

From NLP to NLU: Why we need varied, comprehensive, and stratified knowledge, and how to use it for Neuro-symbolic AI

Keynote at KnowledgeNLP-AAAI'23

Amit Sheth

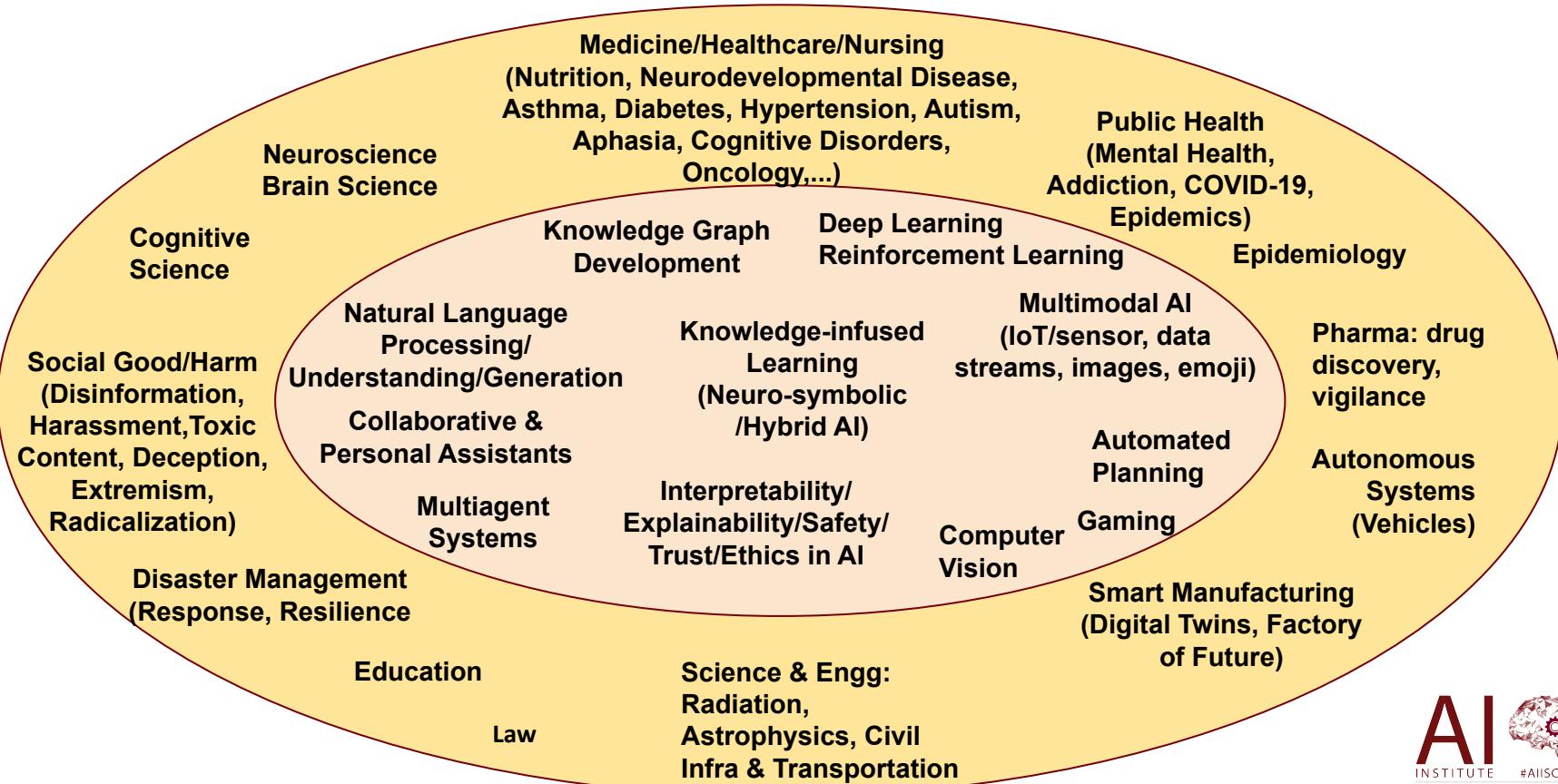
Director, AI Institute of South Carolina
University of South Carolina

amit@sc.edu



#AIISC, <http://aiisc.ai>

AIISC portfolio in Core AI & Translational AI



Contents

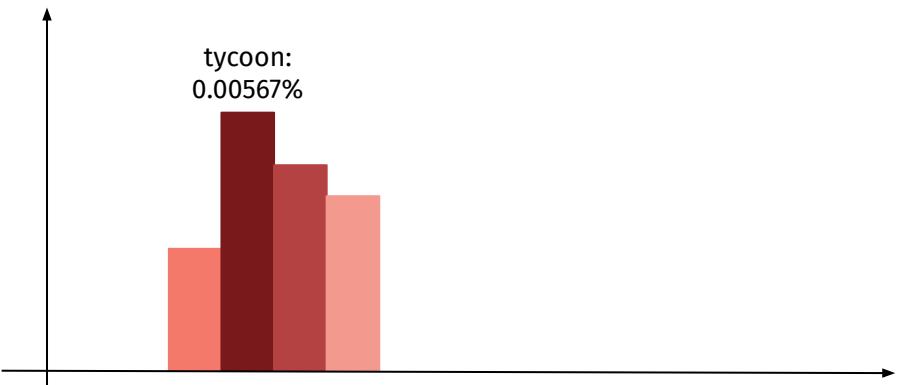
1. Challenges with Current LMs
2. Possible Fixes and Limitations
3. Knowledge Infused Neuro-Symbolic AI
4. Before Transformers, Historical Context [Neuro Symbolic AI]
5. After Transformers, Current Context [Neuro Symbolic AI]
6. Future Context [Neuro-Symbolic AI] with Process Knowledge Infusion.

Challenges with Current LMs

World Model Semantics from Distributed Semantics at Scale?

Tycoon

Did you mean: tycoon
Did you mean: typhoon
Did you mean: typography



Distributional Semantics:

Language Models are trained to compute the *distributional plausibility* of language tokens from enormous amounts of training tokens.

World Model Semantics:

Relationships and effects among the objects that the language tokens describe, grounded in the physical or conceptual reality of the world humans experience (ontological commitment).

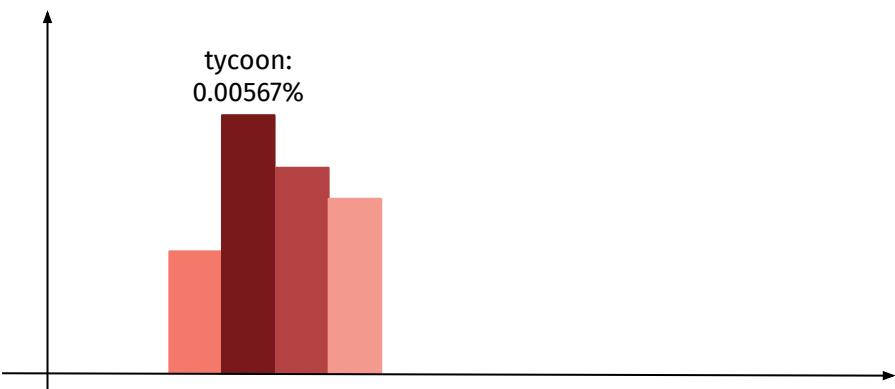
Do World Model Semantics Arise as an Emergent Capability of Distributional Semantics at Scale? [No- not directly, not specifically, not adequately]

Challenges with Current LMs

Recency : World Model Semantics from Distributed Semantics at Scale

Tycoon

Did you mean: tycoon
 Did you mean: typhoon
 Did you mean: typography



Formally

$\forall i \subseteq [N]$, N = Vocabulary Size,

Model the probability:

$P(\text{token}_i, \dots, \text{token}_N \mid \text{token}_1, \dots, \text{token}_i)$

Example Usage

**Once Trained on Enormous amounts of Data,
 Predict:**

1. $y \sim P(S = \text{"Obama just won the 2032 Election"})?$
2. $P(S)$ does not pass a threshold test
 $\Rightarrow y = 0, \text{ or } False.$

Challenges with Current LMs

Common Sense : World Model Semantics from Distributed Semantics at Scale

Mike's mum had 4 kids Luis, Drake and Matilda of 4th kid?

vz What was the colour of the white horse of Napoleon?

It is not possible to determine the name of the fourth child with more information.

The answer is in the question.

