Hybrid Machine Learning/Knowledge Base Systems Learning through Natural Language Dialogue with Deep Learning Models

Sergei Nirenburg[1,*,†], Nikhil Krishnaswamy[2,†] and Marjorie McShane[1,†]

[1]*Rensselaer Polytechnic Institute, 110 8th St. Troy, NY 12180, USA*
[2]*Colorado State University, 1100 Center Avenue Mall, Fort Collins, CO 80523, USA*

Abstract

Neurosymbolic approaches to AI typically involve attempts to reincorporate the structure and speed of symbolic reasoning into the flexible representations of deep learning. "Knowledge," in this understanding, is typically represented in a structured ontology or knowledge base that relies on human expertise and effort to construct. In this paper, we present a vision for "language-endowed intelligent agents," a type of lifelong learner which begins with a hand-crafted knowledge base and a deep language understander and increases it over the course of its life through dialogue and interaction with both humans and other AI systems—generative large language learning models in particular. We discuss the requirements for such a system, present evidence toward the feasibility of the approach, and conclude with future challenges and research directions.

Keywords

Language-endowed intelligent agents (LEIAs), Learning through dialogue between DL model and LEIA, Neurosymbolic AI, Lifelong learning

1. Introduction

The emergence of deep learning made AI today's technology of tomorrow in the eyes of developers, potential users and the general public. Deep learning (DL) models can uncover patterns implicit in enormous collections of data and demonstrate impressive performance on a slew of text processing tasks due largely to improvements in neural language modeling using versions of transformer architectures [1], as implemented in BERT [2], GPT [3, 4] and other model families. Still, they are subject to a number of conceptual and practical limitations related to resource consumption, performance in adversarial settings, and ability to "understand" in a colloquial sense (e.g., [5, 6, 7]). They are simply very efficient methods of mining huge text repositories to find the most probable next word to follow a given word sequence and do not actually understand their inputs or outputs, or the nature of their own processing. This is why

In A. Martin, K. Hinkelmann, H.-G. Fill, A. Gerber, D. Lenat, R. Stolle, F. van Harmelen (Eds.), Proceedings of the AAAI 2023 Spring Symposium on Challenges Requiring the Combination of Machine Learning and Knowledge Engineering (AAAI-MAKE 2023), Hyatt Regency, San Francisco Airport, California, USA, March 27-29, 2023.

*Corresponding author.

†These authors contributed equally.

nirens@rpi.edu (S. Nirenburg); nkrishna@colostate.edu (N. Krishnaswamy); mcsham2@rpi.edu (M. McShane)

CEUR Workshop Proceedings (CEUR-WS.org)

they fail to explain what they are doing and why.[1] This realization – as well as well-known difficulties that DL models experience in adversarial settings [8] and their inability to reason [9] engendered several directions of work to overcome these deficiencies. One example of such a program of work is explainable AI (XAI) [10], a large-scale research program whose aim is to develop models that attempt to explain the decision-making behind the output of DL models by recreating their results using specially developed explanation-oriented models that are based on human-interpretable features.

A general methodology of overcoming limitations of a particular approach is to use it together with another approach in a hybrid system designed to carry out a particular task. A task-oriented methodology involving coexistence of very different processing approaches in a single AI system is more difficult to implement but promises to yield better results than a system that exclusively seeks to exploit a single approach. For example, all pre-semantic processing in our OntoSem language understander [11] is carried out by ML-based subsystems (currently based on Spacy technology[2]). A number of recent proposals have been put forward to combine neural networks and symbolic reasoning. These "neurosymbolic" approaches to AI have so far been understood primarily as a method of using the content, structure and efficiency of symbolic reasoning to boost the performance of DL models [12, 13, 14]. We describe a program of work that also seeks to integrate symbolic and neural net processing. However, the objective of the program we propose is in some sense the inverse of that pursued by the current neurosymbolic approaches.

We propose to use flexible representations, big data orientation and analogical reasoning of DL to overcome the notorious "AI knowledge bottleneck." The complexities and the sheer expense of acquiring knowledge for AI systems have been amply demonstrated and documented [15].[3] Lowering this cost through automation is an attractive option. In what follows, we first describe the infrastructure we developed and demonstrated to facilitate conceptual learning by AI agents capable of understanding the meaning of language inputs. Next, we describe our initial experiments on using DL models to enhance the efficiency of the approach. Finally, we present how we intend to extend the use of DL models in our approach to lifelong agent learning.

