Using Knowledge Graphs and LLMs to Enhance Natural Language Understanding on Voice Assistants

Ashlesha Vishnu Kadam
Amazon.com, LLC
Amazon Music, 2021 7th Ave
Seattle, WA 98121

ABSTRACT
Both, Large Language Models (LLMs) and Knowledge Graphs (KGs) are used in various Natural Language Understanding (NLU) tasks. However, each has some benefits and disadvantages. This paper explores the pros and cons of each, and demonstrates how the two used together can help overcome some of the shortcomings. It also identifies specific applications of KG-enhanced LLMs for music-related user experiences on voice assistants. Finally, it enlists the challenges in KG-enhanced LLM applications.

General Terms
Large Language Models, Natural Language Understanding, Voice Assistants, Voice Technology, Knowledge Graph.

Keywords
LLMs, NLU, NLP, voice assistants, knowledge graphs.

1. INTRODUCTION
Natural Language Understanding (NLU) encompasses machines understanding and interpreting, both syntactically and semantically [1, human language in order to facilitate human machine interaction. Many applications around us rely on NLU, including chatbots, voice assistants, virtual assistants, information retrieval systems, and more. However, NLU is a complex problem, fraught with challenges in syntactic and semantic understanding [2], contextual understanding, tone / accent / speech imperfections (i.e. ASR [3], in case of audio NLU) understanding, human and ambience variability, benchmarking and evaluation [4], fluency and the need for highly scalable and efficient compute. This paper explores how Knowledge Graphs and Language Models can help address some of these NLU challenges, particularly in the domain of audio NLU for music recommendations via voice assistants (VAs).

Knowledge Graphs (KGs) are a relational representation of entities and their attributes. Knowledge graphs came into prominence when Google announced the Google Knowledge Graph [5], followed quickly by other technical companies. A graph consists of nodes that represent entities and edges between nodes that may or may not be directional as relationships between the entities [6]. Nodes can have attributes associated with them. Similarly, edges can have attributes that better establish the relationship between the nodes they connect. At their core, knowledge graphs are a way to encode real world information in a relational, structural manner. Along with direct and unambiguous relations, knowledge graphs also capture semantic meanings and relationships of both nodes and edges to enable better understanding of relationships [7].

Large Language Models (LLMs) are foundational models (FMs), i.e. a type of large neural network that can generate or embed text. LLMs can predict the probabilities of future tokens given an input string of text [8]. LLMs are pre-trained on large-scale unstructured data via self-supervised learning [9]. During inference time, users can provide a snippet of text as a starting point, i.e. a "prompt". This prompt gets converted into an embedding that is used to predict the probabilities of all possible tokens that can follow [10]. LLMs have been successfully deployed to various Natural Language Processing (NLP) tasks like document classification, speech and text generation, sentiment analysis, text summarization, machine translation, and more [11]. LLMs can learn textual representations that capture not just syntactic but also semantic aspects of a language, lending themselves naturally to these NLP tasks.

Both KGs and LLMs have their own drawbacks, as explored later in this article. However, both KGs and LLMs can be used together to vastly improve natural language understanding (NLU), knowledge building and knowledge representation. In this article, the intersection of KGs and LLMs is explored to understand how it can be applied to solve and/or improve real life voice assistant NLU challenges.

2. LLMs FOR NLU: ADVANTAGES AND DRAWBACKS
The biggest advantages of LLMs that lend them to NLU applications are (a) LLMs have the ability to handle popular NLP use cases like speech-to-text, sentiment analysis, spell check, summarization and classification of tokens [11]. They help understand human natural language better and master fluency, (b) LLMs are self-supervised, pre-trained neural networks that can be tuned to handle multiple natural language tasks instead of having to train multiple network models [11], (c) LLMs are able to generate predictions of the next set of tokens based on an input prompt without model training, which allows for faster development and lesser effort [12].

However, the biggest drawbacks of LLMs when it comes to NLU are (a) LLMs hallucinate, i.e. generate text that is incorrect, non-sensical, or not real [13]. This could reduce the applications of LLMs [14] and could have unintended consequences in case users rely on their outputs, (b) Like any other model trained on large existing data sets, LLMs are prone to biases and endorsement of stereotypes, (c) LLMs cannot explain their reasoning for the generated output or provide any insight into their decision making [15], which might limit their applications, especially in cases where explainable AI is desirable, (d) Overall, LLMs can't provide structured responses because of their lack of temporal context and inability to be able to explain their decision-making[15], [16], and (e) The high latency of LLMs is a concern [17] for some applications that require real time or near-real time responses.

