

Perspectives for Leveraging LLMs for Knowledge Engineering

Cogan Shimizu

cogan.shimizu@wright.edu

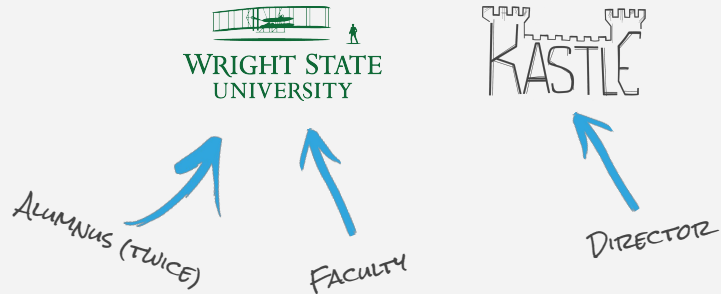
coganshimizu.com

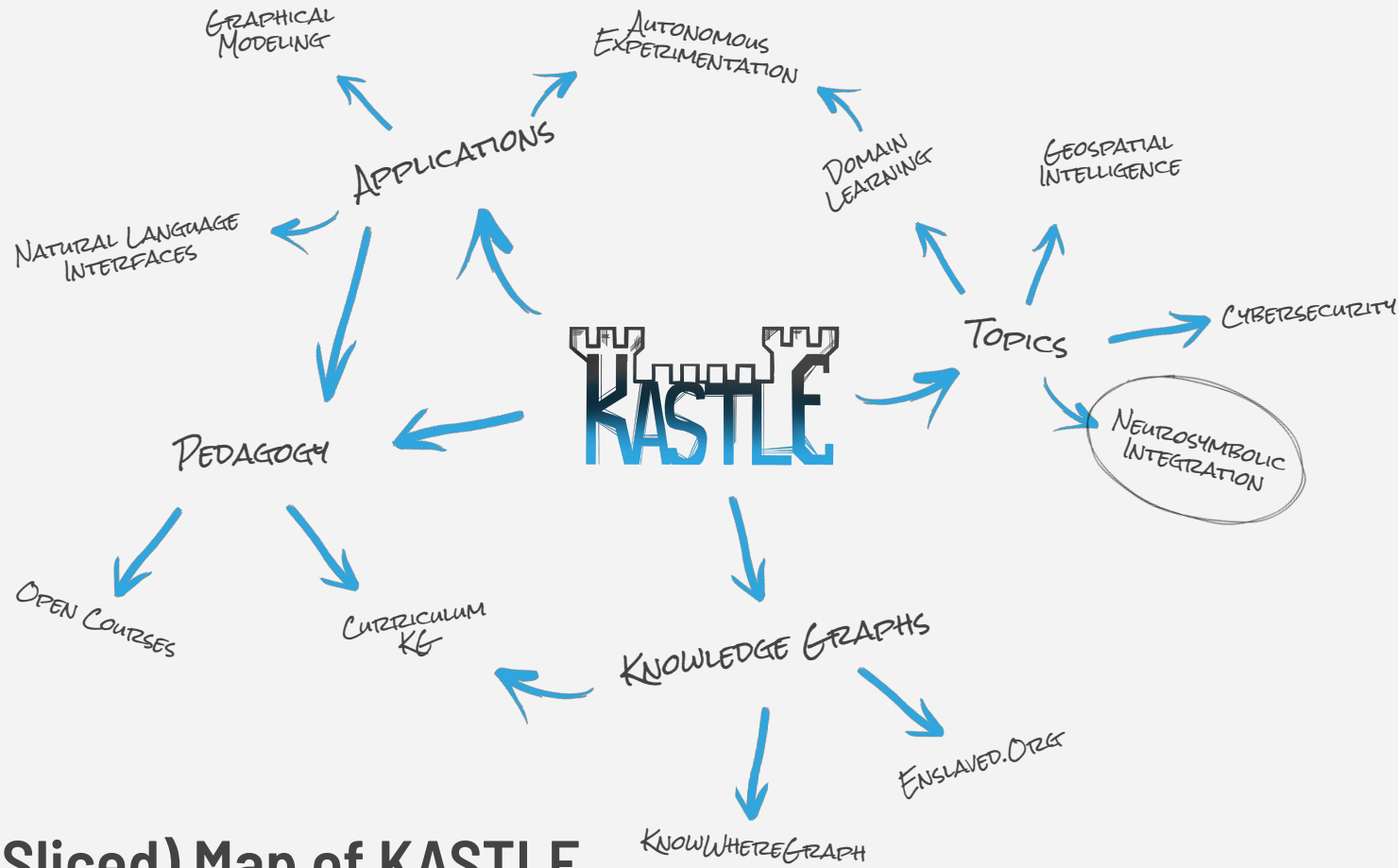
Knowledge and Semantic Technologies Lab
Wright State University

