ML Dublin meets Accenture

accenture

Agenda

Eda Bayram

Incorporating Literals for Knowledge Graph (KG) Completion

Aonghus McGovern

(The) Representation Matters

Paul Walsh

Chains of Thought

Socials

Please feel free to post on

Twitter / LinkedIn / Mastodon / mySpace / Bebo

@DublinML #MLDublin

@Accenture_Irl #Accenture

linkedin.com/school/adapt-centre/

linkedin.com/showcase/accenture-the-dock/

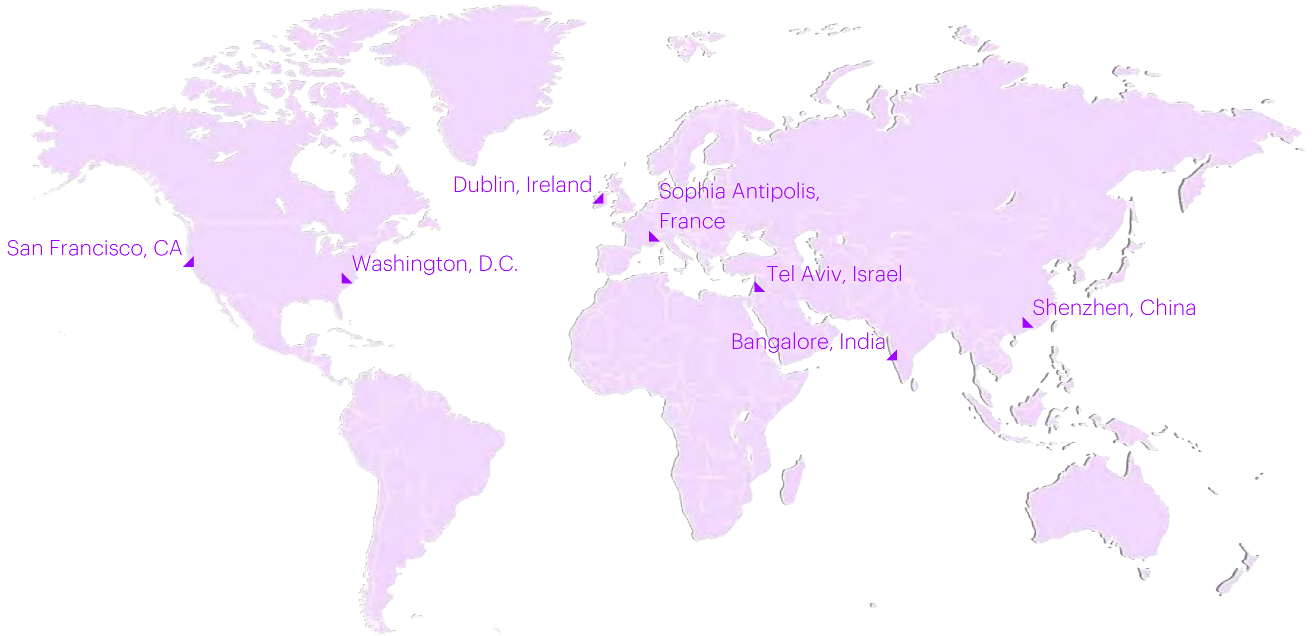
Incorporating Literals for Knowledge Graph Completion

Eda Bayram
Research Scientist
Accenture Labs, Dublin

Accenture **Labs**

Our **7 Accenture Labs** are innovation engines where we harness the most promising breakthrough technologies and trends to shape our clients’ future together.

