Knowledge-Infused Learning: A Sweet Spot in Neuro-Symbolic AI

Manas Gaur, *AI Institute, University of South Carolina, Columbia, SC, 29208, USA*
Kalpa Gunaratna, *Samsung Research America, Mountain View, CA, 94043, USA*
Shreyansh Bhatt, *Amazon Inc., Seattle, WA, 98121-2615, USA*
Amit Sheth, *AI Institute, University of South Carolina, Columbia, SC, 29208, USA*

Deep learning has revolutionized the artificial intelligence (AI) landscape by enhancing machine capabilities to understand data-dependant relationships. On the other hand, knowledge may not directly correlate or depend on the data but represents facts that are true. Combining knowledge with the data-driven deep learning techniques improves upon what can be learned from data alone, resulting in improved performance with reduced training, user-level explainability, modeling uncertainty in deep learning, achieving context-sensitivity, and better control over the behavior of AI systems such as to assure the safety or avoid toxic behavior. We refer to the approach of combining various types of explicit knowledge as knowledge-infused learning (KiL). Knowledge infusion brings symbolic AI into data-driven AI, giving us a class of neuro-symbolic AI methods. The work on KiL has already developed a suite of context-adaptive algorithms that infuses various knowledge into deep learning methods in various ways, broadly categorized as a shallow infusion, semi-deep infusion, and deep infusion. This special issue allows interdisciplinary researchers and practitioners from diverse fields such as natural language processing, recommender systems, and computer vision to contribute their research on the infusion of external and expert-curated knowledge in data-driven learning methodologies for consistency and robustness in outcomes.

WE need to build "AI that captures how humans think." This quote from Gary Marcus is well aligned with Andrew Ng's quote that "the importance of Big Data is overhyped, and AI needs to push for higher levels of machine intelligence," and an even earlier statement by Pedro Domingos, that "Data Alone is Not Enough."[a] These statements recognize that recent advances in AI through scaling parameters (millions to billions) have yielded models that hallucinate when deployed in real-world applications.[b] Lee *et al.* states that gigantic language models or converstational agents have shortcomings.[c] For instance, BlenderBot 1 and BlenderBot 2 have generated factually incorrect responses, that are often unsafe. Policymakers and practitioners assert serious usability and privacy concerns constraining adoption, notably in high-consequence domains, such as healthcare, crisis response, scientific discovery, and others.[d] About 91% of the companies want data-driven AI models to be user-level explainable, and interpretable to gather trust from stakeholders[1e]

The current era and future of AI, as also recognized by DARPA, need hybrid models that combine the respective powers of human knowledge and benefits of AI in a scalable exploration-exploitation strategy for explainable decision making. Human knowledge

[a][Online]. Available: https://tinyurl.com/data-alone-not-enough
[b][Online]. Available: https://tinyurl.com/Google-LaMDA

U.S. Government work not protected by U.S. copyright.
Digital Object Identifier 10.1109/MIC.2022.3179759
Date of current version 19 July 2022.

[c][Online]. Available: https://tinyurl.com/blenderbot-analysis
[d][Online]. Available: https://tinyurl.com/AI-Descartes
[e][Online]. Available: https://tinyurl.com/IBM-AI-Adoption

represents the first phase of AI comprised of handcrafted rules (e.g., Hearst Patterns), which effectively describe underlying patterns in the data. The exploration-exploitation strategy in AI comes from voluminous training data and massive computational power that results in success in narrowly defined tasks. However, data-driven AI models alone are not enough, as their outcomes are subject to varied interpretations by different expert stakeholders. This is because the ground truth datasets constructed by experts require additional efforts, such as: a) They use their past experience in the domain for annotation. b) As an extra pair of eyes, they use web or domain-specific knowledge sources to confirm the quality of labels. c) Expert annotators, in sensitive domains, prepare annotation guidelines to create gold standard datasets. These external pieces of information are not available to data-driven AI models at the time of training or fine-tuning, causing them to approximate their learning behavior and sacrifice consistency and robustness.