Cogan Shimizu

Assistant Professor

Department of Computer Science & Engineering
Wright State University





A (Sliced) Map of KASTLE

Outline & Objectives

An Overview

Introducing the Nesycule 4

Patterns for Neurosymbolic AI

Discussion

An Overview

Introducing the Nesycule 4

Patterns for Neurosymbolic AI

Discussion

Outline & Objectives

Overview

This keynote is (nearly) all positioning.

- Discuss Neurosymbolic AI – and what it comprises for this talk
- Introduce the “Nesycule 4” and some examples of Neurosymbolic AI subdomains
- Propose a framework for using structured knowledge to map between latent spaces
- Discussion

Overview: Large Language Models

Attention Is All You Need

Ashish Vaswani*
Google Brain
avaswani@google.com

Noam Shazeer*
Google Brain
noam@google.com

Niki Parmar*
Google Research
nikip@google.com

Jakob Uszkoreit*
Google Research
usz@google.com

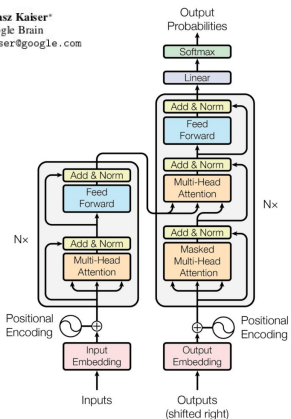
Llion Jones*
Google Research
llion@google.com

Aidan N. Gomez*[†]
University of Toronto
aidan@cs.toronto.edu

Lukasz Kaiser*
Google Brain
lukaszkaizer@google.com

Illia Polosukhin*[†]
illia.polosukhin@gmail.com

Stochastic Parrot or
Natural Language
Understanding?



Stanford
Alpaca

LangChain

ChatGPT

Overview: *Large Language Models*

Internal Knowledge

- An LLM does know things!

Crawling the Internal Knowledge-Base of Language Models

Roi Cohen, Mor Geva, Jonathan Berant, Amir Globerson

- But also, it doesn't: "hallucination"
 - "produce content that is nonsensical or untruthful in relation to certain sources."



neil-gaiman [Follow](#)



hano-dakukita asked:



Hello Mr.Gaiman, I hope you won't find this offensive, but ChatGPT claimed that you are bisexual. I just wonder is that a real thing or it is just another stupid AI made up thing.

It's ok if you don't want to share it, it is a very personal question anyway. Wish you a good day, sir.

Not as far as I know, I'm afraid.

AI isn't about truth though, just truth-shaped sentences.

Overview: *Knowledge Engineering*

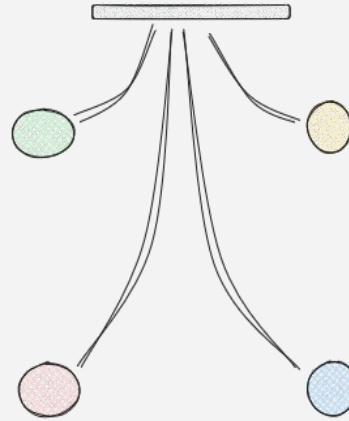
Some Examples...

