International Journal of Political Science and International Relations (IJPSIR)
Volume.14, Number 6; June-2023;
ISSN: 2837-3162 | Impact Factor: 8.25
https://zapjournals.com/Journals/index.php/ijpsir
Published By: Zendo Academic Publishing

UNDERSTANDING ARTIFICIAL INTELLIGENCE: PATHWAYS AND ETHICAL CONSIDERATIONS

[1]Kouassi Konan Jean-Claude

Article Info
Keywords: Artificial Intelligence, Machine Learning, Weak AI, Narrow AI, Artificial General Intelligence, Strong AI, Superintelligence, Machine Intelligence, Theories of Intelligence, Ethics and Risks of AI.

Abstract
As Artificial Intelligence (AI) continues to expand into various fields, it is important that both specialists and beginners have a clear understanding of its diverse approaches, future objectives, potential drawbacks, and issues related to information overload. This paper presents a concise AI approach accessible to both experts and newbies alike, to prevent confusion in the wake of new AI developments. Using comprehensive diagrams, we provide concrete explanations of supervised learning, unsupervised learning, reinforcement learning, and deep learning in machine learning (ML), and show how these technologies have paved the way for current AI. We also demonstrate how theories of intelligence and AI concepts match technological advances and future objectives in AI. Ethical concerns related to AI and risks are also discussed, alongside biases that arise from human data input. This study aims to provide valuable and easily understandable insights to anyone interested in AI, including teachers, decision-makers, developers, students, and practitioners. The analysis of empirical studies and the illustrations of AI concepts in this study contribute to scientific literature and offer practical implications for the global impact of AI

1. INTRODUCTION

The gestation of Artificial Intelligence (AI) started from 1943 to 1955 via various research works that prepared the "birth" of AI, such as the computability of any computable function by some network of connected neurons (1943), the building of the first neural network computer in 1950, and the Turing Test of Alan Turing (cf. [1], section 1.3.1). Then comes the birth of Artificial Intelligence itself during a two-month workshop at Dartmouth College in the summer of 1956, organized by an influential figure in AI, John McCarthy, and assembling ten (10) carefully selected proponent of automata theory scientists. They defined Artificial Intelligence (AI) as a new specific field that embraces the idea of duplicating human faculties such as creativity, selfimprovement, and

[1] Faculty of Computer Science via distance learning, Bircham International University, Madrid, Spain

language use. AI is the only one that *is clearly a branch of Computer Science* (unlike Cybernetics) and that attempts to build machines that will function autonomously in complex, changing environments (cf. [1], section 1.3.2, p.18).

However, in practice the Foundations of Artificial Intelligence are supported by the following eight (08) disciplines that contributed to ideas, viewpoints, and techniques of AI: *Philosophy* (via mind and body theories, etc.), *Mathematics* (via logic, etc.), *Economics* (via decision theory, game theory, etc.), *Neuroscience* (via EEG, etc.), *Psychology* (via the behaviorism movement of input-output stimulus, etc.), *Computer engineering* (via hardware and software that make AI applications possible, etc.), *Control theory* (via the techniques of self-controlling machines that work without software or language, as opposed to digital computers, etc.) and *Linguistics* (via Computational Linguistics or Natural Language Processing, etc.). The association of these disciplines towards the goal of creation of intelligence for machines led to various Concepts of Artificial Intelligence that perfectly match the current technological advances in the field, and all of this is done in adequation with long term objectives too.

In our approach, via comprehensive illustrations, we will first define the pillars of Machine Learning (ML) that lead to current stage of AI. Then, starting by known Theories of Intelligence that led to various AI Concepts (including those supported by nowadays senior AI experts and influential figures in the field of AI), we will discuss the paths towards future goals of Humanity in the creation of Intelligence for Machines. Finally, as the development of AI unfortunately also imply tremendous risks for the Future of Life here on the earth, we will present current Global Solutions intended to overcome the problems of Ethics and Risks of AI. [We provided between squared brackets […] very useful references to support the readers].

2. DEPICTING ARTIFICIAL INTELLIGENCE (AI)

We know that illustrations are always more representative than simple words (cf. simulation or figure animations in [1], section 16.3.4, last paragraph). So, we depict below the whole AI discipline via our own simple illustration for a fast overview and clear insights.

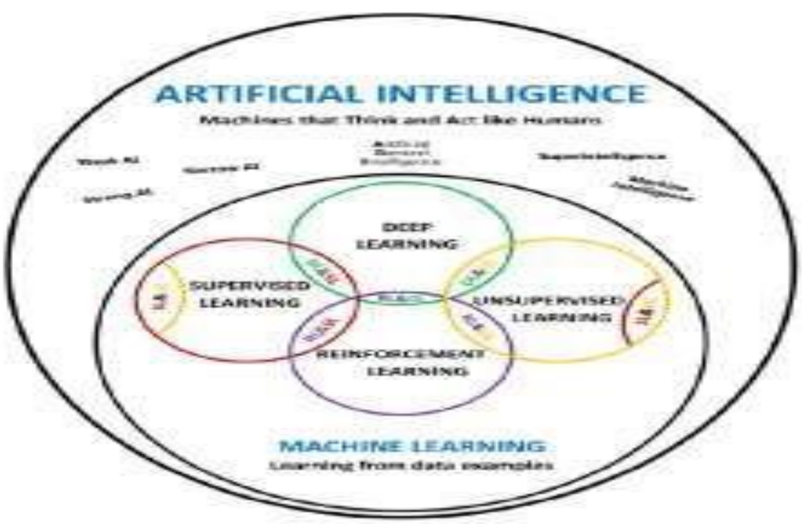


Figure 1. The Artificial Intelligence Realm and the implied Techniques [2]

In Figure 1 above, via a clear illustration we depict the whole AI field with main components and concepts for a fast overview and understanding of this vast discipline. The general realm of ARTIFICIAL INTELLIGENCE (AI) consists of Machines that Think and Act like Humans (cf.
[1], Figure1.1, page 2, our own short reformulation). In practice, AI practitioners (AI Scientists, AI Engineers, AI Technicians, etc.) have to combine several Machine Learning (ML) techniques together in order to make Machines intelligent, and thus realize Artificial Intelligence. Indeed, each learning technique that consists in learning from data examples such as the Supervised Learning technique, when taken individually is a Machine Learning approach. But the crafty combination of Machine Learning techniques that makes machines or systems behave (think and act) like humans is Artificial Intelligence. In addition, depending on philosophical beliefs or feelings of AI researchers, AI has various approaches such as Weak AI, Strong AI, Narrow AI, etc.

3. THE FOUNDATIONS OF MACHINE LEARNING

As illustrated in Figure 1, we define Supervised Learning, UnSupervised Learning,
Reinforcement Learning, and Deep Learning as the four (04) pillars of Machine Learning (ML). As said Professor Harry Surden, ML is the dominant mode of AI today (cf. [3], p.10). Accordingly, in this section we provide a clear overview of each pillar of ML.

3.1. Supervised Learning (SL)

In ML, the Supervised Learning (SL) technique consists in learning with labeled examples. Basically, it consists in teaching a model (as if it was a student) how to identify the correct output or answer for a given input. So, each data set provided by the teacher is curated in two (02) parts as features for the input and labels for the output. Labels are the values we want the model to predict. So, Supervised Learning is similar to a Pattern Recognition model with three (03) possible outputs. When the mapping from input (x) to output (y) via a pattern function f is stochastic (not close to an affine function of elements of the training set), we learn a Conditional Probability Distribution P(Y|X). When the output y is one of a finite set of values, the learning problem is called Classification. And, when the output y is a number, the learning problem is called Regression.

3.2. UnSupervised Learning (UL)

The UnSupervised Learning (UL) technique naturally consists in learning with unlabeled examples using clusters (categorized groups of similar objects or cluster Analysis). The UL model automatically guesses and classifies the given data without any provided label. Nonetheless it has several variants. First, we have the *Self-Supervised Learning* that is a subset of UL that does not directly focus on clustering and grouping, but instead relies on object encoding, and then elements are automatically classified as for the human perception. Next, we have the *Semi-Supervised Learning* that consists in learning with a mixture of unlabeled data and a few labeled data; it is a combination between SL and UL techniques.

3.3. Reinforcement Learning (RL)

Reinforcement Learning (RL) is defined has a Behavior Learning method based on either a Mathematical model (model-free) or a World model (model-based). Fundamentally, it is based on a sort of reward system where the RL model (the agent) always tries to reach the maximal or optimal reward value according to a provided strategy (the policy) that shows the agent how to behave. However, sometimes the best performance of the system is not reached by the maximal value of the defined reward, but by a suboptimal value. In RL we assume no prior knowledge of the environment nor the reward function. We only focus on a fixed policy of state-action pairs for a Passive Reinforcement Learning scheme, and additionally on a self-decision-making strategy about action to take for an Active Reinforcement Learning. No matter if it is Passive or Active, RL could be performed both Online when interacting with the environment or Offline when learning from a fixed dataset (cf. arXiv:2104.06294v1). The mathematical expression of an active RL technique is written as following:

$U(S)=R(S)+\gamma \max_a \sum_{S'} P(S'|S,a)U(S')$ – (without diving deep in technical details, according to our objective: an overview).

3.4. Deep Learning (DL)

The Deep Learning (DL) technique consists in learning with inductive *Connectionist* or *Brainlike Layered* models. The learning process here consists in iteratively passing the data through the DL model with feedback and readjustment of parameters (usually via backpropagation) till the model reaches a steady state where it could be good enough for inference (for its application in the real world). The DL technology is so based on several mathematical concepts including Linear Algebra and Probability [4]. We will dive deep in the DL technology because it is a structural copy of (generally) the human brain functions, and it is highly used with various ML methods.

Several AI ideas adopted by AI scientists was oriented toward creating a brain language through the analysis of its physical aspects in order to solve AI. Let us notify that the term language here stands for the way the brain expresses itself or reacts with the universe via its internal functions. For them, creating a model that mimics all the biological brain functions is creating an artificial model of the brain. So, looking at their various approaches toward the process of creating a language of the brain, we could say that AI researchers, create machine intelligence from observation of brain functions (either brain states or behavioral functions or intentional functions) that are all generated by the same biological brain. Indeed, a thorough analysis of data issued from observation of brain functions could lead to the creation of a reliable enough language from the biological brain (cf. [5], Chapters 15, 16 and 17).