I apologize for the confusion. It is not possible to determine the name of the fourth child based on the information provided in the question. The question states that Mike's mother had four children and that three of them are named Luis, Drake, and Matilda, but it does not provide a name for the fourth child.

Distributional Semantics:

Language Models are trained to compute the *distributional plausibility* of language tokens from enormous amounts of training tokens.

E.g., What was the color of the white horse of Napoleon?

It is not very *distributionally plausible* that a person asks the answer to a question they already know

The answer is abundantly clear in the question

But, (a) distribution semantics has problem with the low probability, and (b) it does not understand! More challenges arise due to lack of context.

Challenges with Current LMs

Challenge - Missing Information: World Model Semantics from Distributed Semantics at Scale

vz What was the name of Napoleon?

It is not possible to determine the name of the white horse because there is no information provided about which white horse Napoleon Bonaparte statesman and general rode during the Napoleonic Wars. It is not clear which specific white horse he was asking about.

Mike's mum had 4 kids; 3 of them are Luis, Drake and Matilda. What is the name of 4th kid?

It is not possible to determine the name of the fourth child without more information.

The answer is in the question

I apologize for the confusion. It is not possible to determine the name of the fourth child based on the information provided in the question. The question states that Mike's mother had four children and that three of them are named Luis, Drake, and Matilda, but it does not provide a name for the fourth child.

Distributional Semantics:

Language Models are trained to compute the *distributional plausibility* of language tokens from enormous amounts of training tokens.

E.g., Mike's mum had 4 kids; 3 of them are X,Y,Z. What is the name of 4th kid?

- Missing information is not clear

Answer is obviously Mike

1. *Mike is not a pet*
2. *More than one child cannot have the same name*
3. *Kid names cannot be any word*
4. *Impossible answers even after the hint:
Mike, Luis, Drake, and Matilda, and All other tokens in the input!*

Distribution semantics does not have commonsense and probability does not help.

Challenges with Current LMs

Challenge - Missing Information: World Model Semantics from Distributed Semantics at Scale

VZ What was the name of Napoleon?

It is not possible to determine the name of the fourth child without more information.

The answer is in the question

I apologize for the confusion. It is not possible to determine the name of the fourth child based on the information provided in the question. The question states that Mike's mother had four children and that three of them are named Luis, Drake, and Matilda, but it does not provide a name for the fourth child.

Distributional Semantics:

Language Models are trained to compute the *distributional plausibility* of language tokens from enormous amounts of training tokens.

E.g., Mike's mum had 4 kids;3 of them are X,Y,Z. What is the name of 4th kid?

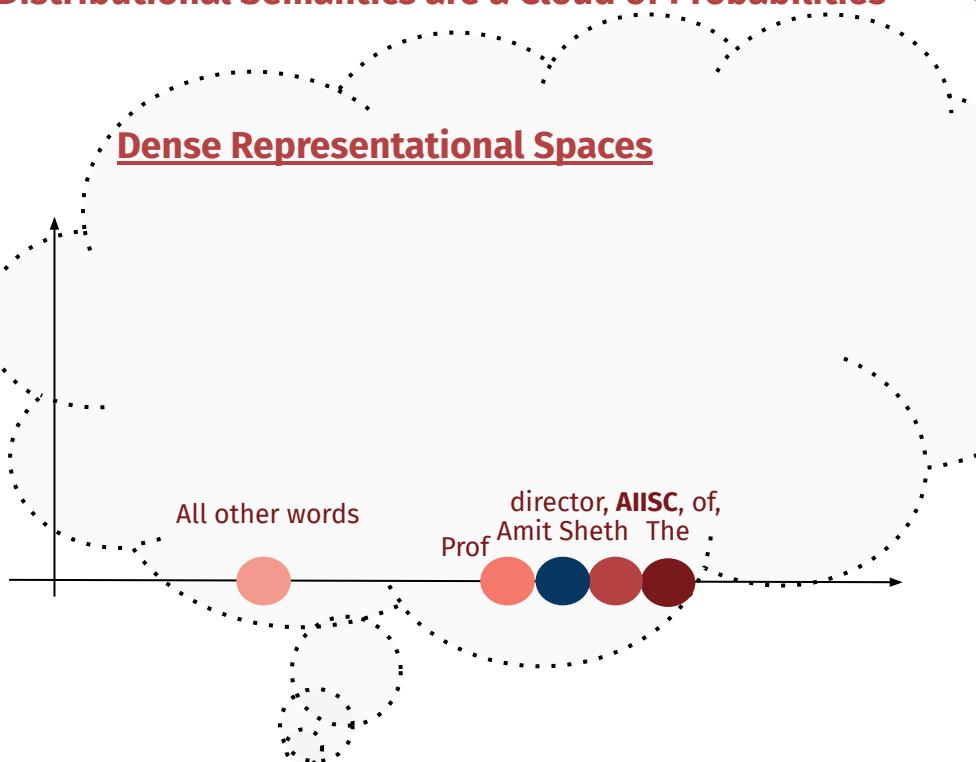
Human's first answers (We asked 10 fellow humans)

1. Mike
2. I think this is a trick question

Distribution semantics does not know to make human-like assumptions about missing information.

Challenges with Current LMs

Distributional Semantics are a Cloud of Probabilities



Hallucinations - It makes up things !

SR Amit Sheth is the director of AIISC. Prof. Amit Sheth leads several research projects at the Institute.

Amit Sheth is a director of AIISC (Artificial Intelligence and Integrated Intelligent Systems Center) and leads several research projects at the institute.

SR Who is Prof. Amit Sheth?

Prof. Amit Sheth is a researcher and professor in the field of artificial intelligence and data science. He is known for his work in the areas of semantic web, linked data, and artificial intelligence. He is also the director of AIISC (Artificial Intelligence and Integrated Intelligent Systems Center) where he leads several research projects.

SR What Institute does Amit Sheth belong to?

I am not aware of the specific Institute that Prof. Amit Sheth belongs to. The

I kind of understand (not really though)