Schema Development

Develop a schema for a KG with a methodology, using available data and subject matter expertise

Schema Learning

Automatically generate a schema for knowledge extracted from unstructured text corpora



Knowledge Alignment

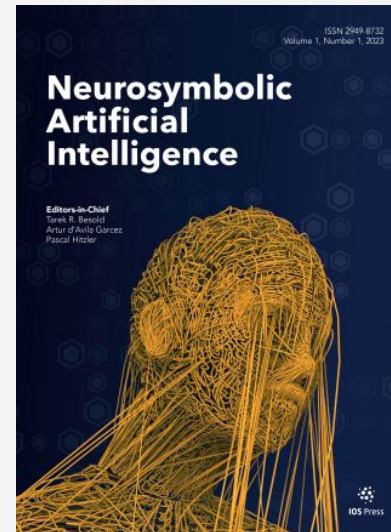
Integrate knowledge fragments modeled in conceptually distinct ways into a singular KG

KG Deployment

Understand the technology stack for efficiently deploying and exposing a knowledge graph

Overview: Neurosymbolic Artificial Intelligence

"Neurosymbolic AI is an emerging field of AI aiming to build rich computational AI models, systems and applications by combining neural and symbolic learning and reasoning."



The Spaghetti Monster

Outline & Objectives

An Overview

Introducing the Nesycule 4

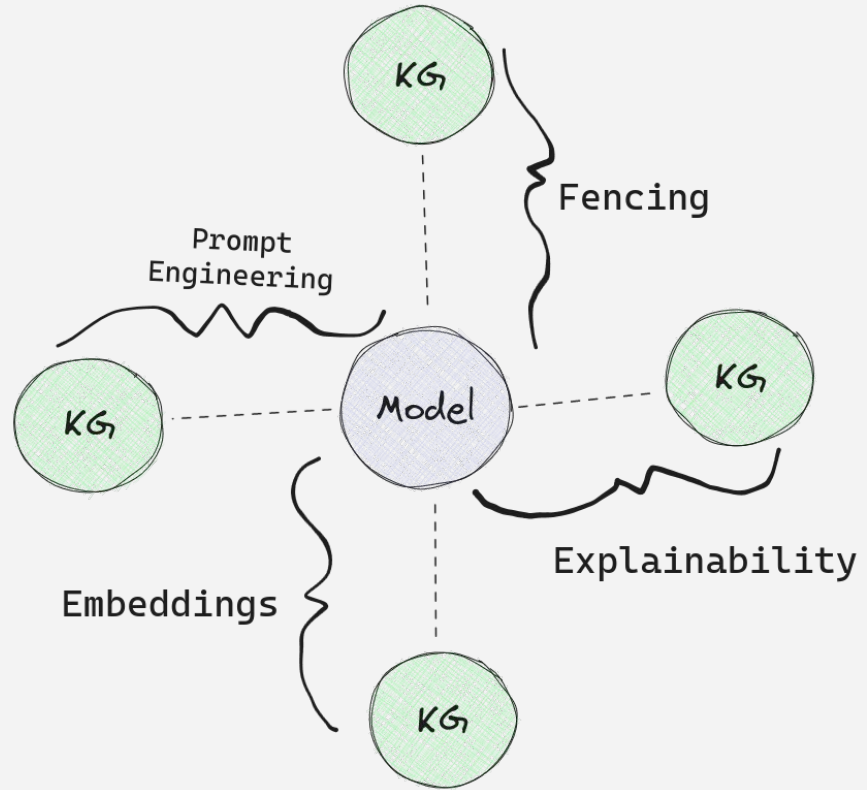
Patterns for Neurosymbolic AI

Discussion

Nesycule 4

What even is this thing?

