

DEPARTMENT: NEUROSymbolic AI

Grounding From an AI and Cognitive Science Lens

Goonmeet Bajaj , The Ohio State University, Columbus, OH, 43210, USA

Valerie L. Shalin , Wright State University, Dayton, OH, 45435, USA

Srinivasan Parthasarathy , The Ohio State University, Columbus, OH, USA

Amit Sheth , University of South Carolina, Columbia, SC, 29208, USA

Grounding is a challenging problem, requiring a formal definition and different levels of abstraction. This article explores grounding from both cognitive science and machine learning perspectives. It identifies the subtleties of grounding, its significance for collaborative agents, and similarities and differences in grounding approaches in both communities. The article examines the potential of neurosymbolic approaches tailored for grounding tasks, showcasing how they can more comprehensively address grounding. Finally, we discuss areas for further exploration and development in grounding.

Robust communication transcends human–human communication settings to include human–machine, machine–machine, and multiagent human–machine teams. *Grounding* fosters a common understanding among agents performing a task—typically in the real world. With the growing number of human–AI interactions, grounding is a fundamentally important capability of AI systems, models, and agents.^{5,8,11,15} Grounding allows AI systems to bridge semantic gaps in the real world, team with other agents in such environments, process inputs from the environment, and learn from interactions. A successful synthetic teammate requires several cognitive capacities, including situation assessment, task behavior, language comprehension and generation,³ and knowledge gap resolution processes. Grounding enables agents with different capabilities to communicate.

Both cognitive scientists and computer scientists have focused on how to make internal mechanisms (or representations) of external entities intrinsic to the agent itself rather than being defined by an external designer or interpreted by an observer.¹⁵ Recent efforts in natural language processing (NLP), computer vision (CV), and human–computer interaction (HCI) improve

the grounding of machine agents. However, it remains a multidimensional challenge, encompassing diverse contexts, abstractions, and modalities of understanding (see [Figure 1](#)). In the absence of a clear definition, we are unable to determine genuine advances or task-specific adjustments.

This article sheds light on different aspects of grounding through the lens of cognitive science and AI, discusses specific neurosymbolic solutions for grounding, and highlights future work.

COGNITIVE SCIENCE LENS

Identification of the symbol grounding problem in cognitive science dates to Harnad. Following the introduction of computation with symbols credited to McCarthy, Searle’s Chinese room problem revealed how amodal symbol manipulation lacks grounding. However, symbol manipulation is often the foundation of contemporary AI systems. Challenges to amodal symbol manipulation include embodied grounding¹² and the recent use of language and simulation to establish grounding.¹³ Given this long-standing research problem, Ziemke¹⁵ groups grounding efforts into two categories: 1) *cognitivist* or 2) *enactivist*.

Cognitivism grounds atomic primitives in sensorimotor invariants.¹⁵ Concepts constructed from these inherit the grounding of their constituents.

1541-1672 © 2024 IEEE

Digital Object Identifier 10.1109/MIS.2024.3366669

Date of current version 10 April 2024.