3. KNOWLEDGE GRAPHS FOR NLU: ADVANTAGES AND DRAWBACKS

KGs have inherent advantages that can help overcome the drawbacks of LLMs. Key advantages of KGs include (a) KGs provide a structured way of building and storing information and knowledge, by having the ability to store data and maintaining a state of their relationships [18] to enable relevant and contextualized responses to queries, (b) KGs can enable the discovery of new insights and patterns in data, improve data quality and accuracy, and enhance data integration and interoperability [19], (c) KGs also enable data governance and explainable AI (XAI) [20]. In other words, the outcomes of the queries to KGs are interpretable, and (d) KGs cannot hallucinate since they are constructed based on existing data and relationships between entities [21]. KGs do not create new data (nodes) or relationships.

However, KGs also suffer from some drawbacks from an NLU application perspective, the top ones being (a) KGs can quickly become very large and unwieldy, making it challenging to further construct and maintain them[19]. Data modeling and graph database management can become complex with increasing size of the KG too, (b) As KGs grow, it might become harder and slower to query for data, resulting in a degradation of the KG's performance, which might be unacceptable to some applications[19], (c) It is hard to construct KGs from unstructured data sources because that data needs to be converted to a structured one for being able to build a KG[7]. This limits their application to natural language processing or understanding, and (d) KGs are often constructed for specific domains, limiting their usefulness in other domains[6].

4. USING KNOWLEDGE GRAPHS TO ENHANCE LLMs

As can be seen from the previous two sections, KGs and LLMs have complementary advantages. Applying KGs to LLMs can help overcome some of the drawbacks that LLMs have, particularly by helping LLMs capture factual knowledge and avoid hallucination, and by becoming more interpretable. Following are some ways in which KGs can be used to enhance the application of LLMs

4.1 Pre-training the language model with KGs

Previous research [**Error! Bookmark not defined.**] proposes adding KG embedding as an input to the transformer of the model during the language model pre-training, which is shown to increase the knowledge that gets added into the transformer parameters. Another research paper [22] presents a knowledge aware language model pre-training method that adds an entity prediction task and an entity-extended tokenizer to the input of the transformer in pretraining and an additional entity prediction task at the output, and demonstrates that adding these entity signals in pretraining packs more knowledge into the transformer parameters.

4.2 Using KGs to improve LLM inference:

There is also research [23] about using KGs during the inference stage of LLMs, instead of the pre-training stage, to augment the LLMs with the knowledge stored in KGs without having to integrate the KG in the pre-training stage

5. APPLICATIONS OF KG-ENHANCED LLM ON VOICE ASSISTANTS FOR MUSIC EXPERIENCES

This section demonstrates the application of KGs to enhance LLMs for the specific use case of music recommendations on voice assistants (VAs).

To demonstrate how a KG-powered LLM can help create music-related experiences on voice, a simplified version of what a music KG might look like has been presented in Fig. (1). The KG has examples from real life, including those related to real world artist (Taylor Swift). The edges indicate the relationship between different entities. Similar entity types are color-coded in the same color (e.g. all artist nodes are orange). This KG has been used as a reference for some of the applications mentioned below.

Fig. (2) demonstrates the end-to-end workflow of one way of implementing the backend of a voice assistant where the LLM is enhanced by a KG.