Hybrid AI has been of interest for consistent and robust AI for the past 54 years, starting from McCarthy and Hayes in 1968, who enumerated some philosophical problems from the standpoint of AI. These were:

- How can AI allow reasoning over the estimation, models and represent world knowledge?
- How can AI capture causality in varied situations?

McCarthy and Hayes emphasized that practical AI systems require epistemological systems that store facts about the world and use it to understand and control decision making of data-driven AI. After that, in 1979, Hofstader discussed the reasons for the question, "Why is AI far from being intelligent?" by pressing on human thinking being artistic and beautiful in his book *Gödel, Escher, Bach: An Eternal Golden Braid*. Cutting through AI winter, where methods like teacher forcing (a part of student-teacher modeling), sequential models based on human's ability to learn by utilization of local knowledge, and convolutional models based on human's ability to examine global knowledge were developed. In 2001, Sheth *et al.* defined the notion of a world model and its utility in enhancing search, personalization, and user profiling.[f] This was an early effort that showcased the efficacy of human-curated knowledge bases in achieving state-of-the-art results in recommender systems with relatively small-scale and interpretable AI models (e.g., Hidden Markov models). Further, it introduced a semantic search engine that utilizes knowledge as graphs (referred to then as world

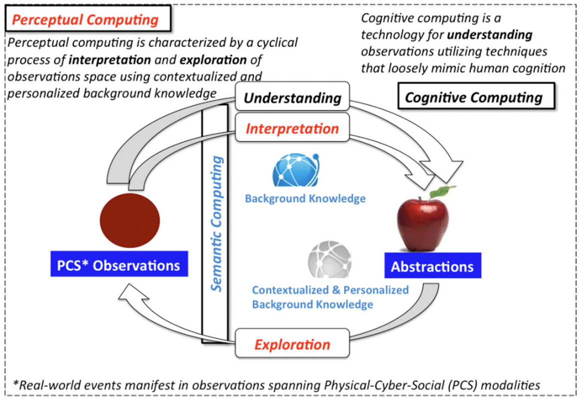

FIGURE 1. The differences between semantic computing (SC), cognitive computing (CC), and perceptual computing (PC) are conceptual. SC provides context to unstructured data that is required to analyze and integrate observations from the physical, cyber, and social (PCS) realms. CC uses SC annotations to interpret observations, and PC looks for observations in the surroundings to acquire relevant data that will help them better comprehend the world. PC works in a cyclical manner, evaluating and examining observations in order to tailor background knowledge to the context and individual of interest.

model or ontologies) combining knowledge bases and statistical classifiers, and enriching content (then called rich media reference) similar to infobox over a decade later. In 2006, Turing laureate Leslie Valiant questioned how machines can acquire and manipulate common sense knowledge, which is passively which is passively addressed in Kahneman's book *Thinking, Fast and Slow* to propose stitching System 1 (best modeled as Statistical AI) and System 2 (Symbolic AI). This was brought out well by Sheth *et al.* by stating that "knowledge will propel machine understanding of content" and the need for hybrid AI to braid semantic, cognitive, and perceptual components in computing (see Figure 1).[2] Unlike data-driven AI, the human brain does not follow a well-defined script, but rather exploits perception (highly statistical) and cognition (based on explicit, structured knowledge) for multifaceted decision-making and communication.[g] A substantial growth in incorporating knowledge with deep learning has pursued.[3]

Humans are capable of handling information at various levels of abstraction, which corresponds to different levels of learning. A knowledge-driven approach in AI should explore stratified knowledge manifested in various human-curated knowledge sources, general-purpose or domain-specific, lexical or graph-based to

[f][Online]. Available: https://tinyurl.com/Taalee

[g][Online]. Available: https://tinyurl.com/Neurosymbolic-wave

find connections between facts and observations, yielding outcomes that humans can relate with their understanding and reason over the model. This is in essence the neuro-symbolic AI, and knowledge-infused learning (KiL) is a class of neuro-symbolic AI technique that emphasize the use of broader forms of knowledge (lexical, domain-specific, common sense, and constraint-based) into addressing limitations of either symbolic or statistical AI approaches, such as model interpretations and user-level explanations. Compared to powerful statistical AI that exploits data, KiL benefits from both data and knowledge.