2. The bootstrapping infrastructure for overcoming the knowledge bottleneck.

AI agents capable of human-level understanding, reasoning, decision-making and action must rely on vast amounts of knowledge about the world, typically in the form of an ontological model. Additional knowledge is necessary to support the agent's ability to interpret percepts (language, images, etc.) in terms of its ontology. Acquiring such knowledge is notoriously difficult. One option is to obviate the need for knowledge (this was the initial hope of empirical methods). The option we are pursuing is to make knowledge acquisition less expensive by

[1] While not yet the topic of substantial academic research due to its novelty, many researchers and some popular media have raised similar concerns about newer models like OpenAI's ChatGPT.

[2] https://spacy.io/

[3] Much less publicized is the fact that the cost of human participation in developing ML models is not at all negligible [16].

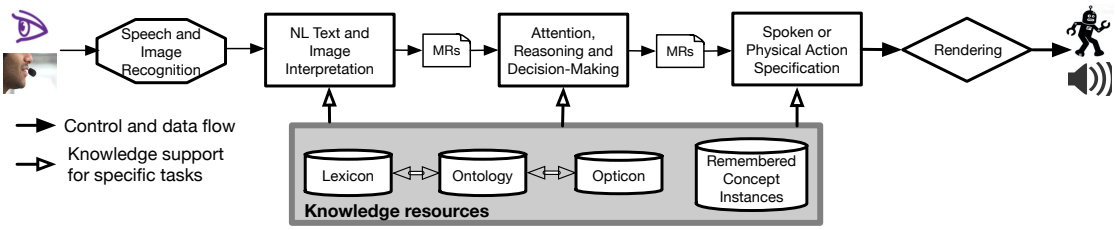

Figure 1: LEIA configuration for routine operation (much detail omitted).

progressively automating it. The architecture of the system (we call such systems "language-endowed intelligent agents," or LEIAs) onto which we want to "graft" this learning capability is illustrated in Figure 1. We develop LEIAs to serve as members of human-AI teams in critical applications where humans must fully trust LEIAs' analyses, conclusions and recommendations, cf. e.g., [17, 18]. As already mentioned, LEIAs include both ML-based and knowledge-based processing modules [19, 20].

The prerequisites to bootstrap the learning capability of LEIAs include the availability of:

- a natural language (NL) understanding system (OntoSem) capable of extracting and representing ontologically grounded meanings of language inputs (text meaning representations, or TMRs) [20, 11, 21, 22];
- an image interpretation system that ontologically interprets results of computer vision, yielding visual meaning representations (VMRs) [23, 24, 25, 26];
- a conceptual learning system that takes TMRs and VMRs as inputs and augments the agent's knowledge resources – an ontological world model, an episodic memory of concept exemplars, a lexicon supporting language interpretation and its counterpart supporting visual perception, the opticon [27, 28].

All three of the above systems used by LEIAs rely on knowledge support – a lexicon for the language analyzer, an opticon (roughly, a set of image-to-concept pairings) for the image interpreter, and an ontology basis for all three. At present, we bootstrap the system with an ontology of ∼160K RDF triples, an English lexicon of ∼30K word senses and a small opticon [29].

The basic idea behind our approach is to exploit the LEIAs' perception interpretation capabilities that are already used in their routine operational configurations and augment their existing reasoning repertoire to include a dedicated learning system that will generate novel, and enhance/modify existing, knowledge elements. This will in turn enhance the coverage and precision of the LEIA's task-oriented perception interpretation and reasoning in a lifelong virtuous circle.

Knowledge acquisition has been a central concern in LEIA development for a long time [30, 31, 32, 33]. Early efforts concentrated on data analytics support, and ergonomics of manual acquisition [34].

Once the minimal bootstrapping prerequisites for learning through language understanding have been developed, we configured several proof-of-concept systems to demonstrate ontology and lexicon learning by reading [35] and through dialogue with a human user [28]. We implemented both a deliberate learning mode, where the LEIA knows its inputs are intended for this