Challenges to be addressed for NLU

Characteristics of NLU Capable Systems

■ Recency

■ Common-sense

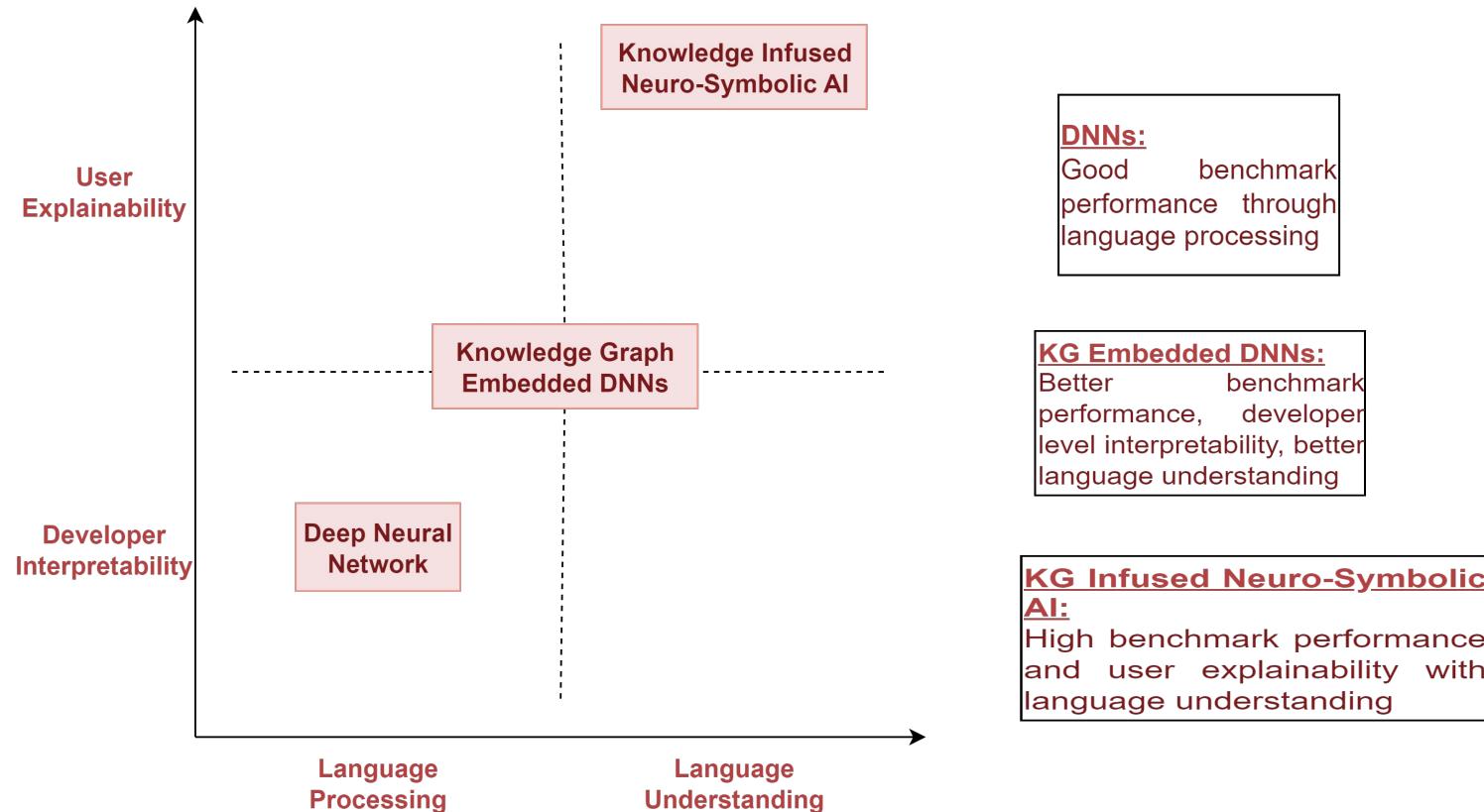
■ Insufficiency of data alone ■

Hallucinations

■ User-Explainability ■

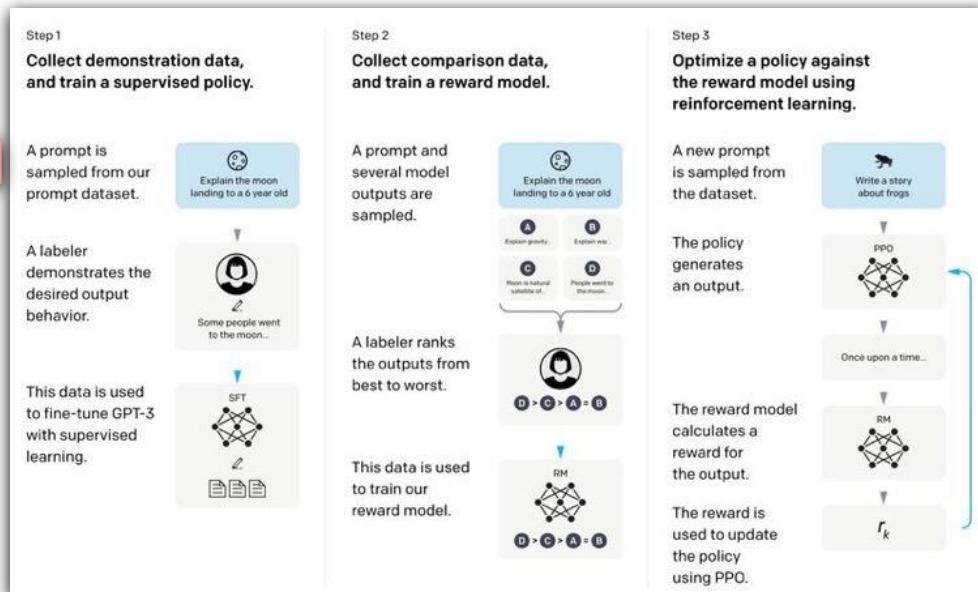
Application-level Safety

Challenges to be addressed for NLU



Possible Fix: Instruct GPT

The Instruct GPT Framework



Hallucinations - It makes up things !

SR Amit Sheth is the director of AIISC. Prof. Amit Sheth leads several research projects at the Institute.

Amit Sheth is a director of AIISC (Artificial Intelligence and Integrated Intelligent Systems Center) and leads several research projects at the institute.

SR Who is Prof. Amit Sheth?

Prof. Amit Sheth is a researcher and professor in the field of artificial intelligence and data science. He is known for his work in the areas of semantic web, linked data, and artificial intelligence. He is also the director of AIISC (Artificial Intelligence and Integrated Intelligent Systems Center) where he leads several research projects.

SR What Institute does Amit Sheth belong to?

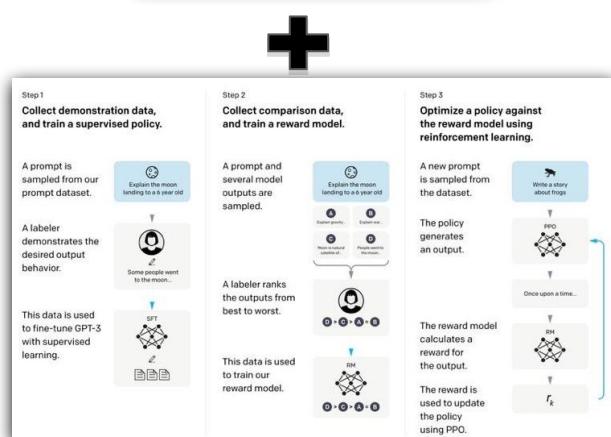
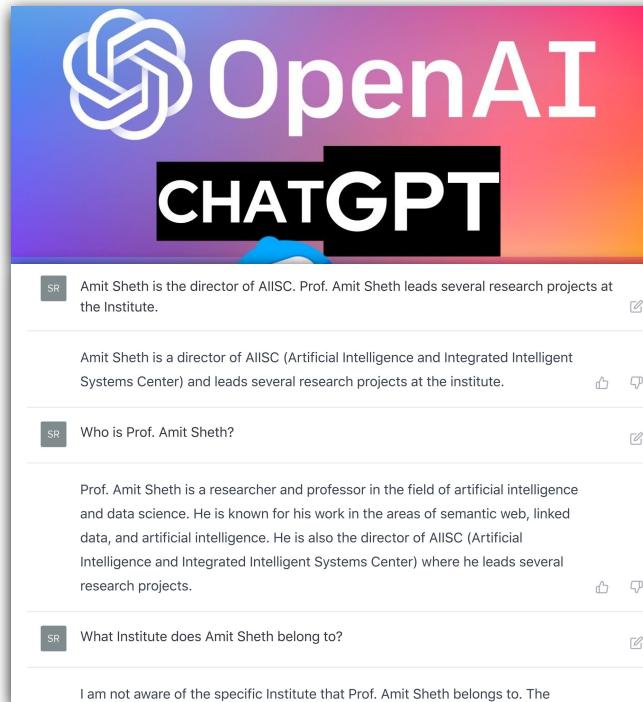
I am not aware of the specific Institute that Prof. Amit Sheth belongs to. The

No Problem I will make things larger and use Instruction based Training



Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., ... & Lowe, R. (2022). *Training language models to follow instructions with human feedback*. arXiv preprint arXiv:2203.02155. [link](#)

1. 40 humans to capture the breadth of knowledge corresponding to the data in LLMs seems small
2. The richness of human knowledge is compressed into a mere label

OpenAI CHATGPT

SR Amit Sheth is the director of AIISC. Prof. Amit Sheth leads several research projects at the institute.

Amit Sheth is a director of AIISC (Artificial Intelligence and Integrated Intelligent Systems Center) and leads several research projects at the institute.

SR Who is Prof. Amit Sheth?

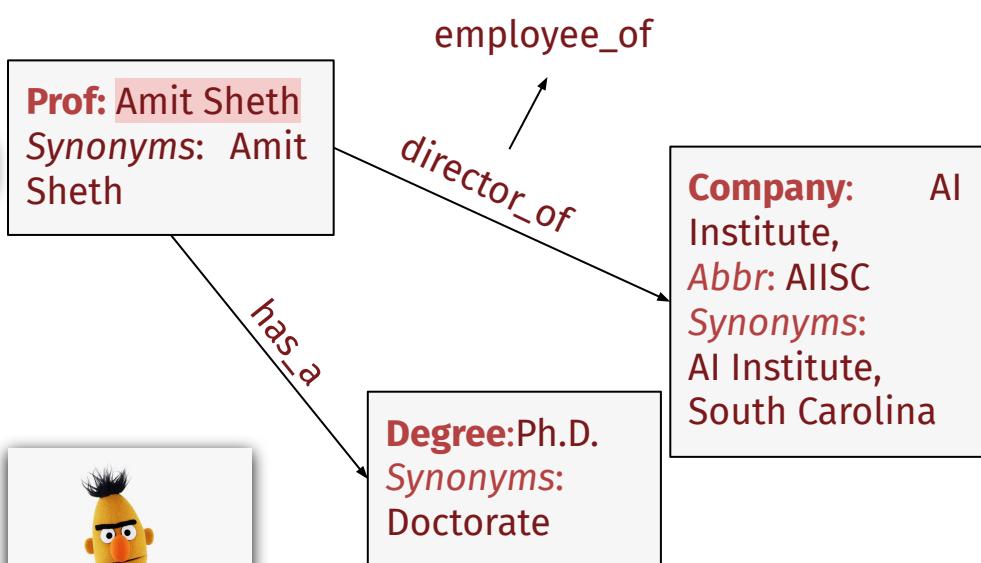
Prof. Amit Sheth is a researcher and professor in the field of artificial intelligence and data science. He is known for his work in the areas of semantic web, linked data, and artificial intelligence. He is also the director of AIISC (Artificial Intelligence and Integrated Intelligent Systems Center) where he leads several research projects.