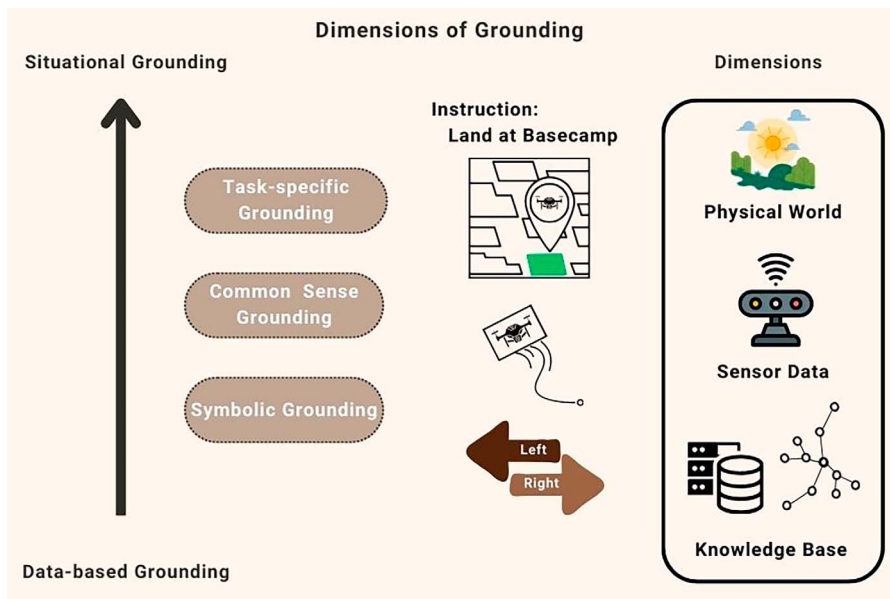


FIGURE 1. Grounding has multiple dimensions and needs knowledge at different levels of abstractions. (This is similar to the need for linguistic, common sense, world model, and domain knowledge for language understanding; see <https://bit.ly/KiLU>, Fig. 3). Grounding may occur at different levels in the task execution process. For example, the drone agent must be able to comprehend the instructions in a natural language format. Next, the instruction may require parsing relevant symbols (in the case of neurosymbolic methods) for reasoning processes. The information extracted must then be grounded to the drone’s capabilities (common sense grounding). Finally, the instruction must be grounded in the specific navigation task.

Nevertheless, different agents may reason with different abstractions, creating a divergence that requires repair. A more recent perspective, *enactivism* values the role of action, embodiment, and environment. Robotic agents can potentially obtain grounding by physically linking to an environment through sensory input and motor output. Agent functions can be either engineered or learned. With meticulously engineered grounding, systems may demonstrate the “correct” behavior, but their internal mechanisms are not inherent to the system. Alternatively, an agent function can be acquired by adjusting connection weights instead of requiring programming. However, the definition of a correct agent function and the question of how to evaluate various agent functions remain.

Given our interest in real-world human–AI interactions and multimodal systems, we emphasize enactivist grounding. However, cognitivist grounding remains essential, as we summarize from Barsalou¹³:

Mental imagery, cognitive grammar, mental spaces, and compositional reasoning support explanations for thought. Neuroimaging shows that higher cognition is realized in the brain’s *model systems*. One theory is that *grounding*

mechanisms serve as an interface, peripheral to core cognitive operations.

Barsalou predicts that future cognitive research will integrate classic symbolic architecture, statistical/dynamic systems, and grounding cognition. Formal and computational accounts of grounding will shift from epiphenomenal to casual. Grounding mechanisms may potentially replace the amodal mechanisms in cognitive architectures. We believe that advancements in neurosymbolic AI will be instrumental here. Next, we describe the grounding problem from a machine learning (ML) and AI perspective and highlight what is missing from a cognitive science lens.

AI LENS

Grounding in AI has typically been referred to as connecting concepts to other knowledge bases (KBs) or world models. We first discuss the different grounding efforts in the AI¹¹ community. We then draw parallels with the cognitive science community^{13,15} and identify some nuanced distinctions.

Chandu et al.¹¹ note the use of *static* versus *dynamic* terminology used in the NLP/CV context, similar to the *cognitivist* versus *enactivist* distinction in