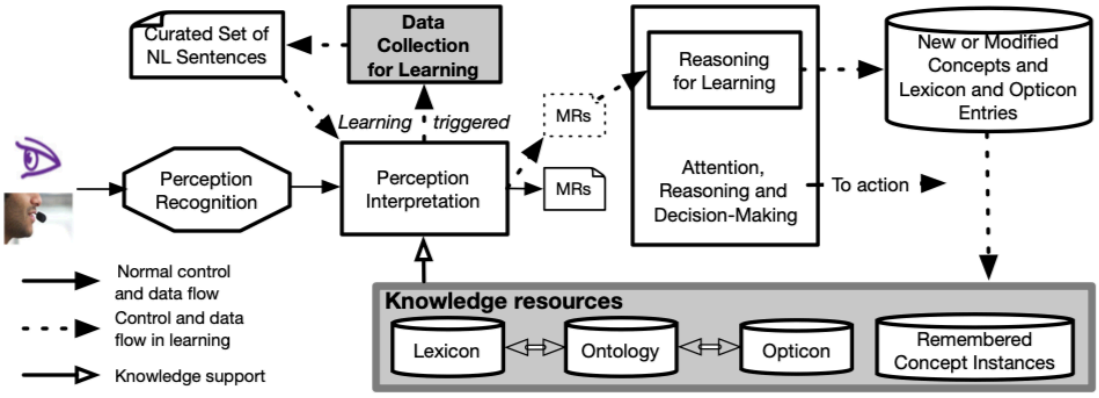

Figure 2: LEIA configuration that incorporates opportunistic learning based on language understanding. Learning is triggered when perception interpretation fails. Up till now, data collection was performed through old-style data analytics and by engaging humans in dialogue. The new configuration will use DL models for this purpose. This learning environment may also be configured with human validation of newly learned content.

purpose, and an opportunistic mode, where learning is co-exists with task-oriented operation. In this mode, if learning is necessary for performing a task at hand, then it is scheduled right away, while the task operation is suspended. If the LEIA derives an actionable TMR for a task-related input even if it cannot interpret the input completely [36], then the learning is postponed until downtime. Figure 2 illustrates a version of LEIA architecture incorporating "opportunistic" learning triggered when LEIAs encounter difficulties in interpreting sensory inputs during routine operation. These difficulties are typically made manifest through lacunae in the lexicon – missing or underspecified lexicon entries, which typically signifies lacunae in the underlying ontological world model [37].

Manual data collection for learning still requires large amounts of expensive human time. In order to gradually alleviate this inefficiency, we have started to experiment with using generative DL models for this purpose instead of either humans or older-style data analytics. As is well known, current DL models may still generate incomplete or fallacious outputs. This is why it might be prudent in a complete system to retain old-style data analytics and human interaction support for the automatic process of learning knowledge content.

3. How LEIA Learning Works: An Example

Suppose, a LEIA receives the input: *Systemic sclerosis is a multisystemic autoimmune disease of unknown origin that affects connective tissue*, where all the the lexical material but the words *systemic* and *sclerosis* are already present in its lexicon. OntoSem creates a skeleton lexicon entry for *systemic* and, because nouns can refer to either objects or events, two skeleton entries for *sclerosis* (Figure 3).

The '?' mark on the entry heads signifies that they are still being learned. The presence of meaning procedure calls signifies that the semantics of their arguments is underdetermined and an attempt must be made to further specify them at runtime. Next, the basic semantics module

of OntoSem generates a set of candidate TMRs (Figure 4 illustrates the correct candidate). A few comments are in order: *Autoimmune disease* is a known collocation, recorded in the lexicon as the 9th verbal sense of *disease*, which requires *autoimmune* as a modifier.

```
systemic-adj1?
  definition
  example          systemic sclerosis
  comments         auto-learned
  syntactic-type   adj-plain
  output-syntax    N
syn-struc
  mods             $var0
  n                $var1
sem-struc
  PROPERTY-1
    DOMAIN         ^$var1
 meaning-procedures
    seek-specification PROPERTY-1
```

```
sclerosis-n1?  ; sclerosis-n2? is also generated
 definition
   example              systemic sclerosis
   comments             auto-learned
   syntactic-type       n-bare
   output-syntax        N
syn-struc
   n                    $var0
sem-struc
  OBJECT-1        ; SCLEROSIS-n2? maps to EVENT
 meaning-procedures
    seek-specification OBJECT-1
```

Figure 3: Initial skeleton lexicon entries learned for the unknown words in the example.

It maps to the concept AUTOIMMUNE-DISEASE. *Of unknown origin* is a construction, recorded as the 16th sense of the preposition *of*, that postmodifies a noun meaning DISEASE, adding the meaning CAUSED-BY IDIOPATHIC-EVENT. *That* is a relative pronoun that establishes an as-yet unspecified CASE-ROLE relation between the preceding noun and the following verb phrase. That relation is specified using a meaning procedure that considers the meanings of the noun and the verb phrase in the input.

```
EVENT-1?                                  ; sclerosis
   IS-A              AUTOIMMUNE-DISEASE-1
   THEME             SET-1                ; multisystemic
   CAUSED-BY         DIOPATHIC-EVENT      ; of unknown origin
   EFFECT            CHANGE-EVENT.1       ; that affects
   from-sense        sclerosis-n1
   PROPERTY-1?                            ; systemic
   DOMAIN            EVENT-1?             ; modifies 'sclerosis'
SET-1                                     ; multisystemic
   MEMBER-TYPE       ANATOMICAL-STRUCTURE
   CARDINALITY       >1
CHANGE-EVENT.1                            ; that affects
  THEME              CONNECTIVE-TISSUE    ; connective tissue
```

Figure 4: A candidate meaning representation (TMR) for the example sentence generated by the basic semantics module of OntoSem.