KiL has seen a continuum of strategies for infusing knowledge, that have been broadly categorized as shallow, semi-deep, and deep infusion. *Shallow knowledge infusion* contextualizes the training examples with expert knowledge to capture meaningful patterns. Some of the shallow infusion examples include contextual modeling,[4] entity normalization,[5] and others.[h] *Semi-deep infusion* attempts to embed expert knowledge in parametric space of the model or guides the model's learning process using constraints.[6–8] Though, this method of infusion has yielded significant gains, it falls short in assisting deep learning models to learn high-level abstractions through its multiple layers. *Deep knowledge infusion* combines the stratified representation of knowledge at varying abstraction levels to be transferred in different layers of deep learning models.[i] A KiL model could provide user-level explanations by mapping the tokens in input data to concepts in the knowledge graph (KG).[9] Further, a KiL model is an interpretable procedure whose learning curve can be monitored by either:

- A weighting function that demonstrates the correlation between the tokens in the input and concepts in knowledge source[j],
- Or logical constraints that monitor the amount of knowledge infusion.[9]

In this article, we briefly discuss three application areas for KiL that are of interest to knowledge discovery and data mining, natural language processing, and computer vision communities.

KiL FOR LANGUAGE MODELING

Understanding language is essential for machines to communicate with humans and achieve trust effectively. Until recently, the focus on improving language models have been purely on improving statistical techniques based on word distribution and frequency. However, symbolic knowledge captured in KGs can help these language models capture entity-specific information better (e.g., Word2Vec, Transformer models, etc.). Further, with knowledge we achieve broader semantics, including named and taxonomic relationships, synonyms, acronyms, and others, which support contextualization and abstraction. As a result, the language model encoding in the embedding space can be made richer. For instance, ERNIE[10] introduces semi-deep infusion in BERT using entity information. This contextualization outperformed prior state-of-the-art data-driven transformer models on general language understanding evaluation (GLUE) tasks. We consider deep infusion of knowledge as a new paradigm that will significantly advance the capabilities and promises of deep neural networks.

One inherent advantage of using knowledge in language models is that they get the additional information through entities or KG triples that otherwise require a lot of data to learn. This knowledge may not be learned from statistical data as cleanly as the direct knowledge infusion techniques. Language models treat each word/subword (or token) without much differentiation from each other. Hence, specific information that an entity or phrase represents in a sentence is difficult to capture without such special processes. For example, the entity "Joe Biden" may not be captured as two words collectively in a language model and hence miss the important knowledge that those two words collectively refer to the US president. Joe Biden will be some common subwords in BERT, for example. It may get to the point that it represents a person when considering the complete sentence but will most probably miss all structural information that the KG will bring in if the model can understand it is an entity. If such language models are used to facilitate a conversation with a user, they end up losing the context, resulting in the generation of factually incorrect responses.[k] Furthermore, the conversations that are not context-controlled result in random conversations and generate sentences that are incoherent and irrelevant for the user. In a complex domain, such as mental health, such models can end up generating unsafe questions or responses that can have a severe consequence on the user's health.[9]

PROCESS KNOWLEDGE IN KiL

Information in KGs can guarantee context capture, but we need another type of knowledge called "process knowledge" to ensure that language models do not hallucinate. Hallucinated model generations are unsafe, incoherent, and irrelevant to the end-user using an