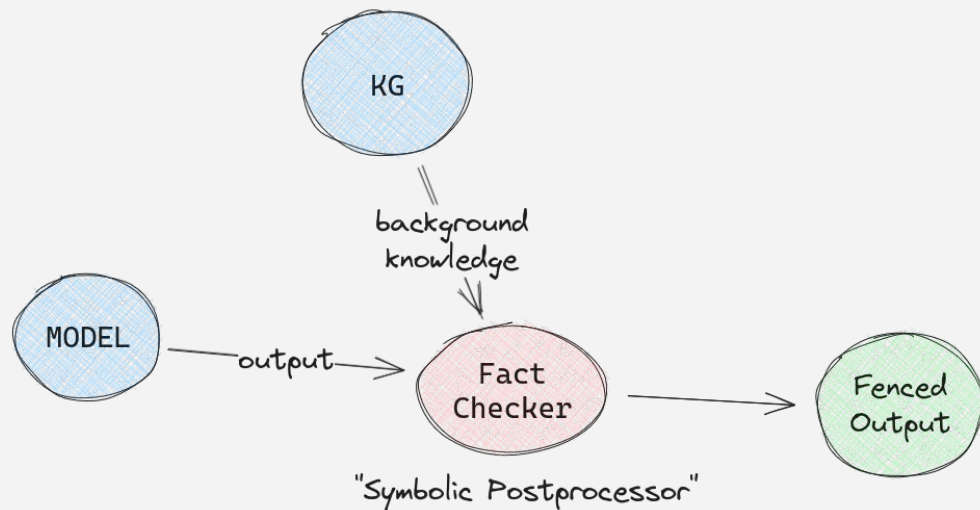
- Imagining relations between an AI/ML/DL Model in an intuitive “spatial” dimension
- KG “over” a model
- KG “alongside” a model
- Model “over” a KG



Nesycule 4: Fencing

Symbolic Post-processing

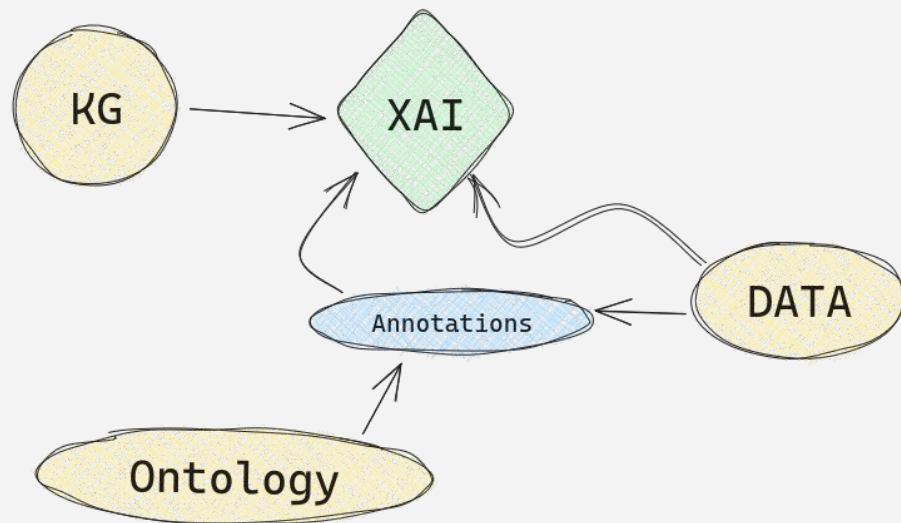
- The process by which output from an LLM (or some model) is fact-checked (common-sense, domain knowledge) for correctness
- Bridging structured knowledge to natural language
 - Lexico-syntactic patterns?
 - Controlled english?
 - Both!?



Nesycule 4: Explainability

KG “alongside” Model (right)

- Learn associations between structured knowledge and raw data, by annotating the raw data according to an ontology (+ common sense)



Nesycule 4: *Prompt Engineering*

KG “alongside” Model (left)

- Controlled Language (redux)
- (OD) Pattern-based language
- Topic Modeling & Extraction



- Keyword extraction as degenerate text summarization?



