AI Teaching and Learning KR, Neuro Symbolism and Reliability Notable Interlinked Gaps

Paola Di Maio[a]

[a]*Center for Systems, Knowledge Representation and Neuroscience*

Abstract

AI is a popular subject in Computer Science covering a wide range of topics from robotics to machine learning. In graduate and postgraduate studies worldwide course content is largely driven by western universities and by the global tech industry mostly due to the fact that the main language for reference materials is English, the lingua franca in science and technology, and because education is typically geared towards professional careers and employment. The topics taught and the choice of resources adopted for each course however can vary from school to school, depending on the degree of autonomy of instructors, In developing an comprehensive global syllabus for teaching Responsible AI, in consideration of epistemic and regional diversity, the need arose for tangible data to understand the state of the art in this field of education. During this work, certain issues in the teaching of Knowledge Representation (KR) typically taught as a subtopic of AI, emerged as worthy of further investigation. KR is a vast subject that lies at the heart of AI and cannot be separated from it. Understanding the breadth of scope of KR and many of its roles not only in devising AI, as well as in other types of systems but is also a key to other spheres of human interest. Adequate KR is also necessary to designing, implementing and evaluating reliable and accountable AI. The paper presents the rationale and preliminary findings of research spanning multiple countries, regions, languages investigating how KR is taught, leading to the hypothesis that this lack of adequate KR education may be contributing to increasing risks in irresponsible AI, as well as to other systemic risks including systemic deviation. One of the early findings of the research so far is that there are notable gaps and inconsistencies in the teaching of Knowledge Representation and in particular, the complete lack of Neuro Symbolic KR and AI ethics in curricular topics and teaching materials and courses, and their relation as cofactors in systemic dysfunction.

1. Introduction

Artificial Intelligence (AI), the combination of technology, software systems and computational techniques resulting in autonomous intelligent functions, is the product of a peak in human evolution, the outcome of synthesis of science, engineering and technology. It can be easily argued that AI is still a work in progress and that it has not yet been truly achieved, nonetheless it receives much attention in graduate and postgraduate education and it is typically taught as an elective in Computer Science, with the main focus is becoming increasingly Machine Learning (ML), with limited or no emphasis on non ML based AI. However the reach of AI is widening, it is increasingly becoming embedded in everyday systems, from banking to voting to all fields of human activity. As such, questions relating to the reliability and fairness of autonomous systems are becoming pivotal to the development implementation and to the teaching of AI [1]. AI can result in skewed, biased and even plain wrong outputs. Given the amount of resources going into AI education and learning and its growing importance, it is surprising to find limited amounts of research literature and evaluation of teaching practices and materials in Knowledge Representation (KR). One of the factors contributing to unreliable AI is that students are not always taught about the importance of the implications and impact of their algorithms when deployed, nor how to reduce and avert the consequences of poor systems development practices using KR. In Systems and Software Engineering there is much emphasis on responsibility and thoroughness of documentation, however the CS and MS community are driven by different priorities such as the quest for novelty, technical prowess and performance over safety and responsibility. When the scope of technical education is limited to the point that technology becomes insulated from socio-technical concerns, such as placing ethics and responsibility at the center of computer science curricula, educational institutions fall short of creating monsters.

2. Motivation

Systemic risks and ethical concerns caused by irresponsible development of autonomous intelligent systems have been at the forefront of systems research since inception, but the adverse consequences and impact of the diffusion of powerful yet inconsiderate AI, have become likely and potentially irreversible and have be-

International Semantic Intelligence Conference (ISIC 2021), Feb 25-27, 2021, New Delhi, India

EMAIL: paola.dimaio@gmail.com (P.D. Maio)

ORCID: 0000-0003-0770-3713 (P.D. Maio)

come inextricable coupled with computation.

