On the relevance of logic for AI: misunderstandings in social media, and the promise of neuro-symbolic learning

Vaishak BELLE [a]

[a] *University of Edinburgh & Alan Turing Institute, UK*

Abstract. Artificial Intelligence (AI) is widely acknowledged as a new kind of science that will bring about (and is already enabling) the next technological revolution. Virtually every week, exciting reports come our way about the use of AI for drug discovery, game playing, stock trading and law enforcement. And virtually all of these are mostly concerned with a very narrow technological capability, that of predicting future instances based on past instances. Although it is now recognized that this type of statistical associations is limited in its ability to understand the world and model its knowledge, there is still a lot of criticism and hesitancy about the use of symbolic logic to accomplish or assist in a broader vision for AI. In this article, we look at some of the assumptions held and circulated in social media about logic and point out that there are deep misunderstandings. By arguing that symbolic logic is more flexible than believed by non-experts, we make a case for Neuro-Symbolic AI offering the best of both worlds.

Keywords. neuro-symbolic AI
logical foundations
GOFAI

They don't have intelligence. They have what I call "thintelligence". They see the immediate situation. They think narrowly and they call it being focused. They don't see the surround. They don't see the consequences. – **Michael Crichton, Jurassic Park**

1. Introduction

Artificial Intelligence (AI) is widely acknowledged as a new kind of science that will bring about (and is already enabling) the next technological revolution. Virtually every week, exciting reports come our way about the use of AI for drug discovery, game playing, stock trading and law enforcement. And virtually all of these are mostly concerned with a very narrow technological capability, that of predicting future instances based on past instances.

Although it is now recognized that this type of statistical associations is limited in its ability to understand the world and model its knowledge, there is still a lot of criticism and hesitancy about the use of symbolic logic to accomplish or assist in a broader vision for AI. In this article, we look at some of the assumptions held and circulated in social media about logic and point out that there are deep misunderstandings about how logic works and how flexible it is. We examine some of the criticisms also raised by Geoff

Hinton. Following these, we make a case for where Neuro-symbolic AI can offer the best of both worlds, but which requires us to acknowledge that logic is a lot more flexible than believed by many of its critics.

2. Thintelligence

Identifying statistical patterns, correlations, and associations are, without doubt, extremely useful. In the first instance, they are needed in numerous applications to inspect features and properties of interest in observed data.

This type of AI is very useful. It drives applications in recommendation systems, for example, and likely is more than sufficient for such applications. For example, while searching for ”how to raise lambs” on an online bookstore, we might be a little disappointed if it suggests ”silence of the lambs” by Thomas Harris, and somewhat annoyed if it suggests cook books on ”how to cook lamb”, but it is unlikely to have long-term effects. This type of AI might also be useful but somewhat nebulous for fast-tracking, say, job applications, provided these models are adjusted for bias, and a human intervenes and interprets the outcome and determines how to act further. This type of AI was largely believed to be sufficient for vision systems [88], until it was realised that self-driving cars fail stupendously, and that the state-of-the-art systems can be fooled in strange and unnatural ways [36].

Be that as it may, this is a very narrow view of AI capabilities. Indeed, AI, as understood by both scientists and science fiction writers, is clearly much broader. In fact, pattern recognition, machine learning and finding associations by mining data are closely related subfields of AI. Put differently, from first principles, what distinguishes big-data analysis from AI is that the set of capabilities we wish to enable with the latter. We are not interested in a “thintelligence”, but rather a general-purpose, autonomous computational entity that, in the very least, has agency.

3. Symbolic logic is old-fashioned?

However, ”modern” AI has moved on, we are told. The idea of using symbolic logic is old-fashioned, and the area of knowledge representation defined over symbolic logic is now affectionately (or perhaps pejoratively) called good old-fashioned AI, or GOFAI short.

In the early days of AI, John McCarthy put forward a profound idea to realise artificial intelligence (AI) systems [68]: he posited that what the system needs to know could be represented in a formal language, and a general-purpose algorithm would then conclude the necessary actions needed to solve the problem at hand. The main advantage is that the representation can be scrutinised and understood by external observers, and the system’s behaviour could be improved by making statements to it. Numerous such languages emerged in the years to follow, but first-order logic remained at the forefront as a general and powerful option [72]. Propositional and first-order logic continue to serve as the underlying language for several areas in AI, including constraint satisfaction, automated planning, database theory, ontology specification, verification, and knowledge representation.

And yet, ”modern” AI has decided that these efforts are superfluous, or at least easily replaceable once a training dataset has been created. An an infamous and inflammatory instance, Turing-award winner Geoff Hinton remarked that fixating on symbols was a waste of time, analogous to funding research on gasoline engines. Implicit here is the argument that we clearly need to be focussing on electric engines, presumably analogous to deep learning.[1] In 2020, he reiterated his position and suggested that:[2] “Deep learning is going to be able to do everything.” Strangely, his position seems to have changed over the years,[3] but it is hard to get a sense of what kind of mixture of symbols vs learning he is advocating for. For example, in a very recent interview after quitting his position at Google, the following transpired:[4]

The dominant idea at the time, known as symbolic AI, was that intelligence involved processing symbols, such as words or numbers.

But Hinton wasn't convinced. He worked on neural networks, software abstractions of brains in which neurons and the connections between them are represented by code. By changing how those neurons are connected — changing the numbers used to represent them — the neural network can be rewired on the fly. In other words, it can be made to learn.

“My father was a biologist, so I was thinking in biological terms,” says Hinton. “And symbolic reasoning is clearly not at the core of biological intelligence.”

“Crows can solve puzzles, and they don't have language. They're not doing it by storing strings of symbols and manipulating them. They're doing it by changing the strengths of connections between neurons in their brain. And so it has to be possible to learn complicated things by changing the strengths of connections in an artificial neural network.”

Almost every scientist loves the appeal of a single model, and hence, the search for the theory that unifies all observational data, as seen in physics with string theory, for example [25]. And by extension, the appeal of purely neural model is attractive indeed. However, there is lots to debate here. Firstly, deep learning models are loosely inspired by the brains but not fully accurate representations (yet) [86,71]. Secondly, there is the notion of innateness [89], and how much evolution might helps the brain in understanding and processing the world in a structured manner. Finally, we must bear in mind that we still lack a complete understanding of how the neurons of a bird (let alone a human) are wired. Merely knowing that neural weights enable birds to solve puzzles and recognize faces does not necessarily imply that our implementation of their neurons should resemble or possess similar properties, at least to a dependable extent.