[h][Online]. Available: https://tinyurl.com/semantics-from-blackbox
[i][Online]. Available: https://tinyurl.com/shades-KiL
[j][Online]. Available: https://tinyurl.com/semantic-autoencoder
[k][Online]. Available: https://tinyurl.com/hallucination-AI

application that employs deep neural language models.[l] This form of knowledge extend other forms of knowledge, specifically: a) KG: They are structured but not ordered. KGs can support context capture but cannot enforce conceptual flow.[8] b) *Semantic lexicons:* They are flattened form of KG that make deep language models context sensitive, add constraints, but cannot enforce conceptual flow.[11] c) *Ontologies:* These are curated schematic forms of KGs having classes, instances, and constraints. Thus, ontologies can provide a stricter control over context and constraints.[m] Using process knowledge, an ontology can enforce order in question generation using deep language models and act as an alternative to process knowledge. For example, Autism Diagnostic Observation Schedule (ADOS) is a diagnostic tool used by a caregiver to improve clinical evaluation and guided therapy. In addition, it is often used in schools having special care facilities for children with autism. Suppose a robotic system is situated in this setting to conduct the activities of a human caregiver. In that case, it needs a set of rules and classes, very much in the form of ontology, to guide its interaction. Such an ontology is called process knowledge-inspired ontology. Likewise, an ontology created using Montreal Cognitive Assessment (MoCA) can support the automation of Aphasia detection methods.

Consider a scenario, where a user asks the following question to a typical agent *without the process knowledge:* "Can you recommend dishes that are calorie sensitive?" If the agent is augmented with search and retrieve capabilities using internet, it would accumulate responses to following related questions listed under *people-also-ask questions*:

- Are restaurants required to put calories on menus?
- Are calorie recommendations accurate?
- Should I eat less than my recommended calories?
- What food can you recommend?

There are two fundamental problems here: (a) The AI system behind these recommendations is confused whether "calorie sensitivity" is a positive or negative concept. (b) The AI system fails to bridge the gap between dishes and calorie sensitivity. Furthermore, a response to such a question is dependent on the time of the day: breakfast, lunch, or dinner. A conversational agent with process knowledge can generate following information-seeking type questions:

1) Do you have a preferred cuisine?
2) Do you want to know about low calorie food in this cuisine for breakfast/lunch/dinner?
3) Do you want me to book reservations for restaurants that serve this cuisine?
4) Do you want me to save your preferences?

If the answer to any question is no, then an alternate path in process knowledge is triggered. Here the process knowledge is the procedure for recommending and ordering food. Moreover, the agent can benefit from 2015–2020 Dietary Guidelines for Americans[n] to emphasize on overall healthy eating patterns supported by five food groups: fruits, vegetables, grains, protein foods, and dairy.

Process knowledge can be used by KiL for additional capabilities such as preventing conversational agents from generating sentences that have severe consequences, such as in mental healthcare (see Figure 2). Another application of process knowledge supported KiL is in personalized recommendation to enhance user engagement, especially when they are interacting with conversational agents.

Identifying entities and linking them to KGs (entity linking) is typically required to enable KiL in language modeling tasks. However, errors in this step may transfer into the KiL process. Let us consider an NLP task of paraphrasing, some of the concerns with regards to knowledge infusion are as follows:

- Extrapolation or Overgeneralization of the Model
  - Sentence: Your **mom and dad** are **toxic** (bold-faced words are target entities for paraphrasing).
  - Paraphrasing without knowledge infusion: **Toxicity** is in your **mom and dad**.
  - Paraphrasing with knowledge infusion: Your **parents** are **radioactive**. (It is intuitive to replace"mom and dad" with parents, however, selection of word "radioactive" in place of "toxic" is *not suitable*).
- Disparity
  - Sentence: She has her **boundaries** for a **reason**.
  - Paraphrasing without knowledge infusion: She has her **borders** for a **factor**.
  - Paraphrasing with knowledge infusion: She has her **bound/limits** for a **purpose/cause**. (Here the knowledge infusion did provide suitable words for paraphrasing but is confused in choosing appropriate word for the sentence).