Next, the extended semantics module of OntoSem uses any number of available disambiguation heuristics to select the most promising candidate TMR and, if possible, improve it. In this example, the most influential heuristic a) detects a *NP is an NP* structure, b) inspects the selectional restrictions on the predicate nominal in the each of the many lexicon senses of *be* in the input and c) filters out all but the definitional sense. As a result, the NP on the left-hand side is hypothesized to be a phrasal, *systemic sclerosis* and the two tentative lexicon entries in Figure 4 give way to a single phrasal entry whose meaning (the filler of its SEM-STRUC zone) is the ontological concept SYSTEMIC-SCLEROSIS (Figure 5).

```
sclerosis-n1
  definition      systemic sclerosis
  example         "Systemic sclerosis is a multisystemic
                  autoimmune disease of unknown origin
                  that affects connective tissue"
  comments        "auto-learned"
  syntactic-type  adj-n
  output-syntax   n
  SYN-STRUC
        adj       $var1 (root systemic)
        n         $var0
SEM-STRUC         SYSTEMIC-SCLEROSIS
```

Figure 5: The improved lexicon entry for the construction *systemic sclerosis*.

The ontological concept in its initial state is created by using the information in the TMR (Figure 6). This is the starting point of the learning cycle that starts with the data collection (Figure 2) whose purpose is to yield a set of sentences that will add and modify the information in the nascent ontological concept. At this point, SYSTEMIC-SCLEROSIS is characterized by the values of four properties – IS-A, CAUSED-BY, THEME and EFFECT. Since the example sentence was interpreted as definitional, SYSTEMIC-SCLEROSIS was made a direct

descendant of AUTOIMMUNE-DISEASE in the ontological graph.

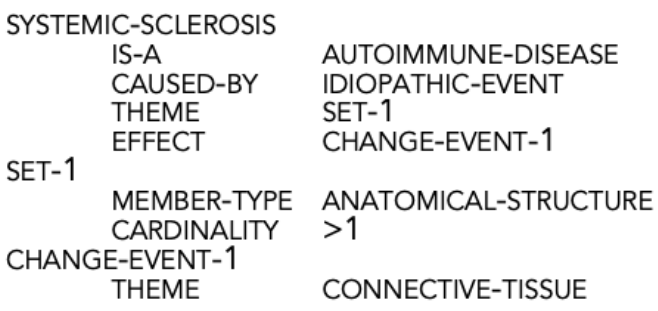

Figure 6: The tentative ontological concept expressing the meaning of *systemic sclerosis*.

This is a windfall because this effectively means that SYSTEMIC-SCLEROSIS inherits all the (many) property-value pairs defined for its ancestors (Figure 7 illustrates one of the inheritance chains for AUTOIMMUNE-DISEASE in the current state of the ontology).

Learning on the basis of TMRs derived from the curated set of sentences assembled using a particular data collection process determines how and whether to merge the local and inherited values for the properties already present in the nascent concept as well as whether to constrain or modify values of inherited properties.

Figure 7: One of the inheritance chains of AUTOIMMUNE-DISEASE in the current state of the ontology.

Before DL models, data collection in our applications was carried out by keyword searches in text corpora [35] or through dialogue with a human [38]. The choice of keywords and the phrasing of human dialogue turns were the means of curating the input to LEIAs' learning. Effective use of DL models for the creation of the curated set of language inputs to LEIA learning requires the creation of a set of natural language prompts expressing the meanings of ontological properties (the current ontology is based on about 300 properties). Once an ontological concept is created the way SYSTEMIC-SCLEROSIS was, prompts for all its (local and, if available, inherited) properties are offered to a DL model to generate text for further learning.[4]

4. Integrating Image Recognition and Learning the Opticon

To support ontological interpretation of visual inputs LEIAs must be equipped with an opticon. The approach to automating its acquisition relies is structurally similar to that of the acquisition of the lexicon in that opticon acquisition may trigger further ontology acquisition. The presence of a lexicon is a prerequisite for our approach to opticon acquisition. This is because the process relies on the image recognition system outputting image representations paired with natural

[4]If the newly learned concept does not specify its ancestors, the latter must be determined; space constraints do not allow us to describe the process we use for this purpose at this time.

language labels. At the most general level, the steps of opticon learning are as follows:

1. the input is a set of DL-generated embeddings of object tokens and English words labeling each of the tokens; the token representations are in terms of (unmotivated, deep-learned) feature vectors generated by a DL model (operating either within a real vision system or a simulated one);
2. The learning system triggers data collection of sentences containing the label
3. Next, the learning system clusters tokens on the basis of the similarity of their embeddings to the embedding generated by processing sentences that contain the natural language label ("target term"), thus effectively disambiguating the label without interpreting them ontologically.
4. The sentences from the cluster that is the best fit for the embedding generated by image recognition are fed into the conceptual learning loop of Figure 2 and result in generating novel skeleton lexicon entries and new or modified existing ontological concepts to formally interpret their meanings.
5. We then can create an opticon entry indexed by the embedding output by the image recognition system and listing the newly created (or modified) ontological concept as its meaning.