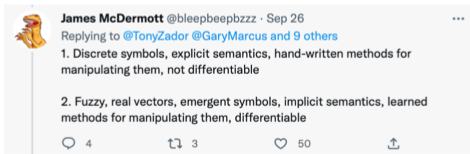
But putting such issues aside, it is also worth noting that proponents of the symbolic approach to AI never explicitly claimed the existence of symbolic representations within our minds. In essence, the symbolic approach offers a coherent strategy for: (a) executing symbolic expressions, and (b) comprehending the (idealized) implications of one's knowledge, as per inference rules stipulated in some logic. As argued by Levesque [60], this is not a novel concept, as Leibniz articulated centuries ago that certain types of

[1] https://twitter.com/tabithagold/status/1070736319901519876

[2] https://www.technologyreview.com/2020/11/03/1011616/ai-godfather-geoffrey-hinton-deep-learning-will-do-everything/

[3] https://www.noemamag.com/deep-learning-alone-isnt-getting-us-to-human-like-ai/

[4] https://www.technologyreview.com/2023/05/02/1072528/geoffrey-hinton-google-why-scared-ai/



thinking adhere to symbolic processing. Hence, why not employ an algebraic treatment for cognition?

Once again, the allure of a purely neural approach is understandable, given its simplicity and the sense of a "unified theory" it evokes.

By taking a step back, we realize that until the past few centuries, our understanding of the brain and neurons was limited. Yet, during this time, we were able to calculate, develop number theory, construct calculators, and ultimately create computers. Imagine if we had solely dedicated ourselves to constructing elaborate brain replicas in the hopes that they could handle tax calculations for us. This underscores the significance of the symbolic approach, which offers an idealized and accurate method for certain forms of reasoning. There is a popular analogy [15] suggesting that we need not build wings and feathers to build airplanes; comprehending the principles of aerodynamics is enough. So, why shouldn't the development of a theory of artificial cognition be just as relevant for a type of AI that is behaviorally similar to humans in some instances, without necessarily resorting to a brain-like architecture?

Goeff Hinton, of course, is not alone is being dismissive about symbols. There are many other more severe views on the relevance of logic for modern AI, but in what follows, we will survey and respond to a selected set of misunderstandings extracted from the social media platform Twitter.

4. There is a dichotomy

A common view held by many in the broader community that there is an inherent dichotomy between symbolic logic and machine learning, the former focused on discrete structures and the latter focused on continuous representations. In fact, even scientists within the AI community make this distinction [79], and suggest that logic is not really appropriate for machine learning. Consider, for example, the tweets by James McDermott.

This, admittedly, is not even necessarily negative on the topic of symbols, but just points out that: (a) symbolic processing is a separate topic of study that can be independently done from symbolic logic, and (b) symbolic logic as used in AI is focused on discrete symbols.

What we are seeing here is a narrowing of the use of "logic" simply as *classical logic* – say, as introduced in [29].

But going back to the history of the use of logic in AI [72], these views do not quite hold. For instance, John McCarthy himself was concerned about probabilities and the use of it in knowledge representation. However, he makes a very salient point that we need to think carefully how numbers and first-order sentences fit together. For example, he argues [69]:

(i) It is not clear how to attach probabilities to statements containing quantifiers in a way that corresponds to the amount of conviction people have.

(ii) The information necessary to assign numerical probabilities is not ordinarily available. Therefore, a formalism that required numerical probabilities would be epistemologically inadequate.

His point, simply, is that we should not be expected to put probabilities on every formula; sometimes it suffices to say that $p \vee q$ holds without saying which, and by how much. Moreover, if we assign a probability of r on that formula, or to, say, $\exists x P(x)$, such an assertion in itself does not provide any additional information on how to further assign a probability to p, q, $P(a)$, and so on. Many popular languages for logic and probability, including Markov logic networks [79], ProbLog [77] and BLOG [70], do not allow this level of flexibility. In fact, this requires a different type of machinery altogether, one which permits multiple prior distributions [12]. In contrast, in ProbLog, it is assumed that there is a single distribution over the model, and not specifying a probability on a disjunction might be interpreted as a hard constraint that is true in all possible worlds.

Admittedly, it is true that the knowledge representation area of AI largely focuses on discrete symbols and a Boolean interpretation. But, on the other hand, it's been close to 60 years since we have fuzzy logic [91], among others [53]. And as suggested above, probability measures [31] on first-order structures and other proposals on logic and uncertainty [79,77,70,12] allow us to mix probability and possibility theory in a logical language in different ways [26].

The one last point on the criticism is that although it is usual to talk about discrete things in logical AI, it is not true that they need to do so. In fact, more generally, although is true that logical formulas are discrete structures, they can very easily also express properties about countably infinite or even uncountably many objects [8]. Reasoning about real numbers have long been an area of interest in mathematical logic [51], going back to Tarksi.

5. Logic only interprets symbols one way

There seems to be a general confusion in the non-AI community between syntax and semantics, or more accurately, that the syntax is tied to semantics in a certain way, as seen in the classical interpretation for propositional logic.

There are many systems for writing down symbols, and interpreting logical symbols and formulas built up these symbols. Classical ideas include propositional logic (Boolean symbols, A and B is true iff A is true and B is true) and first-order logic, which uses quantifiers. In first-order logic, for instance, there is a domain of discourse which stands for the objects in the world. We then say that $\exists P(x)$ is true if and only if there is some individual from the domain of discourse such that the property P is true for that individual. First-order logic can also use functions over reals, as seen in satisfiability modulo theory [5].

However, there is also modal logic [56], which can capture possibilities, beliefs, and intentions [81]. A variant of modal logic with numbers on worlds can lead to probabilistic logics [39]. Fuzzy logics map Boolean symbols to real numbers, leading to real-valued semantics for connectives. For example, if A and B get values between 0 and 1, then A and B gets a value of 1 iff $min(A, B)$ is 1. Moreover, the conjunction could also get a value between 0 and 1.

