 doi:10.20944/preprints202001.0163.v1

Neurosymbolic Knowledge Representation for Explainable and Trustworthy AI

Paola Di Maio

Neuromorphic Systems Research Institute, Doulan, 95941 Taitung, Taiwan

Correspondence: paola.dimaio@gmail.com

Abstract: AI research and implementations are growing, and so are the risks associated with AI (Artificial Intelligence) developments, especially when it's difficult to understand exactly what they do and how they work, both at a localized level, and at deployment, in particular when distributed and on a large scale. Governments are pouring massive funding to promote AI research and education, yet research results and claims, as well as the effectiveness of educational programmes, can be difficult to evaluate given the limited reproducibility of computations based on ML (machine learning) and poor explainability, which in turn limits the accountability of the systems and can cause cascading systemic problems and challenges including poor reproducibility, reliability, and overall lack of trustworthiness. This paper addresses some of the issues in Knowledge Representation for AI at system level, identifies a number of knowledge gaps and epistemological challenges as root causes of risks and challenges for AI, and proposes that neurosymbolic and hybrid KR approaches can serve as mechanisms to address some of the challenges. The paper concludes with a postulate and points to related and future research

Keywords: knowledge representation, artificial intelligence, system level, symbolic, subsymbolic, neuro-symbolism, machine learning, explainability, trustworthiness

1. Introduction

AI is becoming increasingly integrated with software at all levels, to the point that users may never even know, because the architectural layers of most information systems in use, from banking to library systems to public services, from corporate enterprise software and applications to web searches to public services, are largely powered by hidden layers of knowledge, with logical inferences not explicitly represented and therefore not openly scrutinizable. In most cases, users do not want to be exposed to the complexities of many layered computational operations, and may want to avoid having to look into complex hidden logic generating the outcomes of AI systems. Yet these directly or indirectly influence the quality of the information, the way opinions based on such information are formed, down to the selection of practical decisions such as the choice of travel routes and overall consumer and lifestyle choices deeply impacting everyday life and that may become piloted by AI algorithms, with or without user knowledge and/or consent. Without at least some level of explainability (Figure 1) [1], and without being able to know and understand even in part the computations behind a result, it is not possible to evaluate how a system's outputs are constrained, if they are optimal or even just correct, therefore making it hard to evaluate the significance and the consequences of many AI technology mediated systems, and the limitations and impact that these mediations imply. When underlying AI determines the outcomes, functions and results of systems used in everyday lives, it becomes important to have at least some visibility or the workings leading to such outcomes, and knowledge representation techniques can facilitate that. In addition, given that

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 January 2020 doi:10.20944/preprints202001.0163.v1

much 'hype' surrounds real research and science, even within scholarly circles, including some peer reviewed literature, with leading academics and scholars not always prepared or capable to answer questions nor provide explanations of their machine learning work explicit KR is necessary to evaluate research results. This is where rigorous inquiry and explicit knowledge representation can help.

1.1. Motivation and Goal

Thanks to the increased availability of computational power, software and educational and learning opportunities, current generations of computer scientists are delivering novel capabilities in Artificial Intelligence (AI) with promising results, especially using machine learning techniques, in particular various types of Neural Networks (NN)[2]. Given the vast potential and expanding fields of application, there is growing interest not only from academia, with demand for courses in Machine Learning exceeding capacity in many universities worldwide, but also from industry, from the general public, investors and the public sector. The main goal of this paper is to identify some limitations and possible distortions in the way the KR body of knowledge is currently represented and taught, based on the critical analysis of leading AI scholarly references, and to highlight the relevance of explicit KR and its applicability to all types of AI, symbolic, non symbolic, subsymbolic particularly in the context of systemic challenges and threats, and to make the case for better referencing, understanding and for inclusion in teaching curricula of hybrid KR forms, to enable explainability and all its associated desirable qualities - transparency, auditability, reliability, reproducibility - towards the goal of making AI trustworthy [3], irrespective of whether the choice of system design and representation are symbolic or subsymbolic, and to encourage the production of explicit, understandable and shareable KR even when the AI architecture includes machine learning techniques.

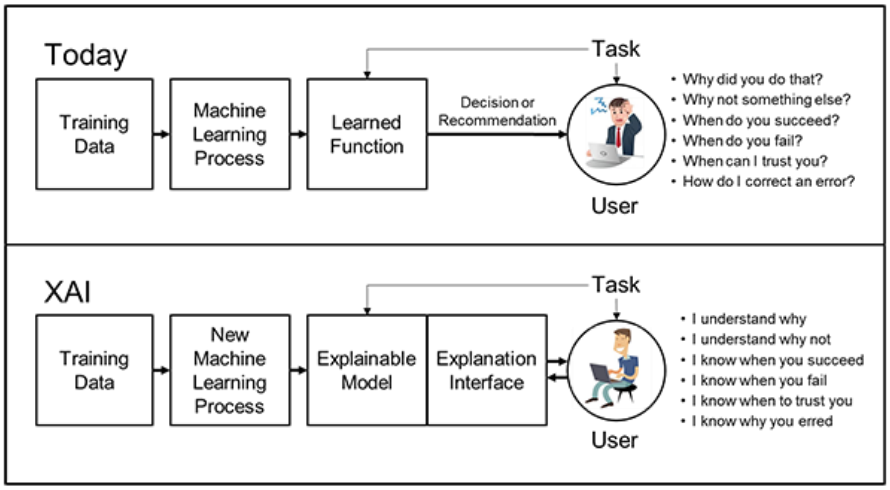

Figure 1 Explainable **AI** [1]

The contribution provides a critical perspective in AI research and teaching, identifies knowledge gaps and current challenges in the domain, and provides guidance on quality indicators for KR It culminates with a postulate and pointers to related work.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 January 2020 doi:10.20944/preprints202001.0163.v1

1.2. Method

This research is the result of domain and knowledge gap analysis, gathered through comparative literature review and direct observations and it is based on a Mixed Method Grounded Theory approach [4]. Methods used in this work include critical review, thought experiment, argumentation and theory formulation.

1.3. What is Knowledge Representation?

This question has been asked more or less on an ongoing basis for decades, and answers inevitably vary from book to book. As technology continually advances, so the meaning of established concepts and theories benefits from being refreshed and re examined in the context of the state of the art. This is non trivial, because contemporary AI research and practice, due to an explosion of interest in the subject together with epistemological fragmentation, new scientific and technological advances, come with new terminology and novel concepts, making it difficult even for experts to identify and evaluate to what extent what is being said corresponds to what is being done. Knowledge Representation (KR) in a broad sense, is what the expression says: a field of study and practice concerned with the representation of knowledge. Considering the universality of 'knowledge' - as a term, as a concept, as a subject of study - KR can be intimately part of every sphere of human interest, from linguistics to logic, computation, cognitive science and epistemology and the professions - as such it can be viewed and understood according to different perspectives. Defining what constitutes knowledge has been always a challenge for logicians and philosophers and more recently in the age of IT, for information technologists. It can be considered an infinitely multifaceted domain, addressing many intractable and open ended questions. The codification and formalization of Knowledge (K) for the purpose of powering intelligent software systems evolved only in the last fifty years or so, relatively recently when compared to other disciplines. In relation to the development of intelligent and autonomous software systems. Many of the early foundations of KR for AI as we know it today were laid by Marvin Minsky [5] Seymour Papert [6] McCarthy [13,14] and their contemporaries. Over the last half century KR has become comparable to a science in itself (as in ontology development) [7] with specific techniques and tools to devise and handle knowledge structures in the form of symbols, natural languages, and images for the purpose of software systems development. In the early days of AI these techniques were particularly applied to expert systems and later to reasoners [8] to enable functional operations, and to provide computable designations for logical constructs, such as the naming of entities, axioms, rules that power computer operations. KR culminated in recent years with semantic web research - starting approximately in the 1990s and continuing up to our days - with the production of an overarching body of knowledge known as ontology engineering and data science, essentially overlapping and extending KR over a much larger and unstructured body of knowledge on the web.