However, does this mean that these created brain languages (models created from human brain analysis) are the real language of the biological brain, the exact way the human brain really communicates? We are not so sure.

Nevertheless, we provide in the following discussion some elaborations about our viewpoint from an analysis based on various results given from related disciplines to AI, including Neuroscience. So, from the Deep Learning (DL) technique that consists in learning with inductive artificially Brain-like Layered models, and from the fact that AI scientists are focused on creating a Brain Language via the DL technique, we had been led to make an original comparison between the Biological Neural Network and the Artificial Neural Network.

First, we know that the number of ways that information flows among neurons in the human brain is so large (10^{11}), it is greater than the number of stars in the entire universe. Also, there are more connections in the human brain than atoms in the Universe [6].

Second, each biological neuron of the human brain can have 10,000 analogical connections on average with 10,000 other neurons via its synapses towards their dendrites. Each neuron can have 50,000 dendrites [7]. There are at least three (03) synaptic connections between two (02) biological neurons. And each synaptic connection of the human brain is not binary, it is one node including up to 5,000 tiny connections between synapses point of contact (Neurotransmitter Postsynaptic Receptors).

Third, and by contrast, the connections with the artificial neuron are binary (not analogical). Between two (02) artificial neurons there is only one binary connection. But each artificial neuron could be connected to several others (very lower than 10,000 due to our current computational power). And nowadays, even Large Language Model (LLM) like Bloom with 176 billion parameters inside the model ($1{,}76.10^{11}$) that are greater than the number of ways that information flows in the human brain (10^{11}) but lower than the number of connections in the human brain (at least 10^{14}), are still unable to solve AI, despite their recognized positive and useful contributions to technological advances worldwide.

We provide below via clear illustrations a summary of our comparison between the Biological Neural Network and the Artificial Neural Network (Figure 2 and Table 1). These recent findings provide a more clear view about neuron connections in a Biological Neural Network.

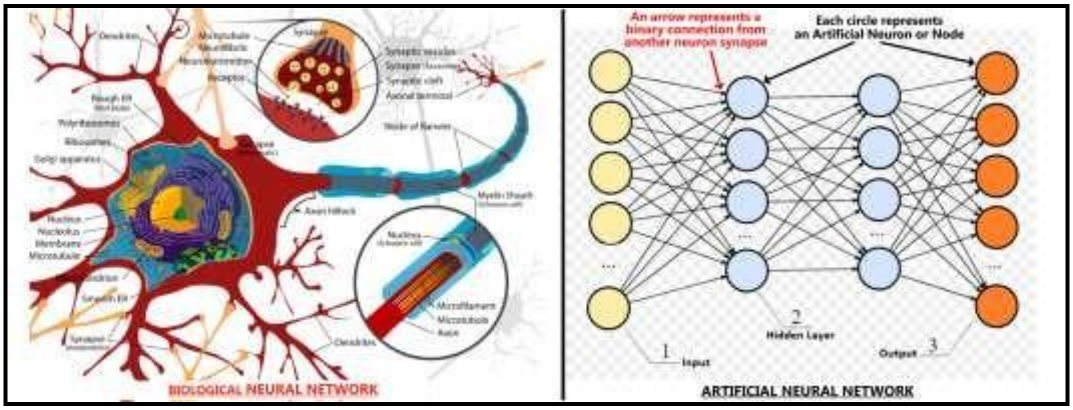


Figure 2. Biological Neural Network vs Artificial Neural Network

The biological neural network is from a Google: Free to use, share or modify, even commercially image (https://en.wikipedia.org/wiki/Neuron and Wikimedia Commons)

Table 1. Comparison of the characteristics of biological and artificial neurons

DESCRIPTIONS	**BIOLOGICAL NEURON**	**ARTIFICIAL NEURON**
Nature of connections	**Analogical** (continuous)	**Binary** (digital)
Number of types of neurons	**10,000 specific types** [8]	**Less than 100**
Average number of possible connections per neuron	**10,000** *(The real number can grow up to 100,000+ synapses. Cf. Wikipedia, Neurone)*	**Very lower than 10,000** *(Depends on the computational power, generally a neuron could be connected to 10 up to 4096 nodes from the next layer in a Fully-Connected Network)*
Average number of distinct active connections per neuron (with its other neighbors)	**10,000** *(Varies between 5,000 and 200,000 other neurons as connected neighbors from one neuron (cf. [8], para. 2))*	**very lower than 10,000** *(Depends on the computational power, generally from 10 to 4096 nodes per layer)*
Number of active connections between 2 neurons	**at least 03** (cf. [9], *Abstract and Figure 1. That is the number of synapses between two neurons. It varies between 3 and 8 synapses)*	**1** *(One unique connection, binary link between two nodes)*
Number of receptors per connection	**up to 5,000+** (cf. [10], *Table 1, Fig 4. Each synapse has up to 5,000+ receptors. Figure 2 contains a zoom over a synapse)*	**1** *(One binary link, not a bunch of wires)*

As a conclusion, we see that Artificial Neural Networks (ANN) are able (via various architectures including Bloom) to mimic the human brain and help in AI development or creation of machine intelligence. We are able today to create awesome applications that could outperform human skills in several domains (e.g. MuZero, AlphaFold, etc.). Moreover, these AI models constitute a sort of artificial brain language which works well and, in the future, will surely lead to true machine intelligence. Creating the artificial language of the brain, let's also quote the Alibaba Cloud Intelligence Brain, an ultra-intelligent AI Platform for solving complex business and social problems (without discrimination, their terminologies better express how AI scientists are focused on creating a brain language). Nevertheless, despite their performance, these applications which mimic the human brain functions (either brain states or behavioral or intentional) are not the true human brain language, because they are not the exact way the human brain communicates. Indeed, due to the nature of the biological brain and its connections described above (Figure 2 and Table 1), the artificial brain language cannot be exactly the same as the biological brain's own language (internal links and communications). Of course, the artificial brain language works well and reproduces significant functions as a human intelligence does, it could and will be able to perform certain tasks even better than humans, but we are not the same species (we think it will remain some specificities only possible for the human race). The nature of the biological human brain imposes unknown and unexplored dimensions of data that the artificial one ignores.

4. HOW THE 04 PILLARS OF MACHINE LEARNING TECHNIQUES LEAD TO THE CURRENT STAGE OF AI

4.1. The process of creating intelligence with Machine Learning

In practice, the ML Framework is made of several packages that imply the use of the four (04) pillars or foundations of ML: Supervised Learning (SL), UnSupervised Learning (UL), Reinforcement Learning (RL), and Deep Learning (DL). In the Alan Turing Institute ML Framework below (written in Julia) we can clearly distinguish various packages that can perform tasks implying the four (04) pillars, such as the *NeuralNetwork* package.

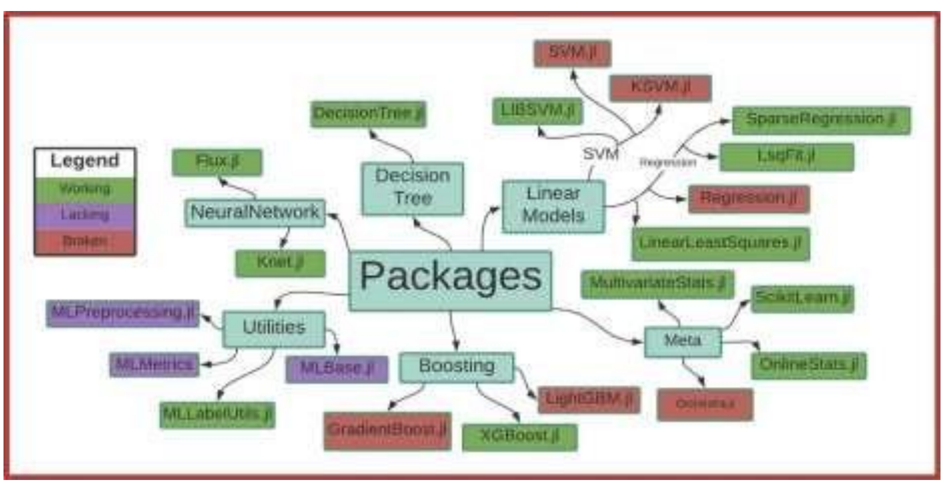


Figure 3. Alan Turing Institute Releases ML Framework Written in Julia (medium.com)

However, these four (04) foundations of Machine Learning techniques could also be combined in order to produce synergetic effects. Through experimentations, AI researchers found six (06) useful possible paired combinations of the foundations of Machine Learning techniques for realworld applications: SL&UL, DL&SL, DL&UL, RL&SL, RL&UL, and RL&DL (cf. Figure 1 above). For example, the object encoding of Self-Supervised Learning could be performed with Reinforcement Learning (RL&UL) or Deep Learning (DL&UL). Then,

looking at the possible packages and ML techniques available, the possible components of any AI agent (or AI model) are made of two (02) main learning approaches depending on the adopted techniques. First, we have the Inductive Learning that is based on exposure to repeated associations between elements (*cf. [1], Supervised Learning, Unsupervised Learning, Deep Learning or Reinforcement*
Learning from chapters 18 to 21 and section 27.1). The second learning method is the Analytical or Deductive Learning that is a logical deduction based on a general mathematical reasoning or a global rule (*cf. [1], Propositional Theorem Proving, First-Order Logic, Higher-Order Logic, Temporal Logic, Probability Theory, and Fuzzy Logic: from sections 1.2.1, 1.2.2, 7.5, 10.4.2, 18.1, Figure 8.1 and the whole part IV that includes chapter 13 to 17*). Therefore, in order to create an Intelligent Agent all the components trained via the inductive and/or deductive learning methods must be presented in a craftily combined form (generally via coding). However, it is also possible to realize an intelligent agent based on only one ML pillar, what matters is that the created AI system should generate (humanlike) intelligence. Anyway, the takeaway is that in the AI discipline (that has ML as a dominant mode) we could create intelligence (i.e., AI agent or AI model) via two (02) methods: Inductive Learning and/or Deductive Learning.