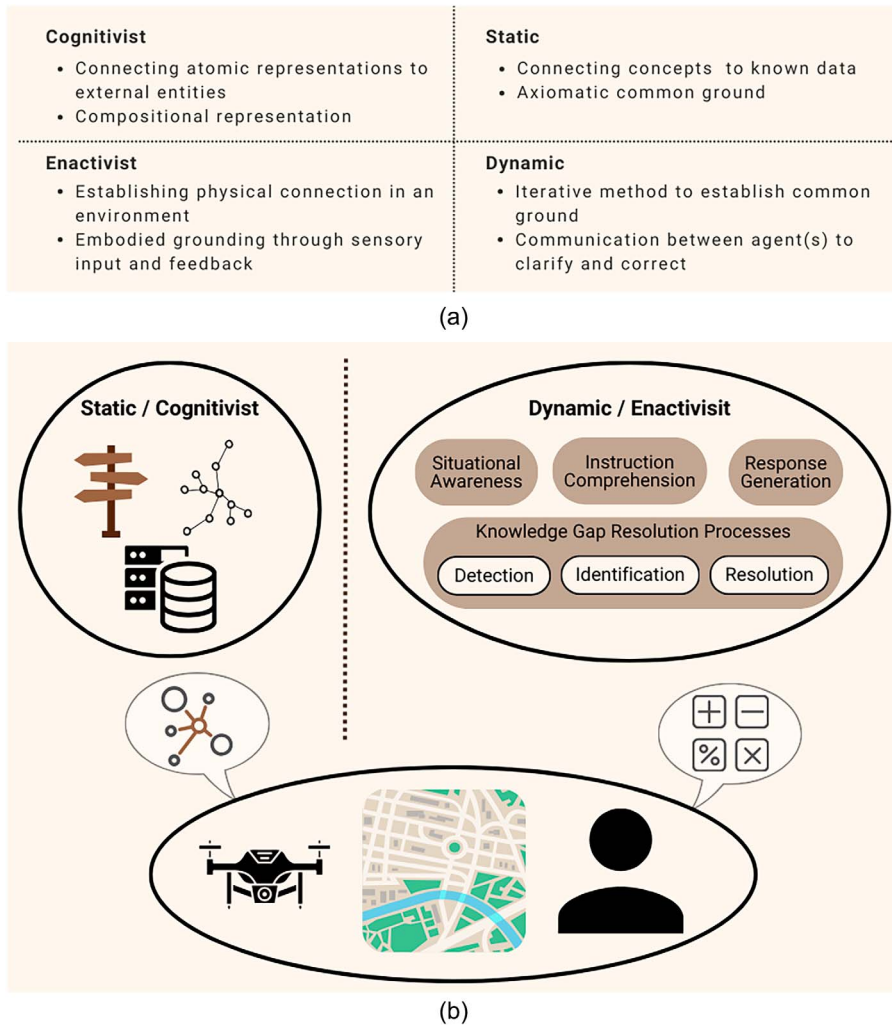


FIGURE 2. Types of grounding. (a) The *static* and *dynamic* definitions in neurosymbolic AI loosely mirror *cognitionist* and *enactivist* grounding in cognitive science, respectively, although there are subtle differences. Cognitive scientists tend to focus more on the specific monitoring and repair processes in the case of different perspectives between interacting agents, whether due to differences in static knowledge or engagement opportunities with an open world. (b) For an AI or cognitive agent to interact effectively with its environment, it must understand the language used by external agents, accurately assess the current situation and context, and identify and address gaps in its knowledge.⁴ This process involves recognizing when it lacks understanding or has difficulty with language comprehension and communicating these knowledge gaps to other agents or external sources. By utilizing these processes, agents can update their knowledge of the task and environment, avoiding catastrophic errors and allowing for the iterative and interactive aspects of dynamic grounding. Consider a machine agent, like a drone, working with a human operator to survey a specific area. The agent can use its background (stored) knowledge to understand the monitored environment and the mission's purpose (static/cognitionist grounding). Suppose the agent comes across something it has not seen before and cannot identify. In that case, it will interactively and iteratively ask the human operator for assistance in closing this gap (dynamic/enactivist grounding).

cognitive science (see Figure 2). Static grounding—the most predominant form—relies on accessible evidence supporting the *common ground* to connect concepts within a given context to the real world.¹¹ For static grounding, common ground is typically through an

agent (e.g., machine) interacting with a static KB to frame a response and deliver it back to the other agent (e.g., human). Both agents may share this common ground by assuming its universality, i.e., no external references. The success of grounding is measured based

on the agent's ability to link the query to the available data. In contrast, dynamic grounding establishes common ground iteratively, where both agents can communicate to seek and provide clarifications, typically in a potentially changing physical environment. This allows corrections to misunderstandings and involves several rounds of clarification and acknowledgments.