[l] [Online]. Available: https://arxiv.org/pdf/2205.13884.pdf
[m] [Online]. Available: https://tinyurl.com/ontology-interpretable-ML

[n] [Online]. Available: https://tinyurl.com/dietary-guideline

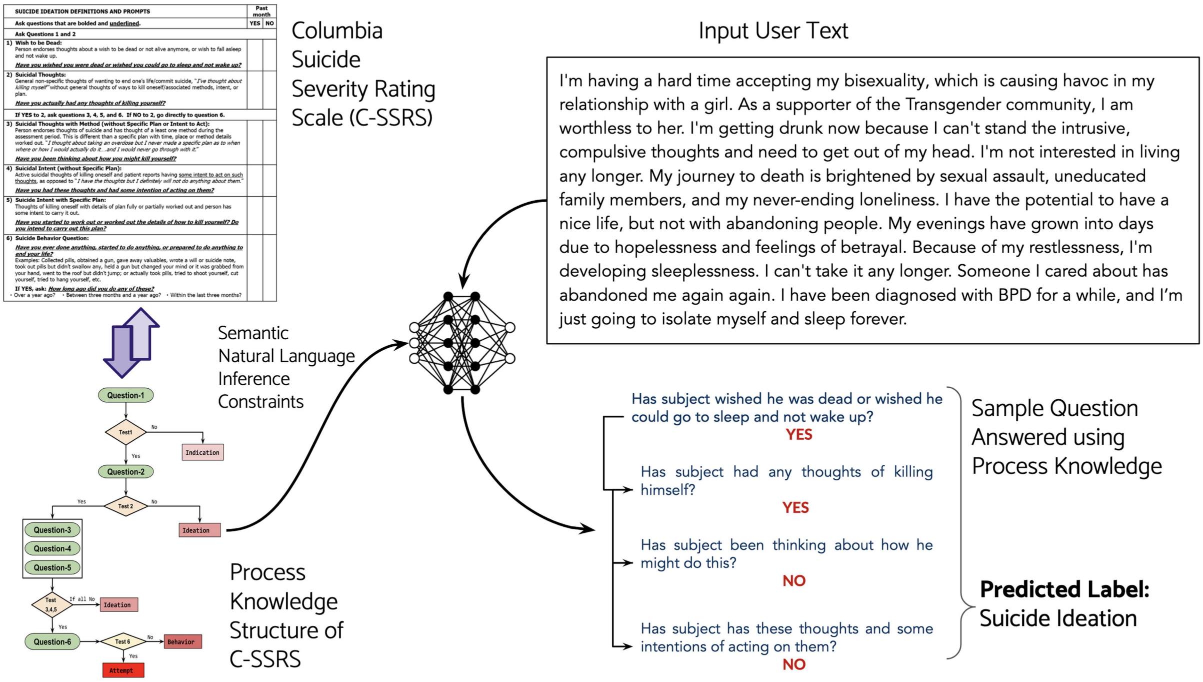

FIGURE 2. A two-stage pipeline comprises creating process knowledge and its infusion into a deep neural network to identify which questions can be answered from the user's input and which questions can be considered as follow-up questions for the complete assessment.[o] The construction of process knowledge is manual for quality and safety purposes. Its infusion in the deep neural network is automated by an interplay of deep language models and an interpretable decision tree algorithm. Note: *Original text modified to ensure anonymity of the user.*

Hence, caution in knowledge processing techniques is beneficial; if not, alternative human-in-the-loop methods should be used to avoid the knowledge infusion step. On the other hand, the vocabulary of a KG is massive compared to a language model word/token vocabulary. Hence, efficient retrieval, encoding, and compression techniques are required for better knowledge infusion.[12,13]

KiL FOR RECOMMENDER SYSTEMS

Data-driven statistical AI-based techniques have shown to be effective for recommending the right product, content, or connecting people to enhance user experience. These recommendation methods rely on understanding a user and recommend product/content based on successful recommendations of the past and aim to develop the recommendation system based on the interaction data. While such purely data-driven techniques can improve success measures, they often lack inherent understanding of user and product/content to recommend, and hence the recommendations may not be personalized.

