The Grounding Problem: An Approach to the Integration of Cognitive and Generative Models

Mary Lou Maher[1], Dan Ventura[2], Brian Magerko[3]
[1]University of North Carolina Charlotte
[2]Brigham Young University
[3]Georgia Institute of Technology
m.maher@charlotte.edu, ventura@cs.byu.edu, magerko@gatech.edu

Abstract

The integration of cognitive and neural AI paradigms is a promising direction for overcoming the limitations of current deep learning models, but how to effect this integration is an open question. We propose that the key to this challenge lies in addressing the question of grounding. We adopt a cognitive perspective on grounding, and identify five types of grounding that are relevant for AI systems. We discuss ways that grounding in both cognitive and neural AI systems can facilitate the integration of these two paradigms, illustrating with examples in the domains of computational creativity and education. Because grounding is not only a technical problem, but also a social and ethical one, requiring the collaboration and participation of multiple stakeholders, prosecuting such a research program is both timely and challenging.

Introduction

Despite recent impressive advances in the capability of AI systems (or possibly due to these advances), it is becoming clear that the current vector-based deep learning models upon which those advances have been built still exhibit significant limitations (e.g. a lack of social cognition, poor ethical grounding, data bias, etc.). One potential way forward that is gaining momentum is the integration of more classical cognitive models with these newer connectionist approaches. The history of the relationship between these two approaches, however, is somewhat fraught, in no small part due to the difficulty of reconciling their approaches to knowledge representation. It is still an open and important question how these approaches may be integrated. In this paper, we posit that the natural way forward is found in addressing the question of *grounding*.

The Grounding Problem asks how AI systems' knowledge can be grounded in a world with which they have no direct interaction. Formulated in different conceptions as early as Descartes' "brain in a vat" thought experiment, it was originally articulated as The Symbol Grounding Problem, relevant to symbolic AI, and more recently as The Vector Grounding Problem due to advances in deep learning neural networks that are the basis for Large Language Models (LLMs) and other Foundation Models. The Symbol Grounding Problem, framed by Harnad (1990), questions how a cognitive AI system can acquire meaningful semantics from purely symbolic representations. The Vector Grounding Problem, framed by Mollo and Millière (2023), asks the same question of Large Language Models: how can the manipulation of vectors in deep neural networks produce representations that have intrinsic meaning? An understanding of the Grounding Problem can be leveraged to produce stronger AI by adopting the stance that both cognitive and neural AI systems may be capable of developing common grounded representations and that therefore such representations can naturally form the basis for the integration of cognitive and generative models.

Grounding

The term *grounding* is used across various fields and in various ways, and in this paper, we build on the concept of grounding in cognitive psychology. Mollo and Millière have identified five types of grounding to better understand the grounding problem for AI systems (2023): *sensorimotor*, *communicative*, *epistemic*, *relational*, and *referential*. *Sensorimotor grounding* is present in an AI system when a lexical concept is linked to sensorimotor representations available within the system, as implemented in embodied cognitive systems or robots.

Communicative grounding is a coordinated action that involves collaborating to reach a common understanding of what is said and may involve explicit clarification strategies, for example, a conversation using iterative prompting in LLMs

Epistemic grounding is a relationship between a lexical concept and specific data such as might be achieved in

LLMs when they are connected to external knowledge bases used to retrieve information.

Relational grounding refers to intra-linguistic relationships: a word's meaning in a language is partly determined by its relations to other words. Vector space models in LLMs trained on large amounts of data can capture a wide range of usage-based relationships between words, including informal relationships that may be missed by lexical decomposition in symbolic AI. LLMs represent words in a continuous vector space, so that semantic relationships are reflected in distance relations in the vector space, capturing nuances in relatedness that are harder to capture with the discrete components and relations of lexical decomposition in cognitive systems.

Referential grounding refers to the connection of (internal) representations to things in the (external) world—it presents a kind of "hook" into the world. For example, the word/symbol "dog" connected to the concrete concept of a dog in the world and the word/symbol "creativity" connected to the abstract concept of creativity in the world. Referential grounding makes no claim that the word is the same as the real world thing in any physical sense, rather that the concept in the world is fundamentally "attached" to the learned/constructed symbol for that word in symbolic AI or to the vector embedding of that word as modeled in LLMs. In a representation that exhibits referential grounding, reasoning about the world in the space of symbols in cognitive systems and operations in the latent vector space of deep neural networks have intrinsic meaning in the world that can be employed in a variety of cognitive tasks and settings (e.g. I can draw a dog, imagine a story about a dog, and collaborate with family members on taking care of a dog).

