From Cogwheels to Concepts: A Perspective on the Interplay of Robotics and Ontology

Mihai Pomarlan [a], Robin Nolte [b], Alberto Olivares-Alarcos [c], Mohammed Diab [d], Daniel Beßler [e] and Stefano Borgo [f]

[a] *Applied Linguistics, University of Bremen, Germany.*
E-mail: pomarlan@uni-bremen.de
[b] *Digital Media Lab, University of Bremen, Germany.*
E-mail: nolte@uni-bremen.de
[c] *Institut de Robòtica i Informàtica Industrial, CSIC-UPC, Spain.*
E-mail: aolivares@iri.upc.edu
[d] *University of Plymouth, Devon PL4 8AA, United Kingdom.*
E-mail: mohammed.diab@upc.edu
[e] *Institute for Artificial Intelligence, University of Bremen, Germany.*
E-mail: danielb@uni-bremen.de
[f] *Laboratory for Applied Ontology, Institute for Cognitive Sciences and Technologies, National Research Council of Italy, Italy.*
E-mail: stefano.borgo@cnr.it

Abstract. The hard problems of robotics occur at the interface of an artificial agent and its surrounding physical world: perception and action. These are often regarded as tasks requiring tacit, unformalized and possibly unformalizable knowledge, making ontology engineering of apparently little value for robotics. We argue instead that an important role has appeared for ontology in current developments in robotics. Recent AI techniques such as foundation models have revealed the potential of explicit knowledge to be useful at all parts of a robotic system, but these techniques will require assistance from formal, verifiable methods to produce trustworthy systems. Further, robotics can contribute to the development of ontology engineering. As robotic agents become more autonomously capable, questions about the meaning of agency and responsibility become increasingly pressing, and will require careful consensus building across a variety of stakeholder groups. Finally, robotics may be a challenging test ground for the philosophical assumptions underlying foundational ontologies, and can thus assist the ontology engineering community in better understanding the consequences of ontological modelling decisions more generally.

Keywords: knowledge representation for robotics, foundation models, robot ethics, empirical philosophy

1. Introduction

In 1988, Hans Moravec observed that cognitive tasks, such as playing checkers, are easier to implement computationally than sensorimotor tasks such as recognizing checker pieces and moving them across the board (Moravec, 1988). This observation became known as Moravec's paradox and remains relevant today: while development in technologies such as generative AI and foundation models is explosive, developments in autonomous robots are lagging. AI is a master of data but flounders when it has to interact with the real-time physical world.

This "paradox" is related to another "paradox" described by Polanyi, summarized in his dictum "we can know more than we can tell" (Polanyi, 1966). Polanyi contrasts explicit knowledge – knowledge which we have verbalized in statements of fact, or rules, or procedures – to tacit knowledge – which we can believe we possess because of our intuitive ability for various tasks, even if we cannot verbalize such knowledge. In such an understanding, an algorithm for addition, or solving a Rubik's cube, are explicit knowledge. The skills of moving a pen on paper or manipulating a Rubik's cube are tacit – while one could describe these skills in words, these descriptions are insufficient to allow implementing them on something resembling a human hand.

Thus, one way to look at Moravec's paradox is that the kinds of tasks that are essential for an autonomous robot to interface with the world – tasks such as perception and manipulation – are tasks for which we humans use a significant amount of tacit knowledge. Since it is hard for us to articulate how we do the basic things we do to cope with the world, our robots will have to discover by themselves via machine learning approaches . To the extent that progress in robotics has occurred in recent years – and it has occurred – it is attributable to the collection and use of more data for robots to learn from. Apart from, perhaps, providing a sophisticated data model, there is apparently nothing for ontology engineering to contribute to the hard problems in robotics. . We disagree with such a view however.

Firstly, explicit knowledge – that has been formally represented and made available for verifiable reasoning – is key to making use of tacit knowledge in a generalizable and trustworthy manner. . This argument also finds support from a somewhat surprising source – the new wonderchild of machine learning, foundation models. We begin from an observation about trends in AI research dubbed "Avenging Polanyi's revenge" (Kambhampati et al., 2024a): a reversal of the trend towards interest in tacit knowledge tasks for AI researchers. Post large language models (LLMs), the focus is again on classically symbolic tasks such as reasoning and planning. Can LLMs perform them? Can they help in performing them? One reason for this swing back is that LLMs provide what at first glance looks like explicit knowledge at a very large scale. LLMs have seen everything – or almost everything on the internet at least – and can "approximately recall" (Kambhampati et al., 2024a) from this massive knowledge base. The implications of this approximate recall are difficult to anticipate, and thus there is a flurry of research into how to make use of LLMs in all manner of applications, including robotics. This goes beyond just using LLMs – and related technologies such as multi-modal (also known as foundation) models – as a tool for user interaction, but rather reaches into ever more finegrained details of robot behavior generation.

However, the key point about approximate recall is that it is approximate. Approximate is good enough for an entertaining chat session, potentially deadly when steering a robot, and it is here, we argue, that techniques from ontology engineering will help tame what is still a very strange and unintuitive technology. Ontology engineering thus can help robotics in establishing methods to interact with the newest and so far best tool we have to tackle open-ended descriptions of situations that a robot may encounter.

Secondly, robotics (especially autonomous robotics) is a domain in which difficult but abstract-seeming philosophical questions get recast as practically relevant in the near future. For example, can we even be sure that autonomous non-human agents will form ontologies compatible with ours? How would we go about ensuring alignment?

Or, to consider other discussions that are increasingly relevant – what should count as an autonomous agent? Assuming, for the sake of argument, that artificial, autonomous agents are possible, how do reliability (on a machine) and trust (on an agent) combine in the case of autonomous agents (Biccheri et al., 2023)? When can we trust an agent with autonomy? What happens when an autonomous artificial agent makes a mistake that leads to human fatalities? Who is an autonomous agent supposed to be loyal

to and to what extent? Can we trust an autonomous agent with the power to purposefully harm human beings? What does an agent need to be like before *we* incur moral obligations to *it*, and how would we know?

It would be tempting to dismiss such problems as falling under the purview of lawmakers, activist groups, roboticists, and/or philosophers, but this would be missing an opportunity to do what ontology engineers are supposed to be doing, i.e. help building a shared consensus among stakeholders about the conceptualization of their domain of activity (Neuhaus, 2023). (Re)building such a shared consensus is needed because the notions used in contemporary ethics do not quite fit the changing technical landscape. Either the concept of moral agency must be expanded to allow for the non-human and non-biological – or the concept of moral responsibility must be expanded to account for attributing moral responsibility to humans for the action of systems whose operation is far from the direct control of said humans. Arriving at a properly informed consensus needs to balance the ethicist's needs for coherence of a theory either of non-biological agents or of how to distribute responsibility for the actions of a complex physical system back to its human creators and/or maintainers, with the roboticist's insight into technological possibilities towards the autonomy of machines.