- Ask it to generate code which constructs KGs according to the use of some pattern

- Ask for output in Controlled Language (e.g., ACE) can help reduce ambiguity in downstream tasks

The key idea is to leverage the internal knowledge base of the LLM in a predictable manner.

Outline & Objectives

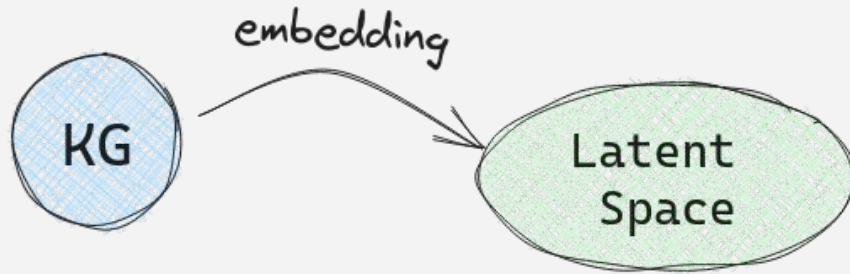
An Overview

Introducing the Nesycule 4

Patterns for Neurosymbolic AI

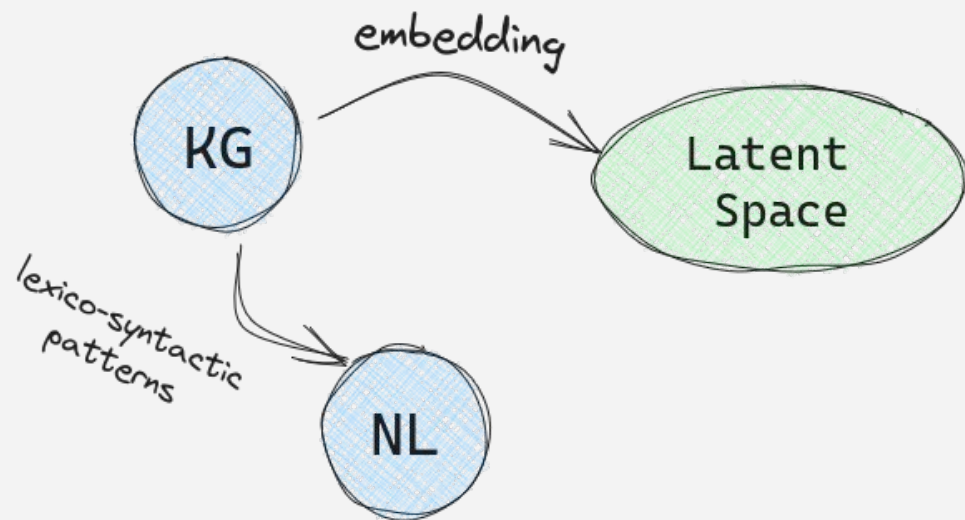
Discussion

Structured Knowledge to Latent Space



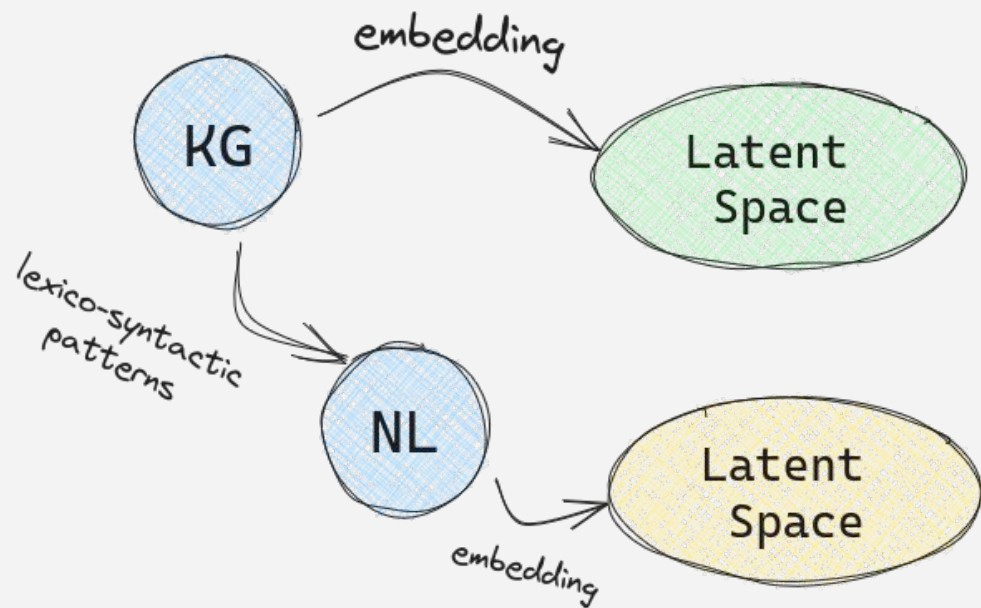
First, start with a KG and embed it into a latent space.