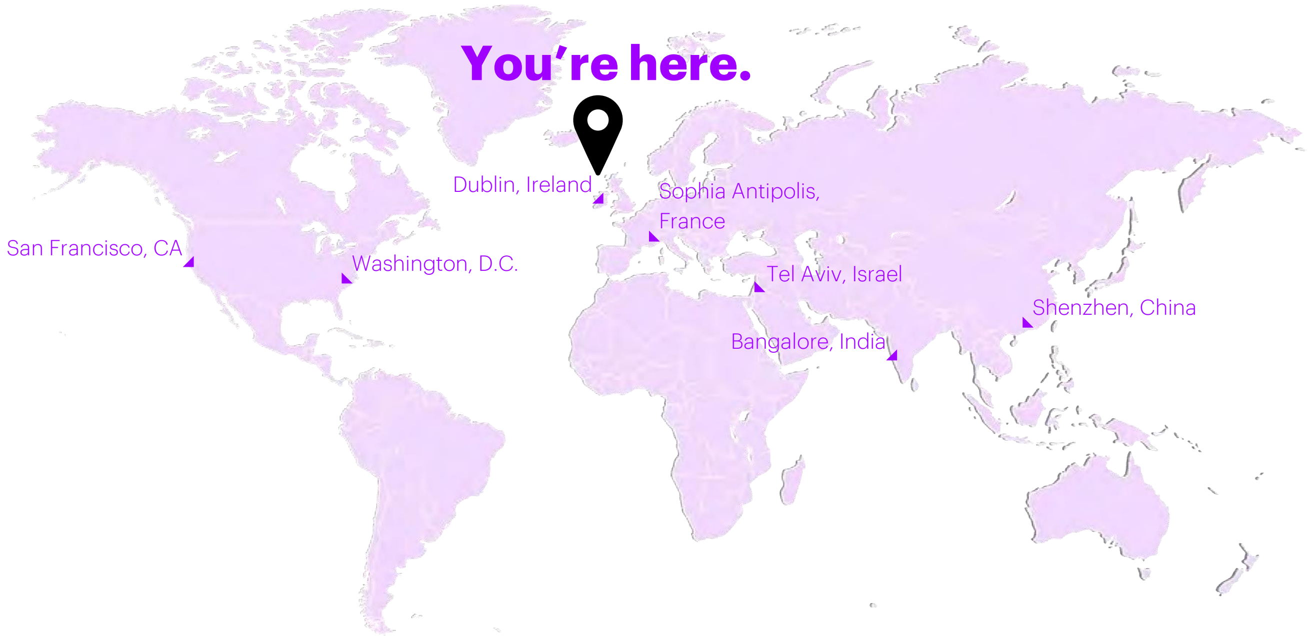
For more information, please visit accenture.com/labs



Accenture **Labs**

Our **7 Accenture Labs** are innovation engines where we harness the most promising breakthrough technologies and trends to shape our clients' future together.