Finally, an attempt to integrate ontology into discussions about the ethics of robotics, or into the practice of robotics itself as an enginering approach, poses one more problem related to the basic philosophical assumptions behind a foundational ontology. Which assumptions are more conducive to useful development in a domain like robotics? Do we require further clarification of the basic elements that constitute an ontology?

In short, while Moravec's paradox may suggest there is little for ontology engineering to do in robotics, we rather argue that:

- *robotics can gain from ontology* methods to treat explicit knowledge tasks in an open world by more reliably making use of the capabilities of foundation models
- *ontology can gain from robotics* a test-bed for philosophical assumptions and arguments

2. Ontologies as Guard Rails for Foundation Models in Robotics

As we previously mentioned, the hard problems of robotics – perception and action – appear not to involve explicit, formally represented knowledge. There is no point in looking for a logical formula that would recognize a cat or control legs for running. However, in recent years AI as well as robotics have shown a renewed interest in "explicit knowledge tasks" (Kambhampati et al., 2024a), and here is an opportunity for ontology engineering to make a contribution.

The trend back to explicit knowledge tasks is powered by current generation large language models (LLMs) and multi-modal (or "foundation") models more generally. While one can have doubts about the staying power of these particular tools, what is likely to remain relevant is what these tools have brought closer to the realm of the possible: the production and use, at large scale, of explicit knowledge (or at least something that looks like it). While classical symbolic AI approaches often suffered from problems of coverage or unsuitability to new situations, current generative AI appears to have an answer ready for every occasion. LLMs are better thought of as an "approximate database" than an actual knowlede store, but when this approximate database is trained on a good part of the whole internet, it may often supply useful answers to an endless variety of queries, which (as we will indicate with citations in the next subsections), may enable generalization in tasks such as scene understanding, action selection, and skill

learning. Generative AI and foundational models themselves may become obsolete as tools; however, the capabilities they enable will remain relevant.

Thus, it turns out explicit knowledge is interesting in robotics after all. However, that requires using and analyzing this knowledge rigorously. When a multi-modal model provides a description of a scene, what exactly is it providing? To ask for a description of a scene is to already make some ontological assumptions, e.g. that there are objects between which relations can exist, that there are events in which these objects may participate and which may be indicated by the relations, and so on. The consequences of such ontological assumptions are not evident, especially if they are made tacitly by an engineering process not informed by an ontological perspective. Are these ontological commitments adequate for the task at hand, and if not how should they be ammended? Are there hard ontological constraints – in which case, given the tendency of generative AI to confabulate[1], do we have ways to check that the model actually sticks to said constraints?

Arguments have already been laid out about how an integration of LLMs and knowledge representation techniques may be helpful for AI in general (Pan et al., 2024). In the next subsections we will describe how ontologies and ontology engineering can contribute to leveraging foundation models to enable the development of generalizable skills for robots. We will focus on interactions between generative AI, ontologies, and algorithmically sound inference, because in the current technological environment and for the foreseeable future it is generative AI that provides the best approximation for the large-scale world knowledge bases that robotics needs.

2.1. Enhancing Perception Capabilities

Foundation models appear suitable to expand the scope of situations that a perception system can "understand". Classically, computer vision pipelines are limited to a usually small set of recognizable objects and relations. Even the object classes that are known to the system are not always reliably recognized when instances of them are encountered.

What foundational models promise to do is to break out of such restrictions. For example, a robot can ask a large multi-modal model about any type of object that can be described in natural language, including parts related to the manifestation of affordances such as blades or handles or buttons (Tong et al., 2024). Actions of an agent can be recognized and described in open-ended ways to reveal whether a robot's performance of a task is inappropriate – this is especially useful when it is difficult to state a priori what counts as failure and the peculiarities of a situation dictate what "appropriate" means (Guan et al., 2024). A robot using foundation models is not even limited to a set of predicates given a priori to it by its designers; it can learn new predicates with which to describe its surroundings in task-relevant ways (Athalye et al., 2024).

However, with foundation models there remains the risk of confabulation. While some creative reinterpretation of a situation may be welcome if it helps a robot better cope with novelty, it should also be constrained by clearly expressed and verifiable ontological commitments. Ontologies provide structured knowledge that helps robots better capture the context of the sensory data they receive, i.e. *contextualized belief-states* (Kümpel et al., 2021; Nguyen et al., 2024). By "belief-state", we refer to the notion used in robotics to name the aggregate of what a robot "believes" about its environment, particularly sub-symbolic aspects such as the estimated poses of physical objects and probability distributions on their types. E.g., as a result of processing video data, a robot may assert with some level of confidence

[1] I.e., the tendency of generative AI techniques to produce assertions of fact that have no basis in reality. The term "hallucination" has also been used to refer to this phenomenon in discussions of LLMs.

that an object of type "cup" is at a particular location and orientation plus/minus some measurement error. By "contextualized belief-state", we aim to describe a notion of belief-state extended with other information that defines a situation, e.g. what time of day it is and what activities are routine at this time, whether the robot was given a task to perform, the relevance of object types for various tasks e.g. that a cup can hold liquids. Ontologies would be necessary in this case because they make explicit what counts as a defining factor for a situation, at least as far as the robot is concerned. We have here suggested that, apart from participating physical objects, also aspects relating to routine activities, norms, affordances would also come into play, and we venture that a capable robot will need to understand and reason with such aspects – and more – in a principled way. In other words, the ontology's role here is to describe what the robot can reason about, and constrain how it should go about said reasoning. As examples of such constraints, we would use again task planning (i.e. the search for a sequence of actions that would drive the world to some desired state) for illustration. Actions involving the manipulation of physical objects can only be feasible if those physical objects actually exist. If the robot were to rely purely on an LLM to find a task plan however, it is quite possible that the generated plan would make use of objects before their presence in the environment is actually established, e.g. the LLM would make a plan to deliver a drink to a human by using a cup, because kitchens typically have cups, even though in this particular kitchen there are no cups to be found.

The *semantic enrichment* obtained from an ontologically-based annotation and categorization of sensory data also provides semantics that foundation models can leverage, via prompt injection, when generating responses or taking actions based on the sensed information. This has other benefits at the robot level as well, since by linking sensory input to predefined concepts and relationships, ontologies enable more accurate interpretation of what the robot senses. For instance, if a robot detects an object, the ontology can clarify what that object is and its significance in a given context.

In short, ontologies provide a formalized structure for representing complex relationships within a domain, i.e. structured knowledge representation. This enables robots to create more precise and accurate models of their environments (Kümpel et al., 2021; Nguyen et al., 2024). While LLMs can process new information and generate insights, ontologies can ensure that the robot's model is updated consistently and logically based on this information. This combination allows for effective adaptation to changes in the environment, i.e. *dynamic updates with contextual insights*.