1.3.1. Background

A detailed chronological account of the evolution of KR in science and philosophy typically includes Aristotle and Plato, as well as countless logicians and mathematicians [9]. In the field of computation, Knowledge representation (KR) consists of the techniques and formalisms to capture and encode systematically into computable logical structures, and they vary according to the problems and challenges that they tackle, and can be sometimes domain specific, that is, the knowledge that is represented can be quite specialised and detailed and relating in particular to a given field. KR in Computer Science evolved through different phases, reflecting respective

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 January 2020 doi:10.20944/preprints202001.0163.v1

paradigms and trends in engineering and systems automation of their time, and includes topics that vary greatly from general knowledge such as classical logic, to basic mathematical concepts including satisfiability and proof, and countless techniques such as description logic, existential and conceptual graphs, non monotonic reasoning, bayesian networks, answer sets, belief revision, temporal and spatial reasoning, situation and event calculus and non monotonic logic [10]. Long before the Greeks tough, KR was also central to the development and implementation of systems of thought in ancient eastern traditions, the known beginnings of which are attributed to the Nalanda School, where it was used as a mechanism to support 'correct thinking' [11]. Contemporary practices in intelligent systems engineering, especially in relation to large and complex systems, KR can serve as a measure of adherence to functional and structural integrity of a system, although this important aspect is being overlooked both in AI research and education, especially in Machine Learning. This paper seeks to address this shortfall. Many types of KR artefacts and considerations address diverse concerns [12]. The thinking behind the beginning of KR as a scientific discipline can be appreciated still today by reading the early AI publications, where questions about the role of symbols and the methods for their representation were addressed for the first time:

This representation is independent of the IBM 704 computer, or of any other electronic computer, and it now seems expedient to expound the system by starting with the class of expressions called S-expressions and the functions called S-functions….where (S stands for symbolic)

……

Second, it is convenient to allow English words and phrases to stand for atomic entities for mnemonic reasons. The symbols are atomic in the sense that any substructure they may have as sequences of characters is ignored. We assume only that different symbols can be distinguished. [11,12] **.....**

Although AI KR so far has produced an immense body of scholarly knowledge, its essence in relation to AI and Knowledge Based Systems (KBS) can be boiled down to a handful of techniques, briefly summarized later in this paper. Key milestones upon which the field consolidated include a global survey of researchers working in early AI implementations [15] Further elaborations and developments delivered increased levels of sophistication, [16] as well as the formal identification of a new architectural layer in computation called the Knowledge Level' [17], which somewhat marked a transition from computation based on data, to knowledge based systems (KBS) using natural language. With the formulation of an explicit Knowledge Level, intelligent architectures could develop new kinds of 'artificial intelligence', as they were no longer limited to handling only numerical operators, and programmes could be written to handle natural language expressions, and by so doing computational and logical operations could be performed, in theory, by referencing entire libraries of facts, information from unstructured language and include various types of dialogues and discourse, the so called 'knowledge bases'. Even today in the age of machine learning, web searches as well as the majority of information retrieval functionalities are carried out using unstructured or semi-structured natural language expressions over datasets which can be open like the web, or closed like library databases. Thanks to the identification, modelling and implementation of the knowledge level new capabilities came into existence for intelligent systems automation and eventually for robotics - but only provided that facts could be represented meaningfully (in a way that they can be identified and interpreted) and correctly (in the way that the facts correspond to truth and can be communicated unambiguously) and that reasoning (inferences) conform to valid logic. The quest for validity in logic is ongoing, and knowledge representation lies at its heart. Hence the importance of the subject to the future of computation, and perhaps, even to the future of humanity itself. The first

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 January 2020 doi:10.20944/preprints202001.0163.v1

generation of intelligent systems using natural language called éxpert systems, consisted of a knowledge base (facts) representing expertise, and an inference mechanism (logical process to navigate the facts) as well as the ability to process an outcome of a query, typically an answer to a logical question, given the facts and the inference. It is not true what is often said that expert systems developed in the seventies ''failed'' because they did not work correctly or they were too expensive to maintain. More likely, their mass production commercialization and use were set aside, because by codifying expertise and standardizing it, making it widely available could impact the structure of society and the economy which are largely constructed on the segregation and the manipulation of knowledge: modern economic systems profit from restricting and channelling access to knowledge. To be fair to experts, expertise is very expensive and takes decades of study practice and dedication to develop. Systems that can capture expertise and give it all away were never looked upon favourably by experts either, who tend to guard what they know with their own lives. Eventually Decision Support Systems (DSS) gained acceptance, and are currently used by experts worldwide in every field to help them navigate complex and large knowledge sets to reach valid conclusions. Today experts in every field, especially in fields where the knowledge required to make informed decisions is vast and complex, such as in the medical to legal professions, rely on DSS - and ADS (automated decision support) systems to help reach decisions, with different degrees of success and reliability. A recent effort in the US Government to review the ADS implementations in public administration [18] has brought to light critical weaknesses of ADS approaches [19]. Much more can be said about what is KR, but due to resource constraints, this paper remands the readers to the extensive references for a more complete and detailed accounts of how the discipline evolved.

1.3.2. Explicit Definitions

The first level of explicit, shared formalization for knowledge and concept is their formal definition.

Explicit intended as 'spelled out', preferably in written form, as opposed to tacit or implicit

Below some definitions of KR excerpted from key references:

- *"concerned with how knowledge can be represented symbolically and manipulated in an automated way by reasoning programs"* [15].
- *- the field of study within AI concerned with using formal symbols to represent a collection of propositions believed by some putative agent*. [85]
- *Knowledge Representation can be defined as the application of logic and ontology to the task of constructing computable models of some domain* [9]

- *In artificial intelligence, knowledge representation is the study of how the beliefs, intentions, and value judgments of an intelligent agent can be expressed in a transparent, symbolic notation suitable for automated reasoning. From a purely computational point of view, the major objectives to be achieved are breadth of scope, expressivity, precision, support of efficient inference, learnability, robustness, and ease of construction. ….Many different general architectures have been used for knowledge representation, including first-order logic, other formal logics, semantic networks, and frame-based systems. - [20]*

As it can be observed in the definitions in use, KR in literature is mostly defined in terms of symbolic AI, where the AI conforms to a paradigm of symbolic computation. KR is also chara*cterized according to 5 roles [20]:*

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 January 2020 doi:10.20944/preprints202001.0163.v1

- *A knowledge representation (KR) is most fundamentally a surrogate, a substitute for the entity itself, used to determine consequences by thinking rather than acting, i.e., by reasoning about the world rather than taking action in it.*
- *Is a set of ontological commitments, i.e., an answer to the question: In what terms should I think about the world?*
- *Is a fragmentary theory of intelligent reasoning, expressed in terms of three components: (i) the representation's fundamental conception of intelligent reasoning; (ii) the set of inferences the representation sanctions; and (iii) the set of inferences it recommends.*
- *Is a medium for pragmatically efficient computation, i.e., the computational environment in which thinking is accomplished. One contribution to this pragmatic efficiency is supplied by the guidance a representation provides for organizing information so as to facilitate making the recommended inferences.*
- *Is a medium of human expression, i.e., a language to describe*

1.3.3. KR Models and Artefacts, methods and tools

Symbolic KR consist of explicit (logically and consistently written, structured, codified, labelled) natural language structures that can serve as descriptions and annotation of facts (the knowledge base) and their logical relations, which sometimes are non linear, sometimes are polivalent, representing more than one thing sometimes may even be hidden, or not formally proven. So to some extent KR, unless formally constrained and heavily axiomatized, as in the representation of exact sciences, can rely on some degree of approximation, it is acceptable to use heuristics and to make some assumptions.Typical knowledge engineering approaches leverage systems modelling methods (as in model driven engineering) as in KADS [21] and MIKE [80] and computational logic. Symbolic KR artefacts are typically based on natural language and, as per the textbooks, can include: Frames, Production Rules, Semantic nets, Bayesian Networks [10]. Non symbolic can refer to knowledge stored as visual imagery and sub symbolic is generally referred to as relating to knowledge of the computational operation, based on mathematical functions. In essence, KR consists of determining what knowledge is required by a system or a process to operate, encoding this knowledge and making it available to the system as a model of the world for the purpose of operating a system. Such explicit representation of processes and artefacts has been largely absent from much of ML driven AI referred to as the leading edge, which focuses on quantitative aspects of computation and on performance. This paper identifies this issue as an epistemological shortfall and seeks to address it. More recently, it has been suggested that ANNs and statistical analysis as well as Knowledge Graphs (KGs) can also be used as mechanisms for Knowledge Representation [22, 23] yet these instruments fail to meet some of the essential quality factors for KR, discussed below, and should not be considered as adequate KR, without at a minimum appropriately identifying their limitations in this respect. In sum, KR can be many things to different people, from maths and logic to representing all types of computation, yet ultimately it is explicit, shared, symbolic AI that makes these systems explainable to humans using natural language. The two ends of the KR spectrum are generally referred to as symbolic vs subsymbolic (IMAGE 2) but there is quite a lot in between - for example Neurosymbolic KR, as mentioned below, that is generally overlooked in education and research, Neurosymbolism is largely absent in AI teaching books, and it was absent from the recent Turing award speech [56]

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 January 2020 doi:10.20944/preprints202001.0163.v1

	Symbolic Approaches	Subsymbolic Approaches
Methods	(Mostly) logical and/or algebraic	(Mostly) analytic
Strengths	Productivity, Recursion Principle, Compositionality	Robustness, Learning Ability, Parsimony, Adaptivity
Weaknesses	Consistency Constraints, Lower Cognitive Abilities	Opaqueness, Higher Cognitive Abilities
Applications	Reasoning, Problem Solving, Planning etc.	Learning, Motor Control, Vision etc.
Relation to CogSci	Not Biologically Inspired	Biologically Inspired
Other Features	Crisp	Fuzzy, Continuous

Figure 2 Symboic vs Subsymbolic [24]

A KR movement referred to as the Anti-representationalists [25] advocates that KR is really not necessary to intelligence, while the abstractist/non conceptualist movement proposes that deep in our brains, there exist non conceptual representation [26]. Both schools follow very interesting, scientifically exciting lines of inquiry at the intersection of the theory of mind, biology and neuroscience, however this paper is limited to addressing functional and structural analysis of KR from a systems engineering perspective. Research in Neuroscience shows that conceptual and linguistic knowledge required for any kind of intelligent reasoning is represented in various regions of the brain [27].