For more information, please visit accenture.com/labs



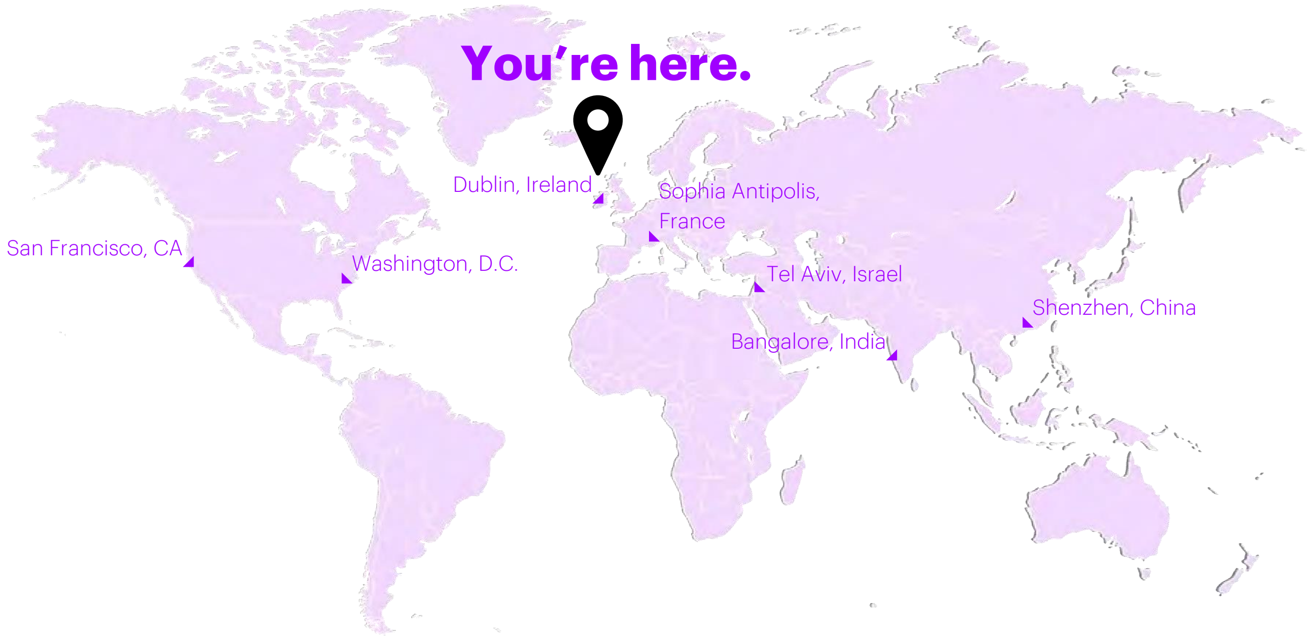
Artificial intelligence

- AI reasoning & inference
- Conversational systems
- Trustworthy AI
- Computational creativity
- Workforce enablement & well-being

Accenture **Labs**

Our **7 Accenture Labs** are innovation engines where we harness the most promising breakthrough technologies and trends to shape our clients' future together.

For more information, please visit accenture.com/labs



Artificial intelligence

- AI reasoning & inference
- Conversational systems
- Trustworthy AI
- Computational creativity
- Workforce enablement & well-being

BioInnovation

- AI for life sciences, healthcare and medicine
- Learning on knowledge graphs of
    - Multi-Omics data e.g., genomics
    - Patients records
- Generate and validate hypothesis

Outline

Incorporating Literals for Knowledge Graph Completion

1. Relational Data and Graph Machine Learning
2. Relational Databases and Knowledge Graphs
3. Link Prediction Task
4. Attribute Prediction Task
5. Incorporating literals for Link Prediction

Relational Data and Graph Machine Learning

Image Credit: Medium

Social Networks

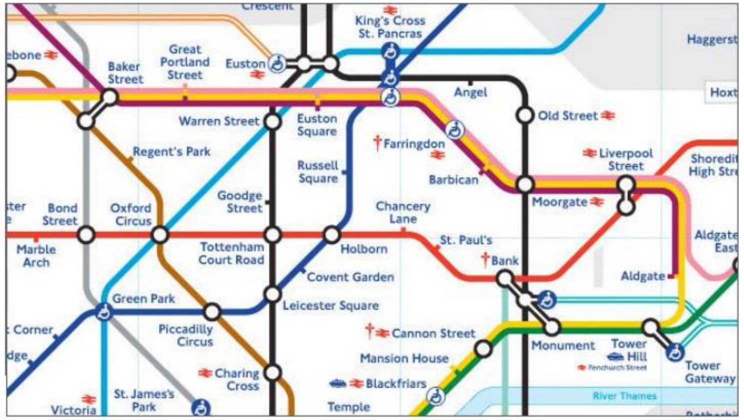
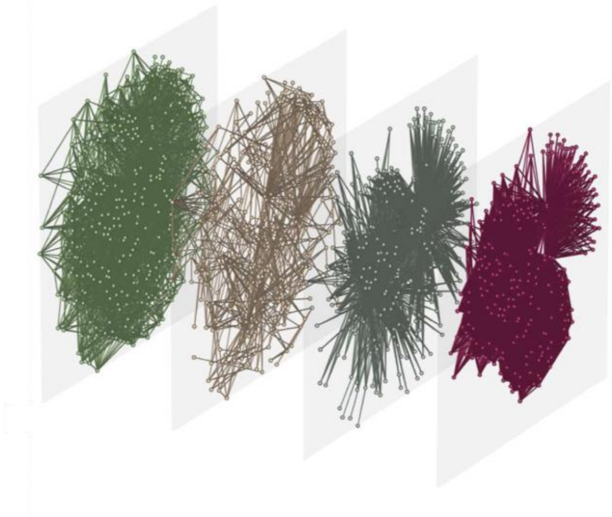