4.2. The Theories of Intelligence

Now, as we are talking about creating intelligence with Machine Learning models, let us elaborate about the nature of intelligence because psychologists defined many types of intelligence [11]. It should be interesting to know what type of intelligence we are creating each time. Moreover, as said Professor Harry Surden, we can get "intelligent" automated results without intelligence (cf. [3], p.15). We can give three (03) examples of "automated results without intelligence". First, in the field of Computer engineering we can create an excellent code for automatically performing a specific task, but without intelligence at all because the code itself doesn't understand the task, it just follows instructions of the programmer without any introspection. Second, in the field of Cybernetics that consists of automated machines without software (no code), there exists watches that provides automated intelligent results about the exact time of the day, but without intelligence at all (https://www.ticwatches.co.uk). Only thoroughly designed calculus is considered, and unlike AI the Cybernetics discipline has no intention to create intelligence at all. Third, in the biological nature we can cite the example of the (unintelligent) dung beetle that continues to try to close the entrance of its nest with a missing ball of dung (removed from its grasp) even after losing it (cf. [1], section 2.2.2, page 39). Indeed, beetles just have innate plans inside their brain that describe how they should live and behave, they are not intelligent at all (no introspection inside their brain).

Therefore, there are two (02) basic types of models of intelligence in psychology (cf. [5], section 15.2). First, we have the Hierarchical models that consist of models of intelligence in which mental abilities are layered and create a hierarchical structure. The *Two-factor theory of intelligence* (corresponding to general intelligence) and the *Gottfredson model of intelligence* (based on reasoning, linguistic and learning abilities) are both hierarchical models of intelligence. Second, we have the Multiple aptitude models that consist of models of intelligence in which mental abilities are treated as a set of equivalent and independent factors. The *Model of primary mental abilities* (based on perception, reasoning, and linguistic abilities), the *Kinesthetic intelligence* (based on both locomotion and manipulation abilities), the *Social intelligence* (based on interpersonal relations and social facility abilities), the *Emotional intelligence* (based on a set of abilities which allow one to perceive others' emotions, to control one's own emotions and to use emotions in mental processes and during problem solving), and the *Creativity* [based on intellect abilities that develop original ideas (intellect theory), on cognitive psychology (mental processes for developing knowledge representation), and on Triarchic theory of intelligence (a distinguished contextual model based on cognitive theories)], are all part of the multiple aptitude models of intelligence. About human faculties that AI is trying to duplicate, Albert Einstein said that "*Creativity is seeing what others see and thinking what no one else ever thought.*" (quote picked on https://analyticsindiamag.com/). So far, creative works (in terms of approach, content, format or setting, etc.) performed by scientists or even advanced students among others cannot yet be handled by machines, only repetitive tasks are easily mastered by Artificially Intelligent Systems.

Below, we also present the philosopher St. Thomas Aquinas (1225 – 1274) point of view about mind and body that is very clearly defined and close to the models of intelligence in psychology (cf. [5], section 15.1.4). He reinterpreted and developed Aristotelian epistemology. He distinguished two (02) great power of the mind. First, we have the intellect which is a *cognitive power* that consists of *three (03) generic operations or acts*: comprehending a concept, pronouncing a judgment, and reasoning. It is based on external senses including sight, hearing, taste, smell, and touch. Second, we have the will which is an *appetitive power* (all forms of internal inclination) that consists of *four (04) internal senses*: Common sense (sensus communis), Imagination, Memory, and the Cogitative power (or *Particular reason*, that can evaluate if a perception is beneficial (useful) or harmful). The Aquinas model of the intellect and will is from our point of view so well defined that it should be considered in AI research for the creation of intelligence for machines (hierarchical and/or multiple aptitude models).

4.3. The Paths towards Artificial Intelligence

We provide in the following table (Table 2), our classification of all Machine Learning categories implied in the creation of AI as **AI Paths**. This table is based on carefully selected empirical research works proving the ten (10) categories. And our classification is established first on the fundamental ML methods (that can also be coded individually to create Intelligent Systems that "think and act" like Humans), second on their possible paired combinations, and third on the learning mode (learning strategy) they use. Additionally, without being exhaustive we provide some good examples for each category. Let us notice that we describe here ***Evolutionary Computation*** [including Evolutionary Algorithms (Genetic Algorithms, Evolutionary Programming and Evolution Strategies) with fixed-size solutions and Genetic Programming with population as a solution] and ***Fuzzy Systems*** as other efficient Supervised or UnSupervised Learning techniques because their training processes are based on input/output examples or variables drawn from the domain data (cf. [12], Chapters 3, 4, 5 | and [13], Fig. 3).

Table 2. Classification of ML Categories (or AI Paths) that lead to Artificial Intelligence (AI)

N°	AI PATH (or ML CATEGORY)	LEARNING MODE OR STRATEGY USED FOR THE CATEGORY	JUST SOME GOOD EXAMPLES
1	Supervised Learning (SL)	Training with labeled examples using *Parametric* and *Nonparametric* Models (In case the SL is done with ANNs, we fall in a Deep Learning case, see line 3 below)	SVM, k-Nearest-Neighbors (k-NN), SVM, Decision Trees (DT), Naive Bayes, Bayesian Networks, Logistic Regression, Fuzzy Systems, Evolutionary Computation (EA, ES, EP, GP), HMM, etc.
2	UnSupervised Learning (UL)	Clustering methods (PCA, KMeans, etc.), Self-Supervised Learning (Encoding methods)	Expectation Maximization (EM), Bayesian Networks with Hidden Variables (cf. [1], *section 20.3.2*), OpenUniverse Probability Models (cf. [1], OUPMs in *section 14.6.3*), Transformers, Self-Organizing Networks, etc.
3	Deep Learning (DL)	Artificial Neural Networks (ANNs)	Fully Connected Network (FCN), Convolution Neural Network (CNN), Optimal Brain Damage and Tiling Algorithms (cf. [1], *section 18.7.5*), etc.
4	Reinforcement Learning (RL)	Model-based and/or Model-free using mathematical functions	Q-learning, TD, ADP, SARSA (cf. [1], *Chapter 21*), etc.
5	SL&UL	Semi-Supervised Learning	Auto-annotation (cf. [1], *section 24.6.1*), Semi-supervised Keypoint Localization [14], etc.
6	DL&SL	ANN + an SL technique (k-NearestNeighbors, SVM, Decision Trees, Naive Bayes, and Logistic regression, etc.)	Deep Learning using Linear Support Vector Machines [15], (cf. [16], Fig. 2), etc.
7	DL&UL	Self-Supervised Learning	Self-supervised Learning for Large-scale Item Recommendations [17], (cf. [16], Fig. 2), etc.
8	RL&SL	An RL model (model-free or world model) + an SL algorithm	RL model associated with SVM or with a Relevance-Based Decision-Tree Learning (cf. [1], *RBDTL in section 19.4.2*), etc.
9	RL&UL	Self-Supervised Learning	Self-Supervised Reinforcement Learning [18], etc.
10	RL&DL	ANN + an RL model (model-free or world model)	Deep Reinforcement Learning (DRL) [19, 20], (cf. [16], Fig. 2), etc.

From these ten (10) ML categories or AI paths depicted in *Figure 1* and *Table 2* of this paper, it is possible to reinvent new combinations with the same intention of creating more capable, more intelligent AI agents (including Expert Systems).

5. THE PATHS TOWARDS SUPER INTELLIGENCE OR REAL MACHINE INTELLIGENCE

As we have seen, nowadays, humans are able to create intelligent machines even if these systems have limited capabilities. However, the ultimate goal of machine intelligence creation is not the current state of AI, the final objective is creating conscious and safe machines that can serve humanity ambitions for a better world (cf. [1], section 26.3, p. 1052, para. 6).

In this section, after a clear presentation of disciplines that have an important influence on the AI methodologies and strategies, we will provide a summary of Views and Concepts that guide worldwide AI researchers in their daily work. Finally, we will show how the combination of the current state of AI paths presented above (*Table 2*) is being continuously improved worldwide towards the goal of creation of true Machine Intelligence, conscious machines that can really behave like human beings.

5.1. The Foundations of Artificial Intelligence

Findings and tools of eight (08) disciplines are used to greatly contribute to ideas, viewpoints, and techniques of AI. We present them all below (cf. [1], section 1.2).

5.1.1. Philosophy

From the Philosophy discipline, several ideas supported various theories about the mind and the body. Good examples of these theories are ***rationalism*** (*expressing how the mind arise from a physical brain*), ***dualism*** (*stipulating that the human mind or soul or spirit is outside of nature, exempt from physical laws*), ***materialism*** (*the brain's operation constitutes the mind*), ***empiricism*** (*knowledge comes from the senses*), ***induction*** (*knowledge comes from exposure to repeated associations between elements*), ***positivism*** (*knowledge comes from observation sentences*), ***confirmation theory*** (*knowledge comes from experience*), and the ***intentionality theory*** (*systems have three (03) levels of abstraction: physical stance, design stance and intentional stance such as beliefs*). All these philosophical theories had been and are still being helpful for the development of the AI technology.

5.1.2. Mathematics

Mathematics led AI to the leap to a formal science in three (03) fundamental areas: logic, computation (calculus and matrix algebra, etc.), and probability.

5.1.3. Economics

The Economics discipline formalized the problem of making decisions that maximize the expected outcome to the decision maker through various theories such as Decision theory, Game theory, Operations research, Markov decision processes, Satisficing, and Decision-theoretic techniques.

5.1.4. Neuroscience

In Neuroscience, the invention of Electroencephalograph (EEG), the recent development of Functional magnetic resonance imaging (fMRI), and advances in single-cell recording of neuron activity, help in significantly improve AI models.

5.1.5. Psychology

The fifth discipline is the scientific Psychology, which contributes to AI through the *Behaviorism movement* (input-output stimulus), the *Cognitive psychology* which views the brain as an information-processing device, and the field of *Cognitive science* which shows how computer models could be used to address the psychology of memory, language, and logical thinking.

5.1.6. Computer Engineering

The Computer engineering discipline provides the ever-more-powerful machines (hardware and software) that make AI applications possible. From here we get computational power and necessary ML/AI packages as software.

5.1.7. Control theory

The seventh discipline, Control theory, deals with designing devices that act optimally on the basis of stable feedback from the environment (self-controlling machines without software or language, as opposed to digital computers, and from which derives the Cybernetics). Its similarities with AI reside in the calculus and matrix algebra tools.

5.1.8. Linguistics

The last discipline is Linguistics, which intersects with AI in a hybrid field called Computational Linguistics or Natural Language Processing (implying syntaxial and semantical understanding of language) and also helped in recent development of Knowledge representation. It provides tools for giving language ability to machines.