Neurosymbolic KR (NSKR) Approaches

Despite a wealth of research and resources produced in the seventies and eighties, and despite being useful and successful, Neurosymbolic approaches have been notably absent from AI text books and educational programmes and have been somewhat marginalized in research by limiting or omitting citations and mentions of early work. An important missing link between symbolic and non symbolic AI, Neurosymbolism denotes autonomous intelligent architectures that adopt both symbolic and subsymbolic methods, in a variety of configurations and following different strategies (Figure 3) [28,29], where r*easoning is performed either in a discrete symbolic space or in a continuous vector space.*

Early examples of NSKR explored neural networks as computational mechanisms to power expert systems [30] and have evolved to become particularly useful in tackling çomplex problems, also called wicked problems, or problems that are considered intractable using one or the other approaches[30,31]. The main strategies adopted in neurosymbolism tend to range between unification vs integration (another metalevel discussion in systems philosophy) across a wide range of problem and solution spaces.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 January 2020 doi:10.20944/preprints202001.0163.v1

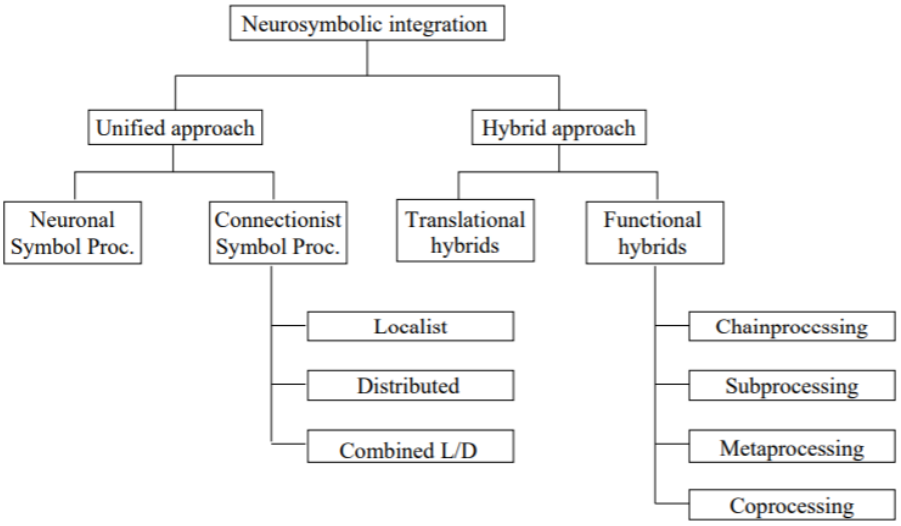

Figure 3 Neurosymbolic Integration[29][1]

From neurosymbolism comes an important construct for KR: **'representational integration',** [31] which can emphasise either the connectivist or symbolic representation or both. Novel types of KR artefacts also emerge, such as **neurules**, defined as 'a type of hybrid rules' which integrate symbolic rules with neurocomputing. [32] In addressing some of the gaps in AI research and education and scholarly publications, Neurosymbolic Integration should be added to the essential body of knowleDge for AI KR

1.3.4. Multiple Perspectives

Of all the fields of knowledge, computer science and informatics represent the peak of cognitive and intellectual achievements for humanity, where science technology and engineering are converging to produce the most innovative advances. Even physical and material sciences are advancing exponentially thanks to the availability of large volumes of data and computational resources Yet, there is a scarcity of depth and qualitative appreciation of the information and knowledge resources available, and sometimes vasts amount of data are mere flat dumps, devoided of semantic structures necessary to facilitate their interpretability. A surprising shallowness in technology applications can be encountered, where intelligence and wisdom remain scarce or biased. This paper contributes two additional perspectives to the already messy epistemological picture of the AI KR landscape: 1) A system view and the systemic perspective 2) their intersection with complexity sciences, briefly discussed below

The Systemic perspective

Certain AI functions, in particular quantitative, computational functions especially in relation to performance, may indeed be carried out solely using standalone Machine Learning algorithms, However in contemporary AI, given the endless plurality of techniques and fields of applications is important to consider the system level, where emergent behaviours are likely to occur. The implication of adopting a system level view of AI, means that the adequacy of the performance of an algorithm is evaluated not only using a closed world, lab constrained environment, but also, and especially, when deployed in real world situations. AI research should be evaluated also in relation

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 January 2020 doi:10.20944/preprints202001.0163.v1

to the socio technical contexts they apply to, rather than solely in isolation. In particular the lack of adherence and poor implementation of adequate KR in AI can lead to systemic deviation [34]

Complexity

Complexity sciences tackle the interrelatedness of the dynamics of systems with different types and scale: from the entire cosmos to sub atomic particles, from engineered technologies, to the human brain, from the animal to the mineral kingdoms, from the macro to the micro, Complexity science identifies and can help to explain systemic similarities and differences, from defining universal laws to randomness and chaos. Complexity is becoming an important factor in AI, because of the quantity and quality of variables, functions, behaviours, methods are becoming virtually limitless. By removing epistemological boundaries and computational limitations disciplines collide and can cause exponential knowledge explosions. While certain quantitative aspects of handling exponentiality are now possible thanks to the increased power of computation, there are still cognitive limits for human developers and users, Complexity is not new, but it is becoming more obvious and increasingly impacting the state of the art in different fields.

Knowledge Engineers and Ontologists - who have the cognitive capacity and technical tools to address large complex knowledge in different domains and data sets, may require some standard KR approaches to be able to contribute to the organization of knowledge domains like Neuroscience, in all their facets (Neuroinformatics, Cognitive Neuroscience, Neuroengineerig etc). In addition to developing increasingly robust and flexible methods for knowledge modelling and computation of complex sciences, to stay sane useful and viable the field of AI needs to encourage the ability of humans with different backgrounds and skillsets to grasp and organise vast amounts of knowledge. AI KR is becoming necessary to the expansion of cognition to large scale, so to speak. Yet, despite recent waves of theoretical interdisciplinarity, in practice science still suffers from the limitations of epistemological and disciplinary segregation. The diagram below (Figure 4) provides an overview of different KR perspectives in relation to different knowledge domains

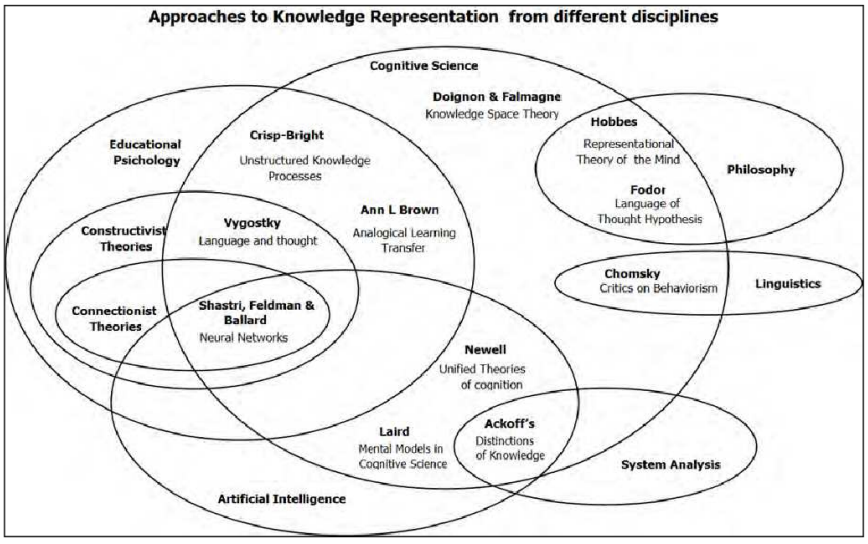

Figure 4 : Muplitple perspectives in KR. [36]

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 January 2020 doi:10.20944/preprints202001.0163.v1