SR What Institute does Amit Sheth belong to?

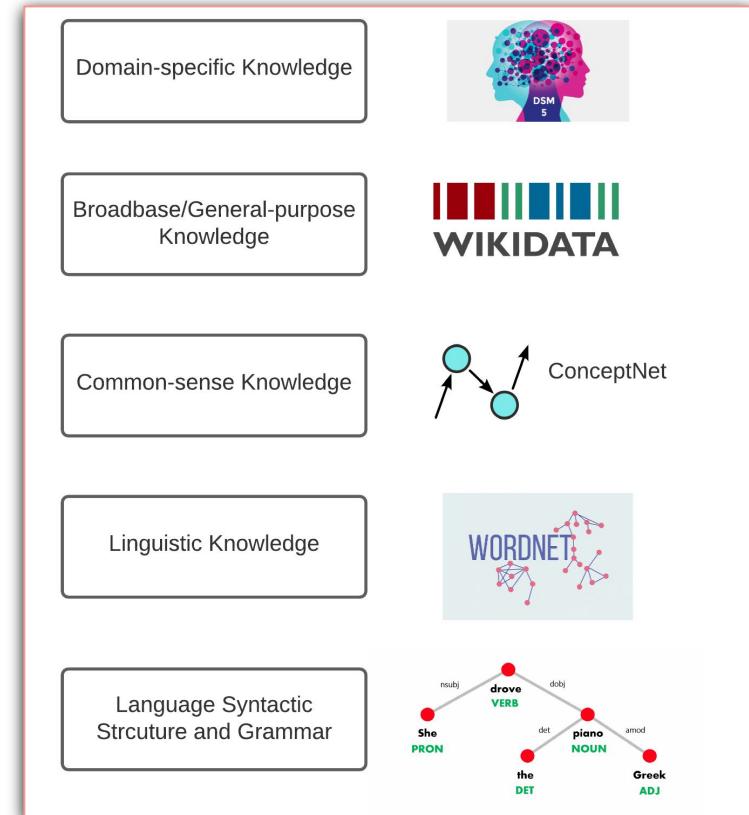
I am not aware of the specific Institute that Prof. Amit Sheth belongs to. The

Knowledge (Graphs) to the rescue

Fix: Addressing Hallucinations: Recency, Common-Sense, and Implicit Entity Mentions, etc.,



Let's see what this model of understanding can yield



Semantics supported by Knowledge (Graphs) AI

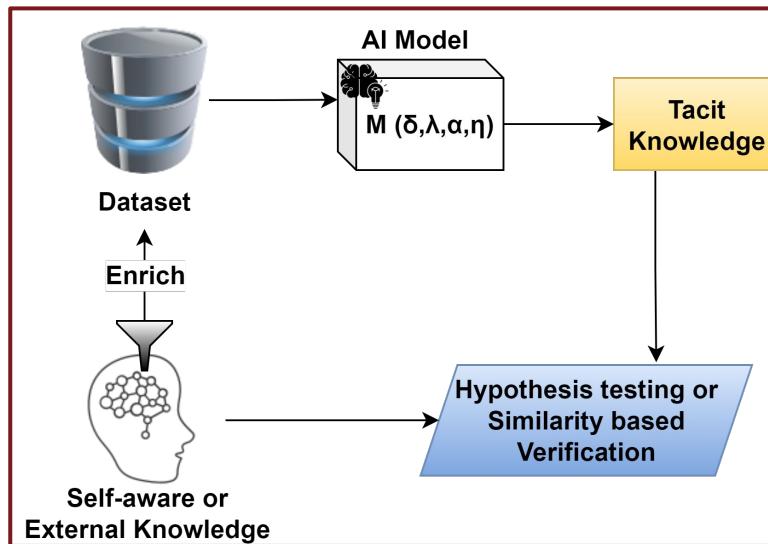
- Explicit model of recency and common-sense.
- Supply missing Knowledge (entities, relationships).
- World as Concepts vs. World as Probabilities.
- User-level Explanations and Safety Constraints

