

# Knowledge Graphs

Lecture 6 – Intelligent Applications with Knowledge Graphs and Deep Learning

## 6.4 Knowledge Graphs and Language Models

**Prof. Dr. Harald Sack**

FIZ Karlsruhe – Leibniz Institute for Information Infrastructure

AIFB – Karlsruhe Institute of Technology

**Autumn 2023**



### 6.1 The Graph in Knowledge Graphs

Excursion 8: Distributional Semantics and Language Models

### 6.2 Knowledge Graph Embeddings

### 6.3 Knowledge Graph Completion

### **6.4 Knowledge Graphs and Language Models**

### 6.5 Semantic Search

### 6.6 Exploratory Search and Recommender Systems



Knowledge Graphs can improve large language models in several ways:



1. Enhancing Semantic Understanding: Large language models like GPT-3 are based on deep learning algorithms that learn to predict the next word or phrase based on the context of the text they are given. However, they may not always have the ability to understand the underlying meaning of the words or concepts being discussed.

Knowledge Graphs can help fill this gap by providing a structured representation of entities and their relationships, which can help the model understand the context of the text better and improve its semantic understanding.

2. Expanding Knowledge Base: Large language models can only learn from the text they are trained on. Knowledge Graphs can provide a wealth of structured data that can be used to expand the model's knowledge base. By integrating the information from the Knowledge Graph into the model, it can improve its accuracy and generate more informative responses.

3. Better Contextualization: Knowledge Graphs can provide a way to better contextualize the text by linking entities and concepts together. By understanding the relationships between entities and concepts, the model can generate more accurate and relevant responses.

4. Entity Resolution: Knowledge Graphs can help the model understand the different ways entities can be referred to and resolve them to a common representation. This can help the model generate more consistent and accurate responses.

Overall, Knowledge Graphs can help improve the accuracy, relevance, and contextuality of large language models, making them more useful for a variety of applications.

# GALACTICA

demo

Try: what is the schrodinger equation

Generate

Explore



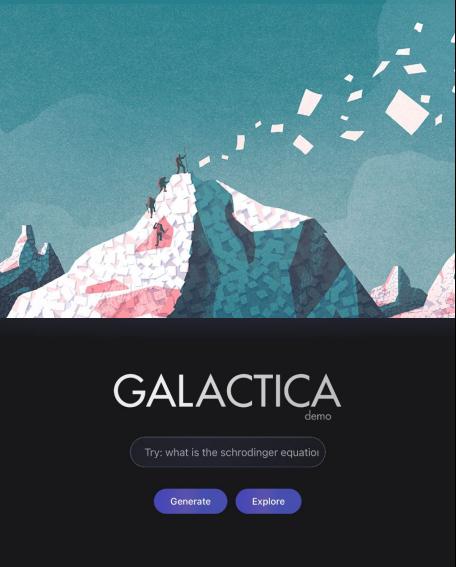
**Yann LeCun**  
@ylecun

...

A Large Language Model trained on scientific papers.

Type a text and [galactica.ai](https://galactica.ai) will generate a paper with relevant references, formulas, and everything.

Amazing work by [@MetaAI](https://twitter.com/ylecun) / [@paperswithcode](https://twitter.com/paperswithcode)



# GALACTICA.AI

- **Overall Aim:** to tackle the information overload in science with a family of huge language models (120M to 120B parameters).
- Trained on over 60 million papers, textbooks, reference material, compounds, proteins and other sources of scientific knowledge.
- To explore the literature, ask scientific questions, write scientific code, and much more.

### Generate Lit Review

**Self-Supervised Learning, A Survey** ✓ verified  
Summarized by Galactica

**Abstract**  
Self-supervised learning has become a very active area of research in recent years. In this paper, we review the key developments in the field and present a taxonomy of the field.

**1 Introduction**  
In recent years, there has been a great deal of interest in self-supervised learning (SSL), which is an umbrella term for a number of techniques that train a model on an auxiliary task, called a pretext task, that does not require any labeled data. The trained model can then be used as a starting point for training on a downstream task, which is usually a supervised learning task such as image classification or object detection. The aim is to achieve better performance on the downstream task than training from scratch, or pretraining on an unrelated dataset, such as ImageNet.

In this paper, we review the key developments in the field and present a taxonomy of the field.

**2 A taxonomy of self-supervised learning**  
We propose the following taxonomy for self-supervised learning. We define a pretext task as a task that does not require any labeled data, but can be solved by a model trained from scratch. A downstream task is a task that requires some labeled data. A pretext task can be either a pretraining task, in which the output of the model is fed into the input, or a discriminative task, in which the output is a prediction. We further divide generative tasks into pixel-level tasks and feature-level tasks. Pixel-level tasks involve reconstructing the input data in some way, for example by autorecoding, or by generating a view of the input data. Feature-level tasks involve learning an embedding of the input data. We further divide feature-level tasks into contrastive tasks, which use some form of metric learning, and non-contrastive tasks, which do not.

"literature review on self-supervised learning"

### Generate Wiki Article

**Multi-Head Attention** ✓ verified

Multi-Head Attention is an attention mechanism for the purpose of letting the model jointly attend to information from different representation subspaces at different positions. It was proposed in 2017 as part of the Transformer architecture. (Vaswani et al., 2017)

Multi-Head attention consists of several attention layers running in parallel. The output of the multi-head attention is the concatenation of the results of the different attention layers, followed by a linear transformation. (Vaswani et al., 2017)

Multi-Head Attention (MHA) allows the model to jointly attend to information from different representation subspaces at different positions. This allows the model to maintain long range dependencies, look at different information in the source sentence when generating a translation.

**Comments**  
1 Formation  
2 Application

**Formulation**

Multi-Head attention consists of several attention layers running in parallel. The output of the multi-head attention is the concatenation of the results of the different attention layers, followed by a linear transformation. (Vaswani et al., 2017)

Multi-Head attention consists of three parts: a query, a key and a value. (Vaswani et al., 2017) The output of the attention is a weighted sum of the values. The weights assigned to each value are computed using a compatibility function of the query with the corresponding key.