2.2. Enhancing Action Capabilities

Multi-modal models (e.g. vision-language models) have shown themselves to be a promising approach to robot policy learning and generalization (Zitkovich et al., 2023; Black et al., 2024). Foundation models have also been fruitfully employed on planning-adjacent tasks such as the creation of planning domains (Oswald et al., 2024), or extracting sub-goals from natural language goals and translating them into symbolic representations (Izquierdo-Badiola et al., 2024).

However, the abilities of LLMs and similar generative AI approaches to do task planning are questionable, as the performance of LLMs degrades on problems where trivial distractions – e.g., additions of entities not needed in a plan, or removal of obstacles that appear in a typical formulation of a problem – or renamings of entities are introduced (Valmeekam et al., 2023; Kambhampati, 2024); even newer "large reasoning models" do not plan reliably (Valmeekam et al., 2024). Similar degradation in performance under trivial distractions is observed for problems of logical reasoning (Dziri et al., 2023; Hoppe et al., 2025). This gives empirical support to the suspicion that LLMs only pattern match, albeit in a massive database, as opposed to figuring something out via some algorithm-like procedure that captures the structure of a problem domain beyond what training examples reveal.

Instead, a promising approach is to use "LLM modulo planners" combinations (Kambhampati et al., 2024b; Aghzal et al., 2025) – i.e., systems in which an LLM and a classical planning algorithm cooperate to construct plans. This combination of LLMs and classical planning can be fairly loosely integrated, e.g. with the LLM creating domains and inputs for a planner (Oswald et al., 2024), or can consist of the LLM providing an initial plan which can then be repaired or at least checked (de la Rosa et al., 2024; Valmeekam et al., 2024; Curtis et al., 2024). In principle it could also be a tight integration of the two, in which the LLM heuristically guides the search (Farrell and Ware, 2024), but to avoid expensive LLM calls during planning it is also possible to have LLMs generate search heuristics (Zheng et al., 2025). An "LLM modulo formal methods" combination would leverage the precise and verifiable character of classic algorithms together with the large-scale "approximate knowledge bases" of LLMs and foundation models.

Another emerging use case for LLMs in robotics is enhancing reinforcement learning of fine manipulation skills. Such skills would be infeasible to program directly, hence the approach for several years has been to use learning for them – informally said, it is hard to program a robot to juggle, but it may be possible to program it so that it learns to juggle. Reinforcement learning however suffers from data efficiency and convergence issues, thus approaches such as reward shaping with LLMs have been recently explored, either by using a linguistic description of the world state to generate rewards (Du et al., 2023), or by using LLMs to write code that computes rewards (Xie et al., 2024; Ma et al., 2024b). Other approaches aim to improve the exploration performed by reinforcement learning by providing it a good initial guess at a control policy from an LLM (Chen et al., 2024), or by using an LLM to reshape the action and observation spaces, i.e. to change the ways in which these are described to the learning process (Ma et al., 2024a).

In all such applications, however, it is important to specify how this plethora of systems – LLMs or other foundation models, planners, reinforcement learners, etc. – can interface. At the very least, this involves defining the kinds of entities that can appear in the representations passed between these systems, and what constraints may apply to the structure of those representations. Interestingly, such a specification cannot be only formal. It has to be formal to be meaningful to the algorithmic parts of an "LLM modulo" system. However, it should also be "informal", in whatever language is likely to prompt the LLM towards activating patterns involving the requested concept. Making good prompts is somewhat of an art form, but a good place to start may be the kind of linguistic descriptions of concepts produced by ontology engineers and documented in research papers and ontology annotation comments. This latter suggestion is of course mere speculation on our part, but we think it a plausible one. Ontology engineers, by profession, have to think about concepts at various levels of formality and also keep in touch with how these concepts are understood, described, and distinguished in a vast body of literature; hence our suggestion that they may be well positioned to define the interfaces of LLM modulo systems in a way that is effective for all components of such systems as well.

2.3. Human-Robot Interaction

The benefits of LLMs to interaction seem quite obvious but nevertheless we should also list them here. LLMs have shown themselves able to provide human-like communication capabilities and thus enable natural language interface with users to explain and improve robot behaviors (Fouilhé et al., 2025; Devarakonda et al., 2024) or receive commands (Bassiouny et al., 2025) – and in the latter citation, the benefits of combining LLMs with explicit knowledge stores are already being explored.

For our own arguments, we note that, especially in more specialized domains, it may be necessary to ensure the meanings behind terms and concepts are clear and consistent. This would require an ontology,

and this combination can enhance interactions between robots and humans, making communication more effective and reducing misunderstandings. Ontologies can further enrich this capability by offering explicit context that LLMs can reference, again via prompt injection. This is especially important in HRI, where detecting the nuances of human interactions can lead to more relevant and appropriate responses. In short, ontologies can help to listen better.

Ontologies can also help to speak better by providing a framework for verifying the information generated by an LLM. This reduces the risk of generating misleading or incorrect content, making interactions with autonomous systems more reliable.

Importantly, ontologies do not limit a robot to what it carries in its head, but rather enable robots to *connect with external knowledge sources on the Semantic Web*, allowing them to enrich their representation of their environment with additional context or information. LLMs can then generate natural language summaries or explanations based on this enriched knowledge.

It should also be noted that the context window of an LLM – the amount of information it can take as input – is limited. Thus it is important to have other stores for relevant data, and ways to decide what to put in or leave out of a prompt. And in any case, some information must remain transparently accessible, such as user preferences. Ontologies to represent user preferences, needs, restrictions can therefore help deliver even more tailored and relevant interactions, enhancing user satisfaction and trust.

Finally, while the discussion above focused on linguistic interactions, it should be noted that ontologies can help bridge different modes of interaction (e.g., text, speech, visual) as well, by providing a unified framework that guides how LLMs and/or multi-modal models interpret and respond to diverse inputs, creating a smoother interaction experience.

3. Robotics as Testing Ground for Ontologies

Autonomous robotics presents a set of challenges that distinguish it from many fields in which ontology engineering has been previously applied. Chief among these challenges is that, by definition, autonomous systems are intended to operate with as little human supervision as possible; further, robotic control loops must be fast to be able to cope with the physical world. This is unlike something like a decision support system or even a personal assistant agent – in which cases there is ample opportunity for a human to intervene and perform corrections. In other words, whatever reasoning is there to be supported by the ontology has to be very precisely described, as there is little room for human intervention to fill in gaps with commonsense judgement. In this chapter however we will explore a couple of different challenges which, while speculative, appear plausible to us. Both of these aspects would be relevant for autonomous systems in general, robotic or otherwise, however robotics put them in starker detail because of the direct connection between a robotic agent and the physical world of human beings. Firstly, it appears quite possible that an autonomous system, especially when operating in the real world, will need to construct at least part of its concepts on the fly, e.g. by defining object types and/or roles as the pragmatics of a situation demand, rather than relying on a slowly-evolving vocabulary. If this were the case, then the job of ontology engineering would not be to just provide a correct conceptual structure to the system, but rather with a correct way to update a conceptual structure. Secondly, autonomous systems raise important ethical concerns such as how should responsibility for a decision of the system be allocated, and, more fundamentally, what it would mean for an autonomous system to act ethically – i.e. what is a "morally good" decision from the autonomous system. It appears plausible we would want our autonomous systems to be able to reason about the ethical character of their decisions, even if said

machines are not full ethical agents, and to do so in a way that we could verify and constrain. If that is the case however, these machines would have to be given reliable theories to reason with. This issue is deeper than "just" implementation concerns such as choosing some formal mechanism for inference, e.g. non-monotonic logic or Petri nets to handle rules and exceptions. Rather, it is conceptual and relates to what ethical inference should be about, and it would be the job of ontology engineering to formalize answers to questions such as what sorts of entities would appear in a machine theory of ethics, how these entities would connect to each other, what kind of statements could be made about said entities and in which conditions.