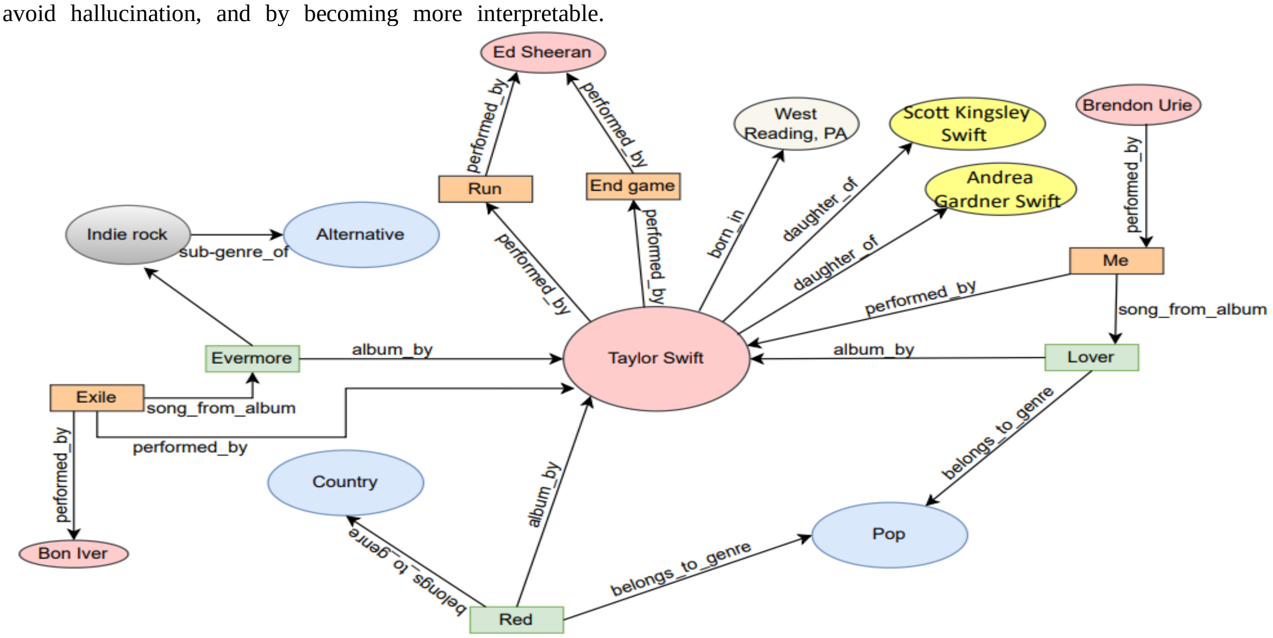

Fig. 1: An example of (part of) a KG based on music data

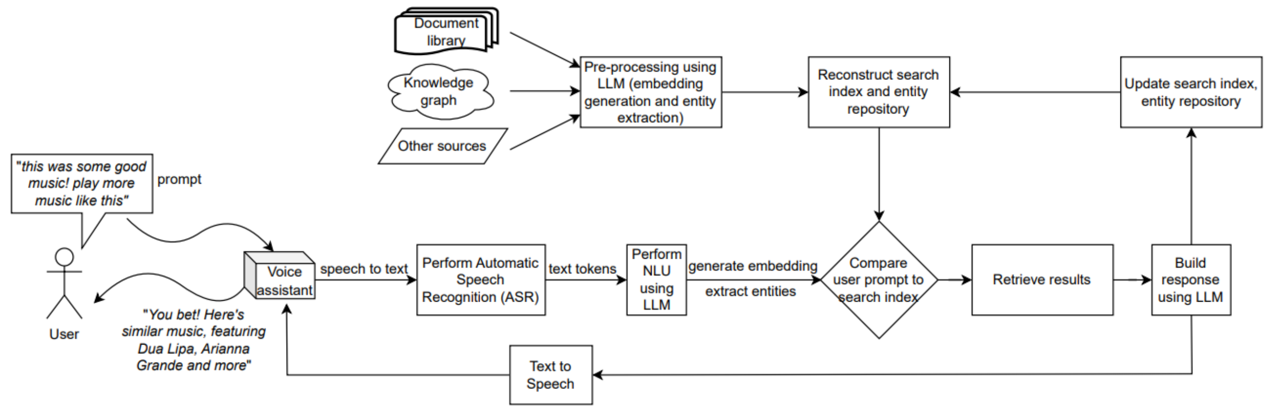

Fig. 2: Journey from user prompt to a voice assistant via LLM based processing to TTS output

Below are some applications of this, or a similar implementation of a KG-enhanced LLM, for music experiences on voice assistants.

5.1 Providing High Precision Responses to Music-related Queries

An LLM enhanced by a KG can accurately respond to specific queries, either for playback or textual information, without hallucinations. For example, if a user is listening to a song by Taylor Swift and makes a vague request to the voice assistant (e.g. "*tell me more about* her"), the LLM is able to understand this contextual and ambiguous request (i.e. ability to disambiguate the reference to "her"). Further, instead of 'hallucinating up' information about the artist, the KG enhanced LLM can provide an accurate response like "*Taylor Swift was born in West Reading Pennsylvania, to Scott Kingsley Swift and Andrea Gardner Swift. Swift's musical creation spans multiple genres like country, pop and alternative genres. She has also collaborated with other music artists like Ed Sheeran, Bon Iver, Brendon Urie, and others. Would you like to learn more about her?"* and continue the conversation (e.g. "*tell me more about her music in the country genre*"), a task well suited for an LLM.