2. KR Challenges

New capabilities provided by AI technology come with different kinds of risks [35]**.** Especially **c**losely related to KR is the risk of **'misrepresentation'** that is, that something (a fact, a function or a process, the result of some computation, in relation to a person or an event or a notion) are represented to be something that they are not. Knowledge Misrepresentations is very widespread, and has different causes, from deliberate misrepresentation, intended to mislead, to honest misrepresentation [36], as the result of some poor logical process or wrong fact, so called error. Both types of misrepresenations may have legal consequences under common law, if they are responsible for causing harm or damage, although in the case of AI, it is not quite clear who carries the liability. Misrepresentation is all too common, used as a tool by media and in politics to leverage conditional responses, It is often hidden in the complexities of the real world and its many languages and cognitive limitations of humans to handle and resolve conflicting or unclear information, and is now becoming transcribed into AI systems. Misrepresentation can cause. trigger and reinforce different kinds of AI Bias [37,38] Another example of new risk from AI is referred to as DEEP FAKES (to represent something which is not so using AI technology) is the direct result of intentional misrepresentation

Deepfake is a term used to describe manipulation of facts and information using AI-based technology to misrepresent a fact, a person but altering images and videos [39]

More general areas of concern and risks associated with AI are:

1) the lack of explainability, accountability, verifiability transparency and reproducibility of AI, which is particularly true in Machine Learning powered AI

2) the lack of understanding among students and faculty members in Computer Science Departments of what really is KR and its role in AI, and their inability to explain the logic and functioning behind their algorithms and justify their results -

3) AI related topics being completely misrepresented and misleading, using the term 'Knowledge Representation' increasingly in ways that do not fit the definition and role as formally defined, and by doing so contribute to promote superficiality disinformation bias and all the AI risks as mentioned so far

KR is generally poorly understood, except for a niche community of researchers, nor it is adequately adopted neither taught in mainstream education and research, to the point of being misrepresented itself. In conclusion, although non- conceptual non- representational AI can be valid constructs with respective acceptable and scientific exciting lines of inquiry however from a systems reliability and accountability point of view, and to evaluate many of the claims made by researchers in relation to the performance of their algorithms, the adoption of shared and explicit representation (referred to as symbolic KR) is useful, possibly even necessary to satisfy essential criteria of representational adequacy.

2.2. Machine Learning

Machine Learning (ML) is driving AI research, a trend typically associated with the successes of the adoption of NN in statistical analysis [40]]. Yet despite its pivotal role, and despite the proven importance of KR to support and make explicit intelligent reasoning, there has not been much

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 January 2020 doi:10.20944/preprints202001.0163.v1

adoption of knowledge analysis and representation techniques in contemporary ML, at least not in the sense KR is used in symbolic AI

In addition, KR has become diluted and even distorted, its meaning and role gradually shifting to imply that even artificial neural networks (ANN) and statistical analysis including various types of code are 'new types of KR [22,41]. In ML NN (neural networks) are considered by some authors as a form of representation, and KR is increasingly used in undefined ways that somewhat depart from the meaning and role of KR as formally known in classical AI. [23]

Intelligence without Representation?

One of the arguments supporting the lack of explicit KR in ML leverages the example that learning is inherently experiential, as in the case of young children who constantly learn without having language to support formal concepts, and thus if a child can learn without KR, so can machine' [42] The idea of Ïntelligence without Representation has not gained widespread acceptance "bottom-up research on mobile robots, although valuable, is neither necessary nor sufficient as a foundation for core AI research. [43,44] Connectionist and Antirepresentationist and non conceptualist arguments aside, there is generally agreement that at a minimum from a systems engineering point of view, KR is necessary to facilitate affordable development, maintenance, the ability to communicate among developers and users. Furthermore it is necessary to achieve explainability, learnability reproducibility verifiability, compliance and other desirable qualities, the lack of which are increasing the risks and challenges for AI adoption. Ultimately 'trustworthy' and ethical AI can only be achieved through shared explicit KR *"There are deep theoretical disputes within AI about how best to model intelligence. Classical (symbolic) AI programs consist of formal rules for manipulating formal symbols; these are carried out sequentially, one after the other. Connectionist systems, also called neural networks, perform many simple processes in parallel (simultaneously); most work in a way described not by lists of rules, but by differential equations. Hybrid systems combine aspects of classical and connectionist AI. More recent approaches seek to construct adaptive autonomous agents, whose behaviour is self-directed rather than imposed from outside and which adjust to environmental conditions. Situated robotics builds robots that react directly to environmental cues, instead of following complex internal plans as classical robots do. The programs, neural networks and robots of evolutionary AI are produced not by detailed human design, but by automatic evolution (variation and selection). [45]*

In children, experiential learning is the primary process of knowledge formation and development directly related to the increase of conceptual cognition, reasoning and language abilities. Research of how knowledge is represented in the brain has only just begun, because relevant neuro imaging and novel sensing techniques and corresponding methods of inquiry into brain level knowledge representation are just being developed - and there is clear evidence to show that knowledge maps exist in some regions of the human brain. [46] Due to epistemological gaps and disciplinary rigidity still influencing much of science, although much of the data resulting from n research in neuroscience may even be publicly available as *open data*, yet it is cognitively impenetrable to anyone but the neuroscientists and biologists themselves, due to the lack of adequate shared and explicit knowledge representation that can be understood by experts in other disciplines. There is mounting scientific evidence that knowledge (facts, images) is represented in corresponding regions the brain: through innovative brain imaging techniques, it is possible to locate signals in the brain where knowledge is stored and processed, yet no actual methodological framework exists yet to explain what type of knowledge (declarative, concepts vs procedural, decisions for example) corresponds to what type of signal. Although research acknowledges that *"Critical to understanding brain and behavior is clarifying the principles of human knowledge representation. To describe selective*

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 January 2020 doi:10.20944/preprints202001.0163.v1

responses to sensory, motor, perceptual, linguistic, or abstract cognitive functions with the dominant model for the representation of semantic knowledge in the human cortex proposes the involvement of multiple spatially localized, functionally specialized but interconnected brain regions" [46]

New Results demonstrate that there are (at least) two forms of knowledge representations in the human brain: sensory-derived and cognitively-derived knowledge, supported by different brain systems. [47]

This is likely to have implications for both KR and Neuroscience (see future work)

2.3. Novel Concepts vs New Names for Old Concepts

In addition to the many and diverse complexities and challenges for AI KR outlined so far, resulting from the data explosion plus interdisciplinary collision, novel concepts and terms describing new research techniques, paradigms, trends in computer science but also in the relevant neighbouring fields such as cognitive psychology, are being introduced everyday. Sometimes. the concepts may indeed be new -pointing to novel discoveries in terms of capabilities or techniques that did not exist before hence the need for new terms to be coined - but sometimes they point to concepts that are just being rehashed, using new names for known constructs yet without referencing the latter, increasing the difficulty to evaluate the research, claims and results being presented. As novel technical capabilities come into existence, new terms are legitimately coined, in some cases so new that no formal definition may yet exist for them. But some of the so called 'new terms' and new capabilities may actually not be new at all, but derived from other domains, or the authors may simply not be familiar with prior art and literature, or simply choose to escape the conventions in use. Two examples follow, Disentanglement and Vector

Disentanglement

A recent line of AI research argues that "representations that are disentangled are an important step towards a better representation learning where disentanglement mans that they should contain all the information present in x in a compact and interpretable structure while being independent from the task at hand" [48]. In software engineering there is a long tradition of disentanglement which can be called either ***decoupling***, [49] or logical ***independence [50]***

They are both well established and considered essential good practices in system design. While affording flexibility of data manipulation in experiments for reasoning and computation [51], in logical functional viable systems, disentangled representations can lead to functional incoherence, and result in various types of malfunctions, or dysfunction. Therefore the role of disentanglement should be disambiguated, well understood and constrained to avoid undesirable systemic consequences which can lead to systemic deviation [33]

Vector

The term vector originates from geometry indicating a spatial relation in a three dimensional place Yet in contemporary ML parlance everything can be a vector and the term is often used to indicate ány relation, in n dimensions [52]

The importance of Using appropriate naming conventions, although generally more appreciated among software engineers than pure computer scientists using mathematical notations, should be practiced in research and added to ML good practices and teaching curricula

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 January 2020 doi:10.20944/preprints202001.0163.v1

2.4. AI and Fallibility

It is important to keep AI potential capabilities 'in context' to avoid being carried away by the hype and speculatory waves, which are aimed to deceive. (At least some) humans are considered to have the most advanced cognitive abilities among all the beings in natural earthly realms, as through the evolution of natural language, they are able not only to formulate and adhere to logic, the structure of intelligent reasoning, but also to express and communicate the product of their intelligent creative processes in using language and various forms of expression, which can include extremely fine crafts. arts and music. Yet despite great intelligence, humans remain highly fallible [53] In using selected aspects of human intelligence as a model for cognitive machines, researchers and developers should at a minimum be reminded that human intelligence comes with limitations: from being error prone, to fluctuate according to physical and psychological conditions, or even be influenced by atmospheric conditions and the weather, and suffering from various kinds of lapsi and biases. And that extrapolating selected features from the human intelligence as a whole, is likely to result in unexpected behaviours. Despite the huge availability of computational resources, skills and education - the world nowadays has the largest numbers of schooled population and Phds - humankind is not yet able to resolve, even partially, fundamental problems at the root of many existential risks such as resource allocation, poverty, pollution, climate change, crime, terrorism, war, famine, food and water shortages, injustice, abuse, disinformation, corruption, maladministration which plague human civilization as a whole, in different countries and society strata to different degrees. In using human intelligence as a model of AI, all the shortfalls of human intelligence are also reproduced, possibly systematized, automated and deployed to mass scale. Each theory of human intelligence comes with corresponding fallacies and unanswered questions. Research demonstrates that errors cannot be avoided totally in complex settings[54]. We better figure out how to conceive risks and limit losses if there is any such thing as collective responsibility of where AI is going. This article would like to make a contribution from the KR point of view to safe AI. It is difficult, if not unimaginable to predict in which direction AI is heading, and with what consequences for humankind, since certain developments take a life of their own, for better or worse. But the cat is out of the bag. Endless novel technical capabilities are blurring the line between reality and fiction, the Deepfake cases being a prime example [see earlier reference]. There are many examples of distortions caused simply by feeding educated intelligent people and advanced systems with incorrect information, either as the result of genuine error or as deliberate strategic manipulation with explicit intent to deceive- where deviation and failure are designed into systems for some malicious reason. Today we call upon KR to help to put that right.