In particular, systemic deviation [2], defined as the drift of a system from an explicitly stated intended goals to tacit unintended outcomes, reflects a representational shift along the life cycle of intelligence systems, has become a subtle yet critical and fairly common built in dysfunction

The relation between AI, knowledge misrepresentation and systemic risks and failures is not generally reflected in Computer Science (CS) curricula, where the main focus of studies remains mostly mathematics based computation and programming. CS curricula do not typically include subjects that teach students how to minimize AI risks by leveraging knowledge and logical integrity and truth preservation and how to design and implement trustworthy systems in accordance with ethics and explicit policies. With few exceptions, typical graduate CS students majoring in AI have limited awareness of issues which are central concerns to contemporary research and practice, such as algorithmic bias, other than notionally. In the context of ongoing research into the advancement of science and education and with awareness of existential risks, an inquiry into qualitative aspects of AI curricula worldwide was started by the Center for Systems, Knowledge Representation and Neuro-Science in 2018. Notable gaps that have been identified include:

1. There is limited emphasis on responsible development practices/ethical implementations, in - terms of roles, good practices, awareness and the role of KR
2. AI KR, bias and AI accountability are not taught as co-factors contributing to the inestimable - risks of unreliable, dangerous intelligent autonomous systems
3. Despite the growing importance of Neuro Symbolic AI and KR, Neuro Symbolism is totally absent from teaching resources (as of January 2020)

KR plays many different roles in AI, CS as well as in other spheres of human thought and activity, as clearly identified in literature [3].

Computationally, KR enables the encoding of logical functions and specific techniques such as logic programming, devised specifically for the purpose of running executable programs, however, KR also plays a role in making explicit, therefore accountable and auditable the validity of system functions throughout the lifecycle and beyond(as in the case of legacy systems), and to make transparent and accountable and above all, consistent their implementation.

The lack of adequate KR especially in large complex systems contributes to hiding flaws and weaknesses eventually resulting in the increased risks of 'irresponsible' AI. This paper's central argument is that the complementarity of these issues should be addressed in education, and presents particle results and conclusions based on the initial phase of the study which is scheduled to continue and more complete results of the analysis of the findings will be reported as they become available.

3. Method and research design

Ongoing reviews of student projects and research publications in recent years, shows worrying gaps in awareness and understanding among the student population of systemic bias, AI ethics, and overall low concern for qualitative aspects of AI development. In planning the content production and publishing of course materials and instructional programs in Responsible AI Systems, it became necessary to characterize and address current gaps and figure out how to fit new topics and novel perspectives into existing CS curricula. A study was devised to acquire actual data and answer the questions:

What is missing to the education of responsible AI?
Is the relationship between AI KR, Technology Ethics and Reliable Systems Development understood by the academic establishment?
What Knowledge Representation topics, techniques and methods are taught in AI courses worldwide?
What books, materials and resources are used?
[1]

A combination of methods is being adopted in this research, partly from existing literature, partly from surveys and a questionnaire. The study is still ongoing. To gauge the needs of instructors and students and to guide the design of didactical resources existing course materials in use are being surveyed. First the main (in rank, size) teaching institutions in a portfolio of target countries and languages are identified through web searches, then web pages of relevant departments and instructors are surveyed. The process is especially time consuming because such lists of resources are not regularly published, and whatever information may be available on websites may be outdated and not correspond to what actually goes on in classrooms. To cross reference and validate the results of the initial survey individual instructors and departments are being contacted. A more targeted questionnaire designed for more detailed data is distributed through the AI Learning community, anticipating a possible skew: respondents from such lists tend to be

[1] Link to questionnaire: http://www.tinyurl.com/y5lqnyfl

English speaking, US based. Current efforts in this study are aimed at ensuring adequate survey samples including representatives of a diversity of countries, languages and types of schools. The overall study is aimed at respondents worldwide and questions were first discussed and reviewed informally with colleagues. The questionnaire may undergo further rounds of validation and refinement and will then be circulated to larger and more comprehensive segments. Responses will be grouped according to role (student, lecturer, other faculty). Finally a cross selection of the respondents will be invited to participate in focus groups to evaluate the recommendations, provide feedback and co-develop possible interventions. Preliminary results and conclusions presented here are based on the early stages of this survey, with the analysis of further outcomes remanded to future work.