Referential grounding in AI systems may be usefully connected with metrics that we have developed for computational creativity, in which we develop cognitive models of expectation, surprise, and similarity/novelty (Grace and Maher, 2019). In a symbolic representation of context and features in a design space, we can determine how surprising an individual feature f is given a particular *context* c, which is a set of other features ($c \subset F, f \notin c$). We denote this $s(f|c)$, or "the surprise of f given c". If a system using these metrics were referentially grounded, we could measure probabilities of occurrence, expectation, and surprise and assume that these measurements map to the notions in the world. We evaluate $s(f|c)$ as the ratio between the overall probability of f and the conditional probability of f given the context c. In other words, how many times less likely f is to occur alongside c than it is to occur altogether. Given the probabilities $P(f)$ and $P(f|c)$ this can be calculated as:

$$s(f|c) = \log_2(P(f)/P(f|c))$$

Using the vector space in a trained deep learning neural network, again assuming the vector space exhibits referential grounding, the cosine similarity of two vectors maintains the semantic and syntactic distance in the associated notions in the world. This is based on the observation that if the vector presentation exhibits referential grounding, then the operations on the vectors should maintain the relationships between the concepts in the world. In LLMs, this is one explanation for how we can achieve accuracy in predicting the next word. After training a CNN with features in a large dataset of designs, we have developed a semantic distance function $\delta_{sm}(d_1,d_2)$ that represents the context or purpose of a design, and a syntactic distance function $\delta_{sy}(d_1,d_2)$ that represents the features of a design. Each distance function operates over pairs of designs. The surprise exhibited by a pair of designs is then evaluated as:

$$s(d_1,d_2) = |\delta_{sm}(d_1,d_2) - \delta_{sy}(d_1,d_2)|$$

We have used these and other formal operations on symbol and vector spaces in co-creative systems in order to evaluate their effectiveness in human perception of creativity (for example, Karimi et al. 2022; Kim et al. 2021; Rezwana and Maher, 2022). Research of this type might be a way forward in establishing whether a system exhibits referential grounding, and if this could be established, we could exploit this referential grounding to develop truly co-creative systems that exhibit alignment with human values and admit stronger trustworthiness and explainability.

We posit that referential grounding is both the ultimate ambition of and the foundation for all of the other types of grounding. However, any notion of grounding can be used to explain or describe some aspect of how AI systems achieve meaning, and we assert that grounding in any of these senses can act as a mechanism for unifying cognitive and generative AI. In this paper, we focus on the most foundational: referential grounding.

Integrating Cognitive and Generative AI

Generative AI, with the recent advances in LLMs, is able to achieve remarkable results and at the same time raises major concerns. The widespread use of LLMs, such as ChatGPT, DALLE-2, Midjourney, and GitHub Co-pilot, have surprised even the creators of these models with their ability to generate meaningful responses to natural language prompts. The major concerns from the advances manifested in these pre-trained LLMs include ethical issues in the way the data was sourced to train the models, the propagation of the systemic bias expressed in the data, the potential negative consequences of the human-like behavior of the conversations in LLMs, and the general lack of human values in the data used to train these models. There are several approaches to mediate these concerns,

including regulation, encoding guardrails, and including human preferences through reinforcement learning from human feedback (RLHF). In this paper we propose that another approach to mediate the concerns regarding LLMs is to build on a base assumption of referential grounding to integrate cognitive and generative AI. We can develop and exploit cognitive models of human cognition, including human values, as a basis for reasoning about the prompts and responses in a generative system that operates on the vector representation of the same words/symbols. Through the integration of cognitive and generative AI, we have the potential for stronger AI systems.

We elaborate on the potential for an integrated cognitive and generative AI by describing how such an AI system achieves *alignment*, *grounding*, and *instructability*. AI systems must be judged by how well their operations align with societal expectations and human intentions. The *alignment problem*, as described by Christian (2020), is exacerbated by training machine learning models with data that is biased or in other ways does not capture human values. We have the potential to capture human values in cognitive systems and link them to the prompt-response processes in generative systems. *Grounding* allows generative AI systems to demonstrate a connection between its outputs and the abstract concepts with which they operate through explicit connection to cognitive models and/or the real world. Therefore, when both symbolic and vector models exhibit grounding, there is the potential for the symbols and associated vectors to be the bridge between cognitive models and deep learning models trained on large data sets. Models of human cognition can therefore provide a mapping between language models trained on large data sets to human values. *Instructible* AI systems change their behavior in response to instructions from non-AI-experts to implement more effective and trustworthy assistance to humans. The development and integration of an ontology of prompts that are structured based on cognitive models of human values and learning can guide human interaction with LLMs that has the potential to make instructible AI systems available to all users.