These are all part and parcel of symbolic logic. The choice of the language, the choice of the semantic rules that we use over the well-defined formulas, along with its computational properties such as decidability are all put together in a logical framework [41]. Moreover, once a logical framework is considered, we could choose to prove logical entailments either by considering assignments to the variables and seeing if the consequent follows, or by applying inference rules established in a proof theory. This leads to the choice of doing one of: theorem proving [42], model checking [4], SAT solving [5], and model counting [35].

6. Monotoncity

Classical logic is monotonic. That is, if $\alpha_1, \ldots, \alpha_k \models \beta$, then it cannot be the case that adding new knowledge, say, α' forces us to retract β: formally, it has to be that $\alpha_1, \ldots, \alpha_k, \alpha' \models \beta$ also.

John McCarthy was concerned about the problem of monotonicity and wondered how we might deal with exceptions and abnormality. The problem of monotonicity is so ubiquitous, it even comes up in the formulation of automated planning [78]. For example, imagine that you have an action to paint a box blue and another action that pushes the object. Let us say we paint the object and next, we push the object. When we execute the second action, it is implicit that the color of the object does not change. So we would have to somehow codify not only what the effects of the push action are, but also what the non-effects are. And if we start writing down all the non-effects, there could be exponentially many. Moreover, there are various preconditions that must hold for us to be able to push the object. For instance, we should be strong enough to push it, we must not be holding other objects, we are presumably operating under reasonable gravity assumptions, and so on. And if we start expressing all of them, it again looks like a hopeless task. Yet under some assumptions, we can assume so-called causal completeness [78] in the sense that the conditions provided are both necessary and sufficient for describing the action.

Of course, if we don't make that assumption, then the only way to deal with this is to allow for a lot of usual cases, but also allow for unusual and exceptional cases using some notion of abnormality. All of this requires notions of non-monotonicity.

There is a general view that non-monotonicity is not needed and thereby a wasted effort, or already solved. Neal Parikh's tweet, for example, has a view that non-monotonic reasoning is a wasted effort.

This seems to be a fairly superficial remark because there is no evidence that the problems identified in the non-monotonic reasoning community have been successfully

addressed using any technique. It is, of course, true that many machine learning models when trained on existing data, they can produce the same patterns and report on the abnormalities seen in that data [55,67]; however there is no general mechanism to deal with such ideas. Moreover, non-monotonic logic reasoning has given us notions like stable model semantics [34] which now powers very recent approaches to neuro-symbolic AI [90]. Interestingly, non-monotonic semantics can also allow us to capture cycles in graphs [23], which ordinarily requires recursion using, say, second-order logic.

7. Differentiability

In so much as recent approaches to machine learning can be summarized, differentiability is a key concept. But it is widely held that logic cannot play a role in this. For example, Turing Award winner Yann LeCunn quips [59]:

How can machine reason and plan in ways that are compatible with gradient-based learning?

Our best approaches to learning rely on estimating and using the gradient of a loss, which can only be performed with differentiable architectures and is difficult to reconcile with logic-based symbolic reasoning.

But as indicated by the points above, this view is simply uniformed. In fact, as already shown, probabilities as well as real arithmetic can be mapped on to logical symbols and this means that probabilistic and real-valued semantics seem to naturally lead to differentiability. There is a longstanding view that logic and probability are compatible with each other [18]. And if you look at recent considerations on reasoning about probabilities using logic-based solvers, such as those based on weighted model counting [3], it is very hard to argue for a dichotomy. Consider the position in Guy Van Den Broeck's tweet, for example.

At this point, there are plenty of approaches that explicitly use logic for the training of neural networks, especially in the context of regularization and differentiability. This started with the work of UCLA's Semantic Loss [32] and KU Leuven's DeepProbLog [66], both of which adjust the loss function of the deep learning model based on a logical encoding of the constraints and program, respectively. There are recent approaches that are based on either proposition logic with possible worlds or real-valued propositions [45]. Providing arithmetic constraints to the training of deep learning networks, and ensuring consistency with the provided domain knowledge is also an important problem for areas like physics [85].

However, it would be remiss not to point out that just because differentiability seems to be an important ingredient in the training of machine learning models, it does not mean that we expect every scientist in the area of logic to play game. There is still deep and rigorous work to be done on the integration of logical querying (e.g., computational effort needed to evaluate queries on a large knowledge basis [64]) and probability [7], for example. On the representation side, there are important issues to grapple with, such as languages that reason about the cardinality of reals and the consistency of the axiomatization of natural numbers [62]. Moreover, modal logics like temporal logics and dynamic logics become useful for deep learning-based endeavors as we navigate to more open-ended problems. For example, in [48], temporal logic formulas are used to train deep reinforcement learning agents. In [83], large language models are used to learn dynamic epistemic properties [10]. And finally, in [49], a temporally extended semantic loss function is considered.

8. What about "human-like" semantic definitions?

The most well-studied semantics, or perhaps more accurately, the most widely-used semantics in computer science, remains classical [17]. That is, atoms are accorded values of either 0 or 1, and so formulas become Boolean functions. If modalities are introduced, such as time and actions [30], then we look at sequences of models: either a linear sequence or a tree-like sequence [78], for example.

But as mentioned above, there are also approaches where a degree of truth is accorded to formulas, either by allowing the atoms themselves to have non-binary values [91] or by according probabilities or other kinds of measures for complex formulas [26].

All of these notions are explored by establishing some kind of well-definedness, and logicians explore the implications of those conditions. For example, intuitionistic logic looks to weaken material implication [27]. Non-classical belief logics control the proof-depth of logical reasoners [63], and so on.

There is an informal argument often made that perhaps these systems are too precise. Perhaps on training neural networks with real-world observations, they might exhibit more human-like reasoning capabilities. But the evidence for this has not yet been established. Moreover, is such a feature desirable? Consider, for example, an informal inquiry by Oxford philosopher Luciano Floridi asks:[5]

what's the name of the daughter of Laura's mother?

ChatGPT (even GPT-4, the latest and best) comes back with something to the effect of: "I'm sorry, I don't have personal information about specific individuals."

Analogously, the computer scientist Vincent Conitzer tried out the below puzzle:[6]

I put one apple in a bag. Then I put either one pear or two bananas in the bag. Then I remove one piece of fruit. Then I remove all the remaining pieces of fruit from the bag.