The case of Tweetie

In explaining first order logic and comparing the reliability of different types of reasoning (deduction vs induction and abduction) various versions of a notable textbook example [55] have been used in lectures and papers, for decades

Given: Every bird flies. Tweety is a bird. Infer: Tweety flies.

But actually, the fact is that not every bird flies. Without fact checking, even a reliable and formally correct deductive reasoning can lead to a wrong conclusion: The assertion *every bird flies* is incorrect (not all birds fly, and not every animal that flies is a bird). The repercussions of incorrect. or even partially correct representation can be endless, and scaled up, could lead to fatalities. Without fact checking, even the most impeccable logical reasoning, can lead to incorrect conclusions, and even

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 January 2020 doi:10.20944/preprints202001.0163.v1

with fact checking, logical reasoning can lead to wrong conclusions, because absolute çorrectness'can be difficult, maybe even impossible to represent economically.

The same however may also be true for mathematics. Not every mathematically correct result guarantees a correct interpretation, function or behaviour. Now consider highly reliable, formally constructed AI powered logical reasoners drawing conclusions on information which is not factual, or simply incomplete (not all knowledge facts may have been known or codified at the time the programme was written or implemented) or partially true or imprecise (as natural language is) or fuzzy. Much of the knowledge used in AI and ML is not fully verifiable, and exact theoretical sciences may not alway always result exact when applied in the complex and fluid real world settings. The prestigious ACM Turing Award provides an excellent example of the state of the art and its limitations: one of the 'takeways' in the recent Turing Award Lecture given by Lecun and Hinton [56], struck a dissonant chord in the KR community, when Lecunn declared that symbolic AI is gone (*final nail in the coffin for symbolic AI* were the words used in the lecture). Later in the talk, the same speaker said that "what is needed now is for neural nets to begin to be able to explain reasoning".[48] There can be limited 'knowability' (except for experiential knowledge perhaps) nor 'explainability' without explicit KR, and KR as we know it, is mostly applicable to symbolic AI. Thus the ACM Turing Award speakers themselves, although leading AI researchers, make logically contradictory statements, somewhat being a very loud and clear testament of their own bias. The lack of logical coherence over large data can gradually lead to huge semantic shifts and lead to misrepresentation. Eventually, if adopted and massively deployed misrepresentation leads to failure. It is going to be difficult to understand and explain reasoning without symbolic KR, because human languages rely on symbols for communication [57]. At least some of the risks associated with developing systems using advanced computation can be mitigated by ensuring that the logic behind their operation can reduce the cognitive load and cost of making the logic visible interpretable and understandable, and KR methods and tools serve precisely that purpose.

3. Quality Criteria for KR

Given the vast universe of discourse which constitutes KR in AI, the main examples of quality criteria applicable to symbolic AI are briefly provided to serve as reference as to what properties and functions should a KR display to fulfil its purpose: to represent knowledge adequately. To address adequacy in KR is not trivial, and the guidelines provided should only be taken as a starting point. In the first instance, KR should do what the term says it does: to represent knowledge. The first and foremost risk of any representational device is to either not represent knowledge (omission) or to mis-represent knowledge (distortion). Both types of misrepresentation can occur very easily and cause malfunctions and result in the incorrect outputs, Engineers who may want to maliciously leverage systemic deviation [33] know how to exploit misrepresentation, and probably do not want the question of knowledge misrepresentation resolved. How to evaluate the adequacy of knowledge representation in AI? Principles and criteria to evaluate KR, are shared with, and sometimes derived from neighbouring disciplines, such as for example Formal Logic, Information Science and Ontology Engineering. They are useful, and should be balanced by heuristic evaluations, such as 'fitness for purpose', a call that AI developers have to learn how to make. This section highlights some criteria and principles in use, intended as pointers to requirements that KR must satisfy in order to be considered **adequate**. Not all the adequacy and fitness for purpose 'principles and criteria for KR listed may be necessary, and the set provided may not fully exhaustive, but are included to guide the evolution of computable quality metrics for KR.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 January 2020 doi:10.20944/preprints202001.0163.v1

3.1. Principles

Knowledge Representation can be described by five fundamental roles, referred to as Knowledge Representation principles [58]:

- A surrogate: symbols are used to represent external things that cannot be stored in a computer, i.e., physical objects, events, and relationships. Symbols are surrogates for the external things. Symbols and links between them form a model of the external system that can be manipulated to simulate it or reason about it.
- A set of ontological commitments: Ontology is the study of existence. Thus, ontology determines the categories of things that exist or may exist in an application domain. Those categories set the ontological commitments of the application designer or knowledge engineer.
- Aa fragmentary theory of intelligent reasoning: to support reasoning about modelled things in a domain, a knowledge representation must describe their behaviour and interactions. The description constitutes a theory of the application domain. It can be stated, for instance, as explicit axioms or compiled into computable programs.
- A medium for efficient computation: besides representing knowledge, an Artificial Intelligence System must encode knowledge in a form that can be processed efficiently by the available computing equipment. Therefore, developments in computer hardware and programming theory have a great influence on knowledge representation.
- A medium for human expression: a good knowledge representation language should facilitate communication between the knowledge engineers who manage knowledge tools and the domain experts who understand the application domain. Domain experts should be able to read and verify the domain definitions and rules written by knowledge engineers

3.2. Levels

When applied in the computer domain, knowledge representations range from computer-oriented forms to conceptual ones nearer to those present in our internal world models. Five knowledge levels are identified in literature using this criterion [59]:

- Implementational: this is the more computer aware level. It includes data structures such as atoms, pointers, lists and other programming notations.
- Logical: symbolic logic is inside this level. Thus, symbolic logic propositions, predicates, variables, quantifiers and Boolean operations are included.
- Epistemological: a level for defining concept types with subtypes, inheritance, and structuring relations.
- Conceptual: the level of semantic relations, linguistic roles, objects and actions.
- Linguistic: the more computers distant level, it deals with arbitrary concepts, words and expressions of natural languages

Overall Adequacy

To be useful, KR needs to adhere to certain criteria and requirements, which at least in part are common to most types of information systems, knowledge based systems and ontologies, such as [60,61]. Several types of **adequacy** criteria for of KR are in use for example: **Terminological adequacy** the ability to form the appropriate kind of technical vocabulary and understand the dependencies among the terms; **assertional adequacy** involves the ability to form the kind of theory appropriate to

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 January 2020 doi:10.20944/preprints202001.0163.v1

the world knowledge of a system and understand the implications of the theory. *[62].* Furthermore, the following characterizations can be made:

1. Representational Adequacy – the ability to represent all the different kinds of knowledge that might be needed in that domain. [63]
2. Inferential Adequacy –the ability to manipulate the representational structures to derive new structures (corresponding to new knowledge) from existing structures.[64]
3. Inferential Efficiency – the ability to incorporate additional information into the knowledge structure which can be used to focus the attention of the inference mechanisms in the most promising directions. [65]
4. Acquisitional Efficiency– the ability to acquire new information easily. Ideally the agent should be able to control its own knowledge acquisition, but direct insertion of information by a 'knowledge engineer' would be acceptable. Finding a single system that optimises these for all possible domains is not always feasible [66], however researchers and developers are encouraged to develop and apply their own evaluation schemas for KR. Other more generic criteria that can be used to guide the development of adequate KR in ML (gathered from various sources) include:

A representation should be sufficiently complete and explicit should be able to satisfy a query it relates to

Should be Interpretable and Human and Machine readable –
Should make the important objects and relations in the system explicit and accessible – so that it is easy to see what is going on, and how the various components interact.
Should suppress irrelevant detail – so that rarely used details don't introduce unnecessary complications, but are still available when needed.
Should expose any natural constraints – so that it is easy to express how one object or relation influences another.
Should be transparent – so you can easily understand what is being said.its implementation needs to be concise and fast to process – so that information can be stored, retrieved and manipulated [67]

In addition, in relation to ontologies, which can be considered as a type of KR, the trade off between five design criteria has been identified which can also apply to most other forms of KR [68]

Clarity should effectively communicate the intended meaning of defined terms. Definitions should be objective, formal (if possible), and documented with natural language.