3.1. Foundational Ontologies for Autonomous Systems

Unlike classical AI, formal ontology engineering commonly makes philosophical commitments explicit and argues about their broad suitability. This is especially true for *foundational ontologies* (FOs), where discussions often result in sharply opposing views. Perhaps the most prominent example is the clash between DOLCE's (Borgo et al., 2022b) conceptualist perspective – its "categories refer to cognitive artifacts more or less depending on human perception, cultural imprints and social conventions" (Masolo et al., 2003, p. 7) – and BFO's (Otte et al., 2022) commitment to Aristotelian-inspired ontological realism that only admits universals actually instantiated by concrete particulars in the empirical world (Dumontier and Hoehndorf, 2010; Smith and Ceusters, 2010). These differences, which occasionally erupted in vehement debates at scientific conferences, mirror the two's contrasting domains of origin. It should come as no surprise that DOLCE, born in cognitive science and linguistics (Masolo et al., 2003), assumes a different attitude towards concepts than BFO with its biomedical background (Grenon et al., 2004). However, their commitments have practical consequences, directly determining the methodology for modeling different types of domain problems, and therefore constituting a key factor when choosing a FO in ontology development (Khan and Keet, 2012; Keet and Khan, 2022; Borgo et al., 2022a). This raises an interested empirical question: Do these constrasting philosophical assumptions have practical consequences on the ontologies for robots and for human-robot interaction?

Empirical ontology evaluation is a rather complicated subject, which has produced few clear-cut techniques, and there is always room for more contention as to which approach is best suited. A well-known approach, *task-based ontology evaluation*, assesses the quality of domain ontologies by the performance of an employing system in solving a problem, such as tagging words with concepts from the ontology (Porzel and Malaka, 2004). Yet this approach struggles with FOs, which operate indirectly, shaping domain models through axiomatic foundations and complementing philosophical frameworks rather than being directly implementated in systems. To exploit robot and interaction ontologies to evaluate FOs, one must ask: Do their philosophical commitments *matter* when robots face real-world tasks? This aligns with William James' pragmatist maxim: If competing philosophical "world-formulas" (James, 2014, p. 50), e.g., BFO's realism vs. DOLCE's constructivism, yield no measurable difference, "all dispute is idle" (James, 2014, p. 45).

A compelling historical precedent supporting James' pragmatic principle is Dreyfus' (1972; 1992) critique of classical AI. Dreyfus argued that classical AI served as an empirical test of — and ultimately an empirical refutation — of Cartesian rationalist philosophy as a way to understand the mind. What Dreyfus meant by this is a view of the universe as consisting of subjects and objects which can be considered independently of each other and anything else, of which context-independent properties can be predicated, and that it is the minds of subjects that are capable of imbuing objects and their properties with context- and situation-dependent significance. Objects, in this view, are like Aristotelean substances

– in principle independent of anything else and each object being, because of its essential properties, an instance of some type. He predicted that such an understanding of the universe and the mind's place in it would not be capable of dealing with problems such as identifying the relevant aspects of a situation (e.g., "by focusing on objects, one loses track of the sense of what counts as an object"). As he expected, classical AI was never able to cope with the real world, but rather had its successes restricted to microworlds where human developers fully specified what was relevant, relieving the AI of most of the work. There is one more aspect to the example of Dreyfus: he was not against Cartesian philosophy; in the way in which he defined it, it underpins the scientific method which has proven succesful at gaining knowledge about the physical universe. But his example does show that even good ideas have limited applicability, and it may help to explore and understand the limits of our conceptualizations.

To benchmark FOs one may rely on a variety of parameters such as measuring the proportion of systems employing domain ontologies, the comparative performance of such systems to perform their prescribed functions, or even numbering relevant publications. Irrespective of the metrics that will be ultimately applied, in the spirit of James, it remains interesting whether FOs perform unevenly across domains — which is not unreasonable, as they were (at least originally) tailored to different ones — or whether employing one over the other just does not make relevant difference. In this sense, we see robotics as a suitable and neutral proving ground. Robotics thus becomes a contemporary analogue, probing foundational ontologies empirically as classical AI once probed Cartesian rationalism in Dreyfus's critique.

Indeed, Dreyfus' critique of classical AI is more pertinent to this discussion as it is related to an old question in philosophy: do natural kinds exist, or are categories pragmatically constructed by human beings? As examples of defenses of categories as pragmatic constructs, we cite Dreyfus (1972) above, and more recently Hacking (2007b,a) and Khalidi (2010); one of the coauthors has already explored, in a previous position paper, the possibility that if an autonomous agent is allowed to develop its own concepts then they need not resemble the ones we would use (Borgo, 2020).

This does not necesserily doom ontological approaches that are realist about natural kinds. Biological taxonomies of species are still used and useful, and will likely remain so even as the concept of species is increasingly problematic (Mishler, 1999). However, this does raise questions as to how an autonomous system should think about concept drift, and what, if anything, should count as a natural boundary between kinds. Of course, ontological approaches that are not realist about natural kinds avoid the latter question, but must then provide answers as to what is a good way to organize a conceptual structure such that it is not completely ad-hoc to the moment and opaque to anyone but the agent using it.

By now, ontologies in robotics have (at least partially) adapted a variety of foundational ontologies such as BFO, Cyc, DOLCE + DnS Ultralight (DUL), OpenCyc, and SUMO (Bateman and Farrar, 2005; Olivares-Alarcos et al., 2019). This way, robotics' interdisciplinary arena is serving as a test bed, and may perhaps become a litmus test, for foundational ontologies.

3.2. Explicitly Ethical Agents

The title of this section makes reference to Moor (2006). Moor distinguishes implicitly ethical, explicitly ethical, and fully ethical agents. An implicitly ethical agent is one that is programmed to do the right thing; e.g., a bank teller machine should not defraud either the customer or the bank. Competent adult human beings are fully ethical agents. The explicit ethical agent category operates as a middle point between the previous two. Its main function is to avoid discussions as to whether machines may be conscious or possess free will or whatever other metaphysical properties one would invoke to justify

the moral value of human beings. Explicitly ethical agents "reason" ethically in the same way that chess computers "play" chess. Such machines might not engage with their domain with true understanding, whatever that is, but would be judged by an outside human observer as competently playing chess or performing, and expressing justifications for, ethical choices.