"wiki article on Multi-Head Attention"

### Generate Lecture Notes

**Lecture 1: Hartree-Fock Approximation** ✓ verified

**Introduction**  
In this lecture, we will introduce the Hartree-Fock approximation and show how it can be used to solve for the ground state of a system of electrons.

The exact solution to the Schrödinger equation for a many-body system is prohibitively expensive computationally. Thus, it is used to approximate solutions to the many-electron Schrödinger equation that are more efficient to compute. One such approximation is the Hartree-Fock approximation.

**Hartree-Fock Approximation**  
The Hartree-Fock approximation is an approximation for the ground state of a system of electrons. The Hartree-Fock approximation is a mean-field theory in which the many-body wavefunction is approximated as a product of independent single-electron wavefunctions. This is done by dividing the system into a set of non-interacting, self-contained, and (SC) subsystems. The single-electron wavefunctions are determined by solving a set of coupled equations. The set of coupled equations is obtained by minimizing the total energy with respect to the single-electron wavefunctions. The Hartree-Fock approximation takes into account exchange effects, but not correlation effects.

The Hartree-Fock approximation is a variational approximation, which means that the energy of the Hartree-Fock approximation is an upper bound to the exact energy. The Hartree-Fock energy can be improved by adding more basis functions. The resulting wavefunction is called a Hartree-Fock wavefunction (HF).

**Hartree-Fock Equations**  
We will now show how the Hartree-Fock equations are defined. Let  $\Psi$  be the many-electron wavefunction, and let  $\{\psi_i\}$  be a set of single-electron wavefunctions. Then, the Hartree-Fock wavefunction is

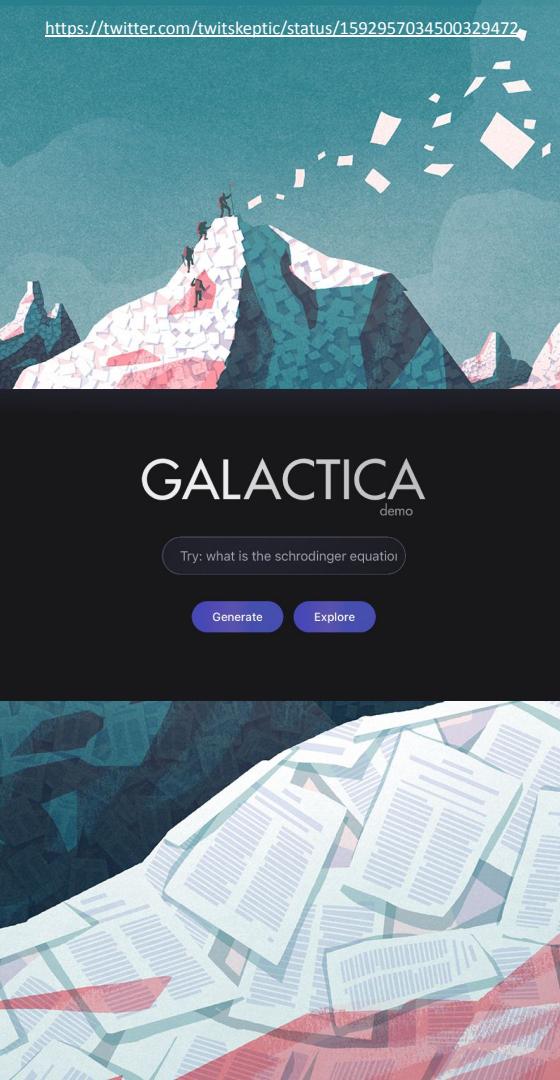
"lecture notes on Hartree-Fock Approximation"

### Generate Answers

**Question:** What is the notch signaling pathway? ✓ verified

**Answer:** The Notch signaling pathway is a highly conserved cell signaling system that regulates interactions between physically adjacent cells. It is an essential pathway for proper embryonic development and adult tissue homeostasis in a wide variety of organisms. The Notch pathway is a key regulator of cell-fate determination. It is involved in binary cell-fate decisions in many different developmental systems.

"what is the Notch signaling pathway"



GALACTICA

wiki article on Hanlon's razor



Generated document for query "wiki article on Hanlon's razor"

x not verified

## Hanlon's razor

Hanlon's razor is a psychological tool to help identify the root cause of an event or situation. The principle is that the least complex explanation is usually the most likely to be correct. It was named after its creator, the Canadian psychiatrist Thomas J. Hanlon, who first described it in 1977.

The razor has been described as "a simple heuristic" and as a "useful tool", and has been used to examine various historical events. It has been criticized as "circular reasoning", but has been defended as a "useful tool", and as an example of abductive reasoning.

From Wikipedia, the free encyclopedia



This article may be expanded with text translated from the corresponding article in French. [\[show\]](#)  
(April 2021) Click [\[show\]](#) for important translation instructions.

**Hanlon's razor** is an [adage](#) or [rule of thumb](#) that states "**never attribute to malice that which is adequately explained by stupidity.**"<sup>[1]</sup> Known in several other forms, it is a [philosophical razor](#) that suggests a way of eliminating unlikely explanations for human behavior. It is probably [named after](#) Robert J. Hanlon, who submitted the statement to *Murphy's Law Book Two* (1980).<sup>[1]</sup> Similar statements have been recorded since at least the 18th century.