Coherence should sanction inferences that are consistent with the definitions.

Extensibility should be structured such that it can be extended and specialised monotonically, without needing to change itself.

Minimal encoding bias should be specified at an appropriate knowledge level, without depending on a particular encoding, for convenience of notation or implementation.

Minimal commitment should only enforce the minimal commitment necessary to support the intended knowledge sharing activities

Several KR evaluation frameworks exist, each analyzing different aspects of KR adequacy according to purpose and other factors, from which a basic set of quality criteria can be summarized at a minimum as:

Adequacy (Variety of Expressiveness, Modularity, Semantics, and Organization of Knowledge and other criteria);

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 January 2020 doi:10.20944/preprints202001.0163.v1

Inference Methods (Reasoning Strategies, Data, Control and Search Strategies);

Inference Requirements (Computational Efficiency, Transparency of line of control, Completeness, and Consistency)

Ability to use a priori knowledge, and update it with newly-acquired knowledge.

Dealing with incomplete and imperfect knowledge.

Correctness Evaluation - the KR should include the ability to estimate its representational correctness [69,70,71,72]

KR Checklist

Although it can be argued that some machine learning algorithms may not be replicable due to the inherent uncertainties of probabilistic causal computation, it should be possible to use a combination of different KR techniques to facilitate the explicit representation and replication, even partia, of artificial neural network computations [86].Therefore a sample KR checklist is synthesized from the literature,

KR be said to be adequate if it identifies and makes explicit:

Individuals/Components/Entities
Axioms/Laws/Constraints/Limitations
Processes/Functions
Inputs/Outputs/Outcomes
Type Of Reasoning/Inference
Relevant Variables
Behaviours
Structure/Patterns
Levels Of Predictability Of The Behaviours
(Probability, Randomness)
Influence Factors
Variables That May Influence Factors
Interactions
Type Of Notation/Encoding
Knowledge Level (Logical, Implementational, Epistemological, Conceptual. Linguistic)
Overall Scientific/Computational Paradigm
Complete/Sufficiently Describing The Function
Should Allow Manipulation Of The Representation For Testing/Evaluation Purposes
Human And Machine Readable

In sum, mathematical and computational notations alone may not adequately satisfy all the requirements for explicit KR. ANNs, statistical analysis and Knowledge Graphs even if valid and useful computational techniques, may not necessarily be adequate to represent the knowledge for the AI systems they power, as they fall short of sufficient criteria necessary for the qualitative aspects for the evaluation of intelligent reasoning processes and query outcomes, with some exceptions [73] including very recent advances which have started to implement the postulate in section 5 of this paper.

4. Symbolic Representation for Deep KR

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 January 2020 doi:10.20944/preprints202001.0163.v1

Symbolic, non symbolic sub symbolic, connectionist AI, are distinguished mainly through the different types of knowledge representation, denoting a broad range. It is interesting to note that many current AI solutions adopt multiple approaches, as discussed in the paragraph devoted to Neurosymbolism in this paper. The adoption of symbolic representation for non symbolic computing can be made on at least two grounds

1. Symbolic KR is still largely necessary for explaining and communicating algorithms
2. Non symbolic computation should be explainable in symbolic terms as a form of justification, providing epistemological evidence for the reasoning,

At the moment limited explicit KR is practiced in ML. Partly because the performace of neural networks relies on their unpredictability. Yet at least from the point of view of systems reliability, unpredictability is perceived as an important weakness in AI. It can be said that even the distinction of symbolic vs subsymbolic AI is just a choice of convention, as most intelligent reasoning, artificial or not, is complex and results from combining different approaches

4.1. Hybrid KR

To communicate, discuss and evaluate the effectiveness and the performance of automated learning functions, as for example in Machine Learning, at least some level of symbolic Knowledge Representation is necessary, or at least greatly beneficial, as it makes its understanding easier, clearer and less ambiguous. KR for example is necessary in systems design processes [74] and where diverse and interdisciplinary teams collaborate to deliver a solution. *Learning can be viewed as a particular type of problem-solving, and thus search and representation schemes used in 'non-learning' systems are also of relevance to machine learning* [75]. Furthermore, representational schemas used in a learning systems play a computational (as well as a semantic) role in determining how knowledge is used. Other examples of use of symbolic KR in machine learning advocate that rule induction systems use simple propositional-like logic representations, enabling search to be adequately constrained, and that learning systems can easily augment such rules to include measures of certainty or probability

In related research, the modelling of classes in knowledge schemas used in Neuroscience (cell, synapses etc, as in the case of NeuroML for example (Figure 5) are evaluated for their suitability and adequacy to represent ML algorithms.[76] By using a working convention that provides a structured conceptual schema for the neurons in the physical brain, using a model driven approach, it should be possible to develop equivalent conceptual schemas for ANNs (artificial neural networks). If machine learning adopts human learning models, then it should be possible to find some correspondence between the two. The discussion of the results of this evaluation are remanded to future work. Many interesting knowledge representation systems exist in neuroscience, and the majority are openly accessible (open source) however what they represent is mostly meaningful to computational neuroscientists and biologists. Anyone else, including ontologists and AI experts may not be able to process the meaning and structure of their content.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 January 2020 doi:10.20944/preprints202001.0163.v1

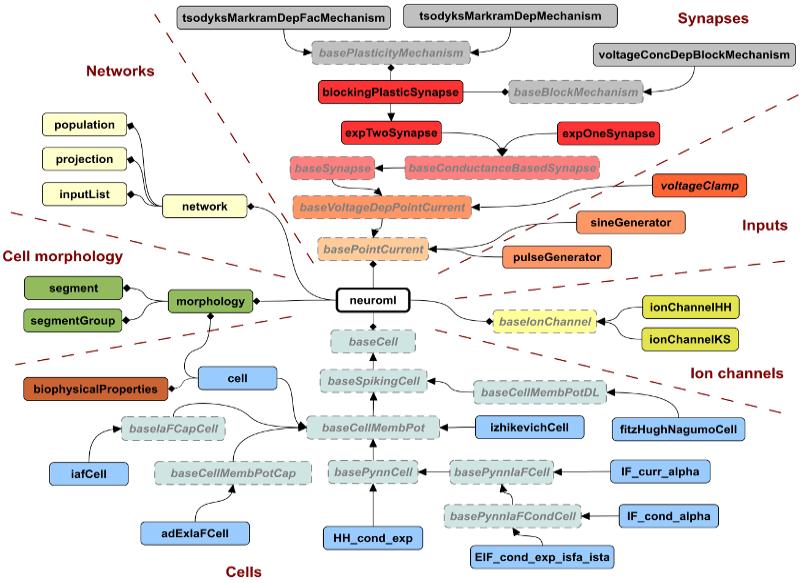

Figure 5. NeuromML top level schema [84]

4.1. Additional Factors Influencing Deep KR and Other Considerations

Contemporary science and technologies are developing exponentially and concurrently in many directions Several factors are contributing to creating a complex landscape where theory and praxis, science and technology are closely intertwined. In transdisciplinary literature, the issue of knowledge representation and communication is recurrent due to a perceived need to communicate increasingly complex and dynamic insights, [78]:

Knowledge representation and mediation become issues when different sources and types of knowledge have to be "integrated in a framework of analysis" or have to be articulated in a same decision space. It often happens that the need to deal with a diversity of knowledge originates from those who are already used to a certain type of framing and the deployment of specific tools of assessment

As AI is applied to a wider range of scientific and pragmatic problems. so KR tools and techniques may have to develop accordingly to take into account emergent perspectives problems and requirements from different disciplines. In some cases. AI based on Machine learning may not include any form of explicit knowledge representation at all, while others may use the term KR with a different meaning [79] and rely on different kinds of formalism for theri communication and dissemination , mostly different types of diagrams and graphical visualizations, sketches and drawings, generally accompanied by narrative annotations. In essence, AI and KR cannot be separated, and KR cannot generally be completely devoided of natural language even in the case of subsymbolic AI. The nature of intelligence is coemergent resulting from processing different kinds of information using different senses, processed by multiple parts of the brain, leveraging a combination of physical and cognitive elements. When creating AI which leverages biologically inspired or subsymbolic approaches, the complexities and compositional aspects of intelligent functions should also be reflected in the computational representation.