Image Credit: visitlondon.com

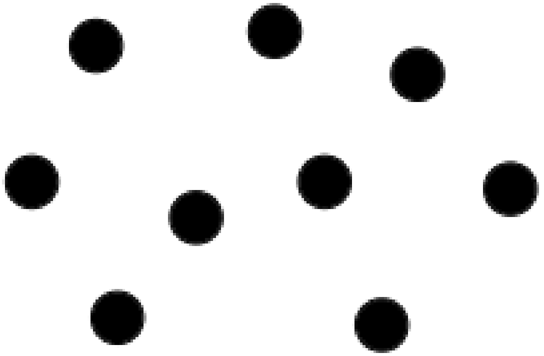
Transportation Networks



Multi-layer connectome of C. elegans
[Bentley et al., 2016]

Biological systems

Relational Data and Graph Machine Learning



Graph (set of Nodes,

Image Credit: Medium

Social Networks

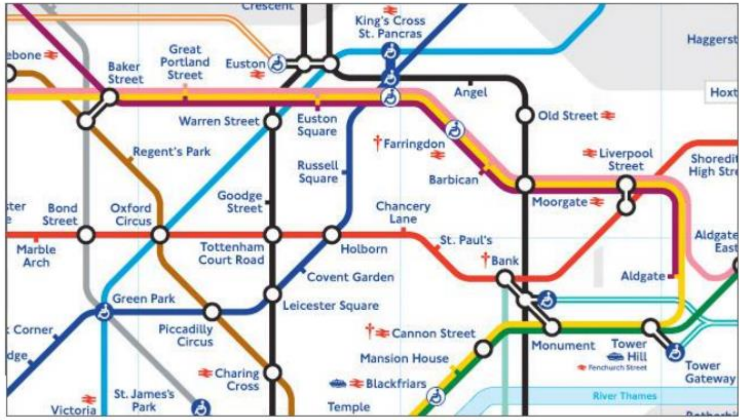
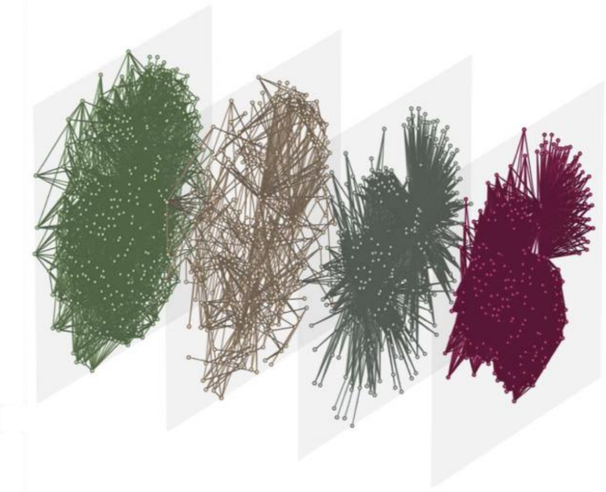


Image Credit: visitlondon.com

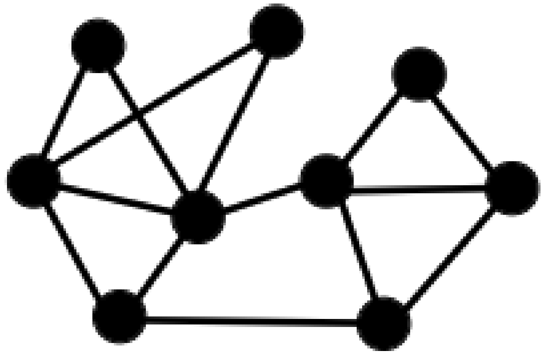
Transportation Networks



Multi-layer connectome of C. elegans
[Bentley et al., 2016]

Biological systems

Relational Data and Graph Machine Learning



Graph (set of Nodes, set of Edges)

How to take advantage of relational structure for better prediction?