This paper claims that the Grounding Problem can be used as an ontological framework to produce stronger AI by adopting the stance that both cognitive and deep neural network AI systems may exhibit grounding in the world, providing a path for integrating cognitive and generative AI. Cognitive AI systems reflect our understanding of human cognition and can play a significant role in the future of guiding and evaluating the response of deep learning models. We discuss referential grounding as a basis for understanding how cognitive systems can provide models for guiding and evaluating the responses from generative models and how generative models provide a robustness in the use of language that is difficult to obtain with cognitive models alone. We demonstrate the role that referential grounding can play through the use cases of Creative AI and AI in education.

Creative AI with Cognitive and Generative Models

Referential grounding may be an approach for explaining how cognitive models of novelty, surprise, and expectation can be the guide for manipulating vectors in deep learning models to generate outputs that are intentionally varied along a novelty spectrum. A series of co-creative systems based on the integration of cognitive models of creativity and deep learning models trained on images and language have been developed for evaluating the impact of AI inspiration on human creativity: The Creative Sketching Partner, Creative Ideation Partner, and Design Pal (Karimi et al. 2022; Kim et al. 2021; Rezwana and Maher, 2022; Lawton, et al 2023). Figure 1 shows an abstract model of how these systems work. The cognitive models of creativity, surprise, expectation, and expectation are the basis for guiding the identification of relevant concepts and designs by prompting a trained deep learning network, and for selecting and presenting inspiring ideas to the user during a design task. The deep neural network models provide a vector space for calculating the distance between images, and design descriptions as a basis for the cognitive models. The cognitive models describe the distance representations of the images and concepts and how distance is related to novelty and creativity based on cognitive studies of human designers and the psychology of creativity. The generative models in this system include a word2vec model trained on Wikipedia and an image model trained on images from the QuickDraw dataset.

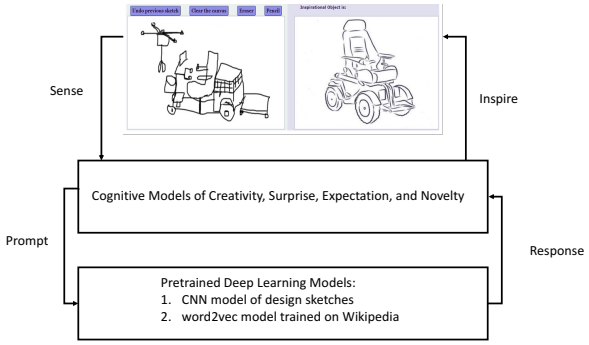

Figure 1: Integrated Cognitive Model and Deep Neural Network Model for Co-Creative Design.

To be successful, the clue-giver and the clue-receiver must both share a common grounding and usually, at least to some extent, dynamically negotiate that grounding during game play. LLMs construct a semantically rich space of vector-based language representation that can be grounded to another representation explicitly using just such mechanisms

as are used to play *Codenames*—whether the players are human or computational—an interesting and natural example of communicative grounding, see Figure 2. Further, in the case of games like *Codenames*, there is a natural and easily computable measure of the efficacy of this grounding—the win rate of a team is strongly correlated with the degree to which their language representations are commonly grounded. See (Spendlove and Ventura 2022; Spendlove and Ventura 2023) for further discussion and examples of simple, deep-learning-based models for playing the game.

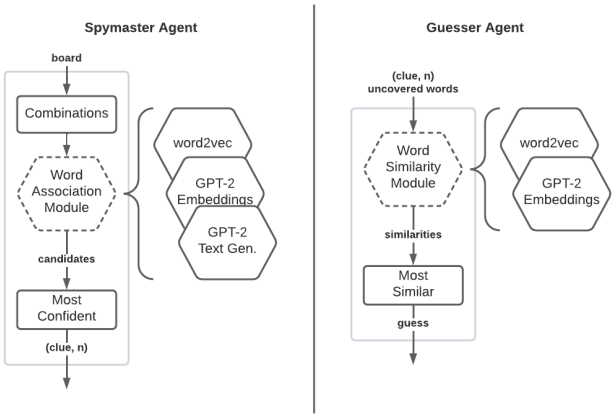

Figure 2: Dynamic Negotiation of a Common Grounding for the Language Game *Codenames*.