5. The Postulate

The analysis of a selection of scholarly articles in the fields of Neuroscience - where research about neural networks originates - and machine learning - shows that AI systems are becoming increasingly complex, and their formulation includes much more than individual algorithms, leading to advanced logical structures supporting vast inference networks. They are also becoming

 doi:10.20944/preprints202001.0163.v1

increasingly hybrid, using a combination of symbolic and non symbolic approaches. Even in the case where a particular ML algorithm is developed and tested in an isolated way, independently of other components and architectures, its explanation in a scholarly communication or teaching/learning context implies an element of narrative and often the use of visual or graphical annotations, the latter being common place in systems and software engineering research and teaching and documentation. Therefore, since AI is becoming increasingly relevant to a wide range of disciplines and contexts, a postulate is proposed to facilitate explainability and trustworthyness and other qualities

for AI to be explainable

each MLA (machine learning algorithm)
there should correspond
A visual, graphical, a NL(natural language) or CL (controlled language)
annotation

5.1 Postulate Validation

Said postulate has been work-in-progress for the last few years , and early versions were shared informally with peers during workshops, in research notes and via mailing lists and discussion groups [1]. It serves as a guiding principle currently being implemented in various forms, with a recent wave of research both in explainable AI [81, 82] and reproducibile Machine Learning [83] adopting it. In addition, a convergence between neural computation and knowledge based systems is leading much of current innovations in AI, making said postulate defacto for hybrid models

6. Conclusion and Related Work

Advances in the automation capabilities of intelligent engineered systems are developing alongside other rapid changes impacting humanity. The universe of discourse is expanding, yet technology development cycles are becoming shorter. Even getting small things wrong, could have long term dramatic consequences and repercussions. This is a time for ' all hands on deck' but limited explainability of much of what is being done and publicly funded in AI, can become a barrier. In related research, a qualitative investigation of AI Educational resources is carried out, where the gaps identified in this paper are evaluated in relation to graduate teaching curricula [87]. In an oral presentation at the Brain Informatics Conference [88] I discuss the need to facilitate the convergence of knowledge representation, Neuroscience and Machine Learning , where I anticipate some knowledge domain synthesis to take place. The importance of AI KR in relation to eGovernance transparency and accountability is work in progress [89] Finally I emphasise the role of KR in Autonomous Instructional Systems to ensure functional ethics [90]

This paper considers the lack of adequate adoption of KR observed in much of machine learning research and education as part of greater epistemological omissions contributing to uncertainties and risks in AI, and notes the convergence of knowledge based systems and machine learning into neuromorphic engineerings and makes the case for adequacy of KR in the light of forthcoming advances.

[1] https://lists.w3.org/Archives/Public/semantic-web/2019Aug/0048.html

References

1. Turek, M - DARPA Expainable AI 2016. https://www.darpa.mil/program/explainable-artificial-intelligence
2. Deng, Jia, et al. "Imagenet: A large-scale hierarchical image database." 2009 IEEE conference on computer vision and pattern recognition. Ieee, 2009
3. Yingtan, Li. "Research On Trustworthiness And Trustworthy Algorithm Of Ai [J]." Information And Control(1999).
4. Mixed Grounded Theory: Merging Grounded Theory with Mixed Methods and Multimethod Research The SAGE Handbook of Current Developments in Grounded Theory https://methods.sagepub.com/book/the-sage-handbook-of-grounded-theory-2e/i3779.xml
5. Marvin Minsky. 1974. A Framework for Representing Knowledge. Technical Report. Massachusetts Institute of Technology, Cambridge, MA, USA.
6. Goldstein, I., & Papert, S. (1977). Artificial intelligence, language, and the study of knowledge. *Cognitive science*, *1*(1), 84-123.
7. Guarino, Nicola. "The ontological level: Revisiting 30 years of knowledge representation." Conceptual modeling: Foundations and applications. Springer, Berlin, Heidelberg, 2009. 52-67.
8. Mac Gregor, Robert. "The evolving technology of classification-based knowledge representation systems." Principles of semantic networks. Morgan Kaufmann, 1991. 385-400.
9. Sowa, John F. Knowledge representation: logical, philosophical, and computational foundations. Vol. 13. Pacific Grove, CA: Brooks/Cole, 2000.
10. Van Harmelen, Frank, Vladimir Lifschitz, and Bruce Porter, eds. Handbook of knowledge representation. Vol. 1. Elsevier, 2008.
11. Di Maio, Mindful Technology Buddhist Door, Online Article Hong Kong 2019 https://www.buddhistdoor.net/features/knowledge-representation-in-the-nalanda-buddhist-tradition
12. Zarri, Gian Piero. "Functional and semantic roles in a high-level knowledge representation language." Artificial Intelligence Review 51.4 (2019):537-575. link.springer.com/article/10.1007/s10462-017-9571-5
13. McCarthy, J. J., M. L. Minsky, and N. Rochester. Artificial intelligence. Research Laboratory of Electronics (RLE) at the Massachusetts Institute of Technology (MIT), 1959. https://dspace.mit.edu/bitstream/handle/1721.1/52263/RLE_QPR_053_XIII.pdf
14. McCarthy, John. "Programs with common sense. Mechanization of thought processes, Vol. I." (1959). https://stacks.stanford.edu/file/druid:yt623dt2417/yt623dt2417.pdf
15. Brachman, Ronald J. "The Future of Knowledge Representation." AAAI. Vol. 90. 1990.
16. Brachman, Ronald J., Hector J. Levesque, and Raymond Reiter, eds. Knowledge representation. MIT press, 1992.
17. Newell, Allen. "The knowledge level." Artificial intelligence 18.1 (1982): 87-127.
18. https://www1.nyc.gov/assets/adstaskforce/downloads/pdf/ADS-Report-11192019.pdf
19. [https://www.fastcompany.com/90436012/the-first-effort-to-regulate-ai-was-a-spectacular-failure
20. R. Davis, H. Shrobe, and P. Szolovits. What is a Knowledge Representation? AI Magazine, 14(1):17-33, 1993.
21. Schreiber, G. 1993. Operationalizing models of expertise. In G. Schreiber, B. Wielinga, and J. Breuker (Eds.), KADS—APrincipledApproachtoKnowledge-BasedSystemsDevelopment, London: AcademicPress, pp.119– 149
22. Davies Knowledge Representation, in International Encyclopedia of the Social & Behavioral Sciences, 2001 https://doi.org/10.1016/B0-08-043076-7/00540-4
23. Bergman M. KR Practionary Springer 2016
24. Kühnberger, K. U., Gust, H., & Geibel, P. (2008). Perspectives of Neuro--Symbolic Integration--Extended Abstract-In Dagstuhl Seminar Proceedings. Schloss Dagstuhl-Leibniz-Zentrum für Informatik.
25. Brooks, Rodney A. "Intelligence without representation." Artificial intelligence 47.1-3 (1991): 139-159.
26. Pylyshyn, Zenon W. Things and places: How the mind connects with the world. MIT press, 2007.

27. Neural Signatures of Compositionality in the Human Brain Elizabeth A. Shay

28 R. Sun, "A discrete neural network model for conceptual representation and reasoning." Proceedings of the 11th Cognitive Science Society Conference. pp. 916-923. Lawrence Erlbaum Associates, Hillsdale, NJ. 1989.

29 Hilario, Melanie. "An overview of strategies for neurosymbolic integration." *Connectionist-Symbolic Integration: From Unified to Hybrid Approaches* (1997): 13-36.

30. Connectionist-Symbolic Integration: From Unified to Hybrid Approaches. Sun, Ron Ed, and Frederic Ed Alexandre. "Connectionist-symbolic integration: From unified to hybrid approaches." IJCAI Workshop on Connectionist-Symbolic Integration: From Unified to Hybrid Approaches, Aug, 1995, Montreal, PQ, Canada; 14th International Joint Conference on Artificial Intelligence. Lawrence Erlbaum Associates Publishers, 1997.

31. Ioannis Hatzilygeroudis Jim Prentzas Neuro-Symbolic Approaches for Knowledge Representation in Expert Systems January 2004International journal of hybrid intelligent systems 1(3):111-126 DOI: 10.3233/HIS-2004-13-401 SourceDBLP

32. INTERNATIONAL CONFERENCE ON INTELLIGENT SYSTEMS (LONDON ENGLAND UK). 1995 1Neuro-symbolic reasoning - a solution for complex. Problemas J. M. CorchadoDepartment of Computing and Information Systems, University of Paisley, High Street, Paisley PA1 2BE, Scotland, U.K.