5.2. Views and Concepts of Artificial Intelligence

In this section, we will provide simple and clear definitions of concepts of AI that lead various groups of AI researchers during their work. They are necessary elements of definition and utility for a clear understanding of the AI field. Indeed, depending on their feelings we could find researchers who are proponent of one or another form of these conceptual AI beliefs.

First, the assertion that machines could act as if they were intelligent is called the Weak AI hypothesis by philosophers. Weak AI is so about a simulation of the human intelligence. But, the assertion that machines that act as if they were intelligent are actually thinking (not just simulating thinking) is called the Strong AI hypothesis. Strong AI is so about true or real intelligence from machines themselves (cf. [1], intro of chapter 26).

Second, Narrow AI is about the development of an AI that only covers a specific domain. It is the mastering of humans' domain-specific knowledge or expertise by machines. An AI agent solving only a chess game or solving only translation of specific languages are good examples of narrow AI. Unlikely, Artificial General Intelligence (AGI) is a General-Purpose Intelligent standalone agent that can perform at human-level in a wide variety of environments such as for example playing games, driving cars, cooking, teaching, etc., all without the need for a full retraining redeployment process for new unseen domains. AGI is also called Human-Level AI (HLAI) and it is a hierarchical model of intelligence (section 4.2 of this paper) while Narrow AI is a multiple aptitude model (cf. [1], section 1.3.9). AGI should not be confused with the term Global Artificial Intelligence (GAI) that is about all AI strategies in all industry verticals (Finance, Retail/ECommerce/M-Commerce, Healthcare/Pharma/BioTech, Energy, Education, Insurance, Manufacturing, Telco, Auto, Hi-Tech, Media, Agriculture, Chemical, Government, Transportation etc.). Generally, these various AI strategies are presented during conferences such as the Global Artificial Intelligence Summit & Awards (GAISA), the Global Conference on Artificial Intelligence and Applications (GCAIA), and the Global Artificial Intelligence Virtual Conference.

Third, according to the way an AI agent provides its results we distinguish the following concepts. The Black-Box AI is an AI model that learns from unknown parameters. It could perform with high performance, but it is not possible to explain the logic behind the given results. A good example of Black-Box AI model is an Artificial Neural Network, because researchers have no way to understand the parameters' values (numbers) in order to know exactly how the ANN model makes its choices. The Interpretable AI is an AI model that provides its results (outputs) in an understandable way by humans. For example, even if an ANN is a black-box model we can make its outputs understandable via coding in order to allow humans to directly recognize the results instead of getting just numbers [21]. Just as an example, the *Input of an ANN* could be: [0.05 1.75 2.87], and its *Output* could be: [0.87 0.13]. So, for detecting between banana and orange, the result of an Interpretable AI will appear directly as ―banana‖ or ―orange‖ and not as the above numbers (via coding). The Explainable AI is an AI model that can explain exactly why an AI model provided a result, throughout a coherent analysis of the decision milestones. For this transparency, it is also called a *White-Box model*. The Decision-Tree (DT) is a good example of explainable AI model because all steps of a DT decision are explainable and understandable. Unlike the DT, the Black-Box ANN cannot be naturally explainable because no one can't know the logic behind its (sometimes) billion parameters' values, even if we can make its outputs interpretable. Moreover, there is a new related concept called Actionable AI that is being developed and proposed by Moto DEI, a Data Scientist and Actuary with more than fifteen (15) years of experience in media, marketing, insurance, and healthcare. A bit like the RBAC and FBAC techniques in Cybersecurity, Actionable AI makes another stride above Explainable AI by filling the holes. Indeed, in addition to explanation of the results of an AI model, the Actionable AI also provides only working and clarified guidelines for getting better performance.

Fourth, as in Artificial Intelligence an immense significance is attached to natural language and an intelligent system's ability to use it (cf. [5], P.7, para. 1), let us define Natural Language Processing (NLP). As stated in its name, NLP is the way machines process humans' natural languages, first in order to communicate or speak with them, and second in order to acquire or receive information from written language. This led to the two (02) subfields of NLP: Natural Language Understanding or Knowing or Interpretation (NLU or NLI) that is the process of reading and interpreting language [22], and Natural Language Generation (NLG) that is the process of writing or generating language [22]. NLU or NLI (when ―I‖ stands for Interpretation) should not be confused with *Natural Language Inference (NLI)* that is *the problem of determining whether a natural language hypothesis h can reasonably be inferred from a natural language premise p* (cf. Bill MacCartney Ph.D. Thesis in June 2009). It is like the generation of a short comprehensive summarization of a statement. NLI (when "I" stands for Inference) is a subtopic of NLU and is used to improve many NLP tasks such as Question Answering and Machine Translation. For example, *Machine Translation (MT)* is naturally a NLU task because it consists in reading and interpreting a language into another one. MT translates into another language (input and output languages are different). NLI infers from a predicate in the same language (input and output languages are the same). But both MT and NLI (when ―I‖ stands for Inference) are NLU tasks.

Fifth, in the following we provide various definitions related to the creation of real Machine Intelligence which is an Artificial Intelligence according to the Strong AI viewpoint (cf. [1], section 26.3 and [23]). An Ultraintelligent Machine is a machine that can far surpass all intellectual activities of any man however clever. The apparition of such systems will lead to an ―machine intelligence explosion‖ so that machines will create themselves better and better new machines, provided that they will be docile enough to tell us (humans) how to keep them under control. The Technological Singularity is another name of the ―machine intelligence explosion‖ defined by the mathematics professor and science fiction author Vernor Vinge (1993). And the Superintelligence is also another name of the Ultraintelligent Machine or Singularity [24].

Looking at their final goal, Superintelligence, Ultraintelligent Machine and Singularity are the same state of conscious machines.

Sixth, in the following we underline various definitions related to the safe use of AI and the possible dangers. The Ethics of Artificial Intelligence is the branch of the ethics of technology specific to artificially intelligent systems. It is sometimes divided into a concern with the moral behavior of humans as they design, make, use and treat artificially intelligent systems, and a concern with the behavior of machines, in machine ethics. It also includes the issue of a possible singularity due to superintelligent AI (cf. Wikipedia, Ethics of AI). The Risks of Artificial Intelligence could be defined as the possible dangers, threats or vulnerabilities that appear during the development of artificially intelligent systems. These risks could lie at different level of the AI system building process (before building, during building or after building and deployment). Some good examples of Risks of Developing Artificial Intelligence are the risk of misconceptions about AI systems' building, the risk of loss of job due to automation, the problem of too much or too little leisure time due to AI, the misuse of AI systems, the possibility of AI system failure, and the risk of human race extinction.

Here, we know that the *Aquinas model of the intellect and will* (cf. section 4.2) implies Generic Operations or Acts, Common sense (sensus communis), Imagination, Memory, and a Cogitative power so that it can synthesize objects into a coherent representation, handle mental image of something in its absence, store perceptions and perform an evaluation of a perception with respect to the interests of the perceiver, i.e., whether it is beneficial (useful) or harmful. So, we think it is *very suitable for dealing with Ethics and Risks of AI when trying to create beneficial AI*. As we are turned towards the creation of Artificial General Intelligence (AGI) and not Artificial General Demolition, we have with this concept an independent and impartial definition of intelligence including all necessary components that match the AI field (the *intellect concept* leading to creativity and the *will concept* leading to consciousness).

5.3. The Turing Test in the Middle

Alan Turing (1912-1954) is one of the fathers of computer science because he provided the characteristics of computable functions which are able to be turned into a computer program (1936). He also provided the most influential work for a solution to the AI problem, through his paper "Computing Machinery and Intelligence" wherein he introduced the Turing Test, Machine Learning, Genetic Algorithms, and Reinforcement Learning.

The **Turing Test**, proposed by Alan Turing (1950), is an interview test that was **designed to provide a satisfactory operational definition of intelligence**. In order to pass the rigorously defined **total Turing Test**, a computer would need to possess the following capabilities which are six (06) disciplines that compose most of AI: Natural Language Processing, Knowledge Representation, Automated Reasoning, Machine Learning, Computer Vision, and Robotics. Indeed, **the total Turing Test includes a video signal** so that the interrogator can test the subject's perceptual abilities, as well as the opportunity for the interrogator to pass physical objects through the hatch.

Conditions for successfully pass the Turing Test: A computer passes the Turing Test if a human interrogator, after posing some written questions, cannot tell whether the written responses come from a person or from a computer. In other words, a machine passes the Turing Test if and only if the human interrogator is not able to distinguish between both the human intelligence and the artificial intelligence. Therefore, the Turing Test is necessary for checking if a created AI model has truly reached Machine Intelligence.

Conceptually, Alan Turing claims that in order to think machines should have **consciousness** (awareness of their own mental states and actions). While Professor Geoffrey Jefferson from its side relates to **phenomenology** (emotional machines), and others focus on **intentionality** pretending that the machine thinks if its purported beliefs, desires and other representations are about something in the real world. **Therefore, Turing concluded that instead of focusing on a specific philosophy, what matters is focusing on experiencing a machine that really act intelligently, such that it can hold a dialog with a true human**. Reaching this goal, **we will no longer distinguish between real and artificial thinking**, exactly like we cannot distinguish today artificial urea (synthesized) from real urea (natural) because both have the same physical (or biological) properties. The issue of the difference between real and artificial thinking will eventually go away by itself **once machines reach a certain level of sophistication**. And consequently, **the difference between weak and strong AI will disappear** (cf. [1], sections 1.1.1 and 26.2). The famous Turing Test remains until today (in 2023) the biggest challenge for the existence of AI; so far, no AI system worldwide has been able to pass the Turing Test. In the meanwhile, researchers still pursue their works towards understanding the human consciousness. Andrew Budson, a researcher and professor of neurology at the Boston University and his colleagues recently developed a new theory of consciousness suggesting that decisions are made unconsciously, then about half a second later, they become conscious [25]. Their theory presents Consciousness as a Memory System, stipulating that consciousness developed as a memory system that is used by our unconscious brain to help us flexibly and creatively imagine the future and plan accordingly. "What is completely new about this theory is that it suggests we don't perceive the world, make decisions, or perform actions directly. Instead, we do all these things unconsciously and then—

about half a second later—consciously remember doing them.". They believe that their theory may have profound implications for understanding intentional action and consciousness in general. With further exploration, their work could provide insight into philosophical issues around free will and moral responsibility, and why not into the development of conscious machines. Here also, the Aquinas model of the will (Common sense, Imagination, Memory, and Cogitative power) in section 4.2 perfectly matches with their strategy.