Based on what I told you, is it possible to know exactly how many pieces of fruit remain in the bag?

Here's what we get:

[5] https://twitter.com/Floridi/status/1635951391968567296?s=20

[6] https://twitter.com/conitzer/status/1636156048347111425?s=20

Based on the information provided, it is not possible to know exactly how many pieces of fruit remain in the bag after removing one piece of fruit and then removing all remaining pieces. This is because there are two possible scenarios, but we don't have enough information to determine which one took place:

1. *One apple and one pear were in the bag. If you removed one piece of fruit, there would be one piece left. After removing all remaining pieces, there would be 0 pieces left.*
2. *One apple and two bananas were in the bag. If you removed one piece of fruit, there would be two pieces left. After removing all remaining pieces, there would also be 0 pieces left.*

In both scenarios, after removing one piece of fruit and then removing all remaining pieces, there are no pieces of fruit left in the bag.

But apparently in a follow-up reply to that tweet, GPT-4 seems to get it correct.

What can we make of all of this? These models are clearly not reliable, or consistent [50]. Perhaps this what prompted Stephen Wolfram to consider a systematic integration of ChatGPT and Wolfram Alpha.[7]

There are, of course, plenty of other examples for integrating LLMs and other symbolic solvers, and this is an exciting area for the future of logic, as seen in neuro-symbolic AI [92]. However, the "native" capabilities are clearly limited. An intriguing hypothesis put forward by eminent AI scientist Subbarao Kambhampati is this:[8]

I think many of the claims about LLM's reasoning capabilities miss the point that LLM's are not just trained on "facts" but also, quite often, the deductive closure of those facts. Thus reasoning becomes (approximate) retrieval.

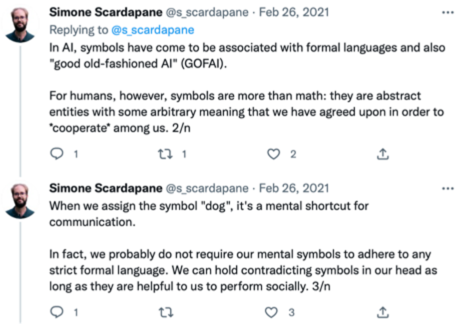
If this is the case, these models do not reason at all, but simply see for patterns of conclusions, which might limit, say, the number of inference steps, or how involved the reasoning is. However, what about consoling ourselves with the idea that the training data might include all such deductions, in which case, LLMs might be sufficient? Sadly, in a critical examination [93], it is shown that LLMs likely pick up unnecessary statistical features of logical inputs, and their logical reasoning abilities may not be sound across different distributions on background theories, and thus, likely not complete.

Nonetheless, it is worth noting that, strictly speaking, we do not require that the semantics be given by humans, or that they be hand-written. Symbols can obtained from low-level data (via symbol grounding). The use of symbols in AI also does not mean that folks using symbolic logic assume humans manipulate symbols in their head.

Consider the observation by Simone Scardapane in his tweet, for example. So the semantics of connectives and formulas may be built up from context, social environment and language use. But of course, there is also work in logic on paraconsistency and inconsistency [14]!

Ultimately, we have a whole range of language choices and semantic choices to work with. But does it still make sense to bother with classical semantics? We would argue yes. For example, (a) it is a well-defined mathematical model, (b) with the use of modalities and/or non-classical semantics, we can relate different systems, (c) we do not really know what semantics best approximates human reasoning, (d) we may not want mathematical truths that play fast and loose with inevitable conclusions just because we think humans might have some cognitive biases and entertain inconsistencies, and (e) the

[7] https://writings.stephenwolfram.com/2023/03/chatgpt-gets-its-wolfram-superpowers/
[8] https://twitter.com/rao2z/status/1666294366720360449?s=20



science of robust AI is still out there. Let us investigate the properties of well-defined objects with patience and rigour.

9. Symbols and deep learning can be complementary

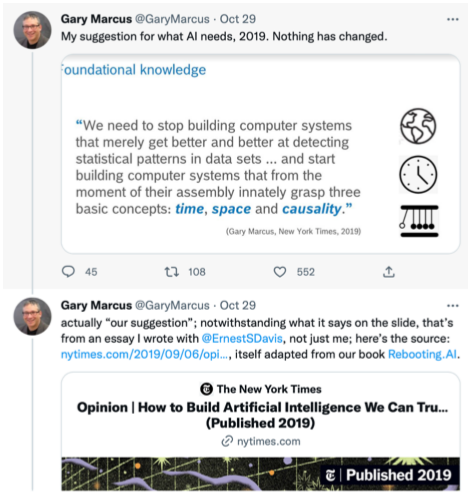
As already hinted above, symbolic logic can play an important role in training deep learning models but also in integrating reasoning as a post-hoc process or as a metalinguistic paradigm. That is, we can ensure that the distribution of the trained network respects domain constraints [45]. We can extract rules from trained models and reason about them outside the framework of the network [75]. Or we can use the outputs of the network as inputs to a computational paradigm such as probabilistic programming [66]. There is very interesting work on the semantics of programs that inherently support some notion of differentiation [1]. This is an object of intense theoretical study that can have consequences on the types of distributions that are expressible in programming languages [84]. So, this theory has far-reaching effects on what type of probabilistic models can be modelled effectively.

Overall, symbols and DL need not compete with each other: they can be complementary. Perhaps the most representative example of this is the burgeoning field of neuro-symbolic AI [33], which has come to encompass things like neural program induction [58], neural theorem improvers, and differentiable logics [92].

10. Symbolic logic as meta-theory

An argument made previously [9] is that symbolic logic can be used to formalize notions currently out of the purview of standard machine learning. These include things like the semantics of involved probabilistic programming languages [84] and understanding the limits of differentiable logics, but it can also pertain to a range of more exotic topics. For example, it is very common in AI applications these days to require frameworks for multi-agent reasoning [2]. In explainable AI [38], in particular, we might require that the robot holds beliefs about the human agent [52]. Modal logics study such phenomena.

Moreover, complex AI systems are not going to be purely based on providing predictions. They will involve search, constraint reasoning, and planning [80]. This has necessitated new approaches for compositionality [84] and modularity [87]. In some recent



work, for instance, it was noted that weighted model counting [35], which provides the foundation for mapping Bayesian inference to SAT solvers, can be upgraded to also reason about maximization and minimization of properties [54], leading to languages where a number of different AI sub-areas can be unified [11].