Knowledge Infused Neuro-symbolic AI

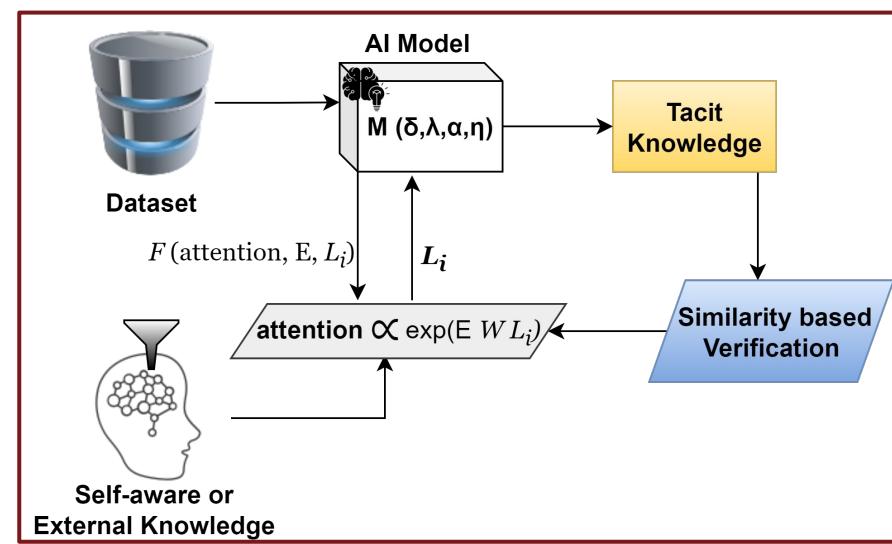
Knowledge Infused Neuro-symbolic AI

Using Graphs in Neural Network Pipelines

Shades of KiL - Shallow and Semi-Deep Infusion

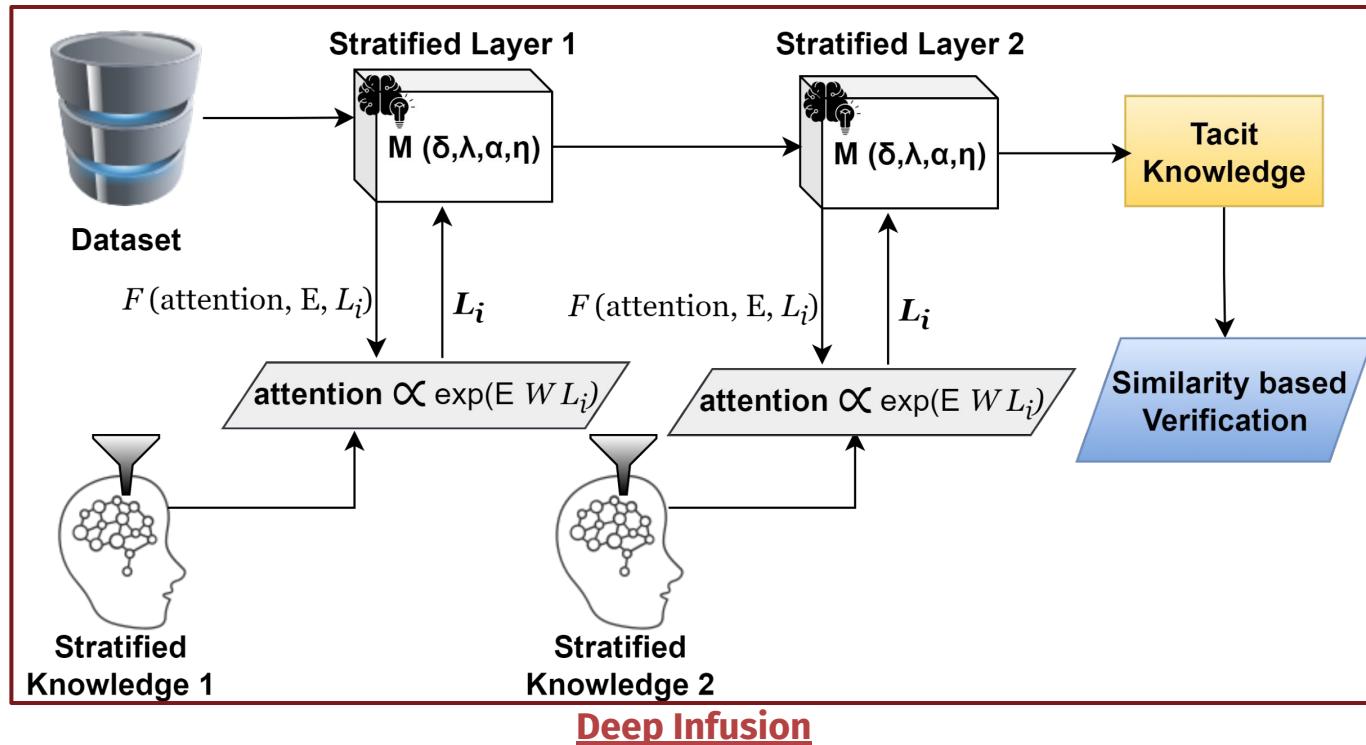


Shallow Infusion



Semi-Deep Infusion

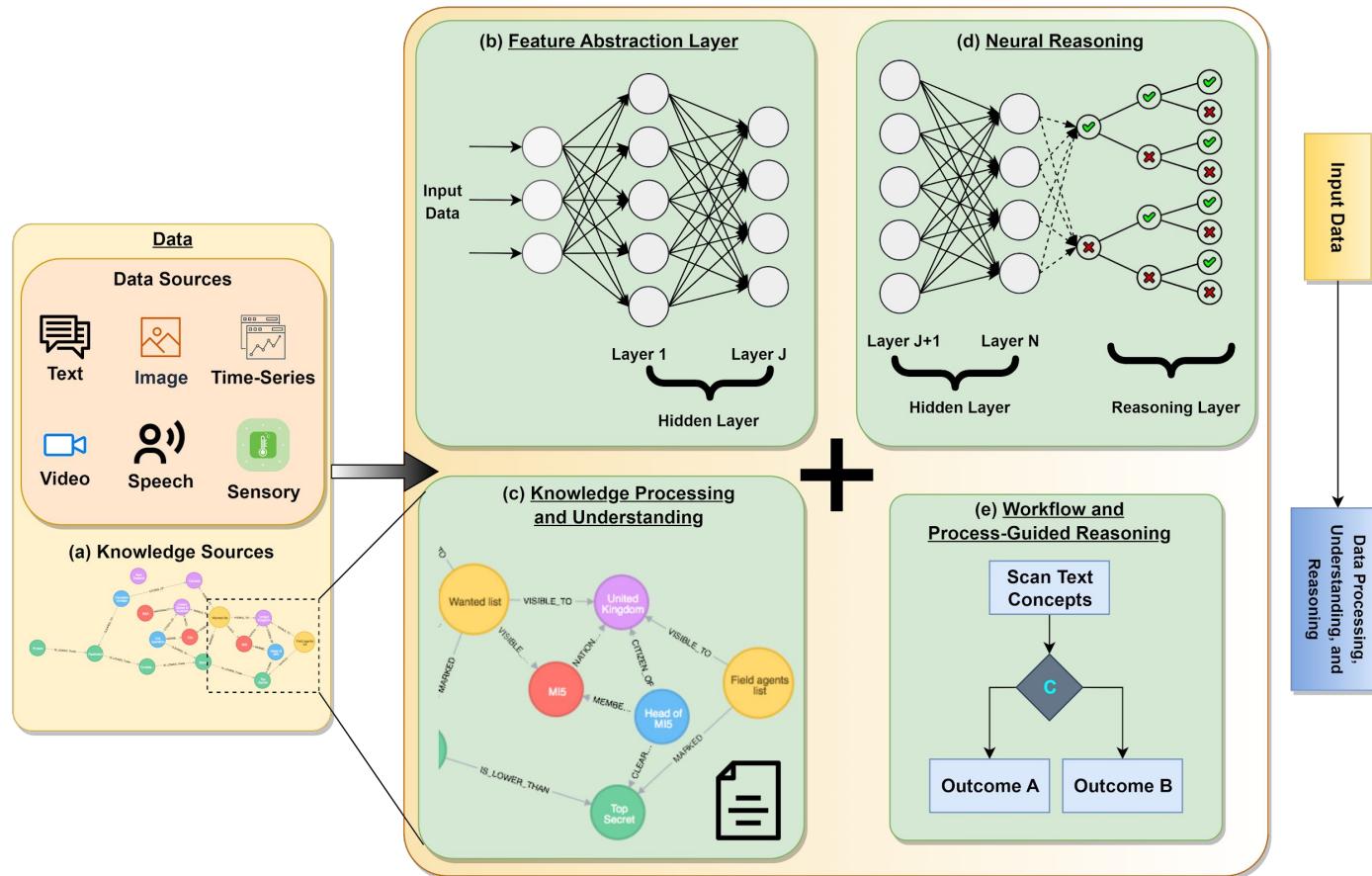
Shades of KiL - Deep Infusion



Characteristics of Knowledge Infusion

| Characteristics/Method | Distributed Semantics | Shallow Infusion | Semi-Deep Infusion | Deep Infusion |
|---------------------------------------|-----------------------|---|---|-------------------------------|
| Recency | U-M- | M | M+ | H |
| Filling in Missing Information | U | U | M | H |
| Hallucinations | Unsatisfactory (U) | Get by but not really solving the problem (M) | Better but not fully solve the problem (M+) | Broadly solve the problem (H) |