5.4. Current paths towards Superintelligence or real Machine intelligence

We know that from categories of ML that we defined as AI paths (Figure 1 and Table 2), it is possible to reinvent new combinations with the same intention of creating more capable and more intelligent AI agents. However, even if the combination of various ML techniques leads to AI, the created AI is nowadays a narrow AI, i.e., current AI machines can only master a specific domain (a specific game only, or vision only, etc.). Narrow AI machines use more powerful, domain-specific knowledge in narrow areas of human expertise, ensuring that all aspects of the domain are covered (cf. [1], section 1.3.5). For humans, the narrow AI corresponds to the narrow content of a particular mental state (cf. [1], section 26.2.1). One AI model cannot yet automatically switch from game playing to car driving or surgery, as a human surgeon is able to do. The capability for machines to aggregate several narrow AI skills, even learn new general tasks and automatically behave like human (using a single AI architecture or model) is called Artificial General Intelligence (AGI) as defined above.

The research toward AGI is still ongoing, but the three (03) Pioneers and Godfathers of AI who won the Turing Award in 2019 (Yann LeCun, Geoffrey Hinton and Yoshua Bengio [26]) propose different paths made of Deep Learning variants and that are promising to possibly lead to AGI. This Award is about significant solutions towards solving the Turing Test. Professor Yann LeCun from its side focuses on UnSupervised Learning (UL) techniques through the path of SelfSupervised Learning [17] (Encoding methods) in order to reach Artificial General Intelligence. Professor Yoshua Bengio vision is the System 2 Deep Learning [27], and Professor Geoffrey Hinton is a firm believer in Deep Neural Networks Systems [28] such as CapsNet, an approach inspired by the structure of the human brain. Recently, in December 2022 he invented the promising Forward-Forward Algorithm that replaces the forward and backward passes of backpropagation by two forward passes (arXiv:2212.13345v1) and is closer to the way the human brain truly processes information.

Additionally, there exist other approaches towards AGI such as OpenAI that among several possibilities focuses on the path of Reinforcement Learning. For example, Universe is an OpenAI reinforcement-based software platform for measuring and training an AI's general intelligence across the world's supply of games, websites and other applications. We also have OpenCog that is an open-source framework for developing AI systems, especially appropriate for ***integrative multi-algorithm systems (Artificial Intelligence systems*** as we defined in Figure 1 and Table 2***)***, and ***Artificial General Intelligence systems***. OpenCog is the framework that runs the android robot *Sofia*. Dr Ben Goertzel, the creator of the android robot Sofia, recently defined ―*The General Theory of General Intelligence: A Pragmatic Patternist Perspective*‖ [29]. Moreover, in [1] section 27.1, page 1062, just before the last paragraph the authors recommend initiatives combining hierarchical representations (closer to human reasoning system, i.e., hierarchical models of intelligence or general intelligence we mentioned in section 4.2) with the probabilistic languages, in order to create more intelligent agents. The probably first initiative of this kind is DeepProbLog [30], a probabilistic logic programming language that incorporates Deep Learning by means of neural predicates (gathering deductive and inductive learnings), and thus combining general-purpose neural networks and expressive probabilistic-logical modeling and reasoning. Inductive learning in a logical setting is a process of gradually eliminating hypotheses that are inconsistent with examples, narrowing down the possibilities (cf. [1], section 19.1.1). DeepProbLog is therefore another path toward AGI. This goes in the same direction than the recent IBM hybrid question-answering system called Neuro-Symbolic-QA (NSQA) that is based on Logical Neural Networks [31] and performs better in real-world situations with State-Of-TheArt (SOTA) performance. This approach is so promising that it has been thoroughly developed by Dr Ben Goertzel at the 15th Annual AGI Conference (AGI-22) that held from 19-22 August 2022 in Seattle, Wa, USA, via the Workshop 1: Scaling up Neural-Symbolic AGI with OpenCog Hyperon.

On another side, the scientist, best-selling author, and entrepreneur Gary Marcus proposed in 2020 four (04) Steps Towards Robust Artificial Intelligence [32] (i.e., AGI) by the next decade. Indeed, he argues for a hybrid approach to AGI as the best way forward, mixing and matching deep learning and knowledge representation as exemplified by Knowledge Graphs. Here, a difference should be made between Knowledge Graph (ontology representation such as Semantic Web) and Knowledge Map (or Bayesian Network including Conditional Probability Table. cf. [1], chap. 14).

Another important initiative towards AGI is with the DeepMind company, a subsidiary of Alphabet that is involved in Solving intelligence to advance science and benefit humanity. Their recent finding of a single Generalist Agent beyond the realm of text outputs called **GATO** (arXiv:2205.06175) is, despite criticism, a proof that they are also clearly involved in solving AGI. With a single set of weights and hundreds of possible tasks, Gato can engage in dialogue, caption images, stack blocks with a real robot arm, outperform humans at playing Atari games, navigate in simulated 3D environments, follow instructions, and more.

We also present from the Allen Institute for AI, **Unified-IO** (arXiv:2206.08916v2) that is a significant milestone in the pursuit of a single unified general-purpose system (AGI) capable of parsing and producing visual, linguistic, and other structured data. It is a single sequence-tosequence Unified Input-Output model that performs tasks across more than 80 diverse computer vision and NLP benchmarks. Unified-IO is the first model capable of performing all 7 tasks on the General Robust Image Task (GRIT) Benchmark and produces strong results across 16 diverse benchmarks with no task-specific fine-tuning. Unified-IO is a pure transformer model largely following the design of T5, while Gato is a Transformer with ResNet patch embedding for multitask, multi-modal behavior cloning.

Without being exhaustive, a last interesting initiative towards AGI is from the Boston Dynamics robotics company that aims to imagine and create exceptional robots that enrich people's lives. One of their AGI-like creation is the Spot flexible platform (including Legged Robots in Public Safety) that can be easily deployed in a variety of environments and hazardous situations. The flexibility of their robots is a hope that we are moving close to AGI.

We stay focused on all of them for solving the Total Turing Test.

Anyway, the future of AI (starting by AGI) will surely emerge from one of the above current paths towards AGI. In case it is something totally new, it will be then classified as a new AI path. And, when we will reach that Human-Level AI (i.e., AGI), it will be the launchpad towards ***Ultraintelligent machines*** also called the ***Superintelligence*** or ***Singularity***, the era of conscious machines.

However, AI practitioners (AI Scientists, AI Engineers, AI Technicians, etc.) cannot separate AI research from its ethical consequences. We will cope with this aspect in the next section.

6. ETHICS AND RISKS OF ARTIFICIAL INTELLIGENCE

6.1. Performance of AI Systems

AI researchers determined that the Performance of AI Systems is provided via three (03) powers: **Hardware** (*providing computational power*), **Software** (*including hard code, algorithms, AI models and frameworks*), and **Data** (*providing the workpiece for AI*) [33, 34, 35, 36]. For this reason, ―three of the main areas where significant innovation will be traced back to AI are hardware, software, and data.‖ [36]. Indeed, ―while hardware accelerators (CPUs, GPUs, TPUs and FPGAs) can deliver impressive AI performance improvements, software

AI accelerators are required to deliver even higher orders of magnitude AI performance gains across deep learning, classical machine learning, and graph analytics, for the same hardware set-up.‖ [34, 37]. Additionally, the improvement in data quality can bring significant increase in the performance of AI models, for the same hardware and software setup [38] (an increase of at least 17% in the model performance for data quality improvement had been proven).

Consequently, in Artificial Intelligence ethical concerns can emerge from these three (03) main aspects. A flaw inside the **data** can be reproduced by the AI and lead to unforeseen behavior of AI models. Also, a bad orientation of the AI development process can be performed at the **software** and **hardware** levels in order to willingnessly or not, harm people. For example, it is possible to fool facial recognition systems via various wearable patterns (shiny fabrics, beanie hats, Hyperface, etc.), but thoroughly designed data may help in overcoming this flaw.

So, in the next section, we will first deal with **Ethical concerns about Data** the workpiece of AI systems, then we will treat Ethical concerns about **Hardware and Software** the mind-body entities of AI systems.

6.2. Ethical concerns due to Data

6.2.1. Human biases

We have seen in Figure 1 that ML and AI are based on data (their workpiece), without data there is no ML, no AI. And as stated on omniscien.com, ―*Data is the Fuel that Powers Artificial Intelligence*‖. Therefore, there is an important attention to put on the nature and quality of the data we use for ML and AI tasks. For example, toxic data could influence the AI models to produce toxic answers. But this is fundamentally due to an influence of human biases. As said Beena Ammanath, the global head of the Deloitte AI Institute, and tech and AI ethics lead at Deloitte: ―*As humans, we are highly biased, and as these biases get baked into the systems, there is very high likelihood of sections of society being left behind—underrepresented minorities, people who don't have access to certain tools—and it can drive more inequity in the world.*‖ [39]. Indeed, **Cognitive Biases** are systematic patterns of deviation from norm and/or rationality in judgment (cf. Wikipedia, List of cognitive biases). These biases affect belief formation, reasoning processes, business and economic decisions, and human behavior in general. ―*For almost a century, scientists have been studying how the brain processes external information to form the basic senses of sight, smell, hearing, taste, and touch that we use to navigate the world. Over time, they have compiled their findings to show how the various sensory areas in the brain are organized to represent different stimuli*‖ [40] (close to the empiricism view of section 5.1.1 and also to the Aquinas model of the intellect of section 4.2).