4. Historical Background in Teaching Knowledge Representation

In software systems engineering there is much emphasis on requirements analysis, design and evaluation, driven by the goal of overall systems efficiency and reliability. By contrast in computer science and especially machine learning, there is much emphasis on devising and testing individual computational techniques and algorithms which are generally evaluated for performance.

High level KR in systems development can be considered as a modelling effort, a prerequisite for any implementation, At the other end KR can also be minted as a language to the level of code starting as implementation and with little or no modelling. KR can be considered as a conceptual blueprint for stakeholders and that ends up being transcribed into actual executable code. Inconsistencies between different levels of system representation results in all sorts of concerns in terms of AI ethics, transparency, accountability, fairness. Through its variety of roles, shared and explicit KR can support the integration of these different levels, yet this important function is not typically taught in AI KR education.

The teaching of KR has historically been challenging, for reasons partially identified [4], our findings so far confirm that the reality has not changed in decades. KR is generally still taught and practiced as a subset of AI, in the same way that software engineering, data science and knowledge modeling are secondary to the teaching of coding in software systems development. Teaching, practicing and learning KR is intellectually and cognitive challenging for reasons mentioned in literature: KR requires a diverse and complex skill set including abstraction, logic, the ability to communicate, express and codify these principles, as well as a familiarity with the countless Knowledge Representation formalisms and their validation. Furthermore as technological development progresses and systems become increasingly complex and more powerful, the role of KR is widening and becoming more central to understanding and addressing the challenges, especially systemic risks. KR also has political implications, because how reality is captured and represented impacts not only the system's outcomes but also people's beliefs, behaviours and decisions, and the public opinion at large. KR may not be entirely an exact science as it is closely related to human cognition, psychology, philosophy and language however it is rooted in formal methods. It is resource intensive, requires a higher order cognition and the ability to grasp complexity. Unlike other aspects of system and software development, KR cannot yet be reliably automated. In addition there are gaps between the theoretical and philosophical aspects of KR and how it is practiced in the 'real world' vs in educational contexts. Learning and teaching KR is becoming more complex as the understanding of human cognition and its applications expand, through evolutionary processes and technological progress. The scope and reach of AI powered systems is widening well beyond the domain of computer science. Engineering autonomous intelligent systems is becoming easier and requires hardly any education at all. KR as modelling activity is rarely or never taught as a standalone curricular subject, and outside learning environments is often limited or not at all included in AI systems documentation. It is not uncommon to find unjustified discrepancies between KR models and their implementations. This is a concern particularly when addressing AI explainability, fairness and accountability, Typically learners cannot even be held accountable for shortfalls and systemic risks because they are never taught how to use KR as a modelling tool to ensure system integrity.

Today, the lack of adequate KR can be considered as one of the contributing factors to the spread of unethical and faulty AI [5] intentionally or accidentally engineered, which are the source of transparency and explainability concerns in AI and causing much alarm in the IT community.

Without explicit, shared. valid and verifiable knowledge models (i.e., models that do not rely on hidden facts and assumptions), capable of covering a broad spectrum of systems. The adequacy and effectiveness

of KR in terms of complexity and entanglement can not be evaluated. In turn, without adequate KR, no trustworthy AI can be developed nor evaluated leading to a vicious circle that ends up as systemic risk and systematic deviation. "KR, as well as its tools and resulting products are permeated by moral values. This gives rise to problems of biases in representation, such as gender issues, dichotomy categorizations and lack of cultural warrant" [6]. Additional factors likely to impact knowledge representation education and training are:

- There is a shift from US centrism and English language to more international, culturally and linguistic diversity in students and developer - base: today many AI programming languages exist which are not based on the English language [7].
- There is an inherent dependence between KR in computation and deontological ethics, such as the role of representation in normative systems, for example juridical systems [8], nonetheless ther role of knowledge representation is missing from evaluations and meta evaluations [9].