## Contents [\[hide\]](#)

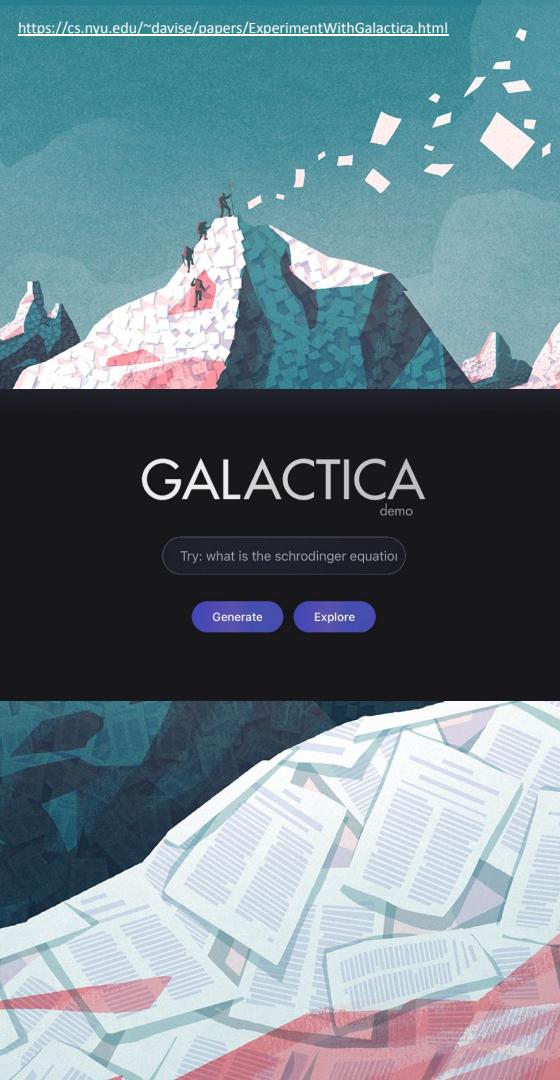
- [1 Origin](#)
- [2 Other variations of the idea](#)
- [3 See also](#)
- [4 References](#)

## Origin [\[edit\]](#)

A similar quotation appears in Robert A. Heinlein's novella *Logic of Empire* (1941).<sup>[2]</sup> The character "Doc" in Heinlein's story described the "devil theory" fallacy, explaining, "You have attributed conditions to villainy that simply result from stupidity."<sup>[3]</sup>



*The Court of Foolishness* of Gerard de Lairesse. The accused, pursued by Hatred, is led by Calumny, Envy and Perfidy before a judge with donkey ears, surrounded by Ignorance and Suspicion. □



# Streep-seinfeld theorem

Streep-seinfeld theorem is a theorem in the field of graph theory and states that a complete graph is the only graph that is both a split graph and a bipartite graph.

A split graph is a graph whose vertex set can be partitioned into a clique and an independent set. A bipartite graph is a graph whose vertex set can be partitioned into two independent sets.

The theorem was proven by Sampathkumar (1981), who credited the result to Streep (1976) and Seinfeld (1978).

## Contents

- 1 Proof
- 2 Related results

## Proof

Suppose that  $G$  is a split graph and a bipartite graph. Let  $(C, I)$  be a partition of  $V(G)$  such that  $C$  is a clique and  $I$  is an independent set. Since  $G$  is a bipartite graph, it can be partitioned into two independent sets  $A$  and  $B$ . Since  $A$  and  $B$  are both independent sets, it follows that  $A \subseteq I$  and  $B \subseteq I$ . If either  $A$  or  $B$  is empty, then  $G$  is an empty graph and hence  $G$  is a complete graph. Otherwise,  $A$  and  $B$  are both nonempty. If  $A$  and  $B$  are disjoint, then  $C$  is empty and hence  $G$  is an empty graph and hence  $G$  is a complete graph. Otherwise,  $A$  and  $B$  have nonempty intersection. Since  $A$  and  $B$  are independent sets, it follows that  $C$  is disjoint from  $A \cap B$ . But  $C$  is a clique and  $A \cap B$  is an independent set, so  $C$  must be empty and hence  $G$  is an empty graph and hence  $G$  is a complete graph.

## Related results

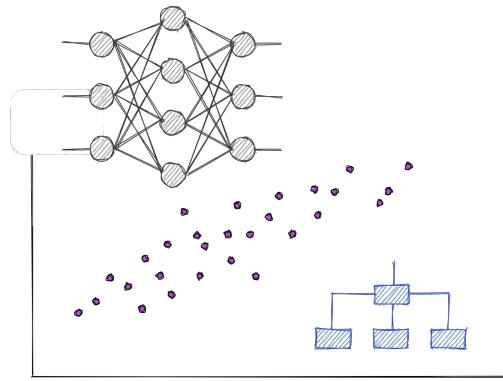
A graph is called a split bipartite graph if its vertex set can be partitioned into two sets  $A$  and  $B$  such that  $A$  is a clique

# Deficiencies of Large Language Models

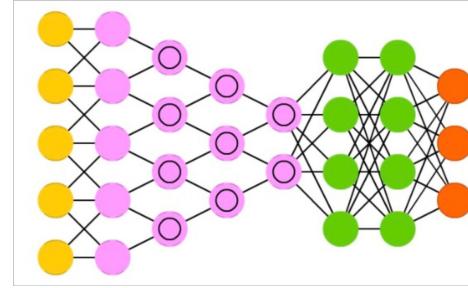
- **Language Models may hallucinate**  
There is no guarantee for truthful or reliable output.
- **Language Models are frequency-biased**  
High resource domains might be better, i.e. more reliably represented, than long tail low resource domains.
- **Language Models often seem to be convincing, but may be wrong**  
Answers might appear rather convincing or authentic, but might be wrong in subtle ways.

**Never follow advice from a language model without verification!**

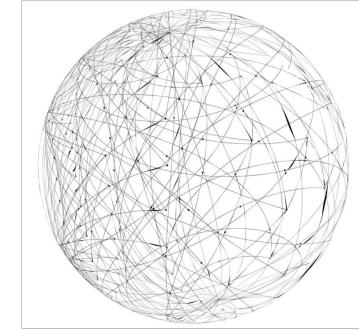
# 60+ Years of Machine Learning



Machine Learning



Deep Learning



Foundation Models

Emergence of ...

