, this including the product's market withdrawal. This obligation relies first on manufacturers and second on authorities.

In order to trial the proposed matrix, we will give an example of a potential qualification of Alexa AI. Alexa, the voice command service that powers Amazon Echo devices, is an AI-based chatbot, a conversational agent, a personal domestic assistant, marketed and available since 2014 as an integrated part of Amazon smart devices series [[5]]. Alexa is a cloud-based service which can operate within the Amazon products ecosystem and with compatible third products.

Alexa is a "pseudo-dialog" assisting technology relying on natural language processing models to "understand" vocal commands (translate basic commands from the user), and speech analysis/synthesis (translate text to speech) generating whole sentences of text in real time. The AI behind Alexa is based on artificial neural networks.

Data collected and recorded by the device and stocked on Amazon's cloud are personal data (e.g. user's voice, personal account data...) which are used to provide the service, to interact with users, to perform tasks and to answer users' questions. Such data are also processed with the purpose of improving Alexa's speaking style and giving more "human-like" voices to the system as well as improving AI models for more reliability in answers.

As technology advances more features are integrated in the device. For instance, an "emotional" program intended to detect human "emotions" through user's vocabulary and intonation (user is unhappy, anxious, stressed…) for enhancing speaking styles and AI responsiveness by using Neural TTS (NTTS) has been patented and made available in 2019 [64, 100]. In addition, an "Alexa Guard" skill allowing domestic sound detection and alarm (footsteps, breathing, door closing, fire emergency, doorbell...) is planned to be deployed in 2021 [79]. Human command is (normally) always required for activating Alexa and realising tasks.

Based on Amazon's publically available data[20] describing this technology (e.g. [4]) and by using the matrix, we propose the following qualification for the Alexa program/device:

> Crystallised AI, Cognitive Level 3 (tends to achieve Level 4), Technology Type 2, Semi-supervised learning, Multitask, Medium previsibility, Executive, Propositional.

> Commercial good, potential professional uses/specifications, hardware (smart speakers) and integrated software (vocal command).

Ethical issues mainly relate to privacy protection and security [66] as well as to technical robustness (due to erroneous results, undesired actions or hazardous behaviours encountered by users) [102]. Constant human monitoring and enhancement of the AI is required, especially when these devices are applied to sensitive domains such as health [31].

Although Amazon's responsibility can be engaged regarding proposed updates, applications and skills boosting Alexa's functionalities or performance, the role of users is to be considered in terms of responsibility sharing. Users are informed by Amazon about the data collected, recorded and processed on Amazon cloud, about possible crossing with other data about processing purposes and limitations. But even if Terms and conditions of use of Alexa are provided online to users [3, 1, 2], the quality and lawfulness of the information provided and the appropriateness of related Amazon data protection practices remain questioned [99]. In addition, profiling and data centralisation on a private cloud can also trigger issues regarding applicable laws and digital sovereignty.

In general, the more conversational AI achieves a high intelligence level, the more human users risk being fooled. This is due to the difficulties to differentiate the voice of the conversational agent from the human voice. Risks of potential confusion of the users should also be addressed, reduced and monitored. This challenge is not specific to Alexa and concerns other similar AI-based programs/devices such as the Google Duplex Assistant technology to which users can delegate some decision-making abilities [63]. Transparent communication is essential to make users aware of the logic and functioning of the device. However, in this case, we found it difficult to understand/find out what type of AI is used for what type

---

[20] In this regard, the public information that we use in this example is often merging technical data with commercial messages which complicate the qualification exercise. Therefore, we assume that the information provided online by the manufacturer is reliable. Retrieved on 20 January 2021.

of Alexa's skill. If necessary, specific qualifications should be considered for improving ethical assessments and public communications on essential product features. Such qualifications should be based on further detailed technical data mainly available to the sole manufacturer's research and development teams.

Conclusion

Advantages of the matrix: The proposed qualification matrix should allow both to reconcile conceptual analyses of intelligence from different disciplines for AI world and allow a deeper analysis of the Crystallised or Fluid AI program or device characteristics. The effects on the environment, on human rights, as well as their potentially broader socio-economic impacts should be assessed in order to attach relevant ethical and legal frameworks, risk-assessment values, methods and scales. The matrix should help understanding AI characteristics and related technological, ethical and legal challenges. This should serve the design of further risk-assessment tools and risk-control measures of any kind. It could also help identifying gaps in regulations which are necessary to fill in as regard to a new AI-based technology and to eventually create new safeguards for research, commercialisation, and use of such technology. Finally, this study could be the basis for further adaptation of foundational concepts of interest for AI in order to (re)tailor basics in the perspective of a consensus building on an intelligible AI classification which would also be useful for communicating with the lay public.