5. The Role of KR in the context of complexity science and AI risks

The research and scientific communities now acknowledge the extent of the risks that can be triggered by AI deployments is inestimable, and that risk reduction strategies have limitations. The top most risk is existential: the risk to life, with other risks likely to cause wide ranging harms which cannot easily be estimated [8]. Many risks in AI can be caused simply by poor design and development, resulting in flaws in reasoning and representational imperfections which generate biased outcomes. Bias is well understood in research and science and thanks to the proliferation of AI and increasing awareness, it is now gaining attention in how it impacts computer science, IT and engineering. There are no simple methods to categorize and prevent bias, because it is inherent to (unenlightened) human nature. Bias can be accidental, caused by limited cognition, poor modeling, flaws in development or by implementation errors. In law this type of bias is referred to as honest misrepresentation (see for example the UK Misrepresentation Act 1967). But bias can also be intentional, deliberately constructed with the intent to deceive. Malfunction by design is a rather common

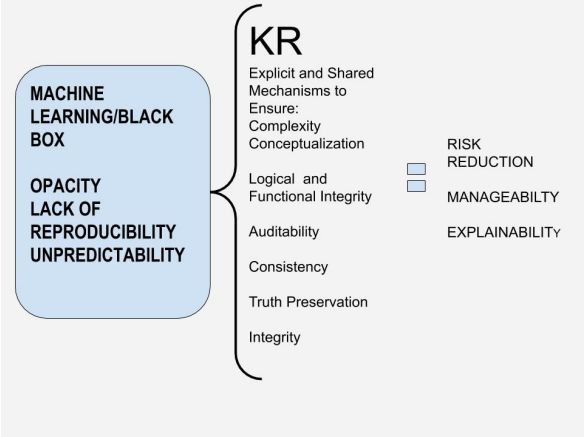

Figure 1: Symbolic KR can reduce AI risks

malicious act and it is aimed at producing misleading functions and results to deliberately manipulate, pilot and deceive, for some gain or to cause harm, as in the case of deep fakes [10].

Deep Fakes are the direct outcome of deliberate intentional knowledge misrepresentation to cause harm, and as such are considered a crime. This type of bias can be unthinkable and inconceivable, technically subtle to identify and expose without thinking with a criminal mind. In addition to known traditional roles for KR in AI development [11] novel roles are here identified which contribute to risk reduction in complex systems design and management and yet are missing from educational literature:

1. KR to support the consistency and integrity of policy throughout the system life cycle: Explicit Knowledge representation is required to make policies shared, verifiable and consistently implemented across systems.
2. To understand and detangle algorithmic Bias in complex systems.

6. Neuro Symbolic Knowledge Representation

Neuro Symbolic approaches in AI have been somewhat lacking mainstream attention, however they are making a comeback on research agendas, although they are not yet integrated in AI KR curricula and teaching materials in use. There can be no shift to Neuro Symbolic AI until Neuro Symbolic KR methods are developed, applied and taught in CS.

By helping to disambiguate and interpret the compositional patterns of the many technologies and methods necessary for complex AI developments, in particular how different algorithms and computational arte facts intersect and combine, neuro-symbolic KR is required to test and assess the consistency and integrity of AI systems and therefore necessary to evaluate the overall trustworthiness, accountability and explainability.

7. Preliminary Findings

The exploratory findings (from direct observations, literature review, knowledge audit of textbooks in use and initial questionnaire responses) so far this study reveal that

- Only one recently published paper in scholarly database searches directly addresses teaching - practices in AI [12], however the paper does not provide shared, reproducible data to support its findings, and it is based on a perspective limited to US institutions. The regional/US centric focus of the survey reported in the paper is not mentioned as a limitation of the research. The shortfall identified in Wollowski Selkowitz et al [[12] contributed to the motivation of this study as it points directly to the need to acquire further data and insights from a more global perspective about how AI and Knowledge Representation are taught.
- Notable gaps and discrepancies can be found in curricula and textbooks:
    1. The definition of what is KR can very considerably among authors
    2. Different textbooks teach different KR topics seemingly arbitrarily with no mention or justification for inclusion and exclusion criteria
    3. Increasingly important techniques such as neuro-symbolism approaches are not generally mentioned nor referenced neither in syllabi nor textbooks in use
    4. Key aspects and roles of KR in ensuring consistency and integrity across the system life cycle are not taught

KR is mostly taught using books written by western authors, and AI curricula are strongly influenced by western Educational system and by technology corporations. None of the major textbooks and resources in use mention Neuro Symbolic AI or Neuro Symbolic AI. The content of teaching curricula are decided by faculty members either collectively or individually, based on personal expertise however it is influenced by the need to pass professional examinations.