11. High-level knowledge

At a number of recent AI events, Daniel Kahneman has been invited to discuss his very famous distinction of the so-called system 1 versus system 2 type cognition in humans.[9] This is owing to the fact that AI scientists, for a very long time, have been deliberating on the appropriate way to abstract away low-level perception data with high-level concept knowledge, perhaps going back to Shakey [57]. Providing mechanisms as well as formal semantics for abstraction remains a topic of theoretical interest even today [46].

In the particular case of deep learning systems, an obvious agenda point is how do we define abstract concepts, either extracted directly from these systems or defined externally to coordinate and interoperate with these systems.

Consider the tweet by Gary Marcus, for example. Indeed, it is widely acknowledged that concepts such as time, abstraction, and causality will play a key role in designing an AI that has a world model that is rich enough to be interpreted in a way we would find reasonable [16].

Although there is some work on providing a causal semantics to deep learning systems [65], it is still in the early years and studied in a limited way. In contrast, we have very well-studied models of time [76] and causality with symbolic calculi [78,40,43]. It seems irresponsible to not utilise these advances simply because they are symbolic, and hence deemed "old-fashioned."

On a related note, symbols can be used as abstract identifiers for human-in-the-loop systems [52], and/or interactive machine learning especially when you have non-expert

[9] https://vimeo.com/390814190

stakeholders engaging with predictors trained on high-dimensional data. See the position in Subbrao Kambhampati's tweet, for example.

12. Symbolic logic can instantiate new methods of inference

One remark that has been left implicit and referred to in the above discussion is the idea that precisely because of the close relationship between logic and probability [18], it is possible to use logic-based solvers for doing probabilistic reasoning.

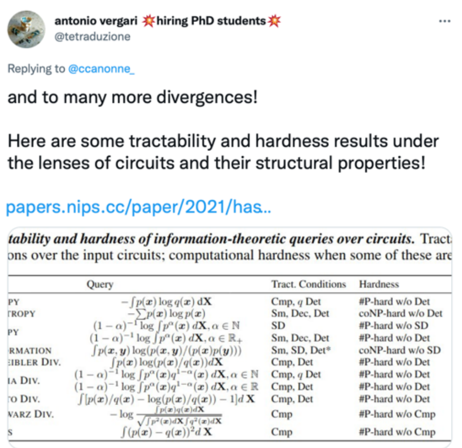
This is primarily seen in weighted model counting [35], which is an extension of SAT solving to identify all possible satisfying assignments [3]. There is also an extension of this strategy to deal with continuous properties via so-called weighted model integration [13]. Interestingly, because it was observed that certain types of these more involved operations beyond SAT exist, it is attractive to construct a data structure for the logical formula at hand. This led to the area of knowledge compilation [22]. A key result of this sub-area is the explicit construction of these data structures or circuits [21].

It turns out that circuits provide a new way of doing inference with probabilistic models with the following properties: you pay a one-time cost of compiling the representation, such as a Bayesian network, into such a circuit, and then every query afterwards can be done in time polynomial in the size of the circuit. See the tweets by Kristian Kersting and Antonio Vergari, for example.

Overall, this is a new and attractive way of doing probabilistic reasoning and has even led to new approaches to inference in probabilistic programming [47]. These circuits can also be learned from data [61], which again leads to a new paradigm of learning probabilistic models. Finally, there is also some work on approximate inference with weighted model counting [19], which provides more certifiable guarantees than variational inference.

13. Symbols do not always need a logic

It is common to come up with a new programming language, defining only the interpreter and the compiler, without an explicit semantics in the sense of formal logic [28]. In this regard, there is a lot of work in Neuro-symbolic AI on learning programs [58] or other kinds of symbolic artifacts without an explicit logic. This is possible, at least until the area matures.



It should be noted, however, that without a clear specification of how compositions and combinations of expressions should be interpreted and evaluated, this type of use of symbols is somewhat ad hoc. However, while the area is still growing, we expect to see more of such constructions. We would then need to consider whether a semantics can be designed post-hoc or synthesized in some fashion by mapping it back to an existing logic [6]. Such a move is especially desirable if we want to check for the internal consistency of the impromptu programming language.

14. Wrapping up

In this article, we looked at a few of the misunderstandings that arise when considering the relevance and use of symbolic AI in modern AI systems. We have only covered some of the ground that we feel frequently comes up, but we have not discussed other relevant issues.

For example, one area where symbolic logic is widely used in many stochastic systems [20] is the verification of safety properties [82], and even ethical principles [24]. This is a vast area, spanning multiple conferences with many scientific bodies dedicated to it. Indeed, in this article, we have mostly focused on the concept of semantics and its importance.

As we discussed, one area where concerns about the use of logic seem to disappear is neuro-symbolic AI. Neuro-symbolic AI holds a lot of promise because it can offer interesting ways to combine symbolic logic and deep learning, and build on the success of both. And like the maxim: "the whole is greater than the sum of the parts," such an integration may not simply be the communication of outputs in a divorced way, but could involve a deeper type of synthesis [44]. Some approaches have dealt with loss functions, while others have focused on post-hoc logical reasoning or extracting rules from networks. All of these approaches are interesting in their own right.

There is also a trade-off between the complexity and level of detail of the logical knowledge and how effectively it can integrate with a learning system. For example, papers focusing on loss functions typically deal with smaller-sized formulas and constraints [45], while works exploring the integration of learning with knowledge graphs often consider ontologies with more than a hundred or even a thousand nodes [73]. There is a range of possibilities, and it may be challenging to determine the correct approach. Perhaps there is no one-size-fits-all solution, and the combination of logic and deep learning can vary depending on the application.

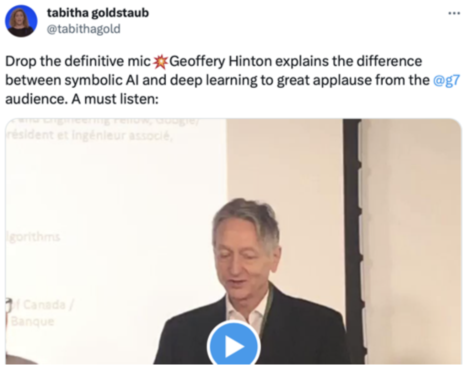
Regardless of the specific approach, it is clear that we need to understand the principles of logical languages and semantics to ensure that resulting mathematical objects are well-defined with desired properties. This appreciation is essential for both theoretical exploration and practical applications.