Limits: While we did our best to cover and sum up a broad domain through an interdisciplinary approach, we are aware of the limitations related to the completeness of this study and the completeness of the concepts studied. Nevertheless, we strived to provide the roots of a joint approach deserving timely rebirth. Relationships between categories and grey areas exist and further specifications would be desirable through additional research and the creation of new assessment tools. The compliance with different National specific legislations, on AI as such, or on human and social rights, and cultural diversity of approaches remain big challenges to achieve a consensus on a commonly accepted academic typology for AI.

References

[1] Amazon, Alexa Terms of Use:
https://www.amazon.com/gp/help/customer/display.html?nodeId=201809740 (Accessed on January

2021).

[2] Amazon, Customer Support, FAQ:
https://www.amazon.com/gp/help/customer/display.html?nodeId=201602230 (Accessed on January 2021).

[3] Amazon, General Conditions of Use of Amazon products and services:
https://www.amazon.com/gp/help/customer/display.html?nodeId=201909000 (Accessed on January 2021).

[4] Amazon jobs, https://www.amazon.jobs/fr/teams/alexa-ai (Accessed on January 2021).

[5] Amazon Smart device series, https://www.amazon.com/smart-home-devices/b?ie=UTF8&node=9818047011 (Accessed on January 2021).

[6] L. W. Anderson, D. R. Krathwohl, B.S. Bloom, A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives: Allyn & Bacon, 2001.

[7] Artificial & Natural Intelligence Toulouse Institute. https://en.univ-toulouse.fr/aniti (Accessed on April 2020).

[8] Association for Computing Machinery's (ACM) Committee On Professional Ethics (COPE). Code of Ethics. Ed. 2018.

[9] Baddeley, A. Working memory. *Current Biology*. **Vol.** 20(4), 2010, p. R136-R140.

[10] B. Barraud, « Le coup de data permanent: la loi des algorithmes », *RDLF*, **chron. 35**, 2017.

[11] BBC Ideas. Could artificial intelligence replace governments? Video. Animated by Kong Studio. 3 October 2019. https://www.bbc.com/ideas/videos/could-artificial-intelligence-replace- governments/p07px5jb (Accessed on April 2020)

[12] P. Besse, A. Besse-Patin, C. Castets-Renard. Legal and Ethical Implications of Artificial Intelligence Algorithms in the Health Field. HAL. 2019.

[13] R. Bhatia, Understanding the difference between symbolic AI & non symbolic AI. *Analytics India Magazine*. Opinions, Epub, 27 December 2017.

[14] Y.C. Blanchon, G. Allouard, Evaluation objective de la dimension handicapante de I'autisme de I'enfant et de I'adolescent. *Neuropsychiatr Enfance Adolesc* **Vol. 46**, 1998, p. 437-445 (in French).

[15] Y.C. Blanchon, G. Gibert, C. D'Hondt, Niveau de dépendance des enfants et adolescents porteurs d'autisme. Suivi longitudinal à cinq ans. *Neuropsychiatr Enfance Adolesc* **Vol.49**, 2001, p. 218-225 (in French).

[16] H. Bruyninckx, L. Preucil, M. Kulich, European Robotics Symposium 2008. Berlin: *Springer Science & Business Media*, ISBN 9783540783152, 2008, p. 23.

[17] G.L Canivez, & E.A. Youngstrom, Challenges to the Cattell-Horn-Carroll Theory: Empirical, Clinical, and Policy Implications. *Applied Measurement in Education*, **Vol. 32:**3, 2019, 232-248, Fig. 2.

[18] J. B. Carroll, *Human Cognitive Abilities: A Survey of Factor-Analytic Studies,* Cambridge University Press, New York, 1993.

[19] CERNA, Commission de réflexion sur l'Éthique de la Recherche en sciences et technologies du Numérique d'Allistene, Rapport n°1, Ethique de la recherche en robotique, 2004, p.12.

[20] C.Q. Choi. Forecast: Sex and Marriage with Robots by 2050. Epub, Livescience.com, 2007.