33. Di Maio, P. Systemic Deviation Research Note ISTCS.org

34. Ramirez, Carlos, and Benjamin Valdes. "A general knowledge representation model of concepts." *Advances in knowledge representation* (2012)

35. Müller, Vincent C. (2016), 'Editorial: Risks of artificial intelligence', in Vincent C. Müller(ed.), Risks of general intelligence (London: CRC Press - Chapman & Hall), 1-8. https://philpapers.org/archive/MLLERO-2.pdf

36 Liability for Honest Misrepresentation Author(s): Samuel Williston Source: Harvard Law Review, Vol. 24, No. 6 (Apr., 1911), pp. 415-440 Published by: The Harvard Law Review Association Stable URL: https://www.jstor.org/stable/1325080 Accessed: 07-12-2019 05:40 UTC

37 Bias in Artificial IntelligenceG S. Nelson, MMCi, CPHIMS⇑ doi: 10.18043/ncm.80.4.220. North Carolina Medical Journal July-August 2019 vol. 80 no. 4 220-222

38. Yu KH, Kohane IS. Framing the challenges of artificial intelligence in medicine. BMJ Qual Saf. 2019;28(3):238-241.

39. Saniat Javid Sohrawardi, Akash Chintha, Bao Thai, Sovantharith Seng, Andrea Hickerson, Raymond Ptucha, and Matthew Wright. 2019. Poster: Towards Robust Open-World Detection of Deepfakes. In Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security (CCS '19). ACM, New York, NY, USA, 2613-2615. DOI: https://doi.org/10.1145/3319535.3363269

40. Arel, Itamar, Derek C. Rose, and Thomas P. Karnowski. "Research frontier: deep machine learning--a new frontier in artificial intelligence research." IEEE Computational Intelligence Magazine 5.4 (2010): 13-18.

41. Li, X., Zhang, S., Huang, R. et al. A survey of knowledge representation methods and applications in machining process planning Int J Adv Manuf Technol (2018) 98: 3041. https://doi-org.nls.idm.oclc.org/10.1007/s00170-018-2433-8

42. Oren etzioni Is There a Future for AI Without Representation?

43. MILLER, D. "INTELLIGENCE WITHOUT ROBOTS." AI MAGAZINE 15.2 (1994): 10-11.

44. Jordanous, Anna. "Intelligence without Representation : A Successful Method of Modelling Intelligence ? Theories of Mind Term Paper 2007." UNPUBLISHED (2007). https://pdfs.semanticscholar.org/6624/3591938bdd352be60515cbfa73cc98c3fe47.pdf

45. Margaret Boden Artificial Intelligence. 10.4324/9780415249126-W001-1 https://www.rep.routledge.com/articles/thematic/artificial-intelligence/v-1

46. Pestilli, Franco. "Human white matter and knowledge representation." PLoS biology 16.4 (2018): e2005758.

47 Wang, Xiaoying, et al. "Two Forms of Knowledge Representations in the Human Brain." bioRxiv (2019): 691931. https://www.biorxiv.org/content/10.1101/691931v1

48. Locatello, Francesco, et al. "Challenging common assumptions in the unsupervised learning of disentangled representations." arXiv preprint arXiv:1811.12359 (2018)

49. VanHilst, Michael, and David Notkin. "Decoupling change from design." ACM SIGSOFT Software Engineering Notes. Vol. 21. No. 6. ACM, 1996.

50. Vardi, Moshe Y. "The universal-relation data model for logical independence." IEEE software 5.2 (1988): 80-85.

51.https://www.researchgate.net/publication/333477500_Are_Disentangled_Representations_Helpful_for_Abstract_Visual_Reasoning

52. https://mathinsight.org/vectors_arbitrary_dimensions

53. Reason, James. "Human error: models and management." Bmj 320.7237 (2000): 768-770.

54. Vester, F. (2007). The art of interconnected thinking. Ideas and tools for tackling complexity. München, Germany: MCB.

55. Sowa, John F. "The challenge of knowledge soup." Research trends in science, technology and mathematics education (2006): 55-90

56. Turing Award Lecture ACM https://awards.acm.org/about/2018-turing Transcript file https://www.zotero.org/groups/2351244/knowledge_representation/items/itemKey/ETCNEQIW

57. Buck, Ross. "Nonverbal communication: Spontaneous and symbolic aspects." American behavioral scientist 31.3 (1988): 341-354.

58. R. Davis, H.S. & Szolovits, P.: "What is knowledge representation?".AI Magazine, Vol. 14, No. 1, pp. 17-33, 1993

59. Brachman, R.J.: "On the Epistemological Status of Semantic Networks". In Findlet, N.V. (ed.): "Associative Networks: Representation and Use of Knowledge by Computers".Academic Press, pp. 3-50, 1979

60 Levesque, Hector J., and Ronald J. Brachman. "Expressiveness and tractability in knowledge representation and reasoning 1." Computational intelligence 3.1 (1987): 78-93.

61.Rich & Knight : Artificial Intelligence, Second Edition, McGraw Hill, 1991.

62. Brachman, Ronald J., and Hector J. Levesque. "Competence in Knowledge Representation." AAAI. 1982. https://pdfs.semanticscholar.org/4436/513cdd7d9d1e0b2d4edf4c16a8b7a5ef57d7.pdf?_ga=2.81184597.354059788.1574996976-335002274.1574246125

63. Ribarić, Slobodan. "Knowledge representation scheme based on Petri net theory." International Journal of Pattern Recognition and Artificial Intelligence 2.04 (1988): 691-700.

64. Inferential Adequacy https://link.springer.com/article/10.1007%2FBF00245941?LI=true

65. Shimojima, Atsushi. "The inferential-expressive trade-off: A case study of tabular representations." International Conference on Theory and Application of Diagrams. Springer, Berlin, Heidelberg, 2002.

66. Ford, Kenneth M., et al. "Knowledge acquisition as a constructive modeling activity." International Journal of Intelligent Systems 8.1 (1993): 9-32.

67 Advani, Aneel, et al. "Developing quality indicators and auditing protocols from formal guideline models: knowledge representation and transformations." AMIA annual symposium proceedings. Vol. 2003. American Medical Informatics Association, 2003.

68. Gruber, Thomas R. "Formal ontology in conceptual analysis and knowledge representation." Chapter" Towards principles for the design of ontologies used for knowledge sharing" in Conceptual Analysis and Knowledge Representation (1993).

69.Bhattacharya, Devanjan & Ghosh, Jayanta. (2008). Evaluation of Knowledge Representation Schemes as a Prerequisite toward Development of a Knowledge-Based System. Journal of Computing in Civil Engineering - J COMPUT CIVIL ENG. 22. 10.1061/(ASCE)0887-3801(2008)22:6(348)

70.Corcho, Oscar, and Asuncion Gomez-Perez. "Evaluating knowledge representation and reasoning capabilites of ontology specification languages." (2000).

71. Bingi, R., Deepak Khazanchi, and Surya B. Yadav. "A framework for the comparative analysis and evaluation of knowledge representation schemes." Information processing & management 31.2 (1995): 233-247. https://doi.org/10.1016/0306-4573(95)80037-T

72. Messina, Elena R., John M. Evans, and James S. Albus. Evaluating knowledge and representation for intelligent control. National Inst Of Standards And Technology Gaithersburg Md Intelligent Systems Div, 2001.

73. Wang, Zhigang, and Juan-Zi Li. "Text-Enhanced Representation Learning for Knowledge Graph." *IJCAI*. 2016.

74. Sattler, Ulrike. *Terminological knowledge representation systems in a process engineering application*. Mainz, 1998.

75. Clark, Peter. "Knowledge representation in machine learning." *Machine and Human Learning* (1989): 35-49.
76. Di Maio Paola, KR A bRIDGE BETWEEN AI AND NS Oral Presentation and Extended Abstract 12th Conference of Brain Informatics Haikou Hainan Dec 2019
78. Pereira, G., and S. O. Funtowicz. "Knowledge representation and mediation for transdisciplinary frameworks: tools to inform debates, dialogues & deliberations." International Journal of Transdisciplinary Research 1.1 (2006): 34-50.
79 Neelakantan, Arvind Ramanathan, "Knowledge Representation and Reasoning with Deep Neural Networks" (2017). Doctoral Dissertations. 1114. https://scholarworks.umass.edu/dissertations_2/1114
80. Angele, J., Fensel, D., Landes, D., & Studer, R. (1998). Developing Knowledge-Based Systems with MIKE. Automated Software Engineering, 5, 389-418.
81.https://towardsdatascience.com/the-how-of-explainable-ai-post-modelling-explainability-8b4cbc7adf5f
82. http://www.semantic-explainability.com/
83. https://www.cs.mcgill.ca/~jpineau/ReproducibilityChecklist.pdf
84. NeuroML.org
85. Levesque, H. J., & Lakemeyer, G. (2001). The logic of knowledge bases. MIT Press.
86. Partridge, Derek, and William B. Yates. "Replicability of neural computing experiments." Complex Systems 10.4 (1996): 257-282.
87. Di Maio, P Accepted Abstract [Special Issue: THE FATE of AIED: Fairness, Accountability, Transparency, and Ethics (abstracts due: Jan 31, 2020; papers due: Mar 30, 2020)] (## Special Issue: The FATE of AIED)
88. P Di M aio Oral Presentation
89 AICIO Conference Taipei 2019
90 P Di Maio Invited Paper, forthcoming ID:1697 Knowledge Representation for Ethical AI in AIS