Image Credit: Medium

Social Networks

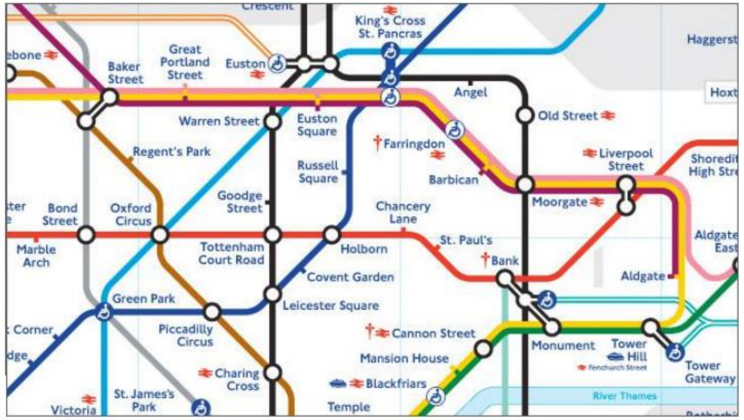
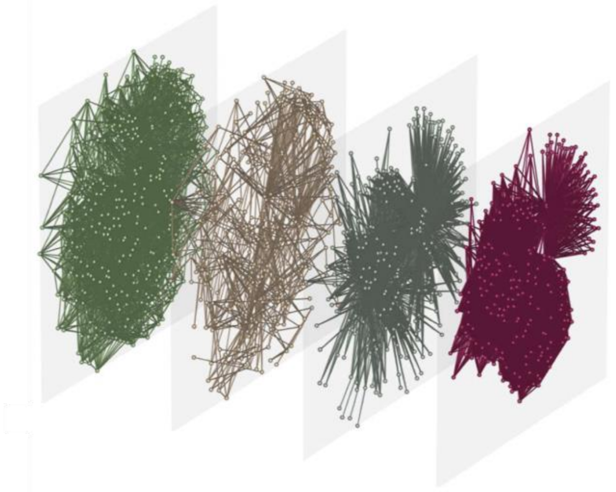


Image Credit: visitlondon.com

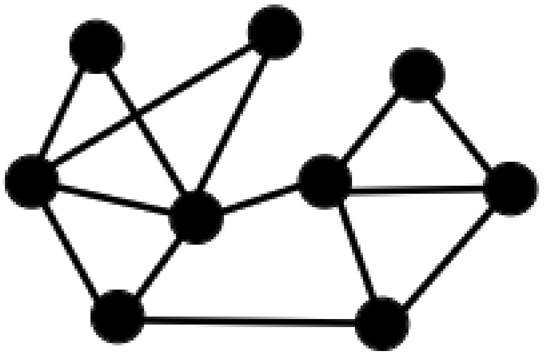
Transportation Networks



Multi-layer connectome of C. elegans
[Bentley et al., 2016]

Biological systems

Relational Data and Graph Machine Learning



Graph (set of Nodes, set of Edges)

How to take advantage of relational structure for better prediction?

Graph Machine Learning

- ML models studying underlying graph structure of complex data and its features
  - Complex/irregular/network structured data rather than regular such as tabular, grid, sequential data

- Exploit background relational information for better performance

Relational Databases

where rich and valuable background relational information available!