[21] K. Cherry, Gardner's theory of Multiple Intelligences. Epub. Verywellmind.com. Updated 3 September 2019.

[22] K. Cherry, How General Intelligence Influences Performance on Cognitive Tasks. Epub. Verywellmind.com. 24 September 2019.

[23] G.N. Chethan Kumar, Artificial Intelligence: Definition, types, examples, technologies. Epub: medium.com. 31 August 2018.

[24] G.N. Chethan Kumar, Learning Types and Algorithms. Epub: medium.com. 6 September 2018.

[25] P. Cihon, AI & Global Governance: Using International Standards as an Agile tool for Governance. United Nations University – Center for Policy Research. *AI & Global Governance Articles & Insights*. Digital Technology and Global Order, 2019.

[26] CJEU, Commission of the European Communities v Italian Republic, Case 7-68. 10 December 1968.

[27] C. Clark, The conditions of Economic Progress, London, 1957.

[28] CNOM, Médecins et Patients dans le Monde des Data, des Algorithmes et de l'Intelligence Artificielle, Analyses et recommandations du Cnom, 2018 (in French).

[29] J. Cohn, Google's Algorithms Discriminate Against Women And People Of Color, Epub: *Bright Magazine*. 16 May 2019.

[30] Conseil d'Etat, Révision de la loi de bioéthique : quelles options pour demain ? 2018 (in French).

[31] N. Couldry and Ulises Ali Mejias, Big Tech's latest moves raise health privacy fears, Special Report Future of AI and Digital Healthcare, Financial Times, 7 December 2020.

[32] Council of Europe, Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems, Adopted by the Committee of Ministers on 8 April 2020 at the 1373rd meeting of the Ministers' Deputies, 2020.

[33] Council of Europe, Study on the human rights dimensions of automated data processing techniques and possible regulatory implications, DGI(2017)12, prepared by the committee of experts of internet intermediaries (MSI-NET), 2018.

[34] N. Cowan, What are the differences between long-term, short-term, and working memory? *Progress in Brain Research*, **169**, 2008, p. 323–338.

[35] CREC Saint-Cyr. Autonomie et létalité en robotique militaire. Cahier de la Revue Défense Nationale (4ème Trimestre 2018), 195-237.

[36] I. Davidson, P. B. Walker, Towards Fluid Machine Intelligence: Can We Make a Gifted AI? TheThirty-Third AAAI Conference on Artificial Intelligence (AAAI-19), Association for the Advancement of Artificial Intelligence, 2019.

[37] T. Daups, Le Robot, bien ou personne, un enjeu de civilisation ? *Gazette du Palais*. LPA du 11 Mai 2017, n° 126a5, 2017, p.7 (in French).

[38] T. Daups, Etat de Droit, Etat Robotique. Synthèse de l'activité Scientifique en vue d'une Habilitation à Diriger des Recherches. Université de Rennes 2, 2019 (in French).

[39] C. De Ganay, D. Gillot, Pour une intelligence artificielle maîtrisée, utile et démystifiée, OPECST, Rapport d'information n°464, 2017, Section II.A.1. (in French).

[40] J. Delcker, Europe divided over robot 'personhood'. *Politico*. Epub, 2018.

[41] M. Delvaux Report, Civil Law Rules on Robotics, European Parliament, PE 582.443v03-00. A8-0005/2017, 2017.

[42] N. Drougard, D. Dubois, J.L. Farges, F. Teichteil-Königsbuch, Planning in Partially Observable Domains with Fuzzy Epistemic States and Probabilistic Dynamics, in: *Scalable Uncertainty management*, Proceedings of the 9th International Conference on Scalable Uncertainty Management (SUM 2015), 16 September 2015 - 18 September 2015 (Quebec, Canada), Eds. C. Beierle, A. Dekhtyar,
Springer, 2015.

[43] E. Dutton, D. Van Der Linden, R. Lynn, The Negative Flynn Effect: A systematic literature review. *Intelligence*, **Vol. 59**, 2016, p.163-169.

[44] EGE, Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems, Brussels, 2018.

[45] R. Epstein, G. Roberts, G. Beber, Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer. Ed. Springer, 2009.

[46] EU, Directive n°85/374/EEC of 25 July 1985 on the approximation of the laws, regulations and administrative provisions of the Member States concerning liability for defective products. *OJ L 210*, 7 August 1985, p. 29-33.