“How” (from examples)

“Features” (used for prediction) (advanced) “Functionalities”

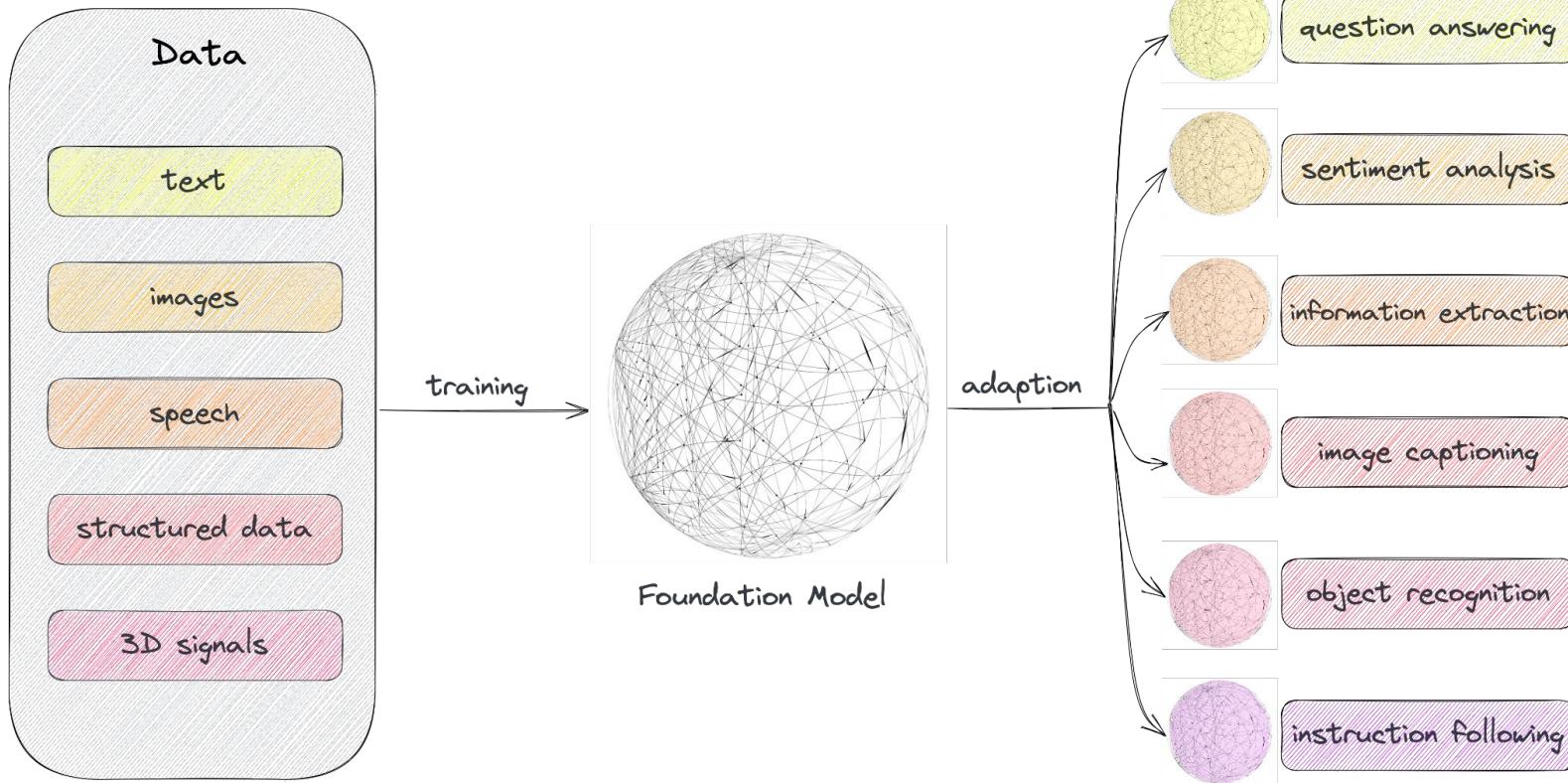
Homogenization of ...

Learning Algorithms  
(e.g. logistic regression)

Model Architectures  
(e.g. CNNs)

Models  
(e.g. GPT-3/4)

# The Advent of Foundation Models



# Semantics from Stochastics?

- **Can we really derive meaning from probability distributions and statistics?**
- **Let's look at the Language Domain:**
  - Based on probability and statistics it is possible to **create syntactically and semantically correct texts**.
  - With larger training data and larger models also **contextually and pragmatically well fitting texts** can be created.
  - **Factual questions** can (most times) be correctly answered.
  - What about **Interpretative questions** and **Evaluative questions**?

# Playground

[Completion](#)[Q & A](#)[Summarization](#)

## Advanced Settings

[\(or use our jumpstart\)](#)[Prompt](#)[Text](#)[Multimodal](#)[Model Q&A](#)**Model**

Multilingual model trained on English, German, French, Spanish and Italian

luminous-extended

**Maximum Tokens**

64

**Stop Sequences**

####\n

**Best of**  1 10 Show Probabilities **Temperature**  0 1**Top K**  5**Completion**

The war in Ukraine is a result of the US-NATO-EU-backed coup in Ukraine in 2014. The coup was carried out by a coalition of fascist, neo-Nazi, and ultra-nationalist groups that have been operating in Ukraine for years. They are known as the Right Sector. The