5.2 Crafting Personalized Music Recommendations

If an LLM informed by a music data rich KG is used to generate recommendations for music requests on voice assistants, it can leverage the additional information available in the KG to provide high quality personalized responses [24]. For example, a user might ask a voice assistant to "*play more songs like this*" when a Taylor Swift song is playing. Along with knowing what song is currently playing and what its characteristics (e.g. lyrical meaning, audio attributes, vocals, etc.), the user's musical preferences, previous listening history and contextual information, the LLM can leverage the relationships in a KG to arrive at tracks that are also similar, lyrically and/or musically, and incorporate those into the recommendation to make it more personalized.

5.3 Assisting in Discovery of New Music

By using the relationships that a KG provides, an LLM can be effective in terms of driving the discovery of music that the user hasn't heard before. For example, a KG informed LLM can derive that when a user makes the request "*play something new by Taylor Swift*" to a voice assistant, based on the music they have already explored from the artist's discography, what are the songs that the user is likeliest to enjoy and haven't heard before. For example, knowing that the user who made the request ("*play something new by Taylor Swift"*) is also a fan of the artist Ed Sheeran, and that there's a song that was recently released by the two artists, Taylor Swift and Ed Sheeran, in collaboration, and that it belongs to the pop genre that the user displays affinity towards, and that the user hasn't heard before, can help craft a suitable response from the voice assistant (e.g. "*here's a new song by Taylor Swift and Ed Sheeran that you haven't heard before, called* Run*"*). See Fig. 1 as a demo. KG-powered LLMs can also help discover new music through questions asked to the voice assistants (e.g. "*is there any collaboration between Ed Sheeran and Taylor Swift?*" can return songs by the duo that the user is likeliest to like). This extends to finding other versions of a song, covers, and more.

5.4 Interpretation of Lyrics

KGs can provide detailed interpretations of lyrics of a song. Further, the relationships of the track whose lyrics the user is interested in with other related concepts, cultural references and events, historical events, and more, can let a KG-powered LLM offer much deeper insights and interpretations of those lyrics. For example, if a user were to ask a voice assistant, "*can you tell me the meaning of the lyrics of the song Run by Taylor Swift and Ed Sheeran",* the voice assistant can explain the deeper meaning behind the lyrics of the song. Further, if the user were to ask for specific / niche music, the KG-powered LLM might be able to present highly relevant and personalized responses. For example, if a user were to say the voice assistant to "*play something that can make me feel better. I just had a fight with my best friend*", by knowing the lyrical interpretations of all songs, the responses could be highly personalized to tracks that the user would find particularly comforting, and are by the artists that they like and at the tempo they prefer given their mood.

5.5 Engaging and Fun Human Machine Interactions

LLMs enhanced by KGs and used by a voice assistant can make the voice assistant a virtual host for entertaining guests. For example, the voice assistant could create competitive trivia competitions based on questions like "*Who was the first record label to sign Taylor Swift?"* or "*What album does the song Exile belong to?"*. The voice assistant can provide an interactive, fun and engaging experience.

6. CHALLENGES OF KG-ENHANCED LLM APPLICATIONS IN MUSIC ON VOICE ASSISTANTS

While a KG-powered LLMs can enable new capabilities on voice assistants that bring the best of both KGs and LLMs to create an enhanced user experience, there are some challenges that continue to persist.

6.1 Latency

Having low latency is critical to applications on voice assistants to avoid user frustration. Along with network latency that a voice assistant anyway has to overcome, the additional time taken for KG traversal for querying and retrieval as well as LLM processing time also add latency to the end-to-end time taken for a response to a user request. Further, KGs need to be updated with new information regularly, especially for KGs from a domain like music where new releases keep happening. All this increases latency to get the most updated information to the user.

6.2 Compute cost

LLMs require significant computational resources to train and execute [25], creating significant compute cost. Further, the cost of running LLMs in cloud can be very expensive especially for a high-volume use case such as asking for music on the voice assistant.

6.3 Data Quality

Creating a KG with music information that has both comprehensive and accurate data requires that the KG has frequent access to high quality, standardized data sources. With multiple different sources of music in different country / region and further made complex with different languages, it is challenging to keep the KG updated.

6.4 Disambiguation of Entities

Even with a KG, an LLM can't at times help with certain queries that have multiple interpretations. For example, if a user asks the voice assistant to "*play Evermore"*, the voice assistant might fail because the LLM starts playing the *album* Evermore while the user was referring to the *song* Evermore.

6.5 Privacy

Unlike mobile devices (e.g. iPhone, Android phone), voice assistants might be located in a place where they are accessible to more than one user. As a result, privacy, especially with respect to any user-specific sensitive / personal data that might be stored in the KG, might result in the LLM crafting a response that inadvertently gives out that personal data.