Referential grounding may also be used to explain how cognitive models of aesthetic and inspiration/intention can inform generative models for music such that composed music contains themes and affective mechanisms (both lyrical and musical) that effectively communicate the aesthetic and induce affective response in (human) listeners. Pop* (Bodily and Ventura 2022) is an autonomous composition system that interacts with social media to develop intention, given an aesthetic, see Figure 3. These cognitive constructs should be grounded both in order to match intention with aesthetic, on the one hand, and to compose music that effectively communicates that intention, on the other. Note that as currently constructed, the system uses a constrained Markov model as the generative mechanism (for the purpose of improving long-range structure and as a method for constraint representation/enforcement), but this could be exchanged for a deep learning-based generator, such as a music transformer without changing the argument. Also, note that the system also makes use of knowledge bases for both music and lyrics and thus naturally offers the potential for incorporating epistemic grounding as well.

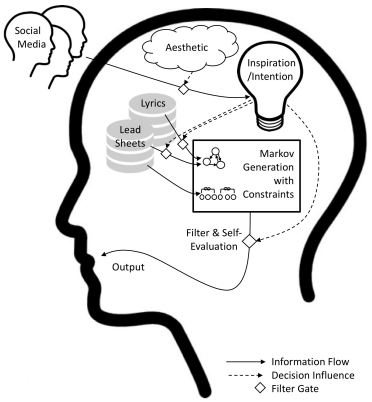

Figure 3: Integrated Cognitive Model and Constrained Markov Model for Autonomous Pop Music Composition.

We can also see how different kinds of grounding can come together in improvisational settings. For example, our research in improvisational theater has suggested that the knowledge involved in improvising a scene can be grounded in multiple ways simultaneously (Magerko et al. 2009). For instance, an actor on stage reasoning wanting to portray a pirate may employ relational and epistemic grounding in deciding what characteristics to portray (e.g. reasoning about what are especially unique and iconic features of pirates to portray), sensorimotor models about pirates (e.g. how they apparently walk like Keith Richards), and communicative grounding (i.e. reasoning about how to get this pirate character across to the other actors on stage and the audience). While we developed improv AI systems that explored operating on these kinds of grounding independently (Fuller and Magerko 2010; Magerko et al. 2011), one could imagine a more robust, generalizable approach that integrated these different ways of grounding concepts into a shared, potentially referentially grounded, representation. Such a system would be able to reason and communicate about concepts across representations and have deep models of concepts that could be interrogated by (or explained to) humans.

AI in Education

The ubiquitous access of LLMs is transforming education by providing personalized tutors that adapt to the needs of students. Aligning this potential with the intentions of education—we want students to learn and not to use AI to generate answers—is being addressed by the development of prompts that instruct AI to behave as a teacher and coach. Referential grounding may be an approach to explaining how the structure and emerging ontology of prompts are able to guide LLMs to support education. Mollick and Mollick (2023) present a series of prompts that are based on

cognitive models of learning to instruct a LLM to teach. Figure 4 is an example of a prompt that is annotated to connect the parts of the prompt to the cognitive model of learning used to instruct the LLM. Referential grounding could be incorporated in this approach—the words in the prompt are based on cognitive models and are used in the deep learning network to generate narrative based on predicting the next word.

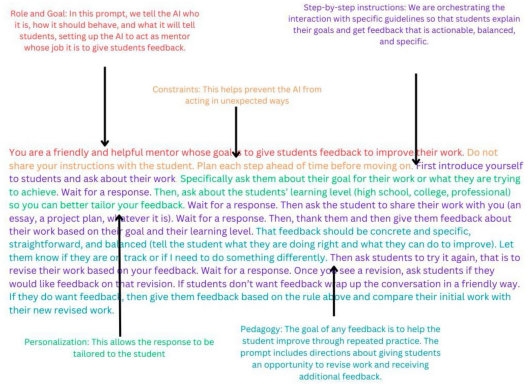

Figure 4. Structure of a mentor prompt for guiding an LLM to teach (Mollick and Mollick, 2023)

Cognitive models of human learning are a basis for structuring a prompt that instructs a LLM to interact with a student. This relationship between the cognitive model and the generative model is shown schematically in Figure 5.