8. Preliminary recommendations

Based on the preliminary findings, considering generally agreed upon good system development practices in education and industry, a set initial recommendations is provided as follows:

1. KR should be taught in relation to extended socio-technical system functions and roles, in consideration of the fact that AI is now being embedded in most engineered systems, in particular to insure the audit-ability of systems consistency, integrity and adherence to policy, as well as its role as a tool for debiasing and algorithmic auditing
2. KR Education should include methods to address the different categories of AI Risks
3. AI KR education should teach how to evaluate the adequacy of KR in relation to systemic risks management and KR should be taught as a co-factor in ethical and trustworthy AI
4. Neuro-symbolic KR and AI methods and models should be included in curricula and teaching resources

9. Conclusion

This paper shares the rationale, research design and preliminary finding of a study aimed at acquiring reliable data about the teaching and learning of Knowledge Representation, in the light of mounting current challenges and risks caused by irresponsible AI, and in consideration of the increasing proliferation of embedded intelligence in automated systems widely in use. This study at the time of writing is a first of its kind. Preliminary findings include that there is limited knowledge about AI curricula and teaching, and limited awareness among the student population of their responsibility and the role of adequate KR in limiting AI risks. Additionally this research identifies that although there is renewed interest in neuro-symbolism, Neuro Symbolic AI/KR techniques are not taught nor are included in the textbooks in use. Additional notable gaps identified include the lack of an internationally balanced, less western centric curriculum and the lack of correspondence between the role of KR and

other aspects of systems reliability and trustworthiness. This research provides preliminary recommendations put forward for consideration by the international AI teaching community. It is expected to have a far reaching impact by benefiting instructors, students,study planners, publishers, academics and institutions at large, as it points out areas of improvement for the curriculum and where future contributions are needed. Future work includes compiling an open access resource based on a comprehensive index of KR topics that reflect the diversity and richness of perspectives to develop a teaching syllabus for Neuro Symbolism in AI, with consideration to increasing requirements for reliable engineering and diversity of views, The study is going to remain work in progress for the foreseeable future, as is the development of AI and responsible systems.

References

[1] E. G. Frankel, Systems reliability and risk analysis, Springer Science & Business Media 1 (2012).

[2] P. D. Maio, Systemic deviation, the evil in the machine, in: Proceedings of the 60th annual meeting of the International Society for the Systems Sciences (ISSS), volume 1, 2016.

[3] C. Ramirez, B. Valdes, A general knowledge representation model of concepts, Advances in knowledge representation (2012).

[4] L. Morgenstern, R. H. Thomason., Teaching knowledge representation: Challenges and proposals., in: Proceedings. KR2000 Conference, 2000, pp. 725–733,.

[5] V. C. Müller, Risks of general intelligence, 2011. URL: https://philpapers.org/archive/MLLERO-2.pdf.

[6] S. O. Milani, Biases in knowledge representation: an analysis of the feminine domain in brazilian indexing languages, NASKO 3 (2011) 94. doi:10.7152/nasko.v3i1.12794.

[7] N. HT, Fligthstar research, 2018. URL: https://github.com/FlightstarResearch/Artificial-Intelligence-in-Non-English-Based\-Programming-Language.

[8] V. M, Towards new ways of modeling deontic concepts in ontologies, JOWO (2019).

[9] T. Hagendorff, The ethics of ai ethics: An evaluation of guidelines., Minds and Machines (2020).

[10] N. D. T. N. S. Nguyen T. T., Nguyen C. M., Deep learning for deep fakes creation and detection (2019).

[11] H. S. Davis, P. Szolovits, What is a knowledge representation?, AI Magazine (1993) 17–33.

[12] M. S. . R. B. Wollowski, A survey of current practice and teaching of ai proceedings of the sixth symposium on educational advances in artificial intelligence, 2016.