Developing and utilizing knowledge bases to augment the data in recommendations can help progress toward more personalized recommendations and also help address one of the intrinsic challenges of recommendation—*cold start recommendations*. Several recent research studies utilize knowledge to complement the data and provide more enhanced, personalized, and meaningful recommendations.[p] Organizing and utilizing the knowledge beyond the available data has become a promising research area to provide a better user experience with enhanced recommendation systems.

KiL FOR COMPUTER VISION

Processing images and videos attempts to give machines the capabilities of human-like sense of vision. However, processing images or video alone has limited contextual information crucial for the machine to make sense to humans.[2] For example, an image may contain a horse

[o][Online]. Available: https://tinyurl.com/Posner-CSSRS

[p][Online]. Available: https://tinyurl.com/KG-for-recommender

and a human. Without additional common sense knowledge, it may be difficult for the machine to understand that humans ride on horses and not the other way round.

A scene graph is a specific representation technique used in image processing to understand and reason over images (and videos using snapshots/frames). This graph representation is infused with subgraphs extracted from KGs to enrich the contextual information of images to improve machine understanding.[14]

Further, visual question answering is an emerging area of interest in the computer vision community. Using KG to better support QA algorithms are preferred because of rich knowledge that they can bring.[15]

IN THIS ISSUE

We present accepted submissions that demonstrate how knowledge infusion in data-driven AI models can alleviate four critical limitations: a) context capture, b) handling uncertainty and risk, c) model interpretability, and (d) model explainability. Further, the knowledge infusion has shown improvement in application performance.

The article "Where Does Bias in Common Sense Knowledge Models Come From?" investigates methods to quantify the source of bias in COMET, a common sense transformer. The novel metrics presented can be considered as a proxy for measuring bias in large KGs. This would regulate knowledge infusion in pretrained language models. The article "Common sense Knowledge Infusion for Visual Understanding and Reasoning: Approaches, Challenges, and Applications" provides in-depth review on knowledge infusion in visual understanding.

An explainable method for question answering is required when the language models are required to generate a paragraph-long answer. It is because the generated paragraph should preserve the context of the scenario. This article "SR3: Sentence Ranking, Reasoning, and Replication for Scenario-Based Essay Question Answering" shows how background knowledge bases can inform natural language questions and answer generations. The next article "Physics-Informed Machine Learning for Uncertainty Reduction in Time Response Reconstruction of a Dynamic System" includes interpretable machine learning, which proposes a method to incorporate a physics-based model to quantify and remove uncertainty when the model is making predictions in dynamical processes.

Today, knowledge infusion is the driving force for next-generation neuro-symbolic AI methods and frameworks. It sees applications in robotics, cognitive science, self-driving cars, personal assistants, etc. With the deep infusion, we envision that the AI models will see a new form of learning, where they will learn by abstraction and control their learning behavior using stratified external knowledge. According to DARPA, the third wave of AI is about contextual adaptation and explanations, where explicit knowledge will have a significant role to play. Moreover, AI will start learning from many disciplines to realize the future potential of engaging and assisting humans in various domain-specific, low resource, and sensitive tasks.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation (NSF) Award #2133842 "EAGER: Advancing Neuro-Symbolic AI with Deep Knowledge-infused Learning." Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF, Samsung Research America, and Amazon.