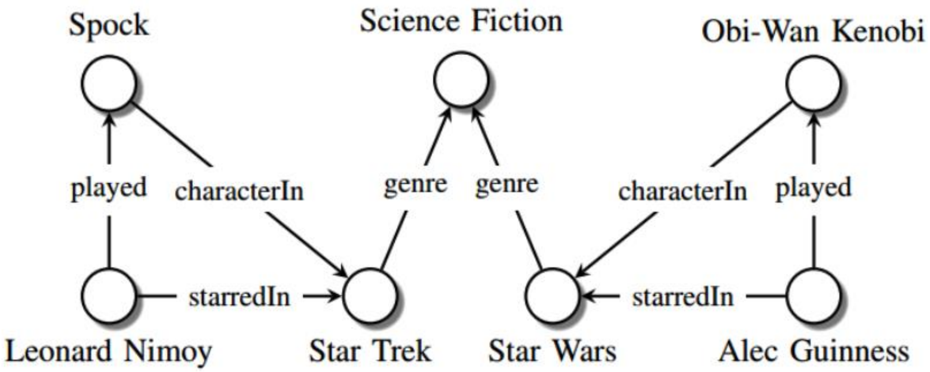
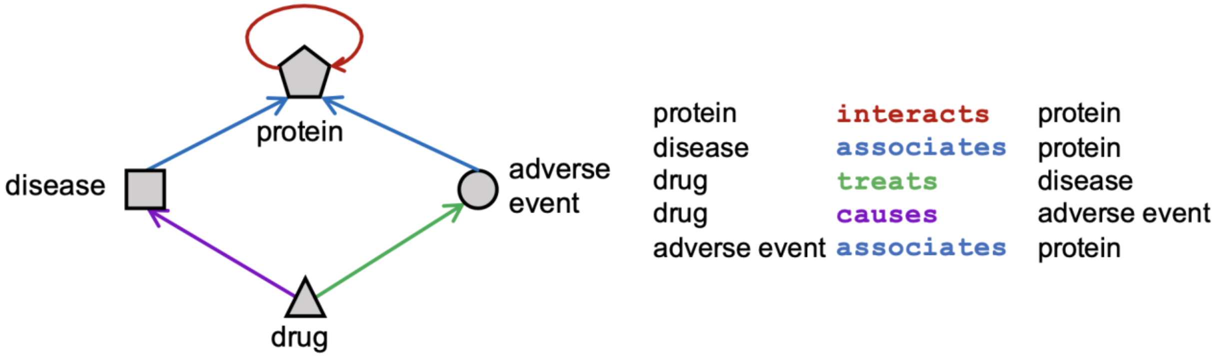