The above process integrates the learning of all the elements of LEIAs' semantic memory – the lexicon, the opticon and the ontology. In the next section we present a method of implementing Step 3 in the above process within a virtual environment, using a simple example.

5. Experimenting with Integrating DL Models

In [37], a simulator based on the VoxWorld platform [39] was used to generate stochastic object placements simulating a stacking task.[5] In this task, the agent attempts to stack objects with different geometric characteristics on top of a cube. These characteristics result in different behavior when stacking is attempted (e.g., a cube placed correctly will remain in place, a sphere will almost always roll off, and a cylinder will roll off if placed horizontally but will remain in place if placed with the flat side oriented downward, etc.). Ontological knowledge, here in the form of VoxML [29] specification of object properties was used to bootstrap the generation of the simulations. Knowledge of the objects' symmetries was used to create perturbations in the virtual environment to cause the objects to behave more realistically after their placement. E.g., objects placed on their rounded edges are more likely to move in directions perpendicular to the object's major axis of symmetry.

The object types explored included *cube*, *sphere*, *cylinder*, *capsule*, *small cube*, *rectangular prism*, *egg*, *pyramid*, and *cone* and we found that neural approaches can successfully classify these geometric features into the different object types. The vector representations developed by these neural models come from information directly grounded to object behavior in the environment.

Our experiment involved the following steps:

[5] In this experiment the input is numerical data drawn directly from object interaction in a virtual environment and does *not* contain visual information, although as demonstrated in [40, 41, 42], the same principles can be applied to representations extracted from visual data.

1. **Prompt a language model**: Generate sentences containing the target term to be grounded;
2. **Extract token-level representations**: For each instance of the target token, extract a contextualized numerical representation from a transformer encoder (e.g., BERT);
3. **Linear regression between paired embedding vectors**: Compute a transformation matrix $\mathcal{M} \in \mathbb{R}^{d_A} \times \mathbb{R}^{d_B}$ using ridge regression;[6]
4. **Dialogue with a language model**: Generate novel sentences containing instances of the now-"grounded" target term, as well as negative examples.
5. **Extract token-level representations**: see step 2;
6. **Transform new tokens into object space**: Multiply extracted novel token embeddings by precomputed matrix $\mathcal{M}$ to transform new tokens into object space.

The language model used here was OpenAI's ChatGPT model. The model was given prompts to generate sentences about objects and their properties in context, e.g., *Write 20 short sentences about how blocks are flat on all sides and can be stacked* or *Write 20 short sentences about how balls are round and cannot be stacked.* Prompts were engineered this way in order to quickly generate sufficient examples, but in a more authentic dialogue setting, could easily be generated one at a time with more naturalistic prompts such as "Tell me why it's hard to stack a sphere on another object."

Here, $\mathcal{M}$ results in a structure-preserving affine transformation between subspaces of each embedding space, where the vectors chosen from the respective embedding space each form at minimum the spanning set of the relevant subspaces. Because the vectors chosen for the computation of $\mathcal{M}$ each represent similar but non-identical individual instances of either a contextualized token or an object under interaction, these vector subspaces define "concept"-level representations within the respective models from which they were drawn. An optimal $\mathcal{M}$ will transform new instances (the test set) of object-denoting tokens into or near the subspace defined by object instances, while non-object-denoting tokens will fall outside of the subspaces of known objects.

Results Figure 8 shows a 2D projection of initial results. The transformed novel word embeddings cluster with the object embeddings such that a K-nearest neighbor classifier will successfully classify these transformed tokens as references to the correct object. The same tokens used in their other senses (e.g., "block" in the sense of a city block or "ball" in the sense of a dance), or completely unconnected tokens will not cluster with any relevant object.

This initial demonstration shows the feasibility of a learning approach mediated by a physics-based interactive event simulator to classify instances of concepts and learn their ontological interpretation. The example given herein is purposely simple for the sake of clarity. When starting with a larger ontology, these specific concepts (*block* and *ball*) already exist and need not be learned, but the same principles apply in an opportunistic-learning setting where the LEIA encounters some novel entity and needs to learn about it and populate its skeleton ontology for it.