7. CONCLUSION

This paper explored the advantages and shortcomings of KGs and LLMs and demonstrated how their usage together can help overcome some of the shortcomings of either. Using a real-life example of an artist (Taylor Swift), it also demonstrated the applications of a KG-enhanced LLM via voice assistants for music experiences. Finally, it articulated some of the challenges that still persist with a KG-enhanced LLM when it comes to providing a superior music experience via voice assistants. These challenges need further research as well as improvements to implementation in order to be overcome

8. REFERENCES

[1] Roberto Navigli, Natural Language Understanding: Instructions for (Present and Future) Use, Proceedings of the Twenty Seventh International Joint Conference on Artificial Intelligence (IJCAI-18)
[2] A Lenci - arXiv preprint arXiv:2303.04229, 2023 - arxiv.org
[3] Manaal Faruqui, Dilek Hakkani-Tür, Revisiting the Boundary between ASR and NLU in the Age of Conversational Dialog Systems, Computational Linguistics (2022) 48 (1): 221–232.
[4] Samuel R. Bowman, George E. Dahl, What Will it Take to Fix Benchmarking in Natural Language Understanding?
[5] Amit Singhal. 2012. Introducing the Knowledge Graph: things, not strings. Google Blog. https://www.blog.google/products/search/introducing-knowledge-graph-things-not/
[6] Aidan Hogan, Eva Blomqvist, Michael Cochez, et all. Knowledge Graphs, arXiv:2003.02320
[7] Shaoxiong Ji, Shirui Pan, Erik Cambria, Pekka Marttinen, Philip S. Yu, A Survey on Knowledge Graphs: Representation, Acquisition and Applications. IEEE TNNLS 2021.
[8] Zhao, W. X. et al. (2023, March 31). A Survey of Large Language Models, https://doi.org/10.48550/arXiv.2303.18223
[9] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed Representations of Words and Phrases and their Compositionality. Advances in Neural Information Processing Systems, 26. https://doi.org/10.48550/arXiv.1310.4546
[10] https://snorkel.ai/large-language-models-llms/
[11] https://www.tasq.ai/blog/large-language-models/
[12] https://www.boost.ai/blog/a-hybrid-llm-chat-experience
[13] https://machinelearningmastery.com/a-gentle-introduction-to-hallucinations-in-large-language-models/
[14] https://spectrum.ieee.org/ai-hallucination
[15] Furkan Ufuk, The Role and Limitations of Large Language Models Such as ChatGPT in Clinical Settings and Medical Journalism, https://doi.org/10.1148/radiol.230276
[16] https://towardsdatascience.com/overcoming-the-limitations-of-large-language-models-9d4e92ad9823

[17] Ying Sheng, Lianmin Zheng, Binhang Yuan, Zhuohan Li, Max Ryabinin, Daniel Y. Fu, Zhiqiang Xie, Beidi Chen, Clark Barrett, Joseph E. Gonzalez, Percy Liang, Christopher Re´, Ion Stoica, Ce Zhang, FlexGen: High-Throughput Generative Inference of Large Language Models with a Single GPU, arXiv:2303.06865v2 [cs.LG] 12 Jun 2023

[18] https://engineb.com/2021/02/8-key-benefits-of-knowledge-graphs/

[19] Ciyuan Peng, Feng Xia, Mehdi Naseriparsa, and Francesco Osborne, Knowledge Graphs: Opportunities and Challenges, doi: 10.1007/s10462-023-10465-9

[20] Bilal Abu-Salih, Domain-specific Knowledge Graphs: A survey, King Abdullah II School of Information TechnologyThe University of Jordan

[21] Mahfouz, Mahmoud & Nourbakhsh, Armineh & Shah, Sameena. (2021). A Framework for Institutional Risk Identification using Knowledge Graphs and Automated News Profiling.

[22] Corbin L Rosset, Chenyan Xiong, Minh Phan, Xia Song, Paul N. Bennett, Saurabh Tiwary, Pretrain Knowledge-Aware Language Models, ICLR 2021 Conference

[23] Yu, Donghan & Zhu, Chenguang & Yang, Yiming & Zeng, Michael. (2020). JAKET: Joint Pre-training of Knowledge Graph and Language Understanding.

[24] Ashlesha V Kadam, "Applications of AI and NLP to advance Music Recommendations on Voice Assistants", International Journal of Engineering, Business and Management (IJEBM), ISSN: 2456-7817 [Vol-7, Issue-3, May-Jun, 2023]

[25] Christophe Van Gysel, "Modeling Spoken Information -eries for Virtual Assistants", https://arxiv.org/abs/2304.13149v1