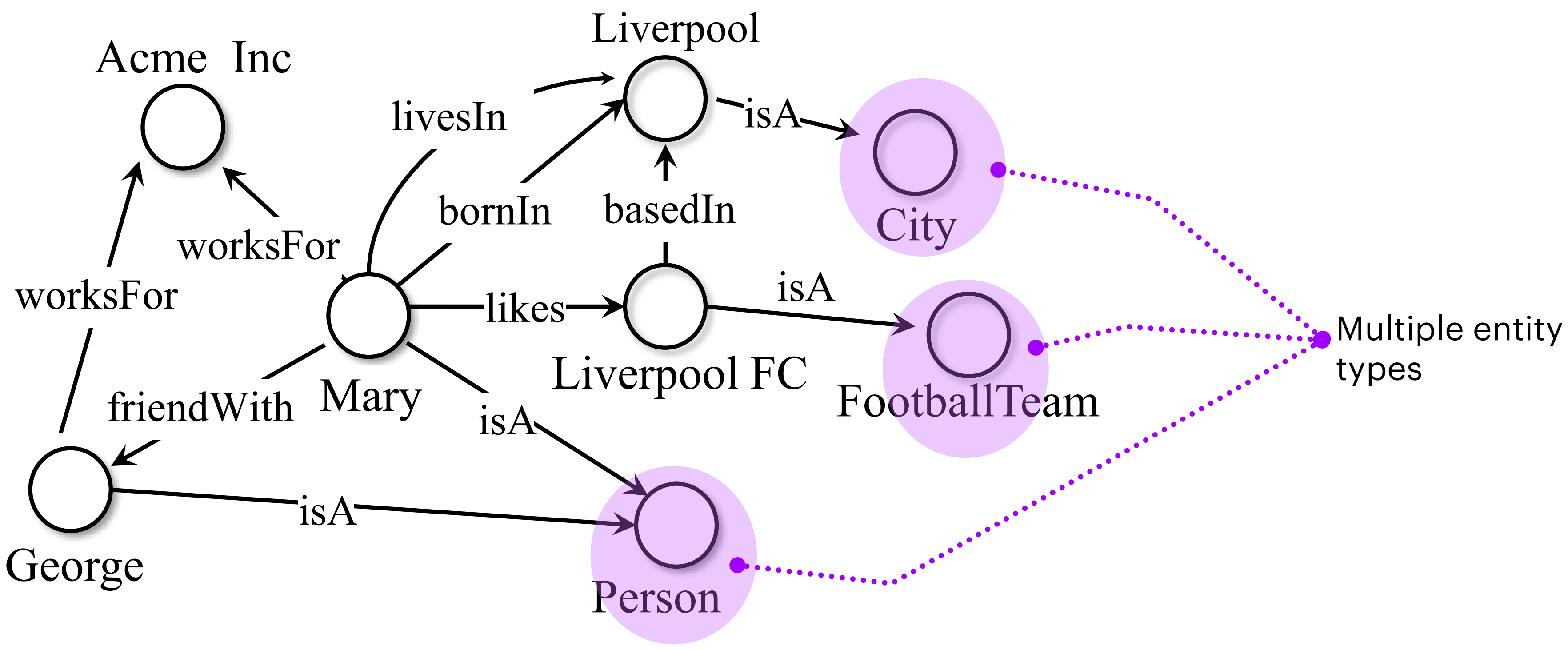
Image credit: Maximilian Nickel et al

Movie Database



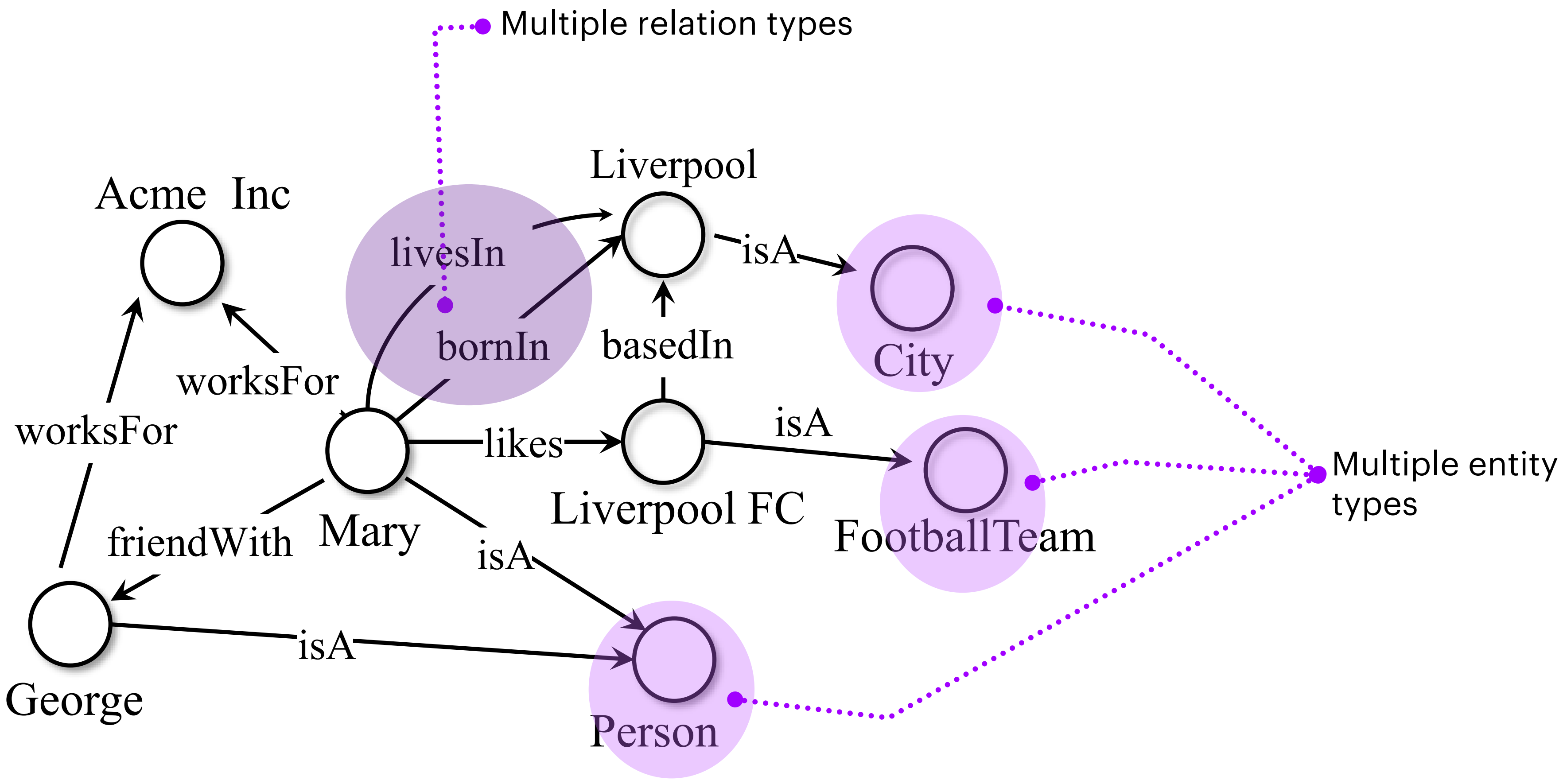
Biomedical Knowledge Base