So, as we saw that AI itself is trying to reproduce the human brain language, let us discuss the biases from the five (05) human senses that could influence our interventions in the AI development process. We know that the **Perception** of humans is based on their five (05) senses: ***sight*** (vision), ***sound*** (hearing), ***smell*** (olfaction), ***taste*** (gustation), and ***touch*** (tactile perception) [41]. First, we denote some biases in the visual perception such as bias towards the center of space (eg. screen center), spatial layout of objects in the environment (occlusion, etc.), parallax effect, etc. Second, there exist biases in the auditory perception such as the bias towards the periphery (exact location of a sound), etc. Third, there are biases in the smell perception including meal-matched odors, subliminal odors influence on preference, etc. We also denote biases in the taste perception such as systematic bias in taste memory (influence of labels on taste recognition related to food quality and preference), etc. And finally, we have some biases in the touch perception such as visual pursuit biases tactile velocity perception, problem of subjective vertical in patients with unilateral neglect, etc. However, even if the human senses are individually highly biased, their multisensory integration via the brain not only benefits perception by improving precision (reducing variability), but also by reducing bias; a win-win approach. Therefore, the multisensory integration leads to the reduction of perceptual biases and so to a better judgment in the decision-making process of humans. That is the reason why this ability is reproduced by AI via the fusion of technologies for better performance. But **from their senses and perception, humans have many other biases that could flaw someway their judgment and lead to biases in the data collection process** (There are biases about what we should remember, biases linked to the amount of information we process, biases about the meaning of things for us, and biases about the reason why we act.). We present all these known human biases in the appendix, and we will discuss the known biases in the data collection process during the next section. As said the AI pioneer Andrew Ng, ―it's time for smart-sized, "datacentric" solutions to big AI issues‖, and these solutions are powered by human intelligence

(implying human involvement) via manual or automated approaches. Moreover, ―*In the world of artificial intelligence, data labeling is the most laborious part of the process and the path to automated data labeling is not clear-cut, and the need for some human oversight is still inevitable"* (marktechpost.com). Therefore, **solutions powered by human intelligence in the creation of better data**, also known as Human-Centered or Human In The Loop (HITL) solutions, **are necessary for the data collection process, while dealing with the biases (due to our cognitive biases)**.

6.2.2. Biases in Data

Now, we know that we must deal with biases during the Data Collection process, partly due to the biases in the perception and behavior of humans themselves, in order to not reproduce them in AI systems. Below we present all known types of possible biases in data that researchers revealed by their work [42, 43]:

1-) *Fallacies in Language resources or corpora*: Bias in Diversity and Inclusion of languages (imbalanced data, out of sample data, etc.), Bias across domain, Bias in annotating social data (invalid data, labeling errors, ambiguities, human input into data, out of sample data, etc.), Gender bias in data/corpus, Bias in syntactic and shallow parsing, etc.

2-) *Fallacies in Word Representations*: Bias in word representations, Representation bias in social data (Danger of proxy variables such as the influence of *level of study*, etc.), Representation bias in applications, Representation bias in language resources, etc.

These biases in the data could lead to various stereotypes based on age, gender, racism, sexism, religion, ideological, political, moral, ethnicity, disability, sexual orientation, user features, and location in AI Models. Also, other biases appear from intersectional bias attributes such as demographic (age, gender, location), gender+racism, and gender+location+job, but are not extensively studied. They have been presented in some works as allocational harm, questionable correlations, or vague/unstated.

Also, as we are talking about the Data Collection process, it is important to know what sort of Data we are collecting. Indeed, we distinguish four (04) types of Data: Raw Data, Dark Data, Dirty Data and Clean Data, that we will define in the following lines.

Raw Data: Also known as *Primary Data*, raw data are data (e.g., numbers, instrument readings, figures, etc.) collected from a computer program or a manual source (e.g., survey) and that have not been subjected to any intended processing or analysis. Raw data is a relative term, because even once raw data have been "cleaned" and processed by one team of researchers, another team may consider these processed data to be "raw data" (i.e., primary data) for another stage of research (cf. Wikipedia).

Dark Data: Dark data is data which is acquired through various computer network operations but cannot be directly used in any manner to derive insights or for decision making. In an industrial context, dark data can include information gathered by sensors and telematics. They are unrefined data and are often stored for regulatory compliance and record keeping (cf. Wikipedia). Unlike Raw Data, Dark Data is not a primary data but the output of a working computer process that is stored, not immediately used, but can possibly be reused later.

Dirty Data: It is a Statistically irrelevant Data containing bad or undesirable patterns, possibly including also some good entries. It is gathered from as many sources as possible. Here, Data quality is not important while Data quantity is important (cf. omniscien.com).

Clean Data: Clean Data is a more statistically relevant Data without bad or undesirable patterns, containing only good entries. It is gathered from a small number of trusted quality sources. Here, Data quality is VERY important while Data quantity is less important (cf. omniscien.com). In a domain-specific expertise, *Dirty Data* and *Clean Data* are both generally collected via a data processing concept called Extract-Transform-Load (ETL) and then put into the *same features*. Only the *amount and quality of entries* are different for both data classes.

Therefore, as we have seen possible biases that are present in these collected data, in the next section we will talk about how to detect biases in Data and solutions for debiasing them.

6.2.3. Debiasing Data

Happily, there exist several Metrics for Evaluating (above) Biases in Data, but the detection and quantification of bias are not always possible by using formal mathematical techniques. Known metrics for bias detection in data are the WEAT metric for SGNS model, the relative norm distance, the WEAT1, WEAT2 metrics for bias b/w embeddings, the SEAT metric for bias in sentence embeddings, the RIPA metric for any embedding model, the MWEAT metric for bias in Machine Translation, the RNSB metric for bias in Sentiment Analysis, the Disparate Impact and the Adjusted Disparate Impact metrics for NLP tasks, the Statistical Parity (SP) Subgroup Fairness metric and the Intersectional Fairness metric for NLP intersectional biases [42, 43]. Consequently, after detecting and measuring biases in data, several Debiasing Techniques had been proposed including Hard-debias, Soft-debias, RBA, Auxiliary dataset, Named Entity anonymization, Data Augmentation, Neutralization and SENT-DEBIAS.

However, in addition to these debiasing techniques, better and sustainable initiatives for debiasing data are those which deal with the data correctness not only at the collection phase, but also during the AI application building and deployment process with feedback-improvement schemes. This means that based on the feedbacks we should be able to recursively reannotate the data for further improvements in the AI models. This viewpoint is also supported by Andrew Ng who said that "*In many industries where giant data sets simply don't exist, I think the focus has to shift from big data to good data. Having 50 thoughtfully engineered examples can be sufficient to explain to the neural network what you want it to learn.*". And data engineers have enough contributions to bring here.

On the other side, there exist several platforms engaged in this direction (*without being exhaustive*) such as Kili-Technology (a Labeling Platform for High-Quality Data), Collibra (a Market Leader in the BARC Score Data Intelligence Platforms Report), Omniscien (a Specialized Machine Translation Platform with suitable tools for Data Creation, i.e. language pairs), Isahit (a Platform for Ethical Data Labeling Solutions for AI and Data Processing, powered by human intelligence), Super AI (combining Manual and Automatic Data Labeling approaches with Customization), Cnvrg.io and Dell Technologies (combining the ML platform, data, and software into a single manageable release cycle), etc.. Standardization and implication of citizens for improving data quality under a review by experts of the field (a Human-Centric approach) is also a good initiative (datainnovation.org).

6.3. Ethical concerns due to Software and Hardware (mind-body view)

6.3.1. The Evidence of Risks in the Development of AI

We already gave some examples of Risks of Artificial Intelligence Development including the misuse of AI systems, the possibility of AI system failure, and the risk of human race extinction. Indeed, Ethical threats that come from the development of the AI technology cover three (03) global areas. First, we have the possibility of *lost in control of AI machines* perhaps due to unforeseen contextual use cases by the app creators including a possible flaw in their code or a revolt of conscious machines if we reach real machine intelligence. Second, there are *crimes that could be willingnessly performed by humans* in order to harm people with the AI technology.

And third, we denote the *natural consequences of the AI technology on public life management* including the risk of loss of job due to automation, the problem of too much or too little leisure time, etc.. All these aspects show a clear Evidence of Risks in the Development of the AI technology. But the infrastructures of these AI systems are mainly made of Software and Hardware, before applying them to the data that can be taken at another location. So, dealing with Risks and Ethics of AI, we should consider Software and Hardware together because there is a close dependency link in their actions (like with the mind-body theory).

6.3.2. Mitigating biases in AI models

Software and Hardware mutually and positively contribute to inventions of new generations of AI technology in a fully integrated way. AI models can be compressed and installed inside chips and provide AI performance at hardware level in AI systems. Also, hardware (including GPU) contributes to creation of highly performant AI model via fast computation and parallelism. A good example of full integration of Software and Hardware is the generation of custom hardwareaware deep learning models with unparalleled performance on any hardware, by Deci's AutoNAC technology. But as the AI Model or Software is the mind and the Hardware is the body (in reference to section 5.1.1), the thinking process is provided by the model or software. Nevertheless, we denote that some vulnerabilities could emerge at Software and Hardware levels no matter the data setting. For these reasons, many current initiatives already try to mitigate biases in AI models (software), mostly inside the promising big models like Transformers. Indeed, Transformers are new generations of AI language models that are trained on a huge amount of data and for which the Pre-Trained models could be reused to solve a great number of downstream tasks generally via Transfer Learning (avoiding a full retraining process each time; only some layers are retrained). However, and irrefutably, there exist fallacies such as systematic biases in the word embeddings from Pre-trained Language Models [42, 43] like GloVe, ELMo, BERT, GPT, etc., that can get baked into the AI systems. Fortunately, AI researchers found some methods for mitigating biases in these AI models (arXiv:2110.08527v3). For example, the GloVe model (Static word embeddings) is debiased using the WordED technique. This method iteratively learns a projection of embeddings that removes the bias information with minimal impact on embedding distances. The BERT and RoBERTa models (Contextualized word embedding) are debiased using the ContextED technique which has been shown to work well at removing gender-bias encoded in embeddings. There are also two (02) alternative debiasing methods for the contextualized word embedding models: Counterfactual Data Augmentation (CDA) and Dropout. The CDA augments the training corpora with counterfactual data so that the language model is pretrained on gender-balanced text. The Dropout technique mitigates gender biases by increasing the dropout rate in the pretrained models.

All these debiasing methods for AI models (and so for Hardware and vice-versa due to their full integration) could be classified at different level of biases mitigation: *dataset level* (CDA), debiasing *during pretraining* (ContextED and Dropout), and *post-tuning* debiasing (WordED). This perfectly matches our definition of Risks of AI in section 5.2, that could lie at different level of the AI system building process (before building, during building or after building and deployment).

6.3.3. Explainable AI and Interpretable AI as solutions to Ethics and Risks of AI

When you understand and can explain the reason why a problem happened, you already solved it by half. This is why the relatively recent concepts of Interpretable AI (directly understandable AI by humans in their natural language) and Explainable AI (fully explainable AI milestones provided for any decision) are now guiding AI practitioners in the development of their applications. Even the naturally Black-Box Neural Network Model is being made explainable using various methods in domain such as Urban Space perception (arXiv:2208.13555v1), Financial Tabular Data (arXiv:2209.10658v1) and Spam detection (arXiv:2209.03166v1). Therefore, Explainable AI and Interpretable AI concepts aim to provide transparent AI solutions for every given real-world problem in order to fully understand any decision of AI systems and deal with Ethics and Risks issues that can occur from their development process.