With progress in AI and robotics making more autonomous agents plausible, Moor's argument for the development of ethical agents is even more relevant now than in 2006. An autonomous system would have to, eventually, cope with situations not described in detail at the time of its creation and thus it would need some kind of ability for ethical inference – and preferably, be able to do so in a manner that is introspectable and with some kind of algorithmic reliability. Must robots always comply with their owners' instructions, or should they retain the freedom to decide otherwise, e.g. if asked to act unlawfully? Who is liable for their actions: the owner, the developer, or somehow the robot itself? How should robots deal with moral dilemmas?

The relatively young field of machine ethics has no definitive answer to many of such questions. Moor identifies as one source of difficulty the fact that humans have limited understanding of what a proper ethical theory looks like, with significant disagreement between individuals and contradictory intuitions within the individual as well.

We see a role for ontology engineering here as a way to guide the construction of *a* theory of ethics for an autonomous agent to reason with. The exact content of the theory would depend on e.g. the developers of a robot and/or their prospective clients, with the ontology engineer acting as a moderator to assist in formalizing the moral intuitions of the stakeholders, akin to the role attributed to ontology engineers by Neuhaus (2023). By way of illustration, we will briefly sketch how one could start from existing foundational ontologies and outline a possible theory of ethics for a robot.

For this illustration, we start from the systemic view inspired by YAMATO (Mizoguchi and Borgo, 2021). Following this view, we would assert that a robot is a system performing a function. In addition to the statements required to describe the robot as a system, our sketched theory would describe the robot as created by, and imbued with a function by, some collective of agents, either natural persons or companies, and that its function impacts some community of morally valuable entities. The moral intuition we aim our sketch to capture is that a robot is a tool, and a tool is built by someone to serve some purpose. The builder of the tool/robot is responsible to build it for a good function and to build it well.

We cannot detail this sketch here, however the reason we offer it is not to create a theory of robot ethics ourselves, but rather to suggest how proceeding from ontological approaches helps focusing questions about how details of the theory should be filled in. Thus, e.g., what is "impact" on a community of morally valuable agents? Do unintended side-effects count? If an agent can adjust its behavior through learning, does that change its function, or is its function to learn how to be good at some set of tasks? And, to strenghten the connection between our sketch and explicit moral agents, what is the connection between the function of the robot and the function of some subsystem of it that employs this theory for moral reasoning?

Answering such questions would, as mentioned, be the job of the other stakeholders in the creation or use of a robot, but the ontology engineer must play, in our view, an important role in defining the questions and the structure in which to attach the answers. Implementing an explicitly ethical agent is not just an issue of choosing the right formal machinery to manipulate some representations of moral values. Before anything like that can be done, one has to have an understanding of what the relationship of an autonomous system is to other, morally valuable entities, but this, on its own, is too large a question to answer. In our illustration above we have split it into aspects such as relationships of creation by,

impact on, and evolving function affecting, some communities of morally valuable entities. This split is not the only possible approach, but it, or other approaches like it, will likely need an ontologist to devise.

Note, we are not claiming that ontology engineering will lead to a definitive, singular and universal theory of robot ethics. It is quite possible that a kind of ethical pluralism may emerge, with different robots operating as explicit ethical agents in the sense of Moor but using different ethical theories. Once made explicit and formalized – an endeavour in which ontology engineering will be of value – the theory of ethics that a robot operates on is part of the robot's specification as a system. This robot ethical pluralism may be a workable, pragmatic compromise until we somehow solve human ethics to everyone's satisfaction. Until then, it would be up to our human judgement as users to employ our tools well – but for that judgement to have a chance at working, it should be clear for us what our tools/robots *do* when they are somehow entrusted with decisions of moral import.

3.3. A Testing Ground for Ontology Construction and Alignment

It is instructive to look back at the purposes that led to the construction of foundational ontologies from the beginning of Applied Ontology in the 1990s. In that period the discussion was primarily around two concepts: philosophical principles and application interoperability. Indeed, we can say today, the community was somehow driven by two main objectives: a) give evidence and formal shape to the idea that one can develop a complete and consistent ontology which conforms with the view of some philosophical school (essentially those within the realist or conceptualist visions since they seemed better suited for the applicative concerns of the community); and b) provide a general understanding of how to make explicit a complete and coherent understanding of the world inspired by cognitive concerns (relying essentially on philosophical principles to avoid the peculiarities of common sense and human perception). The common target, we can say today, was to have a logical theory which should satisfy a few conditions: to be compete (or hypothetically completable for any practical reason), logically and conceptually coherent, philosophically sound, and effectively facilitating interoperability across information systems.

While today we have evidence that most of these conditions have been met (in part due to the lack of an explicit characterization of what interoperability means), we are also aware of the limitations of this endeavor. Let it be philosophical or cognitive, both objectives were human-centric. Our philosophies are highly driven by our way to interact with the physical world we recognize, and our cognitive understanding of the world is an understanding focused on human-cognition. The merging of robotics and artificial intelligence has led to a broadening of the notion of agent and, as a consequence, of an ontology an agent may develop (or assume) to make sense of and successfully interact with the world. This broadening of the perspective pushes ontologists to investigate alternative ways to explain the world and to discuss how ontologies might look, which organization they might have and which kind of interoperability they might allow (Borgo, 2020).

From another perspective, ontological analysis is especially useful when developing an ontology from a foundational point of view, where characterization of core concepts is more important than coverage of the application domain. A foundational viewpoint facilitates the realization of ontological models that capture a general view of the world penalizing subjective perspectives. However, one might wonder if such a general approach is appropriate for the robotics domain as it is today, where robots perceive the world and act in it in ways that differ from how humans do, even though remaining within a setting designed by humans and, to some extent, still controlled by them. One might alleviate this issue by

starting the ontological analysis from a deeper understanding of the target robotic applications, which may be gained from hands-on experience in developing robot perception and decision-making tasks. Hence, ontologists would be able to include the perspective of robots in the analysis and development of an ontology suitable for them.

Note that a foundational perspective helps with the construction of flexible and general ontologies, often small, which can be applied or easily adapted to different domains and applications, thus facilitating the reusability of ontologies across the robotic domain. However, general and reusable concepts might be of little help for the actual reasoning tasks needed in a realistic robotic application. There is a trade-off between flexibility and applicability, which is especially prominent in robotics because of its practical nature. Hence, reconsidering ontological analysis from the target robotic applications may also serve to address this trade-off.