It should be noted that there is a case to be made for expressive representations. For example, some might come away feeling that the best way to approach the future of neuro-symbolic AI is to focus on very limited languages. But such an view may not be fruitful in the long term. For example, it is widely understood that first-order is useful for generalized assertions [60], and modal logics for time and multi-agent beliefs [30]. In general, the language is critical for capturing the domain correctly. In a statement remarkably similar in spirit, Judea Pearl writes [74]:

This is why you will find me emphasizing and reemphasizing notation, language, vocabulary and grammar. For example, I obsess over whether we can express a certain claim in a given language and whether one claim follows from others. My emphasis on language also comes from a deep conviction that language shapes our thoughts. You cannot answer a question that you cannot ask, and cannot ask a question that you have no words for.

And, as with Pearl and the knowledge representation community more generally, we will identify with "representation first, acquisition second."

To sum up, there is a lot to be gained by the relating the mathematical foundations of logic and deep learning. And the benefit is not purely for the logician, but also for the deep learning researcher who wants to think more broadly than prediction with big data. Scientists working on logic and language should be allowed to work on problems that seem scientifically relevant without necessarily linking up to or competing with whatever the zeitgeist of machine learning is.



We should, of course, celebrate successes — its neither an accident nor misplaced opportunism that logic/programming language folks are interested in learning and are eager to understand the latest and best [37]. Moreover, what combination of logic and/or learning we need for general AI is not well-understood yet. We cannot point to the exact approach or balance of innateness vs tabula rasa we need for general AI, because we simply do not know. We can only loosely articulate requirements (e.g., correct, fair and safe by design), capabilities (e.g., ability to reason about causality, time and space models) and desiderata.

Experts can get excited about what works – the success of AlphaGo, as well as large language models, is kind of a success for AI, although of course it opens up questions about generality and correctness. However, there is no need to dismiss other approaches. Indeed, what we do not need are folks (especially Turing award winners like Geoff Hinton) mocking other areas (the gasoline analogy), or others, such as Tabitha Goldstaub, with 16000+ followers sharing derision with conviction.[10]

Acknowledgements

This article builds and extends arguments previously presented in the author's blog: medium.com/@vaishakbelle. The author is grateful to the Royal Society for funding his

[10] https://twitter.com/tabithagold/status/1070736319901519876?s=20

University Research Fellowship.

In this paper, we have freely used screenshots of tweets, but have not reached out to the users for their permission because we have assumed these tweets are public, and hence they can be cited like a website.

Our focus on views of AI from social media may seem unusual; however, Twitter has turned into a dominant space for public statements by leading experts. Although they are not as iron-clad as peer-reviewed position papers, we view them as statements all the same: positions expressed for peers. Of course, it is possible some may wish to retract statements made in tweets, saying that the space limitations forced them to make an informal remark that could be easily misunderstood, or that there was implicit irony. This is why we have included screenshots in most instances, or otherwise linked to them, and admit that we are taking those statements at face value and apologise if we have misrepresented an individual's position.

References

[1] M. Abadi and G. D. Plotkin. A simple differentiable programming language. *Proceedings of the ACM on Programming Languages*, 4(POPL):1–28, 2019.

[2] S. V. Albrecht and P. Stone. Autonomous agents modelling other agents: A comprehensive survey and open problems. *Artificial Intelligence*, 258:66–95, 2018.

[3] F. Bacchus, S. Dalmao, and T. Pitassi. Solving #SAT and Bayesian inference with backtracking search. *J. Artif. Intell. Res. (JAIR)*, 34:391–442, 2009.

[4] C. Baier and J.-P. Katoen. *Principles of Model Checking*. The MIT Press, 2008.

[5] C. Barrett, R. Sebastiani, S. A. Seshia, and C. Tinelli. Satisfiability modulo theories. In *Handbook of Satisfiability*, chapter 26, pages 825–885. IOS Press, 2009.

[6] S. Bartha, J. Cheney, and V. Belle. One down, 699 to go: or, synthesising compositional desugarings. *Proceedings of the ACM on Programming Languages*, 5(OOPSLA):1–29, 2021.

[7] P. Beame, G. Van den Broeck, E. Gribkoff, and D. Suciu. Symmetric weighted first-order model counting. In *PODS*, pages 313–328. ACM, 2015.

[8] V. Belle. Symbolic logic meets machine learning: A brief survey in infinite domains. In *International Conference on Scalable Uncertainty Management*, pages 3–16. Springer, 2020.

[9] V. Belle. Logic meets learning: From aristotle to neural networks. In *Neuro-Symbolic Artificial Intelligence: The State of the Art*, pages 78–102. IOS Press, 2021.

[10] V. Belle, T. Bolander, A. Herzig, and B. Nebel. Epistemic planning: Perspectives on the special issue, 2022.

[11] V. Belle and L. De Raedt. Semiring programming: A semantic framework for generalized sum product problems. *International Journal of Approximate Reasoning*, 126:181–201, 2020.

[12] V. Belle and G. Lakemeyer. Reasoning about probabilities in unbounded first-order dynamical domains. In *IJCAI*, 2017.

[13] V. Belle, A. Passerini, and G. Van den Broeck. Probabilistic inference in hybrid domains by weighted model integration. In *IJCAI*, 2015.

[14] H. A. Blair and V. Subrahmanian. Paraconsistent logic programming. *Theoretical computer science*, 68(2):135–154, 1989.

[15] R. Brachman and H. Levesque. *Knowledge representation and reasoning*. Morgan Kaufmann Pub, 2004.

[16] R. J. Brachman and H. J. Levesque. *Machines like us: toward AI with common sense*. MIT Press, 2022.

[17] A. R. Bradley and Z. Manna. *The calculus of computation: decision procedures with applications to verification*. Springer Science & Business Media, 2007.

[18] R. Carnap. *Logical foundations of probability*. Routledge and Kegan Paul London, 1951.