[6] In this example $\mathcal{M} \in \mathbb{R}^{768} \times \mathbb{R}^{64}$ (768 = BERT-BASE embedding size and 64 = object classifier embedding size). Theoretically the transformation holds without introduction of noise as long as $d_A \geq d_B$. This technique has previously been used to compute transformations between embedding spaces in [40] and [43].

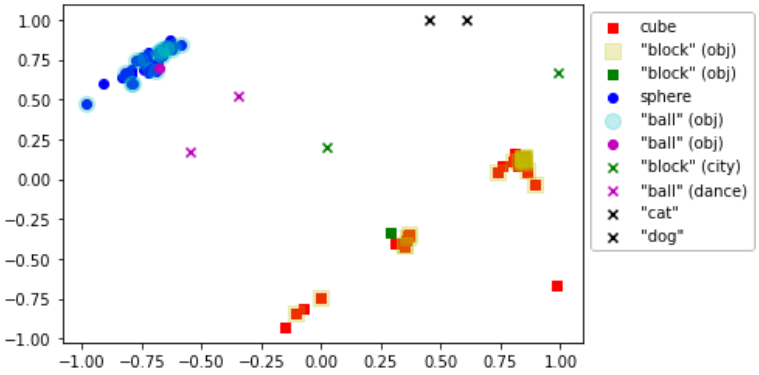

Figure 8: 2D projection of BERT token embeddings to object classifier embeddings. Solid color *cube* and *sphere* points represent object emebeddings. Translucent "block" and "ball" points represent embeddings used to compute mapping $\mathcal{M}$ (shown after transformation). Solid "block" and "ball" points (green and purple) represent transformed token embeddings not used in computing the initial transformation matrix. $\times$s represent tokens identical in form to the grounded tokens, but used in a different sense, or completely extraneous tokens, that cluster with neither object.

6. Discussion

Space constraints do not allow a detailed discussion of the many eventualities that the learning framework we describe will face during its use. Similarly, we do not present here anything like a comprehensive account of the types of elements that LEIAs will learn within this framework. The objective was to give a high-level overview of the approach and to report the the current status of a conceptual learning system based on deep language understanding and the incorporation of the capabilities of DL models with the purpose of making this learning more complete and less expensive.

We contend that the proposed framework will benefit both conceptual learning and DL models. Indeed, the interaction between a DL model and a LEIA can take a push-me-pull-you turn: the DL model output supports automation of ontology, lexicon and opticon acquisition, and the content of LEIA knowledge resources may in turn augment the training datasets of DL models. Retraining the models with LEIAs' knowledge resources will enhance success of the LEIA language interpreter in at least a couple of ways: model outputs can be fomulated using only items in the LEIA lexicon; and model outputs can be at least partially formulated in the LEIAs' ontological metalanguage. One way of doing this is to fine-tune the DL model decoder to output answers using an appropriately restricted vocabulary. That is, in order to instantiate a concept, place it in the correct place in the ontology and specify its properties—many of which will be inherited from its ancestor—*and* make sure that the new concept is different from existing ones, ontological knowledge expressed by the DL model must be expressed in terms already existing within the ontology in order to make it useful. Technically, this is well-within the capabilities of modern language model training, either by outright prohibiting certain tokens from the output (because these terms are not yet present in the ontology), or by manipulating the bias weights in the model's output layer to decrease the bias for token IDs that should appear less frequently.

While generative language models can be useful for generating explanations about phenomena that a LEIA may encounter, they suffer from some limitations resulting in challenges that must be addressed. While a common criticism of large language models is that they are sophisticated "brains-in-vats" that do not possess external understanding of the texts they generate, more prosaically, they are also prone to confidently generating output that is syntactically correct and sounds coherent but is factually incorrect or not self-consistent. These considerations suggest that, at least for the foreseeable future, the results of learning using the proposed framework will have to validated either by humans directly or by analyzing any failures of LEIA functioning due to incorrect or incomplete learning. While we have in the past implemented ergonomic environments to facilitate inspection of LEIAs' knowledge resources and processing results, they remained outside the scope of this paper. We plan to include such a discussion in future reports.

7. Conclusion

This text can be viewed as a methodological position paper describing a novel take on integrating neural net-oriented and symbolic approaches to building artificial intelligent agents, specifically, to the task of overcoming the AI knowledge bottleneck using automatic conceptual learning on the basis of extracting the meaning of natural language texts and interpreting it in terms of a formal ontological world model. This approach builds on our prior work and extends it through the use of DL models. The proposed approach: a) uses deep learning models to create curated collections of sentences; b) uses a bootstrapping knowledge base (ontology, lexicon and opticon) to extract the meanings of these sentences and represent them in an ontologically interpreted metalanguage; and c) uses a dedicated (currently) rule-based learning system to extract from these text meaning representations knowledge elements to enhance and modify the knowledge infrastructure.