Static Grounding

According to Harnad, manipulating symbolic representations without meaning cannot support reasoning. AI researchers responded with different representations for different uses. Typically, these frameworks have the following key components: the *designator*, denoting the name or symbol utilized to identify the category; the *epistemological representation*, employed to recognize instances of the category; and the *inferential representation*, comprising "encyclopedic" knowledge about the category and its members. The *epistemological representations* are termed *concept descriptions*. In CV, these representations are considered object models.⁵

This form of grounding is typically established using deductive learning. It is consistent with cognitivist grounding, with little to no active or online supervision from the environment or an external agent. ML efforts for static grounding include using entity slot filling, adversarial references to grounding visual referring expressions, visual semantic role labeling, and disambiguation of concepts and entities.¹¹ Additionally, methods designed for manipulating representations include fusion and concatenation, representation alignment, and projection of representations into a shared space. Finally, the ML community has designed different learning objectives to address the grounding problem, including multitasking and joint training, the design of new loss functions, and adversarial learning methods.

Dynamic Grounding

Dynamic efforts in ML are typically designed for situations where an entity in an environment is matched with an epistemological representation that activates a larger knowledge structure containing the composite concept representation. Such systems learn to ground their own experience dynamically in the environment, creating more robust capabilities not dependent on preprogrammed representations.⁵ Grounding frameworks, such as *learning from example* and *learning by conversation*, are consistent with enactivist grounding in cognitive science. Efforts in the ML community for dynamic grounding include grounding embodied agents; natural interactions with human-in-the-loop feedback;

and, more recently, grounding large language model-based agents (<https://rb.gy/2pfq1g>). Nevertheless, grounding in mainstream ML exploits *deductive learning*.⁵

Limitations

One notable omission in ML is grounding with sensors or environmental data.¹ Furthermore, the study of *latent pragmatics* is also missing from the work of Chandu et al.¹¹ Pragmatic analysis, pioneered by Austin, Grice, and Searle and extended by others, such as Sperber, focuses on understanding functional intentions and implications based on variations in linguistic content across different contexts.

The lack of a consistent definition of grounding creates considerable ambiguity regarding *how to ground*, *what to ground*, and *where to ground*. To bridge this gap, Chandu et al.¹¹ outline the following grounding stages: stage 1, localization; stage 2, external knowledge; stage 3, common sense; and stage 4, personalized consensus. However, these stages are insufficient. Consider a challenging example of grounding a drone (with an AI-based decision-making model) to ensure safety concerns.

Localization requires that the drone accurately determine its position relative to its surroundings. Next, grounding with external knowledge introduces additional information, such as weather and airspace regulations, that inform subsequent action. However, which external knowledge sources should be used and prioritized in the decision making? Stage 3, common sense, could include avoiding obstacles in its flight path and taking proactive measures to avoid risks or mitigate potential harm. However, it is not easy to quantify and measure this grounding stage. Stage 4, personalized consensus, can help ground decisions to the drone's perception of the environment and prior experiences. However, it is unclear whether the drone must adhere to its prior experiences when they conflict with human instructions that reflect a broader understanding of the situation.

While a helpful initial attempt, these stages are not sufficiently specific to standardize how grounding occurs and at what abstraction levels. The lack of definition allows for creative interpretation of the problem and, therefore, new tasks, datasets, and methods. However, in the future, the community will benefit from a standardized, comprehensive definition.

NEUROSymbOLIC GROUNDING

Neurosymbolic methods can benefit grounding by integrating traditional symbolic reasoning approaches with the generalization capabilities of neural networks.

Neurosymbolic systems integrate the statistical learning capabilities of neural networks with the structured, symbolic representations used in classical AI. Neurosymbolic AI seeks to benefit from the synergy of symbolic and neural methods. Traditional symbolic reasoning methods use formal languages (e.g., Planning Domain Definition Language) for reasoning over knowledge stored in a structured format and represented by “symbols.” These reasoning methods manipulate and infer from structured symbolic representations, such as logic-based rules, knowledge graphs, or ontologies.