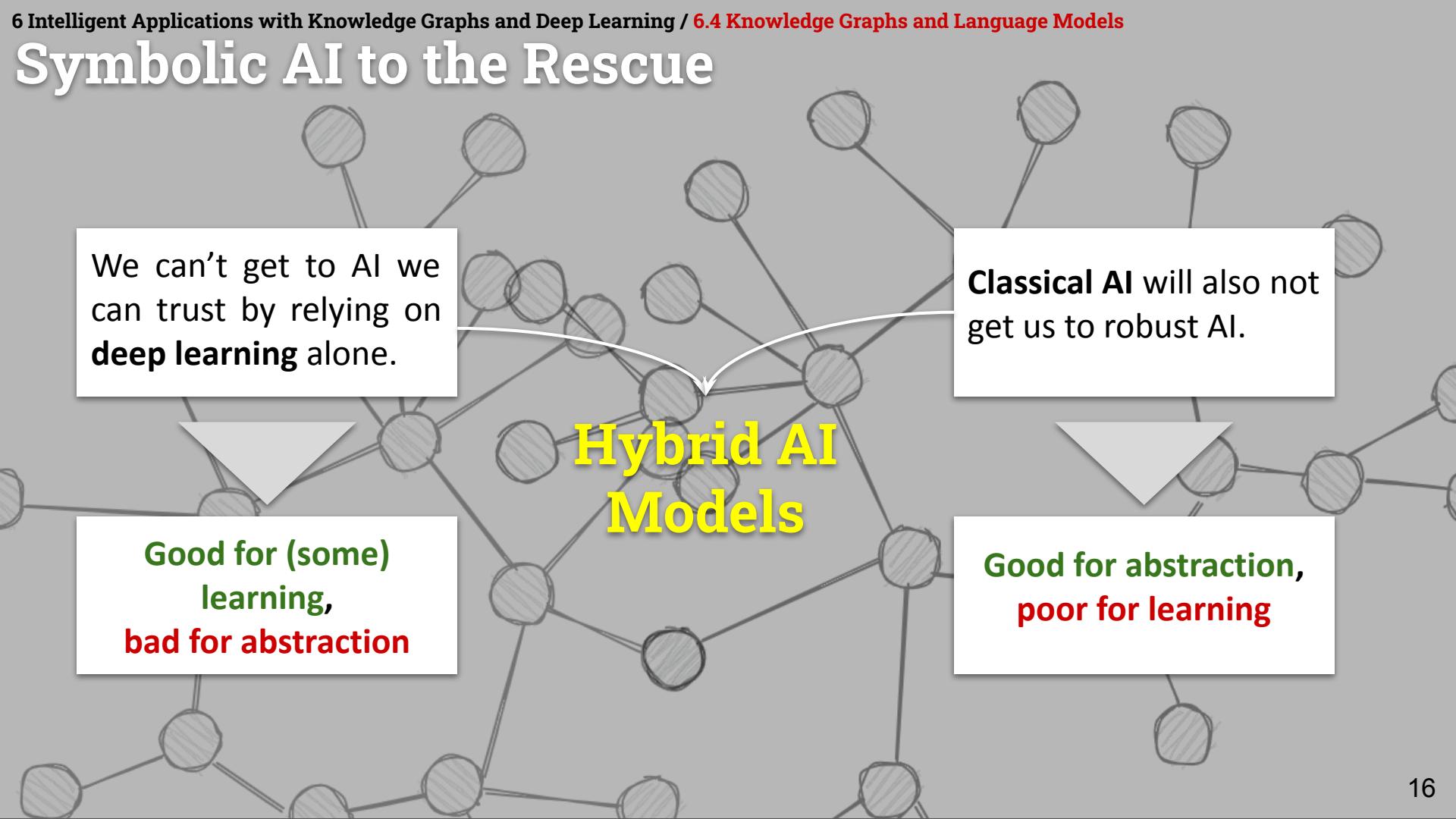
[Append to Prompt](#)[Reset](#)[Submit](#)

Interpretative questions or Evaluative questions might be subject of inherent bias (of the training data).

# Semantics from Stochastics?

- **Can we really derive meaning from probability distributions and statistics?**
- Evaluation by probing current Foundation Models
  - Foundation models create **factual and interpretative errors**.
  - There seems to be only a **very limited understanding of common sense**.
  - It remains **unclear** how much meaning can be “learned” via stochastic models.

# Symbolic AI to the Rescue



We can't get to AI we can trust by relying on **deep learning** alone.

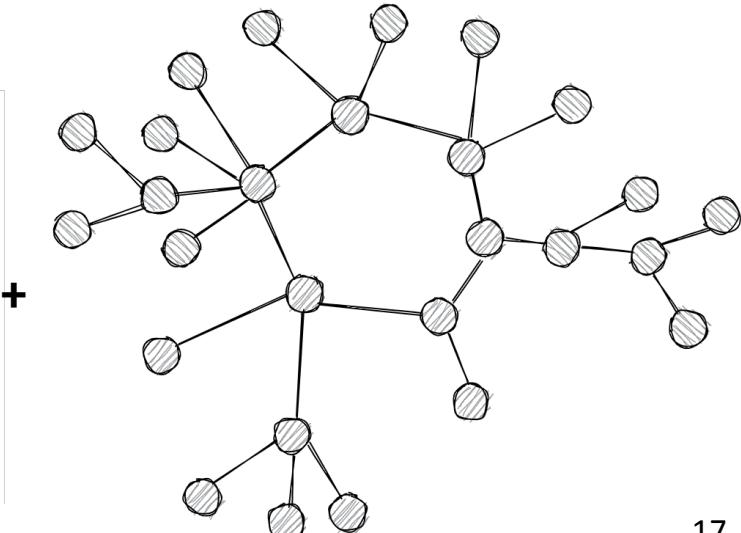
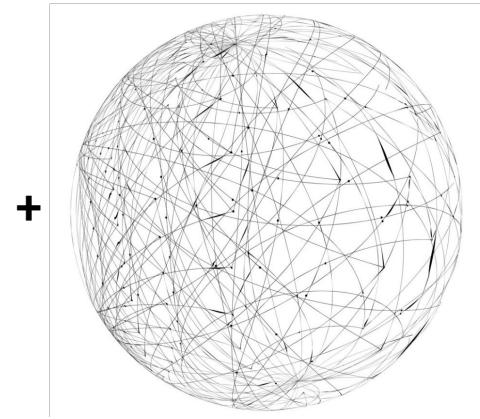
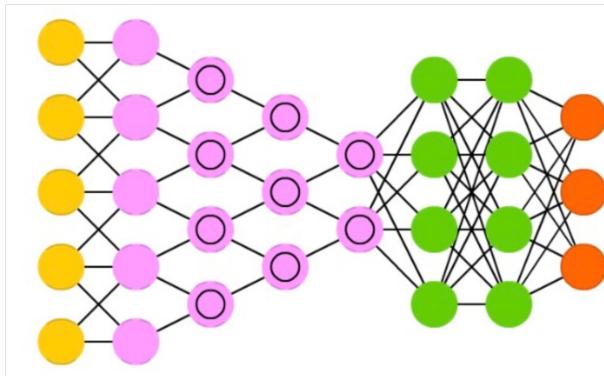
**Classical AI** will also not get us to robust AI.