This approach may be used opportunistically, during the intelligent agent's regular task-oriented operation – triggering learning when the agent receives perceptual input that its existing knowledge substrate does not cover. Alternatively, the learning mode can be deliberate – for example, when a human decides to teach the agent or when the agent initiates learning itself by inspecting its knowledge resources and triggering the learning process when lacunae and/or inconsistencies are encountered.

This paper presents a bird's eye view of the proposed methodology, with many details of the process omitted due to space constraints. Detailed descriptions of the approach, the system under construction and results of experimentation will be presented in future contributions.

Acknowledgments

This work was supported in part by the U.S. Army Research Office on grant #W911NF-23-1-0031 to Colorado State University and by grants #N00014-19-1-2708 and #N00014-23-1-2060 from the U.S. Office of Naval Research to Rensselaer Polytechnic Institute. The positions expressed herein do not reflect the official position of the U.S. Department of Defense or the United States government. Any errors or omissions are the responsibility of the authors.

References

[1] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, Advances in neural information processing systems 30 (2017).

[2] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Association for Computational Linguistics, Minneapolis, Minnesota, 2019, pp. 4171–4186. URL: https://aclanthology.org/N19-1423. doi:10.18653/v1/N19-1423.

[3] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, I. Sutskever, et al., Language models are unsupervised multitask learners, OpenAI blog 1 (2019) 9.

[4] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, et al., Language models are few-shot learners, Advances in neural information processing systems 33 (2020) 1877–1901.

[5] E. M. Bender, A. Koller, Climbing towards nlu: On meaning, form, and understanding in the age of data, in: Proceedings of the 58th annual meeting of the association for computational linguistics, 2020, pp. 5185–5198.

[6] E. M. Bender, T. Gebru, A. McMillan-Major, S. Shmitchell, On the dangers of stochastic parrots: Can language models be too big?, in: Proceedings of the 2021 ACM conference on fairness, accountability, and transparency, 2021, pp. 610–623.

[7] T. Niven, H.-Y. Kao, Probing neural network comprehension of natural language arguments, in: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Florence, Italy, 2019, pp. 4658–4664. URL: https://aclanthology.org/P19-1459. doi:10.18653/v1/P19-1459.

[8] M. M. Waldrop, What are the limits of deep learning?, Proceedings of the National Academy of Sciences 116 (2019) 1074–1077.

[9] N. Sünderhauf, O. Brock, W. Scheirer, R. Hadsell, D. Fox, J. Leitner, B. Upcroft, P. Abbeel, W. Burgard, M. Milford, et al., The limits and potentials of deep learning for robotics, The International journal of robotics research 37 (2018) 405–420.

[10] D. Gunning, Explainable artificial intelligence (XAI), MIT Research Lab Technical Report, Defense Advanced Research Projects Agency (DARPA), 2017.

[11] M. McShane, S. Nirenburg, S. Beale, Language understanding with ontological semantics, Advances in Cognitive Systems 4 (2016) 35–55.

[12] J. Mao, C. Gan, P. Kohli, J. B. Tenenbaum, J. Wu, The neuro-symbolic concept learner: Interpreting scenes, words, and sentences from natural supervision, in: International Conference on Learning Representations, 2018.

[13] A. d. Garcez, M. Gori, L. C. Lamb, L. Serafini, M. Spranger, S. N. Tran, Neural-symbolic computing: An effective methodology for principled integration of machine learning and reasoning, arXiv preprint arXiv:1905.06088 (2019).

[14] A. d. Garcez, S. Bader, H. Bowman, L. C. Lamb, L. de Penning, B. Illuminoo, H. Poon, C. Gerson Zaverucha, Neural-symbolic learning and reasoning: A survey and interpretation, Neuro-Symbolic Artificial Intelligence: The State of the Art 342 (2022) 1.