[47] EU, Directive n°2001/95/EC of the European Parliament and of the Council of 3 December 2001 on general product safety. *OJ L 11*, 15 January 2002, p. 4-17.

[48] EU, Directive n°2009/24/EC of the European Parliament and of the Council of 23 April 2009 on the legal protection of computer programs. *OJ L 111,* 5 May 2009, p. 16–22.

[49] EU, Directive n°2010/63/EU on the protection of animals used for scientific purposes. *OJ L 276*, 20 October 2010, p. 33–79.

[50] EU, Directive n°2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information, *OJ L 172*, 26 June 2019, p. 56–83.

[51] EU, Regulation n°2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), *OJ L 119*, 4 May 2016, p. 1–88.

[52]   Europa portal, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=LEGISSUM%3Ami0016 (Accessed on April 2020)

[53] Europa Portal. https://ec.europa.eu/growth/single-market/ce-marking/manufacturers_en (Accessed on April 2020).

[54] European Commission, Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of

the Regions on Artificial Intelligence for Europe, COM(2018) 237 final, Brussels, 25 April 2018.

[55]   European Commission, White Paper On Artificial Intelligence – A European approach to excellence and trust. COM(2020) 65 final, Brussels, 2020.

[56] European Economic and Social Committee, Opinion on Artificial Intelligence, The consequences of artificial intelligence on the (digital) single market, production, consumption, employment and society, 2017, proposition 3.12.

[57] E. Fernandez, AI Is Not Similar To Human Intelligence. Thinking So Could Be Dangerous. Epub: Forbes, Science, 30 November 2019.

[58] M. Forehand, Bloom's Taxonomy: Original and revised. In: From Emerging Perspectives on Learning, Teaching and Technology. Ed. M. Orey. 2005. https://projects.coe.uga.edu/epltt/index.php?title=Bloom%27s_Taxonomy (Accessed on April 2020)

[59] French Civil Code, Article 515-14.

[60] H.L. Gallager. C.D. Frith. Functional imaging of 'theory of mind'. *TRENDS in Cognitive Sciences*, **Vol.7**(2), 2003.

[61] J.G. Ganascia, Le Mythe de la singularité. *Faut-il craindre l'intelligence artificielle ?* Ed. Seuil. Coll. Science ouverte. 2017, p.51 (in French).

[62] H. Gardner, *Frames of Mind: The Theory of Multiple Intelligences*. 1983. Second edition, 2011.

[63] N. Garun, One year later, restaurants are still confused by Google Duplex, Google Duplex works until it breaks in unexpected ways, TheVerge, Epub. 9 may 2019.

[64] C. Gao, Use New Alexa Emotions and Speaking Styles to Create a More Natural and Intuitive Voice Experience, Amazon.com, Epub, 26 November 2019.

[65] S. Ghaffary, The algorithms that detect hate speech online are biased against black people, Epub: *Vox.* 15 August 2019.

[66] P. Grad, Amazon Alexa bug exposed voice data, Tech Xplore, Epub, 14 August 2020

[67] G20 AI Principles. 9 June 2019.

[68] B. Haas, The Guardian. 4 April 2017. https://www.theguardian.com/world/2017/apr/04/chinese-man-marries-robot-built-himself (Accessed on April 2020).

[69] HAS, Projet de grille d'analyse pour l'évaluation de dispositifs médicaux avec intelligence artificielle, Epub, 2019. https://www.has-sante.fr/jcms/p_3118247/fr/projet-de-grille-d-analyse-pour-l-evaluation-de-dispositifs-medicaux-avec-intelligence-artificielle (in French) (Accessed on January 2020).

[70] T. Hatmaker, Saudi Arabia bestows citizenship on a robot named Sophia. Techcrunch.com. 27 October 2017. https://techcrunch.com/2017/10/26/saudi-arabia-robot-citizen-sophia/ (Accessed on April 2020)

[71] High Level Expert Group on Artificial Intelligence (AIHLEG), A definition of AI: Main capabilities and scientific disciplines, 8 April 2019.

[72] High Level Expert Group on Artificial Intelligence (AIHLEG), Ethics Guidelines for Trustworthy

AI, 8 April 2019.

[73] A. Hintze. Understanding the four types of AI, from reactive robots to self-aware beings. *The Conversation*. 14 November 2016.