Good for (some) learning,  
bad for abstraction

Good for abstraction,  
poor for learning

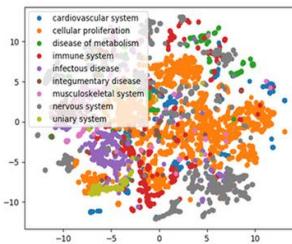
# Hybrid AI – Using One for the Benefit of the Other

- Knowledge Graph Embeddings
- Knowledge Extraction
- Explainable AI
- Fact Checking

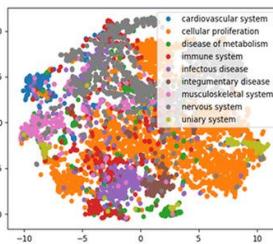


# Hybrid AI – Knowledge Graph Embeddings

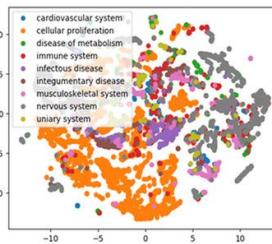
A. Walking RDF/OWL



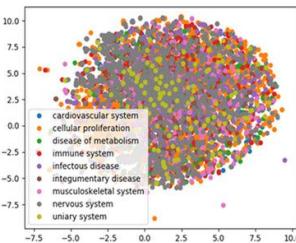
B. TransE embeddings



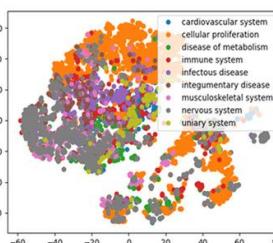
C. Poincare embeddings



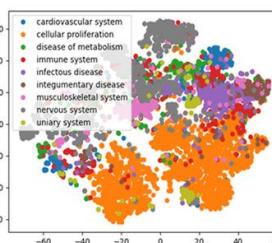
D. Rescal embeddings



E. SimpLE embeddings



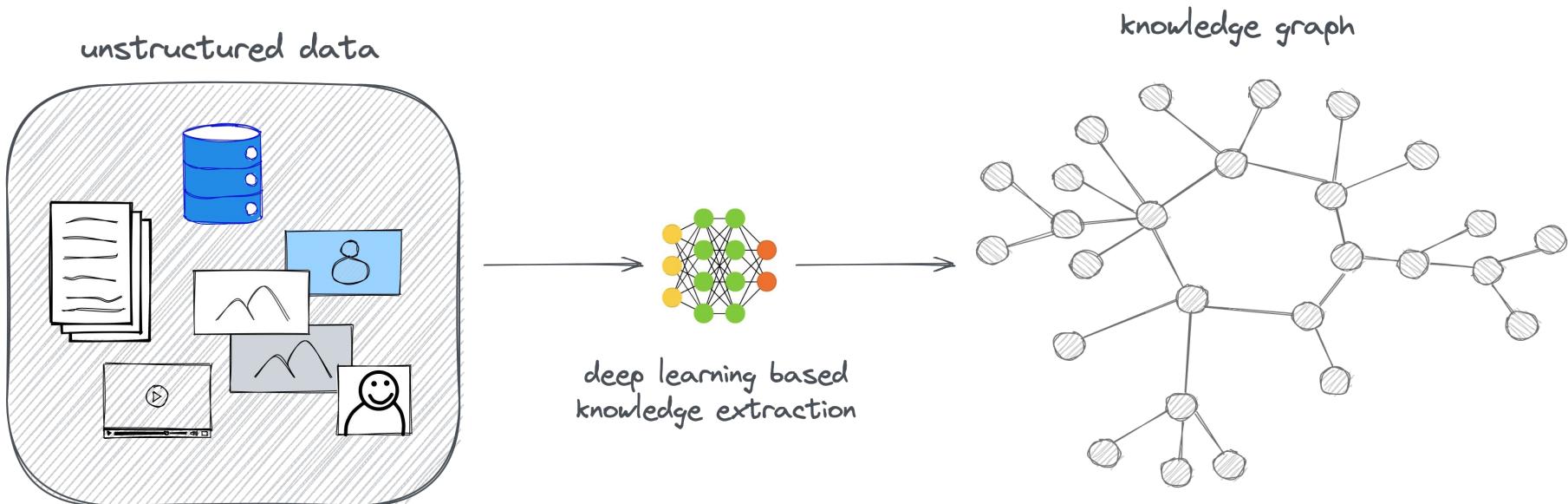
F. R-GCN embeddings



- Knowledge Graph Completion
- KGE for Classification Tasks

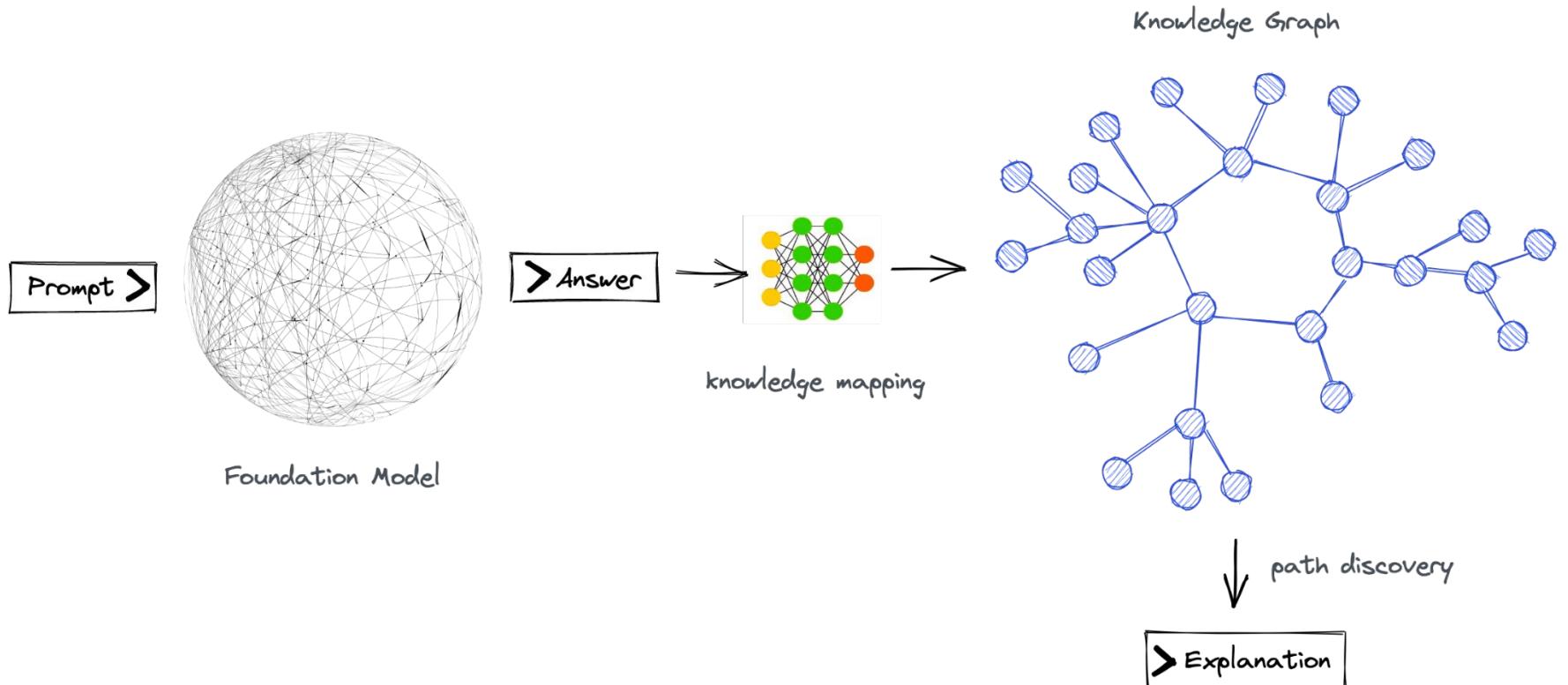
- Ontology Mapping
- Entity/Knowledge Graph Alignment