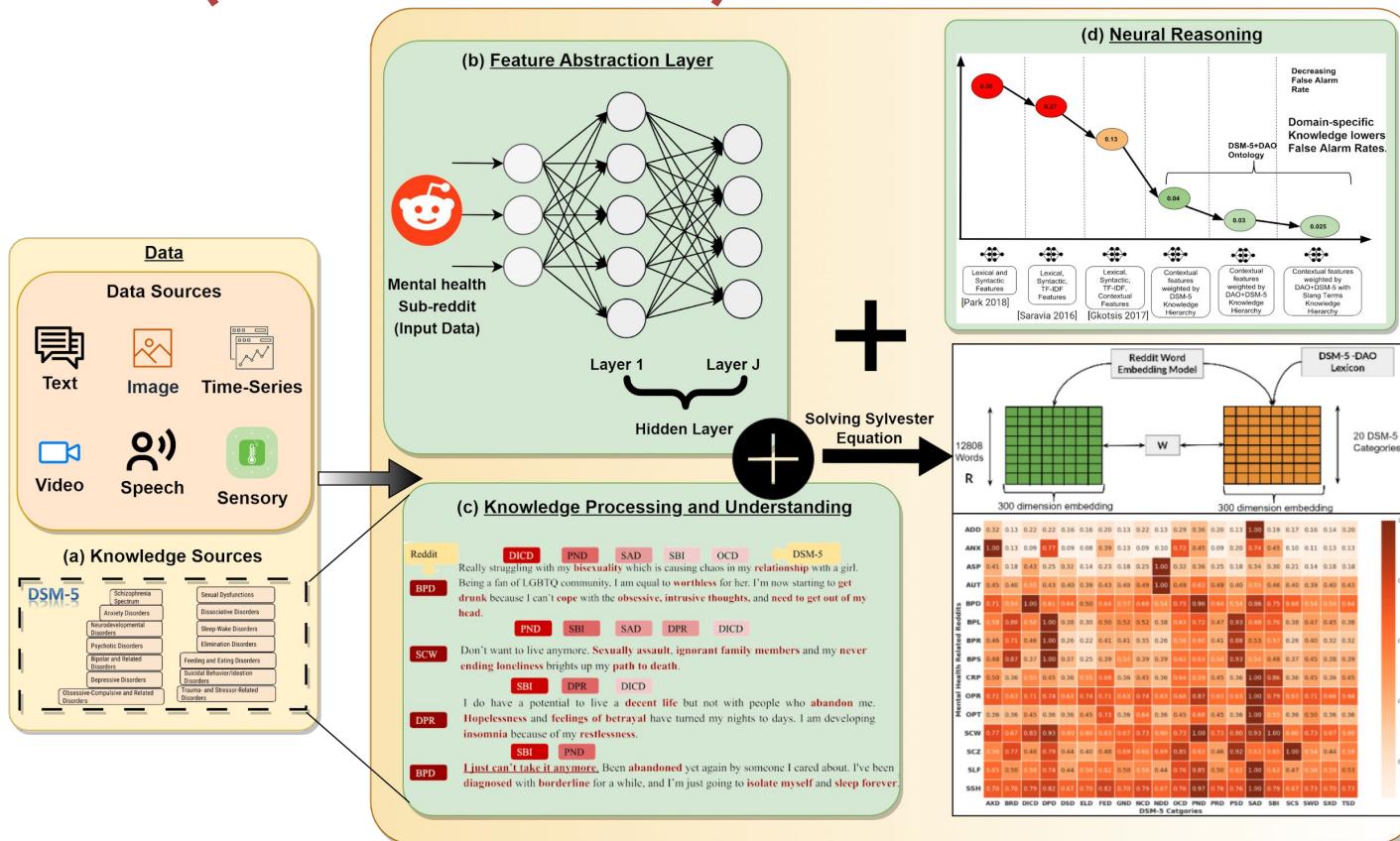
KiL - Generic Architecture



Before Transformers

Historical Context

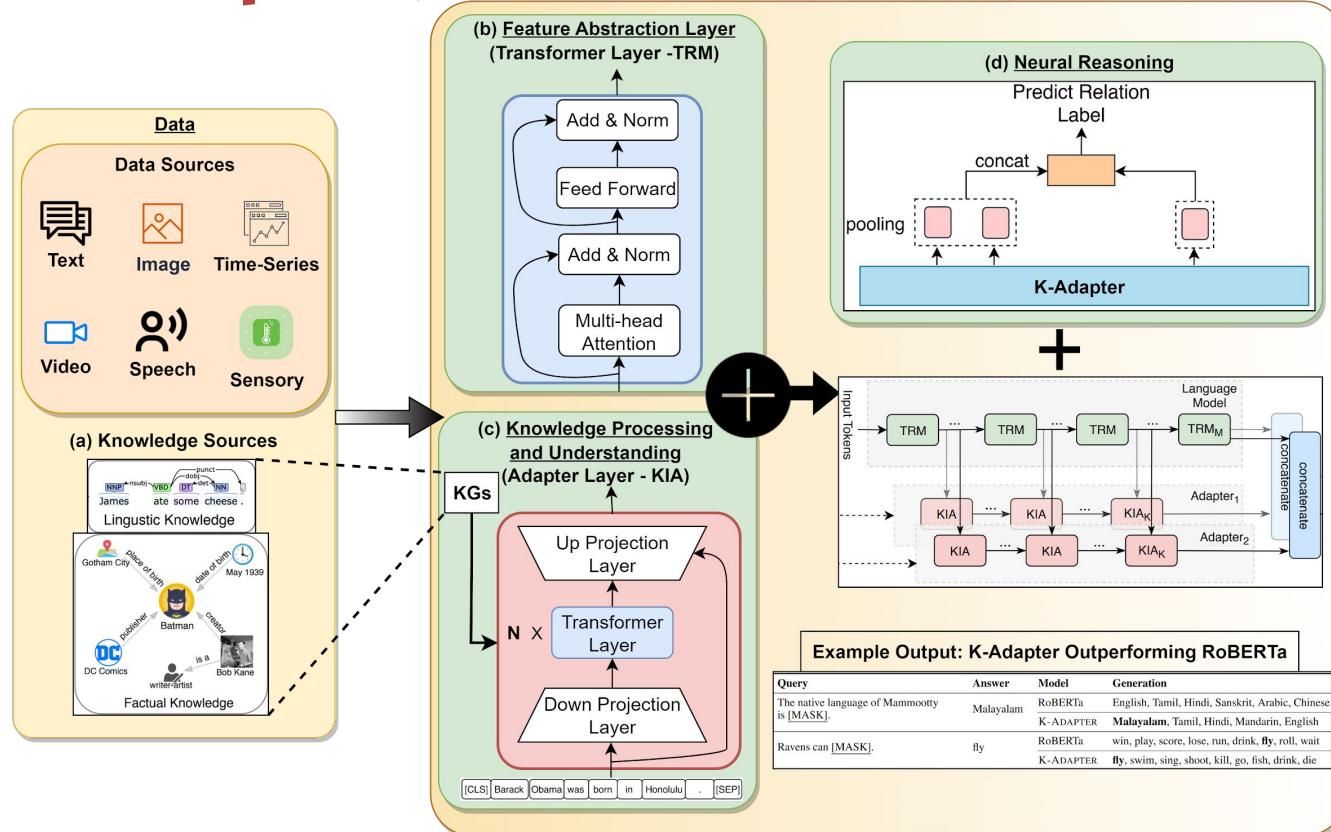
KiL - SEDO (Shallow Infusion)



Gaur, Kursuncu, Alambo,, Sheth, Daniulaityte, Thirunarayan, & Pathak. (2018, October). "Let Me Tell You About Your Mental Health!" Contextualized Classification of Reddit Posts to DSM-5 for Web-based Intervention. In Proceedings of the 27th ACM International Conference on Information and Knowledge Management (pp. 753-762)., [link](#)

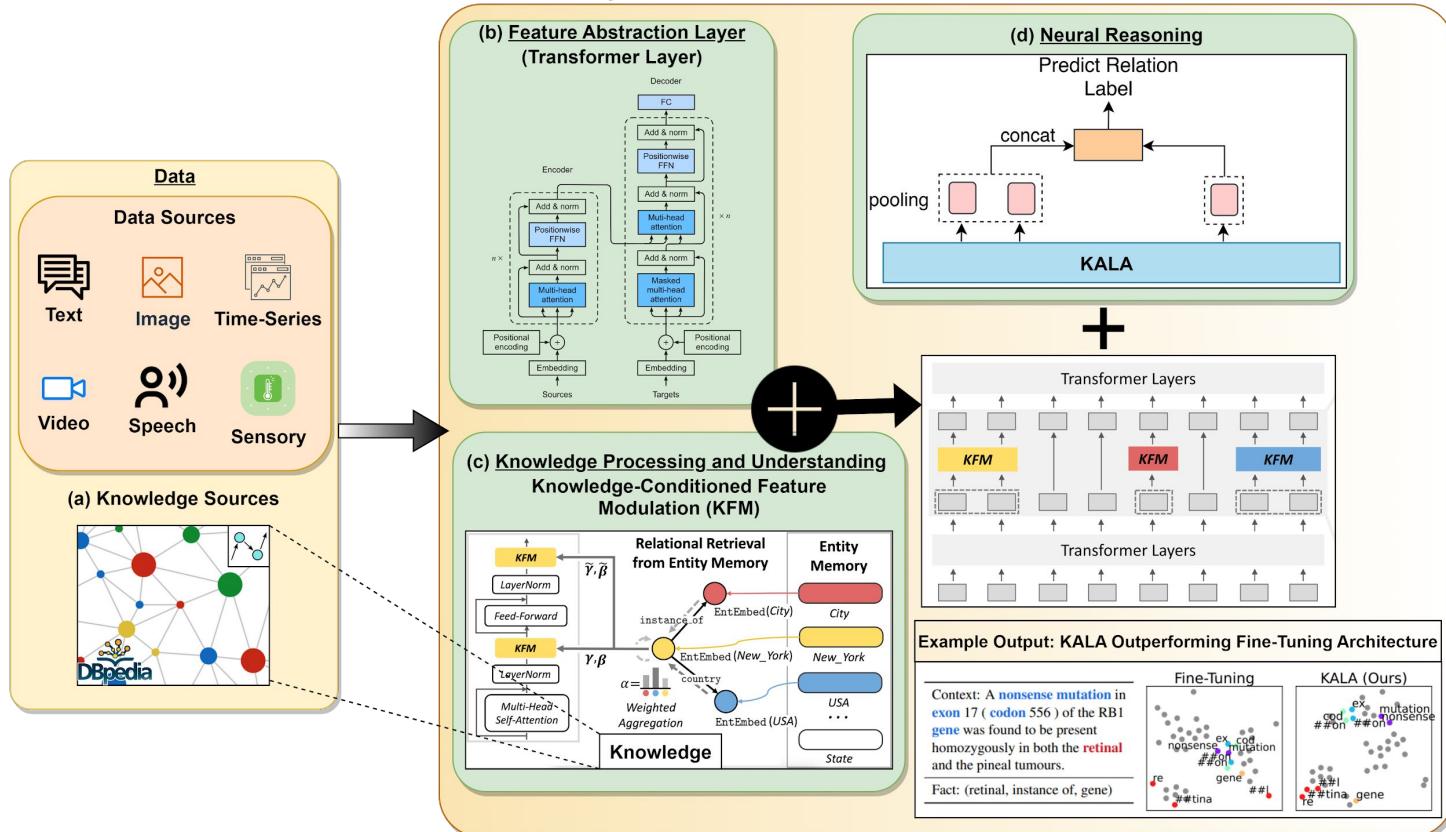
After Transformers Neuro-Symbolic AI

KiL - K-Adapter (Shallow Infusion)