6.3.4. Green AI initiatives

Still for ethical reasons, it has been reported that the AI models' training process consumes tremendous quantity of energy while releasing huge amount of dioxide of carbon (tCO2) in the air, increasing worldwide energy consumption and carbon footprint. For example, the training process of OpenAI's GPT-3 emitted 552 tons of dioxide of carbon (tCO2) against 59 tons for the

Google's Switch Transformer (lemagit.fr). Enters the Green AI concept that combat Climate Change and environmental risks. Green AI is part of a broader, long-standing interest in environmentally friendly and inclusive scientific AI research. Reporting the computational price tag of developing, training, and running models is a key Green AI practice (cacm.acm.org). Unlikely, Red AI refers to AI research that seeks to improve accuracy (or related measures) using massive computational power while disregarding the cost—essentially "buying" stronger results. Therefore, worldwide AI researchers are called to adopt a Green AI strategy during their work for a safe environment and so contributing to the efforts for combating Climate Change.

In this direction, since October 2022, there is exists the Carbon Aware SDK from the Allen Institute for AI, that helps build carbon aware software solutions with the intelligence to use the greenest energy sources. Carbon Aware Software is when software does more when the electricity is clean and does less when the electricity is dirty, or runs in a location where the energy is cleaner. So, this allows software solutions to run at the greenest time, or in the greenest locations, or both. Then, you can capture consistent telemetry and report on your emissions reduction and make informed decisions.

6.4. Global Solutions to Ethics and Risks of AI

Generally, the measures taken for handling Ethics and Risks of AI are open to education (AI teaching and learning) for everybody without restrictions for those who meet these programs' requirements. But for the real-world implementation phase there are some regulations to follow depending on the fields or vertical industries covered by the AI solutions (Health, Army, SelfDriving Cars, etc.), and on the regions (national or international). The Kitchen Knife can hurt you, but as it is useful you don't throw it away, you just clean it and reuse it with more caution. Therefore, accidental mistakes, such as with the Self-Driving Cars (SDC) should not stop the good initiatives towards creating beneficial AI for the Welfare of Humanity and a better Life everywhere on the Earth. Consequently, in order to globally overcome the Ethics and Risks of AI several great initiatives had been adopted worldwide. These initiatives tend to tackle the problem at its source covering problems like AI regulation, AI Safety, AI Development guidelines, etc., no matter the specific frameworks and AI techniques employed. Indeed, these approaches are necessary because the creation of Human-Level intelligence and beyond, would change the lives of a majority of humankind and the future destiny of the human race. AI systems at this level of capability could threaten human autonomy, freedom, and even survival.

One of the first initiatives towards dealing with Ethics and Risks of AI and implying thousands of AI scientists worldwide is the Open Letter from AI & Robotics Researchers from the Future of Life Institute (FLI). They also defined the Asilomar AI Principles for the creation of a beneficial intelligence that helps in avoiding the risk of human race extinction due to AI.

Another big initiative is the Future of Humanity Institute (FHI), a multidisciplinary research institute at the University of Oxford that brings the tools of mathematics, philosophy and social sciences to bear on big-picture questions about humanity and its prospects. The FHI is led by one of the most influential figures in AI, its Founding Director Professor Nick Bostrom.

One more important global initiative is the Montreal Declaration for Responsible Development of Artificial Intelligence, an initiative of Université de Montréal. This declaration has three main objectives: (i) Develop an ethical framework for the development and deployment of AI, (ii) Guide the digital transition so everyone benefits from this technological revolution, and (iii) Open a national and international forum for discussion to collectively achieve equitable, inclusive, and ecologically sustainable AI development. The third principle of this declaration for Responsible AI (also known as Friendly AI) is the Protection of Privacy and Intimacy, that has become a subfield of AI research today (including initiatives from the foundation disciplines of AI such as Computer engineering, Philosophy, Psychology, etc. [44, 45, 46]).

A fourth global initiative is from the Institute for Ethical AI & Machine Learning that proposes 4 phase strategy towards responsible development of AI: by Principle (Empowering individuals through best practices and applied principles), by Process (Empowering leaders through practical industry frameworks and applied guides), by Standards (Empowering entire industries through their contributions to industry standards), and by Regulation (Empowering entire nations through their institute work).

Fifth, we have the Alan Turing Institute that proposes global solutions for safe development of the AI technology. Indeed, proponent of the belief that data science and artificial intelligence will change the world, they are focused on training the next generation of leaders, shaping the public conversation, and pushing the boundaries of these sciences for the public good.

Another interesting global initiative is the Seclea Platform. Seclea de-risks the adoption of advanced AI algorithms for organizations by providing, through its state-of-the-art tools, AI

Regulation, Explainable AI, Responsible AI, AI Risk Management, and Monitoring (accountability, auditability, and trackability). The compliance of the Seclea Platform with any AI Regulation, any AI Model and Any Data makes of it a good candidate for Global Solutions to Ethics and Risks of AI.

An additional good global initiative is the SingularityNET Platform that aims to gather deep selforganizing network of worldwide AI agents running together and producing synergetic effects that will possibly lead to AGI and then to Singularity. While building the foundations for AGI, this Decentralized AI Platform's strategy includes carefully selected vertical markets such as DeFi, Robotics, Biotech and Longevity, Gaming and Media, Arts and Entertainment (Music), and Enterprise-level AI, so that all aspects of the development of AI Technology (including Ethics and Risks Management) are covered.

Without being exhaustive, a last global initiative to overcome the Ethics and Risks of AI is the Booz-Allen-Hamilton AI for Civilian Services, an AI that allows Governments to develop ethical AI that remains human-centered (via their civilians) and protects individual privacy rights consistent with the nation's laws and values. This is a way to ensure AI-driven safe public life by the Public Sector. Of course, an interest for human-centered AI is that (as we demonstrated in Figure 1 & 2 and Table 1 & 2) *although AI could perform better than Humans in several domains, Humans are the best at being Humans. And we should make AI serve Humans' Welfare*.

7. CONCLUSION

As we can see in this paper ―Understanding the Worldwide Paths towards the Creation of True Intelligence for Machines‖, the field of AI is very vast, and the similarity between many aspects of AI could lead to mistakes and/or confusions for newbies and even some practitioners. Our paper proposed an approach that mitigates this struggle by covering the worldwide milestones for the development of the Artificial Intelligence technology, in a very comprehensive way. We provided clear insights about the foundations of Machine Learning, the foundations of Artificial Intelligence, and related Theories of Intelligence that led to current Views and Concepts of Artificial

Intelligence. Also, on the paths towards Superintelligence or real Machine intelligence, we find in the middle the famous total Turing Test that remains until today (in 2023) the biggest challenge for the existence of AI; so far, no AI system worldwide has been able to pass the Turing Test. However, as the development of the AI Technology regrettably implies enormous risks for human species, and to overcome Ethical concerns about AI at the Data, Software, and Hardware levels, exciting worldwide Specific and Global Solutions to the management of Ethics and Risks of AI had been provided.

Even if this comprehensive overview cannot suddenly make of the reader an expert of the AI field due to possible limitations in novelty and technical aspects, it can bring a very good comprehension and invaluable insights. Indeed, based on the analysis of empirical studies that we identified and as a personal and modest contribution to scientific literature, we presented Artificial Intelligence with all elements of definition and utility, and we extracted very useful insights assembled via carefully designed illustrations (including Figure 1 & 2 and Table 1 & 2) so that there is a good knowledge of AI. All these, especially in the situation where the digitization and involvement of AI appears on a global level in all fields of activity, which makes our work have a multiplier effect. We hope our research will shed lights to anyone with interest in the field of AI (including newbies, some practitioners, developers, teachers of various disciplines for a fast and clear overview about AI, students, decision makers, etc.). Future research will focus on the correlation between AI and all other disciplines (i.e., AI strategies in all industry verticals, section 5.2), on the way AI is used in their activities, and so proving the global implication of AI in all fields of activity. All contributive suggestions are welcome.

8. ACKNOWLEDGEMENT

We would like to thank Professor William Martin our Supervisor, Professor ABDO Miled Abou Jaoude and the BIU staff for the reviews of the reports of our Ph.D. program, and for the access to a partial scholarship leading to this highest level of specialization in Artificial Intelligence. Invaluable things we learned there are the supports for writing this paper.

We are also thankful to the Ivoirian Civil Service, as we got access to this Ph.D. program in 2018 as a Computer Scientist Civilian (Civil Servant or Official) from Côte d’Ivoire. Local initiatives encourage commitment to training and to the field of research for civilians.

9. DECLARATIONS

9.1. Funding

No funding was received to assist with the preparation of this manuscript.

9.2. Conflicts of Interests

The authors have no conflicts of interest to declare that are relevant to the content of this article. Anyone interested in getting more knowledge about the (hard) technical details of AI should look at the resources provided in the references or at other alternatives. It is required at least two (02) years of University Mathematics (sophomore level) for a path towards a full understanding in order to reach the highest level of specialization in Artificial Intelligence. Loving Mathematics sooner could help (99math.com).