Structured Knowledge to Natural Language



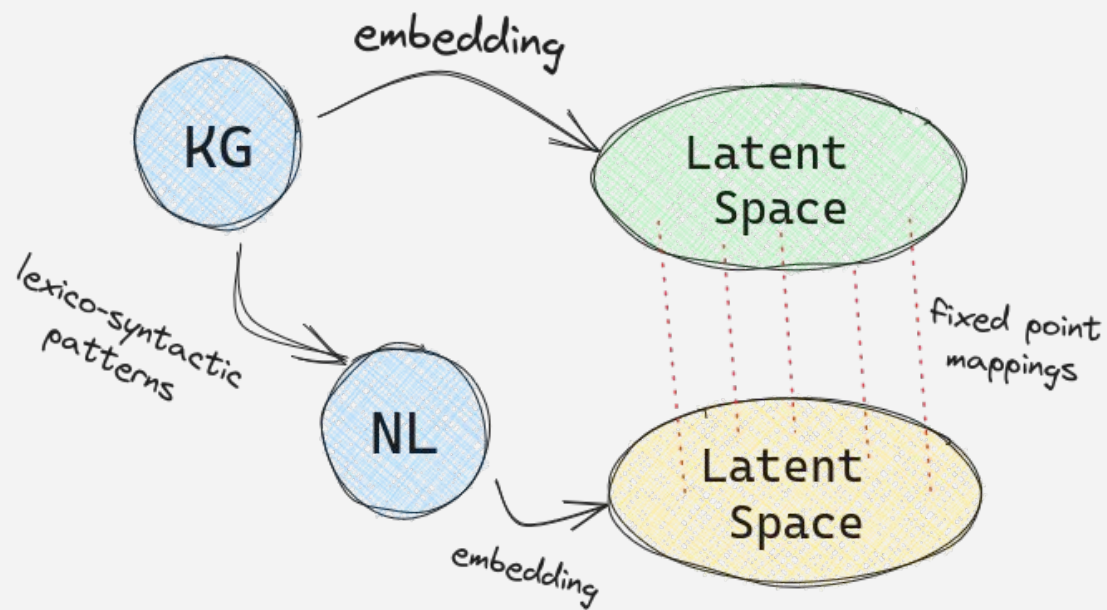
Then, using a library of LSPs, convert the KG into an NL corpus.

Natural Language to Latent Space



Take the NL Corpus and embed that into its own latent space.

Fixed-Point Mapping



Now, train some model to map between the fixed points in the latent spaces.

Follow-up

So what does this get us?

- Extended domain transferability and augmentation
 - Describe an image in Natural Language, do we then have structured knowledge about it?
- An “easy” way to translate natural language back into structured knowledge
 - Leverage metadata and ontology once there?
- What else? (discuss!)



From iStock photos

Outline & Objectives

An Overview

Introducing the Nesycule 4

Patterns for Neurosymbolic AI

Discussion

THANKS!

Cogan Shimizu
cogan.shimizu@wright.edu
coganshimizu.com

Knowledge and Semantic Technologies Laboratory
Joshi 390