The symbolic representations provide a transparent and interpretable framework for knowledge representations and logical reasoning, but these systems are brittle and often cannot be generalized. However, combined with neural methods, traditional methods leverage advancements in deep learning to acquire knowledge representations and enhanced generalization capabilities effectively.¹⁰ Neural networks consist of interconnected layers of *artificial* neurons that use weighted connections, enabling them to learn complex mappings between inputs and outputs. These networks are instrumental in pattern recognition, classification, regression, and sequence prediction. Therefore, neurosymbolic methods leverage the strengths of each paradigm.

As noted earlier, many of the grounding efforts in ML rely on deductive learning and lack active or online supervision from the environment or an external agent. Neurosymbolic methods for grounding can offer several advantages, including compositional reasoning and situational awareness.

One particular asset of neurosymbolic methods is the use of functional modules. Natural language texts (i.e., instructions or queries) are mapped to functional modules that carry out atomic actions. These functional modules can be user defined or learned. This allows agent functions grounded in symbolic representations to complete specific actions or generate responses. Additionally, the compositional nature of these functional programs allows for generalization to new combinations of parsed instructions or queries. The functional modules can be used for both *dynamic* and *static* grounding, as the modules can operate over KBs (including knowledge graphs). This aspect of neurosymbolic methods can help establish common ground, enabling agents to interpret and execute instructions in a manner that aligns with symbolic human reasoning. For example, Mao et al.,¹⁴ designed a new neurosymbolic concept learner that can learn embeddings for symbolic visual inputs. Learning these mappings allows for continual learning and adaptation of new environmental variables while preserving an agent’s task behavior. In this manner, neurosymbolic

approaches can allow for the grounding of new concepts in an environment and facilitate knowledge gap detection, identification, and resolution processes, leading to adaptive and robust models.

CONCLUDING REMARKS

We have focused on explicit notions of grounding. We conclude with remarks on implicit forms of grounding. Recently, the use of digital twins has emerged to augment the performance of ML systems in domains ranging from autonomous transportation⁶ to next-generation wireless communication.⁷ Digital twins can provide a high-fidelity representation of physical entities by accurately modeling the structure, behavior, and characteristics of a real-world system (world model) governed by physical laws. Such ideas can expand the training dataset (out of distribution), allowing models to be generalized. This implicit grounding in physical laws can prevent nonfactual generation, reduce hallucinations, and anchor model responses to specific information, facilitating harmonization with the corresponding world model. We note, regarding using digital twins for grounding parallels, the contention of Pickering and Garrod⁸: that interlocutors implicitly comprehend each other by aligning their models of the discussed situation at various levels of cognitive and linguistic representation. Such implicit alignment processes between agents (akin to digital twins) inspire computational grounding processes.

Finally, we advocate for more knowledge-infused neurosymbolic learning and reasoning systems that naturally integrate linguistic, common sense, general (world model), and domain-specific processes and knowledge to facilitate static grounding. We expect fundamental progress on the synergistic use of neural networks and structured semantics to advance from content processing to content understanding and reasoning with the infusion of symbolic knowledge (<https://rb.gy/67wgx3>). These methods require dynamic knowledge-elicitation strategies integrating multimodal pragmatic context and interactions with domain experts in the loop to achieve dynamic grounding and alignment with user intent. A version of this article with additional references is available in the IEEE Computer Society Digital Library at <https://arxiv.org/abs/2402.13290>.

ACKNOWLEDGMENTS

The authors acknowledge National Science Foundation (NSF) Award 2335967 (to Amit Sheth and Valerie L. Shalin) and NSF Award CNS-2112471 (to Goonmeet Bajaj and Srinivasan Parthasarathy).