Related to the idea of starting from actual robotic tasks, it is easy to run into a challenge that we here refer to as *the relevance problem*. When robots are bombarded with vast amounts of data, both internally and externally, it is challenging to filter out the information irrelevant for abstraction into knowledge principles. A novice ontology developer may attempt to abstract as much data as possible, aiming for large ontologies that will hopefully enable robots to tackle complex reasoning tasks. Those ontologies could be grounded using data coming from robots' perception (e.g., force sensors, cameras, distance sensors, microphones, etc.), and also internal data from the robot (e.g., joints' torque, acceleration, velocity, position, action traces, etc.). However, by trying to represent semantically every piece of data, developers end up capturing vast amounts of seemingly irrelevant knowledge, which does not help the rest of the scientific community. To address these challenges, one approach would be to identify the types of robotic tasks that would genuinely benefit from ontological reasoning, and other types of tasks that can be solved using plain data. We argue that if the community starts investigating this distinction, there would be at the same time better ontologies and better knowledge bases for robotics. For instance, consider a case in which a robot is closely collaborating with a human, working on the same task, sharing a common space. For the task of issuing a safety stop, which requires a fast robot reaction, a robot may rely directly on raw sensor data (e.g., a camera or distance sensor) (Olivares-Alarcos et al., 2023a). However, to report or explain to a human supervisor how many times the robot adapted its plan due to a risk of collision, a robot will indeed benefit from ontology-based reasoning about collaborative and adaptive events. A robot can store the relevant knowledge about how risk events affected its plans' execution, for example for later memory interrogation and explanation generation (Olivares-Alarcos et al., 2023b).

4. Conclusion

In their everyday activities humans use a large body of knowledge based on what they recognize in reality and their understanding of it. When we want to analyze this knowledge we can list several types: knowledge of physical laws, of material objects, of biological systems, of artifacts, of social scenarios, of social expectations, of causality and so on. Yet, in our own experience all these kinds of knowledge form a striking unity: everything goes together. Admittedly, we do make mistakes, forget things and even fail to evaluate rules. But these cases are exactly that: "mistakes", "slips" and "misapplications". To ourselves our overall knowledge remains convincing, solid and fully integrated notwithstanding the faults.

It seems natural to also provide artificial agents with integrated knowledge. After all, to behave as expected in a socio-technical environment a robot needs to perceive and find its way in the environment,

manipulate objects, consider what other agents are doing, anticipate problems, develop expectations, etc. In this context, applied ontology emerges as a perfect candidate to be the integrative knowledge framework for robots. Applied ontology builds on philosophy, cognitive science, linguistics and logic with the purpose of understanding, clarifying, making explicit and communicating people's assumptions about the nature and structure of the world (Oltramari, 2019).

In this paper, we have discussed and made explicit an apparent symbiotic relationship between ontologies and robotics. Initially, we argued that ontologies can facilitate robot explicit knowledge tasks in an open world by more reliably making use of the capabilities of foundation models. This is analyzed according to the main and most challenging robot capabilities, such as perceiving the environment, acting in it, and interacting with other agents (e.g. humans). Finally, the paper argued that robotics offers an ideal test ground for the philosophical assumptions and arguments around ontologies. For instance, robotics can be of use in testing foundational ontologies, in the development of ethical ontologies, and in widening ontology experts' perspective during ontology construction and aligment. These ideas set the basis for our belief that the future of robotics and applied ontology are tightly inter-connected.

Acknowledgements

This work was partially supported by the German Research Foundation DFG, as part of Collaborative Research Center (Sonderforschungsbereich) 1320 Project-ID 329551904 EASE - Everyday Activity Science and Engineering, subprojects "P01 – Embodied semantics for the language of action and change: Combining analysis, reasoning and simulation" and "P05-N — Principles of Metareasoning for Everyday Activities", and also by the European Union under the project ARISE (HORIZON-CL4-2023-DIGITAL-EMERGING-01-101135959).

References

Aghzal, M., Plaku, E., Stein, G.J. & Yao, Z. (2025). A Survey on Large Language Models for Automated Planning. https://arxiv.org/abs/2502.12435.

Athalye, A., Kumar, N., Silver, T., Liang, Y., Lozano-Pérez, T. & Kaelbling, L.P. (2024). Predicate Invention from Pixels via Pretrained Vision-Language Models. https://arxiv.org/abs/2501.00296.

Bassiouny, A., Elsayed, A., Falomir, Z. & Pobil, A. (2025). UJI-Butler: A Symbolic/Non-symbolic Robotic System that Learns Through Multi-modal Interaction. *International Journal of Social Robotics*, 1–21. doi:10.1007/s12369-025-01234-5.

Bateman, J. & Farrar, S. (2005). Modelling Models of Robot Navigation Using Formal Spatial Ontology. In C. Freksa, M. Knauff, B. Krieg-Brückner, B. Nebel and T. Barkowsky (Eds.), *Proceedings of the 4th International Conference on Spatial Cognition 2004*. Lecture Notes in Computer Science (LNCS) (pp. 366–389). Springer. doi:10.1007/978-3-540-32255-9_21.

Biccheri, L., Borgo, S. & Ferrario, R. (2023). On the relation of instrumental dependence. In *Formal Ontology in Information Systems* (pp. 47–61). IOS Press.

Black, K., Brown, N., Driess, D., Esmail, A., Equi, M., Finn, C., Fusai, N., Groom, L., Hausman, K., Ichter, B., Jakubczak, S., Jones, T., Ke, L., Levine, S., Li-Bell, A., Mothukuri, M., Nair, S., Pertsch, K., Shi, L.X., Tanner, J., Vuong, Q., Walling, A., Wang, H. & Zhilinsky, U. (2024). π_0: A Vision-Language-Action Flow Model for General Robot Control.

Borgo, S. (2020). Ontological challenges to cohabitation with self-taught robots. *Semantic Web*, *11*(1), 161–167.

Borgo, S., Galton, A. & Kutz, O. (2022a). Foundational ontologies in action. Understanding foundational ontology through examples. *Applied ontology*, *17*(1), 1–16.

Borgo, S., Ferrario, R., Gangemi, A., Guarino, N., Masolo, C., Porello, D., Sanfilippo, E.M., Vieu, L., Borgo, S., Galton, A. & Kutz, O. (2022b). DOLCE: A descriptive ontology for linguistic and cognitive engineering. *Applied Ontology*, *17*(1), 45–69. doi:10.3233/AO-210259.

Chen, L., Lei, Y., Jin, S., Zhang, Y. & Zhang, L. (2024). RLingua: Improving Reinforcement Learning Sample Efficiency in Robotic Manipulations With Large Language Models. *IEEE Robotics and Automation Letters*, *9*, 1–8. doi:10.1109/LRA.2024.3400189.

Curtis, A., Kumar, N., Cao, J., Lozano-Pérez, T. & Kaelbling, L.P. (2024). Trust the PRoC3S: Solving Long-Horizon Robotics Problems with LLMs and Constraint Satisfaction. https://arxiv.org/abs/2406.05572.

de la Rosa, T., Gopalakrishnan, S., Pozanco, A., Zeng, Z. & Borrajo, D. (2024). TRIP-PAL: Travel Planning with Guarantees by Combining Large Language Models and Automated Planners. https://arxiv.org/abs/2406.10196.