[74] J.L. Horn, & R.B. Cattell, Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology,* **Vol. 57**(5), (1966), 253-270.

[75] R.L. Hudson. Step by step, world leaders strive to find consensus on AI development. *Science Business*. 18 July 2019.

[76] R.L. Hudson**.** France and Canada move forward with plans for global AI expert council. *Science Business*. 19 November 2019.

[77] IEEE Global Initiative for Ethical Considerations in Artificial Intelligence (AI) and Autonomous Systems (AS). https://standards.ieee.org/news/2017/ieee_p7004.html (Accessed on April 2020).

[78] Institute for Applied Psychometrics (IAP), Webpage on Cattell-Horn-Carroll (CHC) Theory and evolutions: http://www.iapsych.com/CHCPP/CHCPP.HTML (Accessed on April 2020)

[79] K. Jonhson, Alexa Guard alerts you when it hears glass breaking or the smoke alarm, Venturebeat.com, Epub, 20 September 2018.

[80] A. Kapoor. Deep Learning vs. Machine Learning: A Simple Explanation. Epub, Hackernoon.com. 25 February 2019.

[81] J. Kerns. What's the Difference Between Weak and Strong AI? Epub: Machine design.com. 16 February 2017.

[82] P. Kraikivski, Seeding the Singularity for A.I. Cornell University, Computer Science, Artificial Intelligence, Epub in *ArXiv*, abs/1908.01766, 2019.

[83] Larousse.fr (25/10/2019).

[84] G. Lea. Why we need a legal definition of artificial intelligence. *The Conversation*. 2 September 2015.

[85] Le modèle d'intelligence de Cattell: l'intelligence fluide et cristallisée. *Nos pensées.fr*. 10 February 2018.

[86] A. Mardan. Blog post. Programmer vs. Software Engineer vs. Software Developer vs. Coder. https://webapplog.com/programmer-vs-software-engineer-vs-software-developer-vs-coder/ (Accessed on 25 October 2019)

[87] B. Marr. The Key Definitions Of Artifical Intelligence (AI) That Explain Its Importance. Epub: Forbes. 14 February 2018.

[88] J. McCarthy, Recursive functions of symbolic expressions and their computation by machine [archive]. *Communications of the ACM* **3(4)**:184-195, 1960.

[89] A. Mendoza-Caminade, Le droit confronté à l'intelligence artificielle des robots : vers l'émergence de nouveaux concepts juridiques ? *Recueil Dalloz* (8 / 7677), 25 Février 2016, p. 445 (in French).

[90] J.H. Moor. *The Turing Test: The Elusive Standard of Artificial Intelligence*. Kluwer Academic Publishers, 2003.

[91] P. Nayak. Understanding searches better than never before. Blog post. 25 October 2019. https://www.blog.google/products/search/search-language-understanding-bert/ (Accessed on April 2020)

[92] H. Nguyen, S. Bailey, Use of Artificial Intelligence for Smart Contracts and Blockchain. Thomson Reuters' *FinTech Law Report* **Vol.21**(2), March/April, 2018, p.1-7.

[93] Nobaproject.com: https://nobaproject.com/modules/intelligence (Accessed on April 2020).

[94] R.B. Parikh, S. Teeple, A.S. Navathe, Addressing Bias in Artificial Intelligence in Health Care, *JAMA*. 2019;**322**(24):2377–2378. doi:10.1001/jama.2019.18058, 2019.

[95] J. Parker, S. Jaeger, Learning in Artificial Intelligence: Does Bloom's Taxonomy Apply? 28 December 2016. Available at SSRN: https://ssrn.com/abstract=2891191 or http://dx.doi.org/10.2139/ssrn.2891191 (Accessed on April 2020).

[96] D. Premack, G. Woodruff, Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences,* **Vol.1**(4), 1978, p. 515–526.

[97] Retresco.de, https://www.retresco.de/en/encyclopedia/natural-language-understanding-2/ (Accessed on April 2020).

[98] A. Robert. Google Assistant et Alexa: qu'écoutent-ils vraiment, que font-ils de vos données personnelles? CNET, Epub, 4 June 2020.

[99] S. Roques, J. Langevin, C. Drouin, J. Faille, *Déficience intellectuelle*. Ed. Nouvelles. Montréal. ISBN 2-921696-50-9, 1999 (in French).

[100] V. Rozgic, Using adversarial training to recognize speakers' emotions, Amazon.science, Epub, 21 May 2019. (Accessed on January 2021).