[19] S. Chakraborty, D. J. Fremont, K. S. Meel, S. A. Seshia, and M. Y. Vardi. Distribution-aware sampling and weighted model counting for sat. *AAAI*, 2014.

[20] T. Chen, M. Diciolla, M. Z. Kwiatkowska, and A. Mereacre. A simulink hybrid heart model for quantitative verification of cardiac pacemakers. In *Proceedings of the 16th international conference on Hy-*

brid systems: computation and control, HSCC 2013, April 8-11, 2013, Philadelphia, PA, USA, pages 131–136, 2013.

[21] A. Choi and A. Darwiche. On relaxing determinism in arithmetic circuits. *arXiv preprint arXiv:1708.06846*, 2017.

[22] A. Darwiche, P. Marquis, D. Suciu, and S. Szeider. Recent trends in knowledge compilation (dagstuhl seminar 17381). In *Dagstuhl Reports*, volume 7. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2018.

[23] M. Denecker, M. Bruynooghe, and V. Marek. Logic programming revisited: Logic programs as inductive definitions. *ACM Transactions on Computational Logic*, 2(4):623–654, 2001.

[24] L. Dennis, M. Fisher, M. Slavkovik, and M. Webster. Formal verification of ethical choices in autonomous systems. *Robotics and Autonomous Systems*, 77:1–14, 2016.

[25] K. R. Dienes. String theory and the path to unification: A review of recent developments. *Physics Reports*, 287(6):447–525, 1997.

[26] D. Dubois and H. Prade. Representation and combination of uncertainty with belief functions and possibility measures. *Computational Intelligence*, 4(3):244–264, 1988.

[27] M. Dummett. The philosophical basis of intuitionistic logic. In *Studies in Logic and the Foundations of Mathematics*, volume 80, pages 5–40. Elsevier, 1975.

[28] K. Ellis, A. Albright, A. Solar-Lezama, J. B. Tenenbaum, and T. J. O'Donnell. Synthesizing theories of human language with bayesian program induction. *Nature communications*, 13(1):5024, 2022.

[29] H. Enderton. *A mathematical introduction to logic*. Academic press New York, 1972.

[30] R. Fagin, J. Y. Halpern, Y. Moses, and M. Y. Vardi. *Reasoning About Knowledge*. MIT Press, 1995.

[31] H. Gaifman. Concerning measures in first order calculi. *Israel Journal of Mathematics*, 2(1):1–18, 1964.

[32] K. Gajowniczek, Y. Liang, T. Friedman, T. Zabkowski, and G. Van den Broeck. Semantic and generalized entropy loss functions for semi-supervised deep learning. *Entropy*, 22(3):334, 2020.

[33] A. S. d. Garcez, K. Broda, D. M. Gabbay, et al. *Neural-symbolic learning systems: foundations and applications*. Springer Science & Business Media, 2002.

[34] M. Gelfond and V. Lifschitz. The stable model semantics for logic programming. In *ICLP*, pages 1070–1080, 1988.

[35] C. P. Gomes, A. Sabharwal, and B. Selman. Model counting. In *Handbook of Satisfiability*. IOS Press, 2009.

[36] I. J. Goodfellow, J. Shlens, and C. Szegedy. Explaining and harnessing adversarial examples. *arXiv preprint arXiv:1412.6572*, 2014.

[37] S. Gulwani. Dimensions in program synthesis. In *PPDP*, pages 13–24. ACM, 2010.

[38] D. Gunning. Explainable artificial intelligence (xai). Technical report, DARPA/I20, 2016.

[39] J. Y. Halpern. *Reasoning about Uncertainty*. MIT Press, 2003.

[40] J. Y. Halpern. *Actual causality*. MiT Press, 2016.

[41] J. Y. Halpern. *Reasoning about uncertainty*. MIT press, 2017.

[42] J. Y. Halpern and M. Y. Vardi. Model checking vs. theorem proving: A manifesto. In J. F. Allen, R. Fikes, and E. Sandewall, editors, *KR*, pages 325–334. Morgan Kaufmann, 1991.

[43] C. Hitchcock. Causality: Models, reasoning and inference, 2001.

[44] P. Hitzler. Neuro-symbolic artificial intelligence: The state of the art. 2022.

[45] N. Hoernle, R. M. Karampatsis, V. Belle, and K. Gal. Multiplexnet: Towards fully satisfied logical constraints in neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 5700–5709, 2022.

[46] T. Hofmann and V. Belle. Abstracting noisy robot programs. *AAMAS*, 2023.

[47] S. Holtzen, G. Van den Broeck, and T. Millstein. Scaling exact inference for discrete probabilistic programs. *Proceedings of the ACM on Programming Languages*, 4(OOPSLA):1–31, 2020.

[48] R. T. Icarte, T. Q. Klassen, R. Valenzano, and S. A. McIlraith. Reward machines: Exploiting reward function structure in reinforcement learning. *Journal of Artificial Intelligence Research*, 73:173–208, 2022.

[49] C. Innes and S. Ramamoorthy. Elaborating on learned demonstrations with temporal logic specifications. *arXiv preprint arXiv:2002.00784*, 2020.

[50] M. Jang and T. Lukasiewicz. Consistency analysis of chatgpt. *arXiv preprint arXiv:2303.06273*, 2023.

[51] D. Jovanović and L. De Moura. Solving non-linear arithmetic. *ACM Communications in Computer Algebra*, 46(3/4):104–105, 2013.

[52] S. Kambhampati. Challenges of human-aware ai systems. *AI Magazine*, 41(3), 2020.

[53] M. Katz. Łukasiewicz logic and the foundations of measurement. *Studia logica*, 40:209–225, 1981.

[54] A. Kimmig, G. V. den Broeck, and L. D. Raedt. Algebraic model counting. *CoRR*, abs/1211.4475, 2012.

[55] V. Kocijan, E. Davis, T. Lukasiewicz, G. Marcus, and L. Morgenstern. The defeat of the winograd schema challenge. *arXiv preprint arXiv:2201.02387*, 2022.

[56] S. Kripke. A completeness theorem in modal logic. *Journal of Symbolic Logic*, 24(1):1–14, 1959.

[57] B. Kuipers, E. A. Feigenbaum, P. E. Hart, and N. J. Nilsson. Shakey: from conception to history. *Ai Magazine*, 38(1):88–103, 2017.