Devarakonda, V.N., Kaypak, A.U., Yuan, S., Krishnamurthy, P., Fang, Y. & Khorrami, F. (2024). MultiTalk: Introspective and Extrospective Dialogue for Human-Environment-LLM Alignment. https://arxiv.org/abs/2409.16455.

Dreyfus, H.L. (1972). *What Computers Can't Do*. New York, N.Y., USA: Harper & Row.

Dreyfus, H.L. (1992). *What computers still can't do: A critique of artificial reason*. The MIT press.

Du, Y., Watkins, O., Wang, Z., Colas, C., Darrell, T., Abbeel, P., Gupta, A. & Andreas, J. (2023). Guiding Pretraining in Reinforcement Learning with Large Language Models. https://arxiv.org/abs/2302.06692.

Dumontier, M. & Hoehndorf, R. (2010). Realism for scientific ontologies. In *Proc. of the 6th International Conference of Formal Ontology in Information Systems (FOIS 2010)* (pp. 387–399). IOS Press.

Dziri, N., Lu, X., Sclar, M., Li, X.L., Jiang, L., Lin, B.Y., Welleck, S., West, P., Bhagavatula, C., Bras, R.L., Hwang, J.D., Sanyal, S., Ren, X., Ettinger, A., Harchaoui, Z. & Choi, Y. (2023). Faith and Fate: Limits of Transformers on Compositionality. In *Thirty-seventh Conference on Neural Information Processing Systems*. https://openreview.net/forum?id=Fkckkr3ya8.

Farrell, R. & Ware, S.G. (2024). Large Language Models as Narrative Planning Search Guides. *IEEE Transactions on Games*, 1–10. doi:10.1109/TG.2024.3487416.

Fouilhé, G., Eifler, R., Thiébaux, S. & Asher, N. (2025). Conversational Goal-Conflict Explanations in Planning via Multi-Agent LLMs. In *AAAI 2025 Workshop LM4Plan*.

Grenon, P., Smith, B. & Goldberg, L. (2004). Biodynamic Ontology: Applying BFO in the Biomedical Domain. In D.M. Pisanelli (Ed.), *Ontologies in Medicine*. Studies in Health Technology and Informatics (Vol. 102, pp. 20–38). doi:10.3233/978-1-60750-945-5-20.

Guan, L., Zhou, Y., Liu, D., Zha, Y., Amor, H.B. & Kambhampati, S. (2024). Task Success is not Enough: Investigating the Use of Video-Language Models as Behavior Critics for Catching Undesirable Agent Behaviors. In *First Conference on Language Modeling*. https://openreview.net/forum?id=otKo4zFKmH.

Hacking, I. (2007a). The Contingencies of Ambiguity. *Analysis*, *67*, 269–277. doi:10.1111/j.1467-8284.2007.00690.x.

Hacking, I. (2007b). Natural Kinds: Rosy Dawn, Scholastic Twilight. *Royal Institute of Philosophy Supplement*, *61*, 203–239. doi:10.1017/s1358246100009802.

Hoppe, F., Ilievski, F. & Kalo, J.-C. (2025). Investigating the Robustness of Deductive Reasoning with Large Language Models. https://arxiv.org/abs/2502.04352.

Izquierdo-Badiola, S., Canal, G., Rizzo, C. & Alenyà, G. (2024). PlanCollabNL: Leveraging Large Language Models for Adaptive Plan Generation in Human-Robot Collaboration. In *2024 IEEE International Conference on Robotics and Automation (ICRA)* (pp. 17344–17350).

James, W. (2014). *Pragmatism: A New Name for Some Old Ways of Thinking*. Cambridge Library Collection - Philosophy. Cambridge, GB: Cambridge University Press. Reprint; first published in 1907. doi:10.1017/cbo9781107360471.

Kambhampati, S. (2024). Can large language models reason and plan? *Annals of the New York Academy of Sciences*, *1534*(1), 15–18. Publisher Copyright: © 2024 The New York Academy of Sciences. doi:10.1111/nyas.15125.

Kambhampati, S., Valmeekam, K., Guan, L., Verma, M., Stechly, K., Bhambri, S., Saldyt, L. & Murthy, A. (2024a). LLMs Can't Plan, But Can Help Planning in LLM-Modulo Frameworks. https://arxiv.org/abs/2402.01817.

Kambhampati, S., Valmeekam, K., Guan, L., Verma, M., Stechly, K., Bhambri, S., Saldyt, L.P. & B Murthy, A. (2024b). Position: LLMs Cant Plan, But Can Help Planning in LLM-Modulo Frameworks. In R. Salakhutdinov, Z. Kolter, K. Heller, A. Weller, N. Oliver, J. Scarlett and F. Berkenkamp (Eds.), *Proceedings of the 41st International Conference on Machine Learning*. Proceedings of Machine Learning Research (Vol. 235, pp. 22895–22907). PMLR. https://proceedings.mlr.press/v235/kambhampati24a.html.

Keet, C.M. & Khan, Z. (2022). Foundational Ontologies: From Theory to Practice and Back. *Journal of Knowledge Structures and Systems*, *3*(1), 67–71.

Khalidi, M.A. (2010). Interactive Kinds. *The British Journal for the Philosophy of Science*, *61*(2), 335–360. doi:10.1093/bjps/axp042.

Khan, Z. & Keet, C.M. (2012). ONSET: Automated Foundational Ontology Selection and Explanation. In A. ten Teije, J. Völker, S. Handschuh, H. Stuckenschmidt, M. d'Acquin, A. Nikolov, N. Aussenac-Gilles and N. Hernandez (Eds.), *Knowledge Engineering and Knowledge Management* (pp. 237–251). Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-642-33876-2_22.

Kümpel, M., Mueller, C.A. & Beetz, M. (2021). Semantic Digital Twins for Retail Logistics. In M. Freitag, H. Kotzab and N. Megow (Eds.), *Dynamics in Logistics: Twenty-Five Years of Interdisciplinary Logistics Research in Bremen, Germany* (pp. 129–153). Cham: Springer International Publishing.

Ma, R., Luijkx, J., Ajanović, Z. & Kober, J. (2024a). ExploRLLM: Guiding Exploration in Reinforcement Learning with Large Language Models. In *RSS 2024 Workshop: Data Generation for Robotics*. https://openreview.net/forum?id=NT2zgVXHn5.

Ma, Y.J., Liang, W., Wang, G., Huang, D.-A., Bastani, O., Jayaraman, D., Zhu, Y., Fan, L. & Anandkumar, A. (2024b). Eureka: Human-Level Reward Design via Coding Large Language Models. In *The Twelfth International Conference on Learning Representations*. https://openreview.net/forum?id=IEduRUO55F.