9.3. Availability of Data and Material

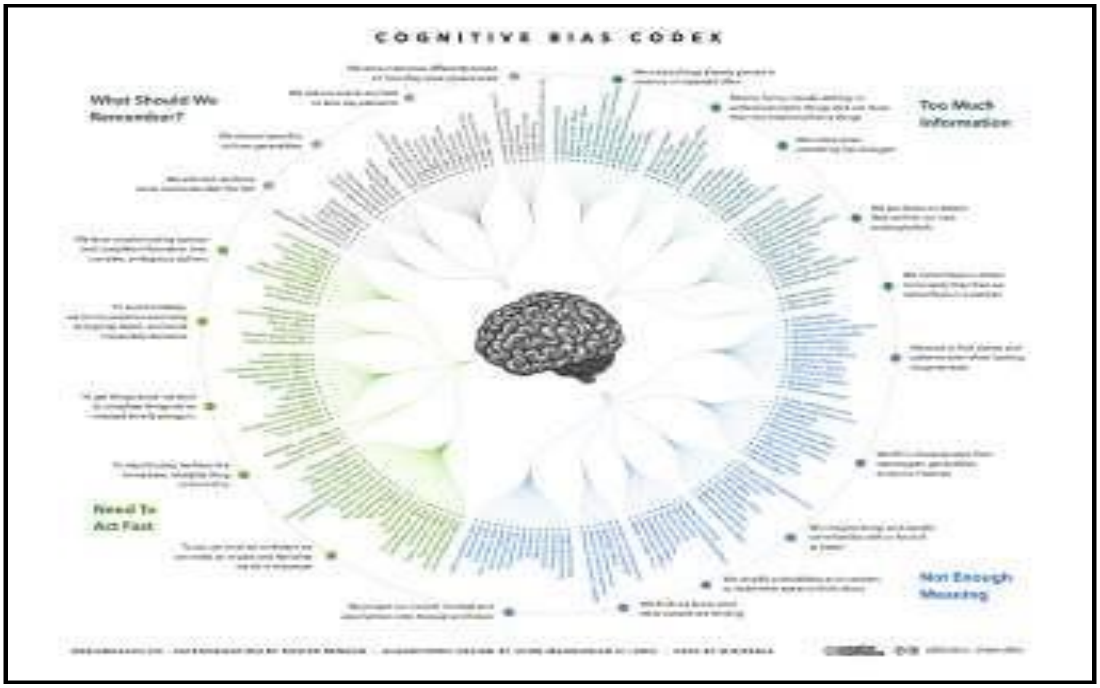
Not applicable

9.4. Code Availability

Not applicable.

APPENDICES

The Cognitive Bias Codex, designed by John Manoogian III



Credit: From Wikipedia, Cognitive Bias Codex, designed by John Manoogian III

REFERENCES

Stuart J. Russell and Peter Norvig, ―Artificial Intelligence: A Modern Approach,‖ Pearson India Education Services Pvt. Ltd., India, Third Edition, twelfth Impression 2018.

Kouassi Konan Jean-Claude, Ph.D. Student in Artificial Intelligence at Bircham International University (BIU) – Madrid since 2018. (2021, February 3). *Case Study of the 4th Book Report, 35 pp*. The Artificial Intelligence Realm and the implied Techniques [Figure], p.26.

Harry Surden, ―Artificial Intelligence Overview,‖ Associate Professor of Law, University of Colorado Law School, Oct. 2017.

Heaton, Jeffrey. (2017). Ian Goodfellow, Yoshua Bengio, and Aaron Courville: Deep learning: The MIT Press, 2016, 800 pp, ISBN: 0262035618. Genetic Programming and Evolvable Machines. 19. 10.1007/s10710-017-9314-z.

Mariusz Flasiński, ―Introduction to Artificial Intelligence,‖ Springer Nature, Switzerland, 2016.

―The Astronomist: A Cubic Millimeter of Your Brain.‖ The Astronomist, http://theastronomist.fieldofscience.com/2011/07/cubic-millimeter-of-your-brain.html. Accessed 14 Aug. 2018.

La Communication Entre Les Neurones : La Transmission Synaptique - Assistance Scolaire Personnalisée et Gratuite - ASP. https://www.assistancescolaire.com/eleve/4e/svt/reviser-unenotion/la-communication-entre-les-neurones-la-transmission-synaptique-3sad06. Accessed 14 Aug. 2018.

Robert Stufflebeam. (n.d.). Neurons, Synapses, Action Potentials, and Neurotransmission. The Mind Project. https://mind.ilstu.edu/curriculum/neurons_intro/neurons_intro.html.

Jeff Beck, Editor. (2015, January 15). ―The Formation of Multi-synaptic Connections by the

Interaction of Synaptic and Structural Plasticity and Their Functional Consequences,‖ Ncbi.Nlm.Nih, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4295841/.

Fabien Tell, Editor. (2015, June 24). ―The Influence of Synaptic Size on AMPA Receptor Activation: A Monte Carlo Model,‖ Ncbi.Nlm.Nih, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4479604/.

Shane Legg et al., ―A Collection of Definitions of Intelligence,‖ arXiv:0706.3639v1, Jun 2007.

H.-P. Schwefel et al., Advances in Computational Intelligence – Theory and Practice, Springer, Germany, 2010.

Songbai Liu, ―A Survey on Learnable Evolutionary Algorithms for Scalable Multiobjective Optimization,‖ arXiv:2206.11526v1, Jun. 2022.

Olga Moskvyak et al., ―Semi-supervised Keypoint Localization,‖ arXiv:2101.07988v1, Jan. 2021.

Yichuan Tang, ―Deep Learning using Linear Support Vector Machines,‖ arXiv:1306.0239v4, Feb. 2015.

Chaoyun Zhang et al., ―Deep Learning in Mobile and Wireless Networking: A Survey,‖ arXiv:1803.04311v1, June 2022.

Tiansheng Yao et al., ―Self-supervised Learning for Large-scale Item Recommendations,‖ arXiv:2007.12865v4, Feb. 2021.

Xin Xin et al., ―Self-Supervised Reinforcement Learning for Recommender Systems,‖ arXiv:2006.05779v2, Jun. 2020.

Yuxi Li, ―DEEP REINFORCEMENT LEARNING: AN OVERVIEW,‖ arXiv:1701.07274v6, Nov. 2018.

Hassam Sheikh, Mariano Phielipp, ―Maximizing Ensemble Diversity in Deep Reinforcement Learning,‖ International Conference on Learning Representations (ICLR), ICLR 2022.

Shivani Choudhary et al., ―Interpretation of Black Box NLP Models: A Survey,‖ arXiv:2203.17081v1, Mar 2022.

Hanaki, Jalaj. ―Differences between NLU and NLG - Python Natural Language Processing [Book].‖ O'Reilly Online Learning, Packt Publishing.

Shane Legg et al., ―Universal intelligence: A definition of machine intelligence,‖ arXiv:0712.3329v1, Dec 2007.

Nick Bostrom and Eliezer Yudkowsky, ―The Ethics of Artificial Intelligence,‖ Future of Humanity Institute and Machine Intelligence Research Institute, June 2014.

Gina, DiGravio. ―A New Explanation for Consciousness.‖ Neuroscience News, 3 Oct. 2022, https://neurosciencenews.com/consciousness-theory-21571/. Accessed 03 Oct. 2022.

The New York Times Company. (2019, March 27). Turing Award Won by 3 Pioneers in Artificial Intelligence. Nytimes. https://www.nytimes.com/2019/03/27/technology/turing-award-ai.html.

Ben Dickson. (2019, December 23). System 2 deep learning: The next step toward artificial general intelligence. TechTalks. Retrieved March 2, 2021, https://bdtechtalks.com/2019/12/23/yoshuabengio-neurips-2019-deep-learning/.

Synced. (2019, May 10). Google I/O 2019 | Geoffrey Hinton Says Machines Can Do Anything Humans Can. Medium. Retrieved March 2, 2021, https://medium.com/syncedreview/google-i-o2019-geoffrey-hinton-says-machines-can-do-anything-humans-can-460dff834ae2.

Ben Goertzel, ―The General Theory of General Intelligence: A Pragmatic Patternist Perspective,‖ arXiv:2103.15100v3, Apr 2021. [30] Robin Manhaeve et al., ―DeepProbLog: Neural Probabilistic Logic Programming,‖ arXiv:1805.10872v2, May 2018.

Ryan Riegel et al., ―Logical Neural Networks,‖ arXiv:2006.13155v1, Jun 2020.

Gary Marcus, ―The Next Decade in AI: Four Steps Towards Robust Artificial Intelligence,‖ arXiv:2002.06177v1, Feb 2020.

Zhen Zhang, ―MiCS: Near-linear Scaling for Training Gigantic Model on Public Cloud,‖ arXiv:2205.00119v4, May 2022.

Wei_Li. ―Software AI Accelerators: AI Performance Boost for Free.‖ CodeProject - For Those Who Code, 29 Apr. 2022, https://www.codeproject.com/Articles/5330215/Software-AI-Accelerators-AIPerformance-Boost-for. Accessed 18 Sept. 2022.

Brian Lester et al., ―The Power of Scale for Parameter-Efficient Prompt Tuning,‖ arXiv:2104.08691v2, Sep 2021.

Ofer, Fryman. ―AI Is Driving the Future of Hardware, Software, and Data.‖ Syte, SyteVisualAI, 13

Dec. 2017, https://www.syte.ai/blog/retail-innovation/ai-hardware-software-data/. Accessed 18 Sept. 2022.

NVIDIA CORPORATION. (2022, September 20). GTC Sept 2022 Keynote with NVIDIA CEO Jensen Huang [Video]. YouTube. https://www.youtube.com/watch?v=PWcNlRI00jo.

Edouard d'Archimbaud co-founder and CTO of Kili Technology. (2022, June 28). Kili Technology & Jellysmack at the WAICF 2022 [Video]. YouTube. https://www.youtube.com/watch?v=92L5A9XhgYQ.

MIT Technology Review Insights. (2022, June 7). Building a better society with better AI. MIT Technology Review. https://www.technologyreview.com/2022/06/07/1053031/building-a-bettersociety-with-better-ai/amp/.

Dennis Nealon – Harvard. (2022, September 3). How the Brain Processes Sensory Information From Internal Organs. Neuroscience News. Retrieved September 12, 2022, from https://neurosciencenews.com/sensory-processing-internal-organs-21356/.

Visible Body. (n.d.). Sight, Sound, Smell, Taste, and Touch: How the Human Body Receives Sensory Information. VisibleBody. Retrieved September 16, 2022, from https://www.visiblebody.com/learn/nervous/five-senses.

Lalor, John P. et al. ―Benchmarking Intersectional Biases in NLP.‖ NAACL (2022).

B, Senthil Kumar et al. ―An Overview of Fairness in Data – Illuminating the Bias in Data Pipeline.‖ LTEDI (2021).

Zheng, Hui et al. ―A mathematical model for intimacy-based security protection in social network without violation of privacy.‖ Int. J. High Perform. Comput. Netw. 15 (2019): 121-132.

Ziefle, Martina et al. ―Medical Technology in Smart Homes: Exploring the User's Perspective on Privacy, Intimacy and Trust.‖ 2011 IEEE 35th Annual Computer Software and Applications Conference Workshops (2011): 410-415.

Toscano, Manuel. ―On the Concept of Privacy: the Relation between Privacy and Intimacy.‖ Isegoria (2017): 533-552.