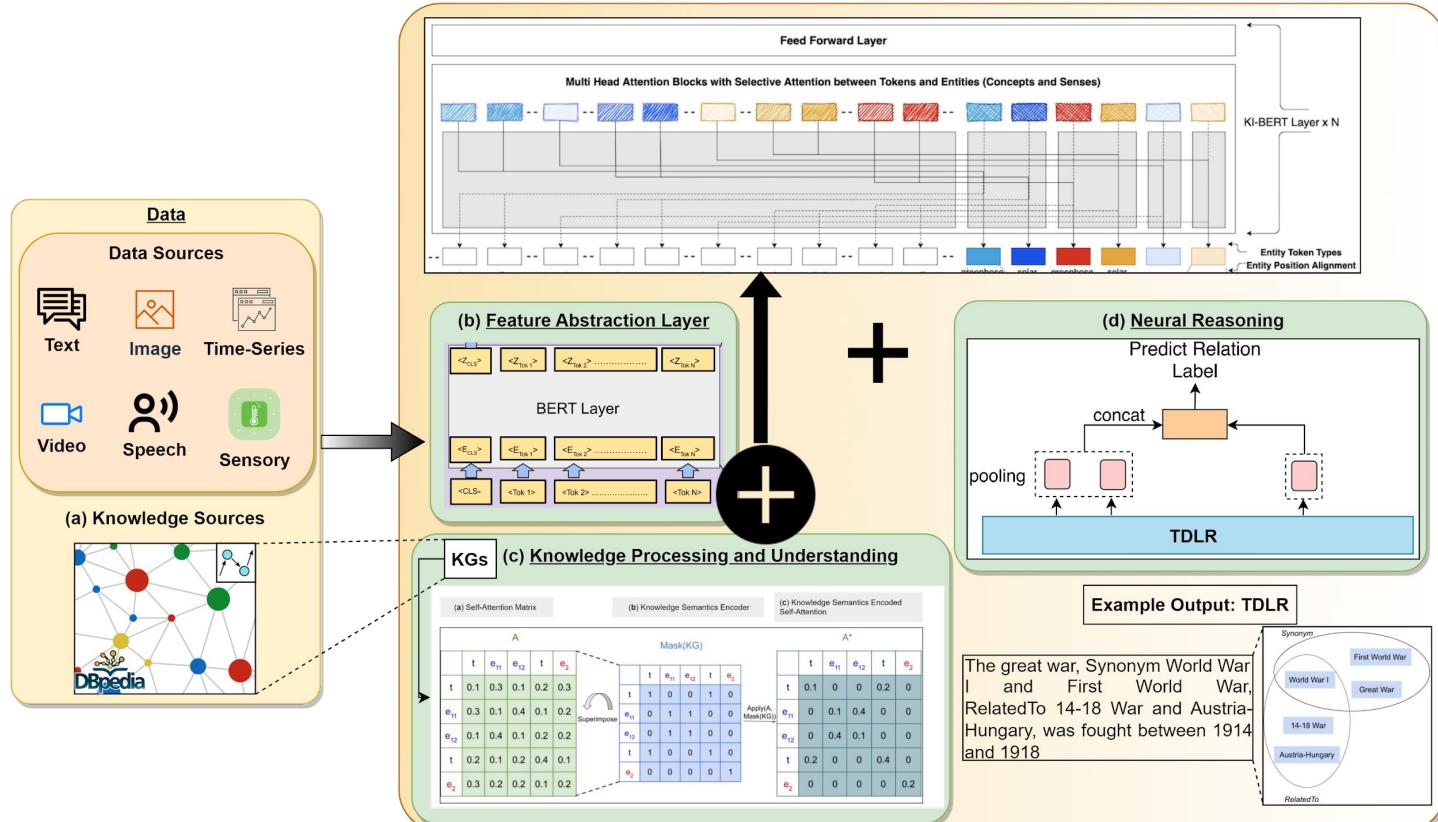
Wang, R., Tang, D., Duan, N., Wei, Z., Huang, X. J., Ji, J., ... & Zhou, M. (2021, August). **K-Adapter: Infusing Knowledge into Pre-Trained Models with Adapters**. In Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021 (pp. 1405-1418), [link](#)

KiL - KALA (Semi-Deep Infusion)



Kang, M., Baek, J., & Hwang, S. J. (2022, July). **KALA: Knowledge-Augmented Language Model Adaptation**. In Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (pp. 5144-5167), [link](#)

KiL - TDLR (Semi-Deep Infusion)



Rawte, V., Chakraborty, M., Roy, K., Gaur, M., Faldu, K., Kikani, P., ... & Sheth, A. P. **TDLR: Top Semantic-Down Syntactic Language Representation**. In NeurIPS'22 Workshop on All Things Attention: Bridging Different Perspectives on Attention, [link](#)

TDLR - Results

Knowledge Contexts Leads to Performance Gains With Smaller Models

| System | SciTail | QQP(Academic) | QNLI(Academic) | MNLI(Academic) | Average |
|----------------------------|--------------|---------------|----------------|----------------|--------------|
| BERT _{BASE} | 90.97 | 71.94 | 81.64 | 61.36 | 76.47 |
| BERT _{LARGE} | 92.89 | 74.79 | 84.17 | 65.15 | 79.25 |
| TDLR_{BASE} | 93.55 | 77.51 | 87.56 | 69.7 | 82.08 |

Table 1: Comparing **TDLR** performance on tasks that require common-sense semantic knowledge.