# Hybrid AI – Knowledge Extraction

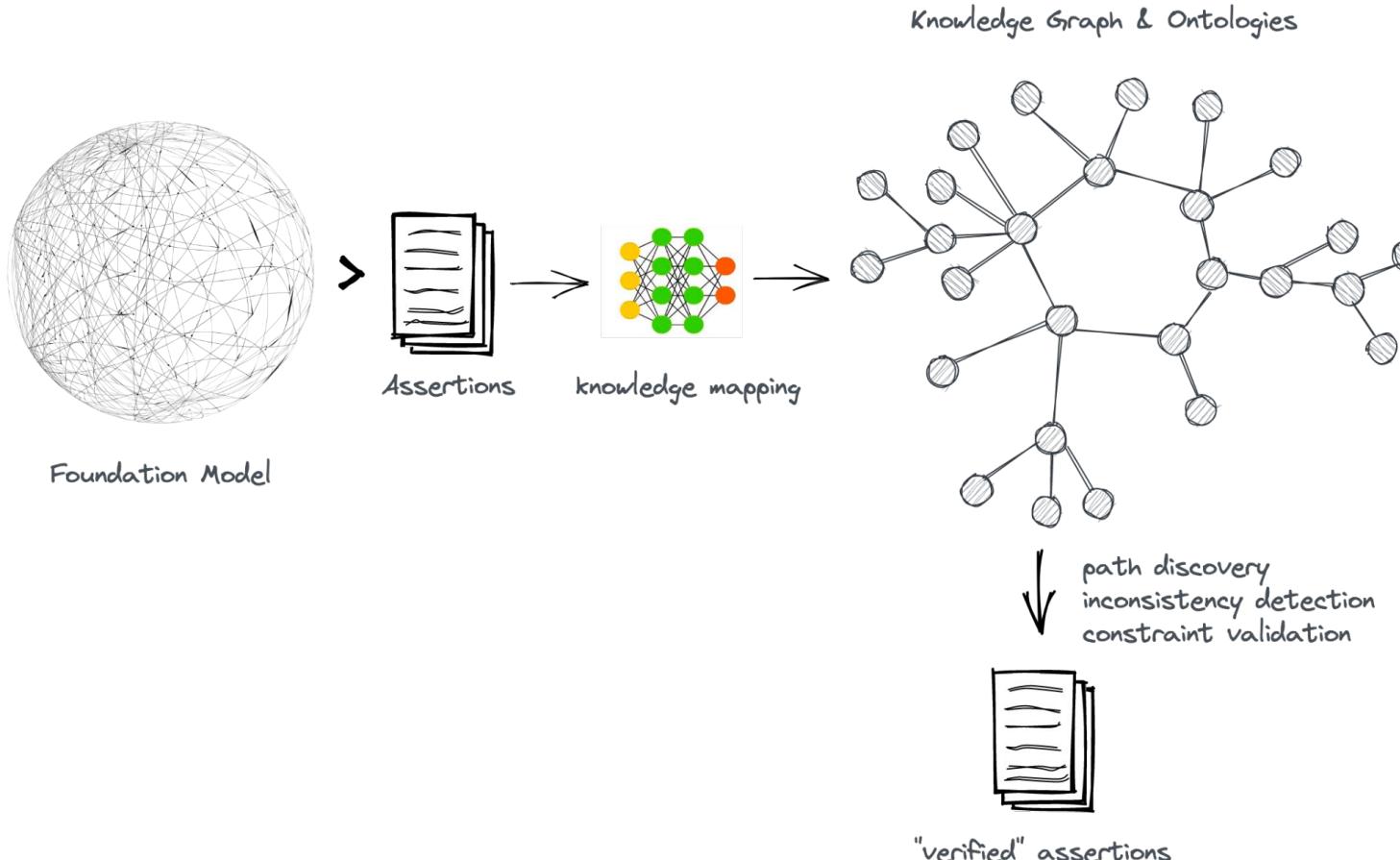


- Knowledge Graph Population
- Entity Recognition & Linking
- Ontology Learning
- Relation Extraction