REFERENCES

1. A. Cangelosi and T. Riga, "An embodied model for sensorimotor grounding and grounding transfer: Experiments with epigenetic robots," *Cogn. Sci.*, vol. 30, no. 4, pp. 673–689 2006, doi: [10.1207/s15516709cog0000_72](https://doi.org/10.1207/s15516709cog0000_72).
2. W. Horton and B. Keysar, "When do speakers take into account common ground?" *Cognition*, vol. 59, no. 1, pp. 91–117, Apr. 1996, doi: [10.1016/0010-0277\(96\)81418-1](https://doi.org/10.1016/0010-0277(96)81418-1).
3. J. Ball et al., "The synthetic teammate project," *Comput. Math. Org. Theory*, vol. 16, no. 3, pp. 271–299, 2010, doi: [10.1007/s10588-010-9065-3](https://doi.org/10.1007/s10588-010-9065-3).
4. G. Bajaj et al., "Knowledge gaps: A challenge for agent-based automatic task completion," *Topics Cogn. Sci.*, vol. 14, no. 4, pp. 780–799, 2022, doi: [10.1111/tops.12584](https://doi.org/10.1111/tops.12584).
5. P. Davidsson, "Toward a general solution to the symbol grounding problem: combining machine learning and computer vision," in *Proc. AAAI Fall Symp.*, 1993, pp. 157–161.
6. A. McClellan et al., "A physics-based digital twin for model predictive control of autonomous unmanned aerial vehicle landing," *Philos. Trans. Roy. Soc. A*, vol. 380, no. 2229, 2022, Art. no. 20210204, doi: [10.1098/rsta.2021.0204](https://doi.org/10.1098/rsta.2021.0204).
7. A. Banerjee et al., "HORCRUX: Accurate cross band channel prediction," in *Proc. Int. Conf. Mobile Comput. Network.*, New York, NY, USA: ACM, 2024.
8. M. Pickering and S. Garrod, "Toward a mechanistic psychology of dialogue," *Behav. Brain Sci.*, vol. 27, no. 2, pp. 169–190, 2004, doi: [10.1017/S0140525X04000056](https://doi.org/10.1017/S0140525X04000056).
9. H. Branigan et al., "Linguistic alignment between people and computers," *J. Pragmatics*, vol. 42, no. 9, pp. 2355–2368, 2010, doi: [10.1016/j.pragma.2009.12.012](https://doi.org/10.1016/j.pragma.2009.12.012).
10. A. Sheth et al., "Neurosymbolic artificial intelligence (Why, What, and How)," *IEEE Intell. Syst.*, vol. 38, no. 3, pp. 56–62, May/Jun. 2023, doi: [10.1109/MIS.2023.3268724](https://doi.org/10.1109/MIS.2023.3268724).
11. K. Chandu, Y. Bisk, and A. Black, "Grounding 'grounding' in NLP," in *Proc. Findings Assoc. Comput. Linguistics (ACL-IJCNLP)*, 2021, pp. 4283–4305.
12. R. Nuñez et al., "Embodied cognition as grounding for situatedness and context in mathematics education," *Educ. Stud. Math.*, vol. 39, pp. 45–65, Jun. 1999, doi: [10.1023/A:1003759711966](https://doi.org/10.1023/A:1003759711966).
13. L. Barsalou, "Grounded cognition: Past, present, and future," *Topics Cogn. Sci.*, vol. 2, no. 4, pp. 716–724, 2010, doi: [10.1111/j.1756-8765.2010.01115.x](https://doi.org/10.1111/j.1756-8765.2010.01115.x).
14. J. Mao et al., "The neuro-symbolic concept learner: Interpreting scenes, words, and sentences from natural supervision," in *Proc. Int. Conf. Learn. Representations*, 2019.
15. T. Ziemke, "Rethinking grounding," in *Understanding Representation in the Cognitive Sciences: Does Representation Need Reality*, A. Riegler, M. Peschl, and A. Stein, Eds., Boston, MA, USA: Springer-Verlag, 1999, pp. 177–190.

GOONMEET BAJAJ is a Ph.D. candidate at The Ohio State University (OSU), Columbus, OH, 43210, USA. Contact her at bajaj.32@osu.edu.

VALERIE L. SHALIN is a professor of psychology at Wright State University, Dayton, OH, 45435, USA. Contact her at valerie.shalin@wright.edu.

SRINIVASAN PARTHASARATHY is a professor of computer science and engineering and codirects the data analytics program at OSU, Columbus, OH, 43210, USA. Contact him at srini@cse.ohio-state.edu.

AMIT SHETH is the founding director of the AI Institute and NCR professor of computer science and engineering at the University of South Carolina, Columbia, SC, 29208, USA. Contact him at amit@sc.edu.