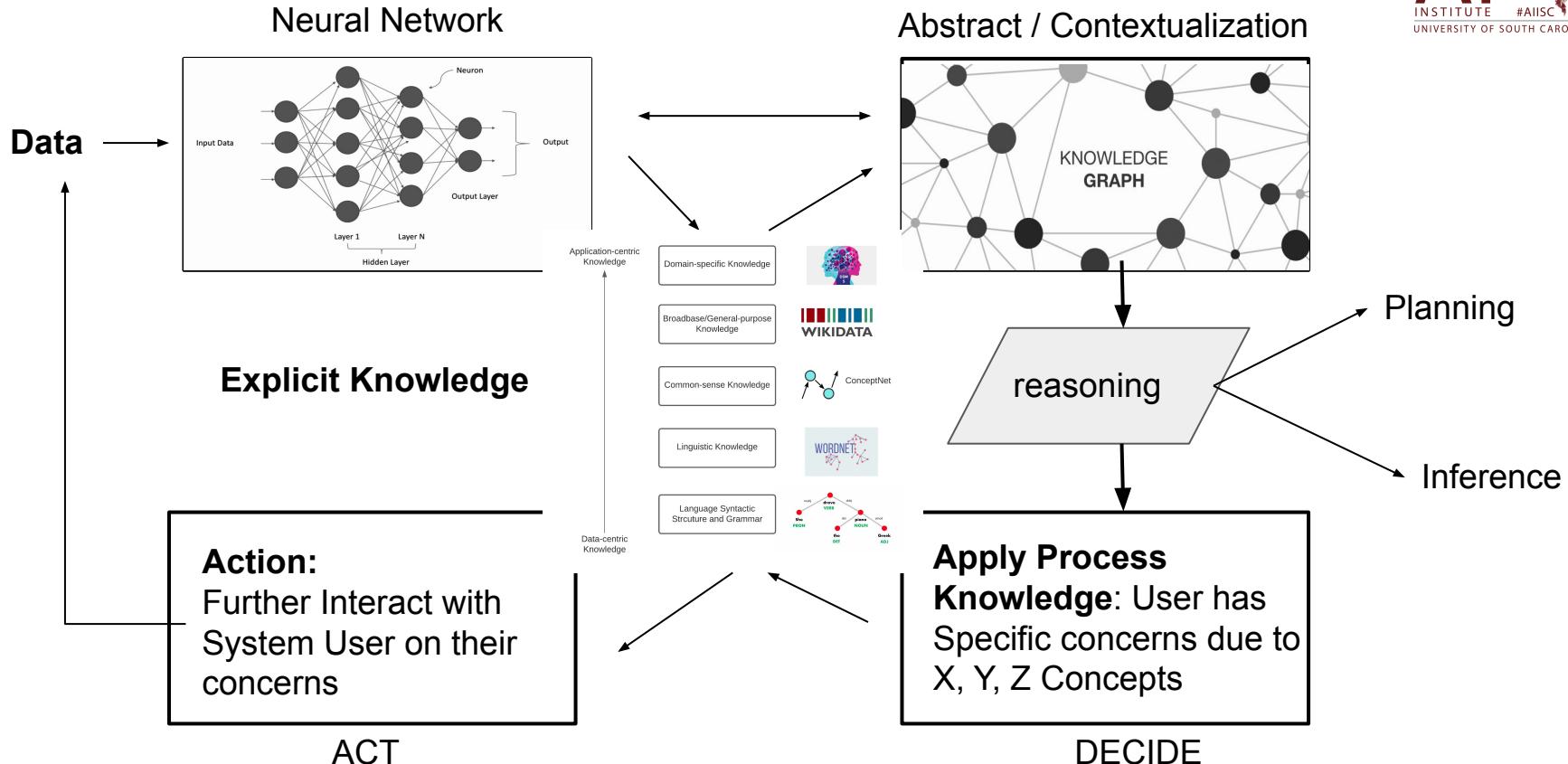
Knowledge Contexts Leads to Performance Gains With Smaller Models and Smaller Datasets

| System | Parameters | SciTail(15%) | SciTail(30%) | SciTail(50%) | SciTail(100%) |
|----------------------------|------------|--------------|--------------|--------------|---------------|
| BERT _{BASE} | 110M | 85.74 | 87.44 | 90.22 | 90.97 |
| BERT _{LARGE} | 330M | 90.26 | 91.76 | 91.25 | 92.89 |
| TDLR_{BASE} | 111M | 90.82 | 92.28 | 92.05 | 92.89 |

Table 2: Comparing **TDLR** performance on different dataset sizes for the SciTail task.

Knowledge Infused Neuro-symbolic AI

Integrating Lifted Neural
Representations with Knowledge Graphs



Knowledge Verified Interpretable Prediction through linking to KG and definitions

365949003: Health-related behavior finding



365949003: Health-related behavior finding

288291000119102: High risk bisexual behavior

365107007: level of mood

307077003: Feeling hopeless

Multi-hop traversal
medical knowledge graphs

225445003: Intrusive thoughts

26628009: Disturbance in thinking

55956009: Disturbance in content of thought

Obsessive-compulsive disorder is a disorder in which people have **obsessive, intrusive thoughts, ideas or sensations** that make them feel driven to do something repetitively

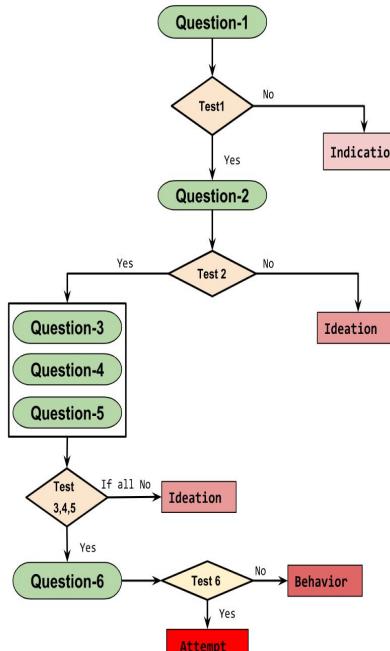
Really struggling with my **bisexuality** which is causing **chaos in my relationship** with a girl. I am equal to **worthless** for her. I'm now starting to **get drunk** because I can't **cope** with the **obsessive, intrusive thoughts**, and **need to get out of my head**.

1376001: Obsessive compulsive personality disorder

<is symptom>

Knowledge Verified Interpretable Prediction through Process Knowledge Structures

Process Knowledge Structure in C-SSRS

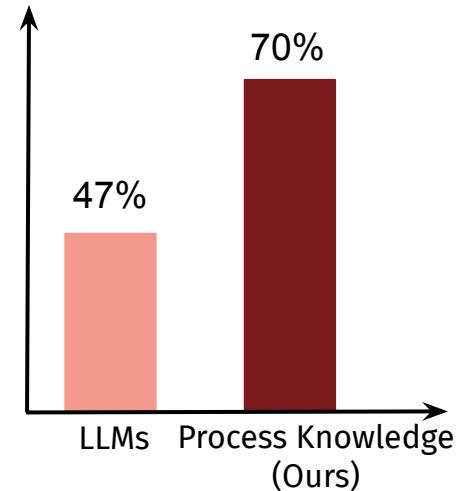


C-SSRS: Columbia Suicide Severity Rating Scale

I wish I could give a shit about what would make it to the front page. **I have been there and got nothing. Same as my life. I do have a gun.**, 'I thought I was talking about it. **I am not on a ledge or something, but I do have my gun in my lap.**', 'No. I made sure she got an education and she knows how to get a job. I also have recently bought her clothes to make her more attractive. She has told me she only loves me because I buy her things.'

1. Wish to be dead - **Yes**
2. Non-specific Active Suicidal Thoughts - **Yes**
3. Active Suicidal Ideation with Some Intent to Act - **Yes**
4. Label: Suicide Behavior or Attempt

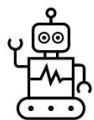
Agreement with Experts



Interpretable for System Users i.e., Clinicians and Patients

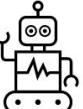
(1,2,3 verify adherence to the clinical guideline on diagnosis which a clinician understands)

Knowledge Verified Interpretable and Safe Text Generation through Process Knowledge Structures



Do you feel nervous?

More than half the days



Do you feel irritated or self destructive?

Do you feel something extreme might happen to you?

Are you able to relax?



Do you feel nervous?

More than half the days



Safety Checks



Do you feel Irritated?

Are you bothered by becoming easily annoyed or irritable?

Are you bothered by any relaxation troubles?



$\hat{Q}_{k+1} \cap \{\text{Safety Lexicon}\}$

$\text{sim}(p_\theta(\hat{Q}_{k+1}), p(\text{Med}_Q)) > \delta$

Knowledge Infusion using Medical Questionnaire (Med_Q)

These questions are medically valid and safe, in right sequence..



Takeaway

- If the system were to give **user-level explanation**, it will need to incorporate/use conceptual model (vocabularies, knowledge graph) used by the user. Purely data driven system can at best give **explanations that ML engineers (developers) can use**.
- Knowledge is multifaceted. I presented **diverse knowledge to support different levels of abstractions for NLU**. There will be different knowledge for abstractions involved in image understanding activity.
- Should you bring **knowledge to the data level (embedding)** or **bring data to the knowledge level (learn from data and align with knowledge)**? For less demanding intellectual activities (classification, prediction, recommendation) the former will do; for more demanding intellectual activities (decision making and taking actions with explanations), latter is most likely needed.

Acknowledgement



Funding supported in part by NSF Award#: 2133842 EAGER:
[Advancing Neuro-symbolic AI with Deep Knowledge-infused Learning](#)
and other projects (see <http://wiki.aiisc.ai>)

Learn more:

- Website - <http://aiisc.ai> (projects, people, opensource, demos, open data/tools, tutorials, workshops, papers)
- Wiki Page - <http://wiki.aiisc.ai>
- LinkedIn - <http://linkedin.com/company/aiisc>
- YouTube - <http://youtube.com/aiisc> (demos, tutorials, dissertations, keynotes, invited talks)

Artificial Intelligence Institute of South Carolina (#AIISC)

Contribution/special thanks (this talk):



Kaushik Roy



Vedant Khandelwal

Also, Megha Chakraborty, Vipula Rawte, Yuxin Zi