[15] D. B. Lenat, Cyc: A large-scale investment in knowledge infrastructure, Communications of ACM 38 (1995).
[16] For ai, data are harder to come by than you think., The Economist (2020).
[17] M. McShane, S. Beale, S. Nirenburg, B. Jarrell, G. Fantry, Inconsistency as a diagnostic tool in a society of intelligent agents, Artificial Intelligence in Medicine 55 (2012) 137–148.
[18] S. Nirenburg, M. McShane, S. Beale, P. Wood, B. Scassellati, O. Magnin, A. Roncone, Toward human-like robot learning, in: International Conference on Applications of Natural Language to Information Systems, Springer, 2018, pp. 73–82.
[19] M. McShane, S. Nirenburg, J. English, Multi-stage language understanding and actionability, Advances in Cognitive Systems 6 (2018) 1–20.
[20] M. McShane, S. Nirenburg, Linguistics for the Age of AI, MIT Press, 2021.
[21] S. Nirenburg, V. Raskin, Ontological semantics, Mit Press, 2004.
[22] J. English, S. Nirenburg, Ontoagent: implementing content-centric cognitive models, in: Proceedings of the Annual Conference on Advances in Cognitive Systems, 2020.
[23] N. Krishnaswamy, P. Narayana, I. Wang, K. Rim, R. Bangar, D. Patil, G. Mulay, R. Beveridge, J. Ruiz, B. Draper, et al., Communicating and acting: Understanding gesture in simulation semantics, in: IWCS 2017—12th International Conference on Computational Semantics—Short papers, 2017.
[24] N. Krishnaswamy, S. Friedman, J. Pustejovsky, Combining deep learning and qualitative spatial reasoning to learn complex structures from sparse examples with noise, in: Proceedings of the AAAI Conference on Artificial Intelligence, volume 33, 2019, pp. 2911–2918.
[25] J. Pustejovsky, N. Krishnaswamy, Embodied human computer interaction, KI-Künstliche Intelligenz 35 (2021) 307–327.
[26] N. Krishnaswamy, J. Pustejovsky, Affordance embeddings for situated language understanding, Frontiers in Artificial Intelligence 5 (2022).
[27] M. McShane, S. Nirenburg, B. Jarrell, G. Fantry, Learning components of computational models from texts, in: 6th Workshop on Computational Models of Narrative (CMN 2015), Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2015.
[28] S. Nirenburg, P. Wood, Toward human-style learning in robots, in: AAAI Fall Symposium on Natural Communication with Robots, 2017.
[29] J. Pustejovsky, N. Krishnaswamy, Voxml: A visualization modeling language, in: Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16), 2016, pp. 4606–4613.
[30] E. Viegas, S. Nirenburg, The ecology of lexical acquisition: Computational lexicon making process, in: Proceedings of Euralex, volume 96, 1996.
[31] S. Nirenburg, V. Raskin, S. Sheremetyeva, Lexical acquisition, NATO Science Series Subseries III Computer and Systems Sciences 188 (2003) 133–172.
[32] S. Nirenburg, M. McShane, J. English, Content-centric computational cognitive modeling, Adv. Cogn. Syst (2020).
[33] S. Nirenburg, M. McShane, J. English, Overcoming the knowledge bottleneck using lifelong learning by social agents, in: International Conference on Applications of Natural Language to Information Systems, Springer, 2021, pp. 24–29.
[34] S. Nirenburg, P. Shell, A. Cohen, P. Cousseau, D. Grannes, C. McNeilly, The translator's workstation, in: Proceedings of the 3rd Conference on Applied Natural Language

Processing. Trento, Italy, April, 1992.
[35] S. Nirenburg, T. Oates, J. English, Learning by reading by learning to read, in: International Conference on Semantic Computing (ICSC 2007), IEEE, 2007, pp. 694–701.
[36] M. McShane, S. Nirenburg, J. English, Multi-stage language understanding and actionability., Advances in Cognitive Systems 6 (2018) 119–138.
[37] S. Ghaffari, N. Krishnaswamy, Detecting and accommodating novel types and concepts in an embodied simulation environment, arXiv preprint arXiv:2211.04555 (2022).
[38] S. Nirenburg, M. McShane, J. English, Artificial intelligent agents go to school., in: 34th International Workshop on qualitative reasoning, IJCAI-2021, ????
[39] N. Krishnaswamy, W. Pickard, B. Cates, N. Blanchard, J. Pustejovsky, The voxworld platform for multimodal embodied agents, in: Proceedings of the Thirteenth Language Resources and Evaluation Conference, 2022, pp. 1529–1541.
[40] D. McNeely-White, B. Sattelberg, N. Blanchard, R. Beveridge, Canonical face embeddings, IEEE Transactions on Biometrics, Behavior, and Identity Science 4 (2022) 197–209.
[41] J. Merullo, L. Castricato, C. Eickhoff, E. Pavlick, Linearly mapping from image to text space, arXiv preprint arXiv:2209.15162 (2022).
[42] J. Pustejovsky, N. Krishnaswamy, Multimodal semantics for affordances and actions, in: Human-Computer Interaction. Theoretical Approaches and Design Methods: Thematic Area, HCI 2022, Held as Part of the 24th HCI International Conference, HCII 2022, Virtual Event, June 26–July 1, 2022, Proceedings, Part I, Springer, 2022, pp. 137–160.
[43] A. Nath, R. Ghosh, N. Krishnaswamy, Phonetic, semantic, and articulatory features in assamese-bengali cognate detection, in: Proceedings of the Ninth Workshop on NLP for Similar Languages, Varieties and Dialects, 2022, pp. 41–53.