[58] B. M. Lake, R. Salakhutdinov, and J. B. Tenenbaum. Human-level concept learning through probabilistic program induction. *Science*, 350(6266):1332–1338, 2015.

[59] Y. LeCun. A path towards autonomous machine intelligence version 0.9. 2, 2022-06-27. *Open Review*, 62, 2022.

[60] H. J. Levesque. *Thinking as Computation: A First Course*. MIT Press, 2012.

[61] Y. Liang, J. Bekker, and G. Van den Broeck. Learning the structure of probabilistic sentential decision diagrams. In *Proceedings of the 33rd Conference on Uncertainty in Artificial Intelligence*, pages 134–145, 2017.

[62] D. Liu, Q. Feng, V. Belle, and G. Lakemeyer. Concerning measures in a first-order logic with actions and meta-beliefs. In *20th International Conference on Principles of Knowledge Representation and Reasoning*, 2023.

[63] Y. Liu, G. Lakemeyer, and H. J. Levesque. A logic of limited belief for reasoning with disjunctive information. In *KR*, pages 587–597, 2004.

[64] Y. Liu and H. J. Levesque. Tractable reasoning in first-order knowledge bases with disjunctive information. In *Proc. AAAI*, pages 639–644, 2005.

[65] Y. Luo, J. Peng, and J. Ma. When causal inference meets deep learning. *Nature Machine Intelligence*, 2(8):426–427, 2020.

[66] R. Manhaeve, S. Dumancic, A. Kimmig, T. Demeester, and L. De Raedt. Deepproblog: Neural probabilistic logic programming. *Advances in Neural Information Processing Systems*, 31, 2018.

[67] G. Marcus. Am i human? *Scientific American*, 316(3):58–63, 2017.

[68] J. McCarthy. Programs with common sense. In *Proceedings of the Symposium on the Mechanization of Thought Processes*, National Physiology Lab, Teddington, England, 1958.

[69] J. McCarthy and P. J. Hayes. Some philosophical problems from the standpoint of artificial intelligence. In *Machine Intelligence*, pages 463–502, 1969.

[70] B. Milch, B. Marthi, S. J. Russell, D. Sontag, D. L. Ong, and A. Kolobov. BLOG: Probabilistic models with unknown objects. In *Proc. IJCAI*, pages 1352–1359, 2005.

[71] P. P. Mitra. The circuit architecture of whole brains at the mesoscopic scale. *Neuron*, 83(6):1273–1283, 2014.

[72] L. Morgenstern and S. A. McIlraith. John McCarthy's legacy. *Artificial Intelligence*, 175(1):1 – 24, 2011.

[73] F. Niu, C. Zhang, C. Ré, and J. W. Shavlik. Deepdive: Web-scale knowledge-base construction using statistical learning and inference. *VLDS*, 12:25–28, 2012.

[74] J. Pearl and D. Mackenzie. The book of why.

[75] C. Persia and A. Ozaki. Extracting rules from neural networks with partial interpretations. *arXiv preprint arXiv:2204.00360*, 2022.

[76] A. Prior. *Past, present and future*. Oxford University Press, 1967.

[77] L. D. Raedt, A. Kimmig, and H. Toivonen. Problog: A probabilistic prolog and its application in link discovery. In *Proc. IJCAI*, pages 2462–2467, 2007.

[78] R. Reiter. *Knowledge in action: logical foundations for specifying and implementing dynamical systems*. MIT Press, 2001.

[79] M. Richardson and P. Domingos. Markov logic networks. *Machine learning*, 62(1):107–136, 2006.

[80] S. Russell and P. Norvig. *Artificial Intelligence: A Modern Approach*. Prentice-Hall, Englewood Cliffs, NJ, 2nd edition edition, 2003.

[81] S. Sardina and Y. Lespérance. Golog speaks the bdi language. In *Programming Multi-Agent Systems*, volume 5919 of *LNCS*, pages 82–99. Springer Berlin Heidelberg, 2010.

[82] A. Shih, A. Darwiche, and A. Choi. Verifying binarized neural networks by angluin-style learning. In *Theory and Applications of Satisfiability Testing–SAT 2019: 22nd International Conference, SAT 2019, Lisbon, Portugal, July 9–12, 2019, Proceedings 22*, pages 354–370. Springer, 2019.

[83] D. Sileo and A. Lernould. Mindgames: Targeting theory of mind in large language models with dynamic

epistemic modal logic. *arXiv preprint arXiv:2305.03353*, 2023.

[84] S. Staton, H. Yang, F. Wood, C. Heunen, and O. Kammar. Semantics for probabilistic programming: higher-order functions, continuous distributions, and soft constraints. In *Proceedings of the 31st Annual ACM/IEEE Symposium on Logic in Computer Science*, pages 525–534, 2016.

[85] R. Stewart and S. Ermon. Label-free supervision of neural networks with physics and domain knowledge. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31, 2017.

[86] L. W. Swanson. *Brain architecture: understanding the basic plan*. Oxford University Press, USA, 2012.

[87] E. Ternovska and D. G. Mitchell. Declarative programming of search problems with built-in arithmetic. In *Proc. IJCAI*, pages 942–947, 2009.

[88] S. Thrun, W. Burgard, and D. Fox. *Probabilistic Robotics*. MIT Press, 2005.

[89] J. Tooby, L. Cosmides, and H. C. Barrett. Resolving the debate on innate ideas. *The innate mind: Structure and content*, pages 305–337, 2005.

[90] Z. Yang, A. Ishay, and J. Lee. Neurasp: Embracing neural networks into answer set programming. In *29th International Joint Conference on Artificial Intelligence (IJCAI 2020)*, 2020.

[91] L. A. Zadeh. Fuzzy sets. *Information and control*, 8(3):338–353, 1965.

[92] H. Zhang, J. Huang, Z. Li, M. Naik, and E. Xing. Improved logical reasoning of language models via differentiable symbolic programming. *arXiv preprint arXiv:2305.03742*, 2023.

[93] H. Zhang, L. H. Li, T. Meng, K.-W. Chang, and G. V. d. Broeck. On the paradox of learning to reason from data. *arXiv preprint arXiv:2205.11502*, 2022.