REFERENCES

1. L. Longo, R. Goebel, F. Lecue, P. Kieseberg, and A. Holzinger, "Explainable artificial intelligence: Concepts, applications, research challenges and visions," in *Proc. Int. Cross-Domain Conf. Mach. Learn. Knowl. Extraction*, 2020, pp. 1–16.
2. A. Zareian, Z. Wang, H. You, , and S.-F. Chang, "Learning visual common sense for robust scene graph generation," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 642–657.
3. T. Dash, S. Chitlangia, A. Ahuja, and A. Srinivasan, "A review of some techniques for inclusion of domain-knowledge into deep neural networks," *Sci. Rep.*, vol. 12, no. 1, pp. 1–15, 2022.
4. S. Bhatt *et al.*, "Knowledge graph enhanced community detection and characterization," in *Proc. 25th ACM Web Search Data Mining*, 2019, pp. 51–59.
5. M. Gaur *et al.*, "Characterization of time-variant and time-invariant assessment of suicidality on Reddit using C-SSRS," *PLoS One*, vol. 16, no. 5, 2021, Art. no. e0250448.
6. K. Faldu, A. Sheth, P. Kikani, and H. Akbari, "Ki-BERT: Infusing knowledge context for better language and domain understanding," 2021, *arXiv:2104.08145*.
7. K. Gunaratna, Y. Wang, and H. Jin, "Entity context graph: Learning entity representations fromsemi-structured textual sources on the web," 2021, *arXiv:2103.15950*.
8. M. Gaur, K. Gunaratna, V. Srinivasan, and H. Jin, "ISEEQ: Information seeking question generation using dynamic meta-information retrieval and knowledge graphs," 2021, *arXiv:2112.07622*.
9. K. Roy, M. Gaur, Q. Zhang, and A. Sheth, "Process knowledge-infused learning for suicidality assessment on social media," 2022, *arXiv:2204.12560*.

10. Z. Zhang, X. Han, Z. Liu, X. Jiang, M. Sun, and Q. Liu, “ERNIE: Enhanced language representation with informative entities,” 2019, *arXiv:1905.07129*.
11. G. Manas *et al.*, “Knowledge-infused abstractive summarization of clinical diagnostic interviews: Framework development study,” *J. Med. Internet Res. Mental Health*, vol. 8, no. 5, 2021, Art. no. e20865.
12. N. De Cao *et al.*, “Multilingual autoregressive entity linking,” *Trans. Assoc. Comput. Linguistics*, vol. 10, pp. 274–290, 2022.
13. Q. Liu, G. Cheng, K. Gunaratna, and Y. Qu, “Entity summarization: State of the art and future challenges,” *J. Web Semantics*, vol. 69, 2021, Art. no. 100647.
14. J. Gu, H. Zhao, Z. Lin, S. Li, J. Cai, and M. Ling, “Scene graph generation with external knowledge and image reconstruction,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 1969–1978.
15. S. Monka, L. Halilaj, and A. Rettinger, “A survey on visual transfer learning using knowledge graphs,” 2022, *arXiv:2201.11794*.

MANAS GAUR is a graduate researcher with Artificial Intelligence Institute, University of South Carolina, Columbia, SC, 29208, USA. He will be joining University of Maryland Baltimore County as Assistant Professor. He has been Data Science fellow with the University of Chicago and Turing fellow with Alan Turing Institute. Contact him at mgaur@email.sc.edu.

KALPA GUNARATNA is a staff research scientist at AI Center of Samsung Research America, Mountain View, CA, 94043, USA. Contact him at k.gunaratna@samsung.com.

SHREYANSH BHATT is a machine learning scientist with Amazon.com, Seattle, WA, 98108-1226, USA. Contact him at bhattshr@amazon.com.

AMIT SHETH is the director of Artificial Intelligence Institute, University of South Carolina, Columbia, SC, 29208, USA (http://aiisc.ai). He is a Fellow of AAAI, AAAS, IEEE, and ACM. Contact him at amit@sc.edu.