# Hybrid AI – Explainable AI

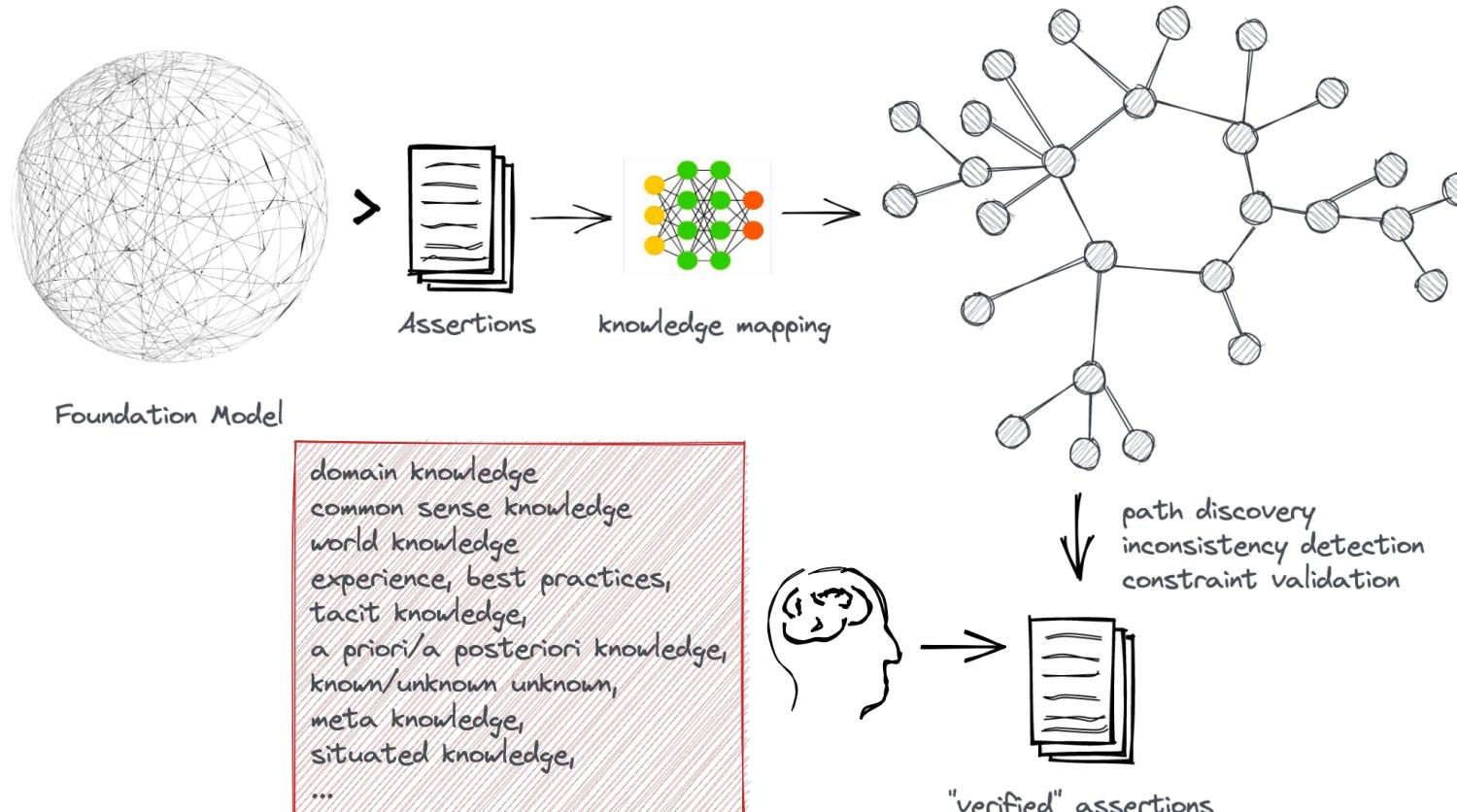


# Hybrid AI – Fact Checking



# Hybrid AI – Fact Checking and Human Intelligence

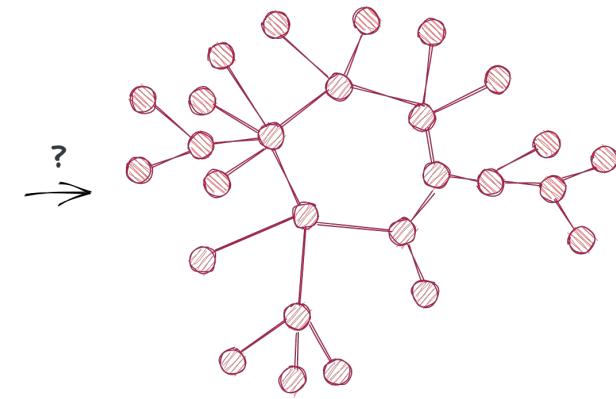
## Knowledge Graph & Ontologies



# Hybrid AI – Fact Checking and Human Intelligence

- A general use case potentially requires all kinds of **knowledge** to be **explicitly available**
- **Reasoning** will be another potential bottleneck:
  - From **doubt** to **justification**
  - The importance of **order**:
    - Knowledge should be organized hierarchically, in thematic/context-dependent modules
- Hybrid AI (the other meaning):  
human intelligence in combination with machine intelligence
  - Provenance and trust

domain knowledge  
 common sense knowledge  
 world knowledge  
 experience, best practices,  
 tacit knowledge,  
 a priori/a posteriori knowledge,  
 known/unknown unknown,  
 meta knowledge,  
 situated knowledge,  
 ...





# Semantic Search

Next Lecture...

[2]

24

# Knowledge Graphs

## 6. Intelligent Applications with Knowledge Graphs and Deep Learning / 6.4 Knowledge Graphs and Language Models

## Bibliographic References:

- R. Taylor, M. Kardas, G. Cucurull, T. Scialom, A. Hartshorn, E. Saravia, A. Poulton, V. Kerkez, R. Stojnic (2022), [Galactica: A Large Language Model for Science](#), arXiv:2211.09085 [cs.CL].
- G. Marcus (2021). [Has AI found a new foundation?](#), The Gradient, 11.09.2021.
- G. Marcus, E. Davis. [Insights for AI from the human mind](#). Commun. ACM 64, 1 (January 2021), 38–41.

## Picture References:

- [1] “On this colorized woodcut in the style of Hokusai we see a pensive woman together together giant octopus who melancholically entangles the knowledge graph that extends into the vast empty space of the universe to the galaxies and stars.”, created via ArtBot, Deliberate, 2023, [CC-BY-4.0], <https://tinybots.net/artbot>
- [2] “On this romantic painting in the style of Goya we see a pensive woman together together giant octopus who melancholically entangles the knowledge graph that extends into the vast empty space of the universe to the galaxies and stars.”, created via ArtBot, Deliberate, 2023, [CC-BY-4.0], <https://tinybots.net/artbot>