[101] S.J. Russell, P. Norvig, *Artificial Intelligence: A Modern Approach,* 2nd edition, Upper Saddle River, New Jersey: Prentice Hall, ISBN 0-13-790395-2, 2003.

[102] S. Sermondadaz, Amazon veut faire taire les rires diaboliques de son assistant vocal Alexa, Science & Avenir, Epub, 8 March 2018.

[103] A.K. Seth, B.J. Baars, D.B. Edelman. Criteria for consciousness in humans and other mammals. *Conscious Cogn.* **Vol.14**(1), 2005, p. 119-39.

[104] W. Stach, L. Kurgan, W. Pedrycz, M. Reformat, Genetic learning of fuzzy cognitive maps. *Fuzzy sets and systems*, **Vol.153**(3), 2005, p. 371-401.

[105] N. Steiwer, Données de santé : comment l'Allemagne tente d'échapper à l'emprise américaine. Enquête. *Les Echos*, Epub, 2020 (in French).

[106] C.D. Stylios, P.P. Groumpos, Modeling complex systems using fuzzy cognitive maps. *IEEE Transactions on Systems, Man, and Cybernetics*, **Vol. 34**(1), 2004, Part A: Systems and Humans, p.155-162.

[107] Techopedia, https://www.techopedia.com/definition/31622/strong-artificial-intelligence-strong-ai (Accessed on April 2020).

[108] Télécom ParisTech. Algorithmes : biais, discrimination et équité. Section 3. Epub. Février 2019 (in French). https://www.telecom-paris.fr/algorithmes-biais-discrimination-et-equite (Accessed on April 2020)

[109] C. Tessier, V. Bonnemains, C. Saurel. Mettre l'éthique dans l'algorithme ? Blog Binaire du Monde, 2018.

[110] A.M. Turing, Computing Machinery and Intelligence, *Mind*, New Series, **Vol. 59**, No. 236, 1950, p. 433-460.

[111] UN OHCHR, Recommendation on the protection and use of health-related data, Annual Report to the Human Rights Council and the UN General Assembly, 5 December 2019.

[112] UNESCO, Working document: Toward a draft text of a Recommendation on the Ethics of Artificial Intelligence: working document, Virtual discussion of the Ad Hoc Expert Group (AHEG) for the preparation of a draft text of a recommendation on the ethics of artificial intelligence, SHS/BIO/AHEG-AI/2020/3 REV, 2020.

[113] Université de Montréal. Déclaration de Montréal pour un développement responsable de l'intelligence artificielle, 2017.

[114] R.D. Jr. Van Valin, From NLP to NLU, Heinrich Heine University Düsseldorf and University at Buffalo, The State University of New York. Epub, http://www.isi.hhu.de/fileadmin/redaktion/Oeffentliche_Medien/Fakultaeten/Philosophische_Fakultaet/Sprache_und_Information/Van_Valin_From_NLP_to_NLU.pdf (Accessed on December 2019), 2016.

[115] M. Velmans. How to Define Consciousness – And How Not To Define Consciousness, *Journal of Consciousness Studies*, **Vol. 16**(5), 2009, p.139-156.

[116] M. Velmans, Understanding Consciousness, Edition 2, London: Routledge/Psychology Press. 2009.

[117] Wikipedia, https://fr.wikipedia.org/wiki/Autonomie (Accessed on April 2020).

[118] Wikipedia, https://en.wikipedia.org/wiki/Natural-language_understanding (Accessed on April 2020).

[119] Wikipedia, https://en.wikipedia.org/wiki/Intelligent_agent (Accessed on April 2020).

[120] WISC-V, Second Edition, 2019.

[121] Yonhap News Agency. Interview. Go master Lee says he quits unable to win over AI Go players. Epub, yna.co.kr, 27 November 2019.

[122] Youtube Videos, Lessons on types of AI agents from Education4u and Gate Smashers channels, (Accessed on April 2020)

[123] L. Zaval, Y. Li, E. J. Johnson, E. U. Weber, Chapter 8 - Complementary Contributions of Fluid and Crystallized Intelligence to Decision Making Across the Life Span, in: *Aging and Decision Making*, Academic Press, Editor(s): T.M. Hess, J.N. Strough, C. E. Löckenhoff, ISBN 9780124171480, 2015, p. 149-168.