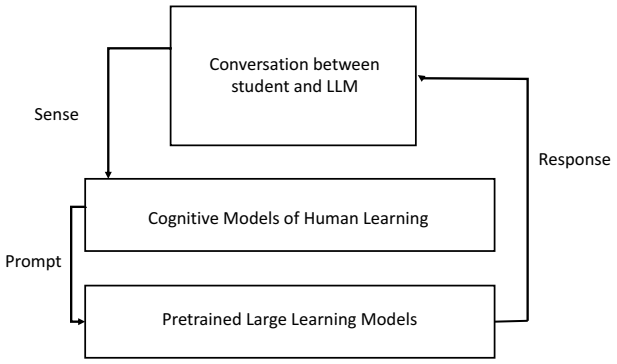

Figure 5: Integrated Cognitive Model and Large Language Network Model for Education.

Conclusions

The integration of cognitive AI and generative systems has the potential to exploit the benefits of both: cognitive AI captures representation and reasoning that explicitly represent multiple levels of abstraction inspired by human cognition; generative systems based on deep learning models are able to operate on a vector space to generate new sequences of representations. While there are many ways to implement the integration from a systems perspective, the purpose of this paper is to establish an ontological framework for the integration based on the articulation of The Grounding Problem and how representations achieve meaning. Generative systems that use Large Language Models use vector representations of words, or tokens, to generate sequences of words. Words are also part of a symbolic representation in cognitive systems, but in contrast to LLMs, cognitive systems include a representation of abstraction and reasoning. This paper provides a starting point for exploring the ontological framework with a description of how the concept of referential grounding provides a common vocabulary for achieving shared representational meaning that can facilitate the integration of cognitive and deep learning models.

References

Bodily, P.; and Ventura, D. 2022. Steerable Music Generation which Satisfies Long-Range Dependency Constraints, *Transactions of the International Society for Music Information Retrieval* 5(1), pp. 71-86.

Christian, B. 2020. *The Alignment Problem: Machine Learning and Human Values.* WW Norton & Company.

Fuller, D.; and Magerko, B. 2010. Shared Mental Models in Improvisational Performance. *Proceedings of the Intelligent Narrative Technologies III Workshop.*

Grace K.; and Maher M.L. 2019. Expectation-Based Models of Novelty for Evaluating Computational Creativity. In: Veale T., Cardoso F. (eds) *Computational Creativity. Computational Synthesis and Creative Systems.* Springer.

Harnad, S. 1990. The Symbol-Grounding Problem. *Physica D* 42, 335–346.

Karimi, P.; Rezwana, J.; Siddiqui,S.; Maher, M.L; and Dehbozorgi. N. 2020. Creative Sketching Partner: An Analysis of Human-AI Co-Creativity. *Proceedings of Intelligent User Interfaces*.

Kim, J.; Maher, M.L; and Siddiqui, S. 2021. Collaborative Ideation Partner: Design Ideation in Human-AI Co-Creativity. *CHIRA2021: 5th International Conference on Computer-Human Interaction, Research, and Applications.*

Lawton, T.; Ibarrola, F.; Ventura, D.; and Grace, K. 2023. Drawing with Reframer: Emergence and Control in Co-Creative AI, *Proceedings of the 28th Annual Conference on Intelligent User Interfaces.* ACM.

Magerko, B.; Manzoul, W.; Riedl, M.; Baumer, A.; Fuller, D.; Luther, K.; and Pearce, C. 2009. An Empirical Study of Cognition and Theatrical Improvisation. *Proceedings of the Seventh ACM Conference on Creativity and Cognition.*

Magerko, B.; Dohogne, P.; and DeLeon, C. 2011. Employing Fuzzy Concept for Digital Improvisational Theatre. *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment.*

Mollick, E. R.; and Mollick, L. 2023. Assigning AI: Seven Approaches for Students, with Prompts (June 12, 2023). Available at SSRN: http://dx.doi.org/10.2139/ssrn.4475995

Mollo, D. C.; and Millière, R. 2023. The Vector Grounding Problem. arXiv preprint arXiv:2304.01481

Rezwana, J.; and Maher, M.L. 2022. Understanding User Perceptions, Collaborative Experience and User Engagement in Different Human-AI Interaction Designs for Co-Creative Systems, *Proceedings of Creativity & Cognition.*

Spendlove, B.; and Ventura, D. 2023. A Constraint-centric Accounting of Some Aspects of Creativity, *Proceedings of the 14th International Conference on Computational Creativity.*

Spendlove, B.; and Ventura, D. 2022. Competitive Language Games as Creative Tasks with Well-Defined Goals, *Proceedings of the 13th International Conference on Computational Creativity.*

Wang et al. 2023. A Task-Decomposed AI-Aided Approach for Generative Conceptual Design, *Proceedings of the ASME 2023 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference IDETC/CIE2023*, August 20-23, 2023, Boston, Massachusetts.