Masolo, C., Borgo, S., Gangemi, A., Guarino, N. & Oltramari, A. (2003). Wonderweb deliverable d18. Technical report. https://www.loa.istc.cnr.it/old/Papers/D18.pdf.

Mishler, B.D. (1999). Getting Rid of Species? In *Species: New Interdisciplinary Essays*. The MIT Press. doi:10.7551/mitpress/6396.003.0020.

Mizoguchi, R. & Borgo, S. (2021). The Role of the Systemic View in Foundational Ontologies. In *proceedings of the 5th Workshop on Foundational Ontology*.

Moor, J.H. (2006). The Nature, Importance, and Difficulty of Machine Ethics. *IEEE Intelligent Systems*, *21*(4), 18–21. doi:10.1109/MIS.2006.80.

Moravec, H. (1988). *Mind children: The future of robot and human intelligence*. Harvard University Press.

Neuhaus, F. (2023). Ontologies in the era of large language models – a perspective. *Applied Ontology*, *18*, 399–407. 4.

Nguyen, G.H., Beßler, D., Stelter, S., Pomarlan, M. & Beetz, M. (2024). Translating Universal Scene Descriptions into Knowledge Graphs for Robotic Environment. In *2024 IEEE International Conference on Robotics and Automation (ICRA)* (pp. 9389–9395). doi:10.1109/ICRA57147.2024.10611691.

Olivares-Alarcos, A., Foix, S. & Alenyà, G. (2023a). *Time-to-contact for robot safety stop in close collaborative tasks*. In *Human-Robot Collaboration: Unlocking the potential for industrial applications*. Control, Robotics and Sensors (pp. 87–104). Institution of Engineering and Technology.

Olivares-Alarcos, A., Beßler, D., Khamis, A., Goncalves, P., Habib, M.K., Bermejo-Alonso, J., Barreto, M., Diab, M., Rosell, J., Quintas, J., Olszweska, J., Nakawala, H., Pignaton, E., Gyrard, A., Borgo, S., Alenya, G., Beetz, M. & Li, H. (2019). A review and comparison of ontology-based approaches to robot autonomy. *The Knowledge Engineering Review*, *34*. doi:10.1017/S0269888919000237.

Olivares-Alarcos, A., Andriella, A., Foix, S. & Alenyà, G. (2023b). Robot explanatory narratives of collaborative and adaptive experiences. In *2023 IEEE International Conference on Robotics and Automation (ICRA)* (pp. 11964–11971).

Oltramari, A. (2019). Artificial Intelligence Within the Bounds of Ontological Reason. In S. Borgo, R. Ferrario and C. Masolo (Eds.), *Ontology Makes Sense* (pp. 37–48). IOS Press.

Oswald, J., Srinivas, K., Kokel, H., Lee, J., Katz, M. & Sohrabi, S. (2024). Large language models as planning domain generators. In *Proceedings of the Thirty-Fourth International Conference on Automated Planning and Scheduling*. ICAPS '24. AAAI Press.

Otte, J.N., Beverley, J. & Ruttenberg, A. (2022). BFO: Basic Formal Ontology. *Applied ontology*, *17*(1), 17–43. doi:10.3233/AO-220262.

Pan, S., Luo, L., Wang, Y., Chen, C., Wang, J. & Wu, X. (2024). Unifying Large Language Models and Knowledge Graphs: A Roadmap. *IEEE Transactions on Knowledge and Data Engineering*, *36*(7), 3580–3599.

Polanyi, M. (1966). *The tacit dimension.* Garden City, NY: Anchor.

Porzel, R. & Malaka, R. (2004). A Task-Based Approach for Ontology Evaluation. In *Proceedings of the Workshop on Ontology Learning and Population at the 16th Eureopean Conf. Artificial Intelligence (ECAI)*.

Smith, B. & Ceusters, W. (2010). Ontological realism: A methodology for coordinated evolution of scientific ontologies. *Applied Ontology*, *5*(3-4), 139–188. doi:10.3233/AO-2010-0079.

Tong, E., Opipari, A., Lewis, S., Zeng, Z. & Jenkins, O.C. (2024). OVAL-Prompt: Open-Vocabulary Affordance Localization for Robot Manipulation through LLM Affordance-Grounding. https://arxiv.org/abs/2404.11000.

Valmeekam, K., Stechly, K. & Kambhampati, S. (2024). LLMs Still Can't Plan; Can LRMs? A Preliminary Evaluation of OpenAI's o1 on PlanBench. https://arxiv.org/abs/2409.13373.

Valmeekam, K., Marquez, M., Sreedharan, S. & Kambhampati, S. (2023). On the Planning Abilities of Large Language Models - A Critical Investigation. In *Thirty-seventh Conference on Neural Information Processing Systems*. https://openreview.net/forum?id=X6dEqXIsEW.

Valmeekam, K., Stechly, K., Gundawar, A. & Kambhampati, S. (2024). Planning in Strawberry Fields: Evaluating and Improving the Planning and Scheduling Capabilities of LRM o1. https://arxiv.org/abs/2410.02162.

Xie, T., Zhao, S., Wu, C.H., Liu, Y., Luo, Q., Zhong, V., Yang, Y. & Yu, T. (2024). Text2Reward: Reward Shaping with Language Models for Reinforcement Learning. In *The Twelfth International Conference on Learning Representations*. https://openreview.net/forum?id=tUM39YTRxH.

Zheng, Z., Xie, Z., Wang, Z. & Hooi, B. (2025). Monte Carlo Tree Search for Comprehensive Exploration in LLM-Based Automatic Heuristic Design. https://arxiv.org/abs/2501.08603.

Zitkovich, B., Yu, T., Xu, S., Xu, P., Xiao, T., Xia, F., Wu, J., Wohlhart, P., Welker, S., Wahid, A., Vuong, Q., Vanhoucke, V., Tran, H., Soricut, R., Singh, A., Singh, J., Sermanet, P., Sanketi, P.R., Salazar, G., Ryoo, M.S., Reymann, K., Rao, K., Pertsch, K., Mordatch, I., Michalewski, H., Lu, Y., Levine, S., Lee, L., Lee, T.-W.E., Leal, I., Kuang, Y., Kalashnikov, D., Julian, R., Joshi, N.J., Irpan, A., Ichter, B., Hsu, J., Herzog, A., Hausman, K., Gopalakrishnan, K., Fu, C., Florence, P., Finn, C., Dubey, K.A., Driess, D., Ding, T., Choromanski, K.M., Chen, X., Chebotar, Y., Carbajal, J., Brown, N., Brohan, A.,

Arenas, M.G. & Han, K. (2023). RT-2: Vision-Language-Action Models Transfer Web Knowledge to Robotic Control. In J. Tan, M. Toussaint and K. Darvish (Eds.), *Proceedings of The 7th Conference on Robot Learning*. Proceedings of Machine Learning Research (Vol. 229, pp. 2165–2183). PMLR.