Knowledge Graph



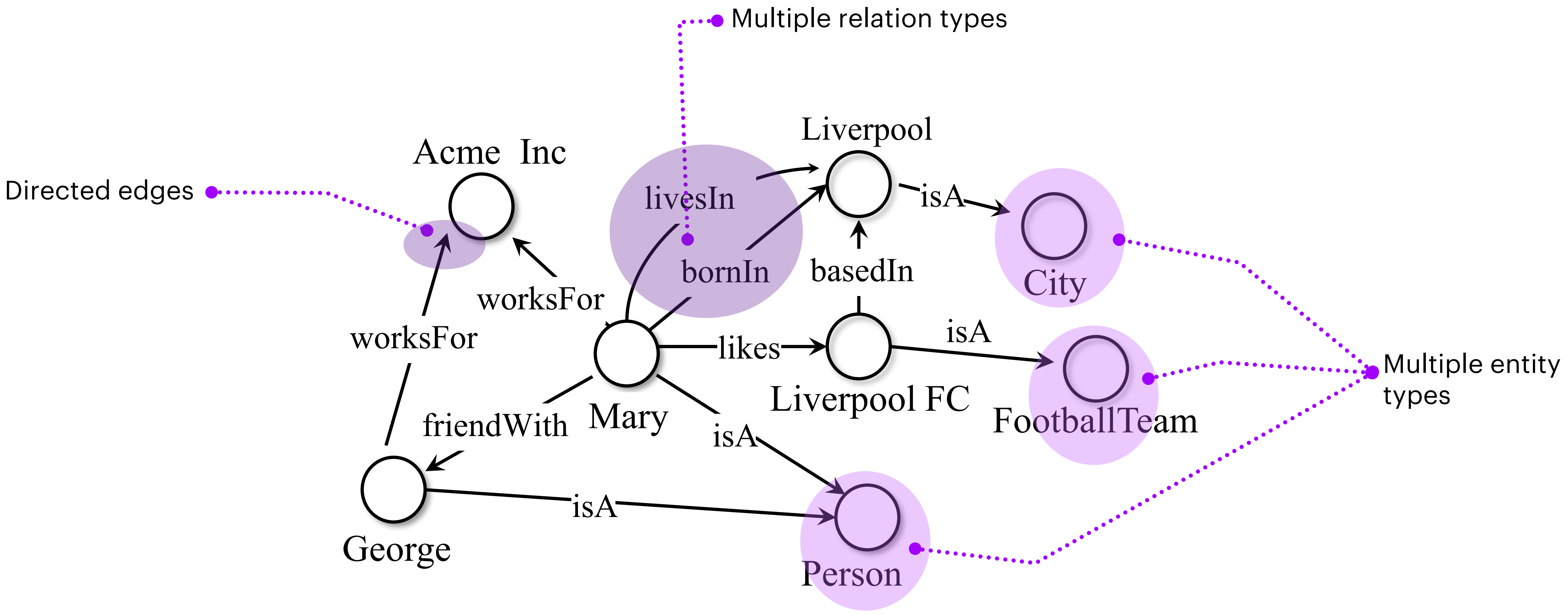
In-depth overview of Knowledge Graphs in [Hogan et al. 2020]

Knowledge Graph



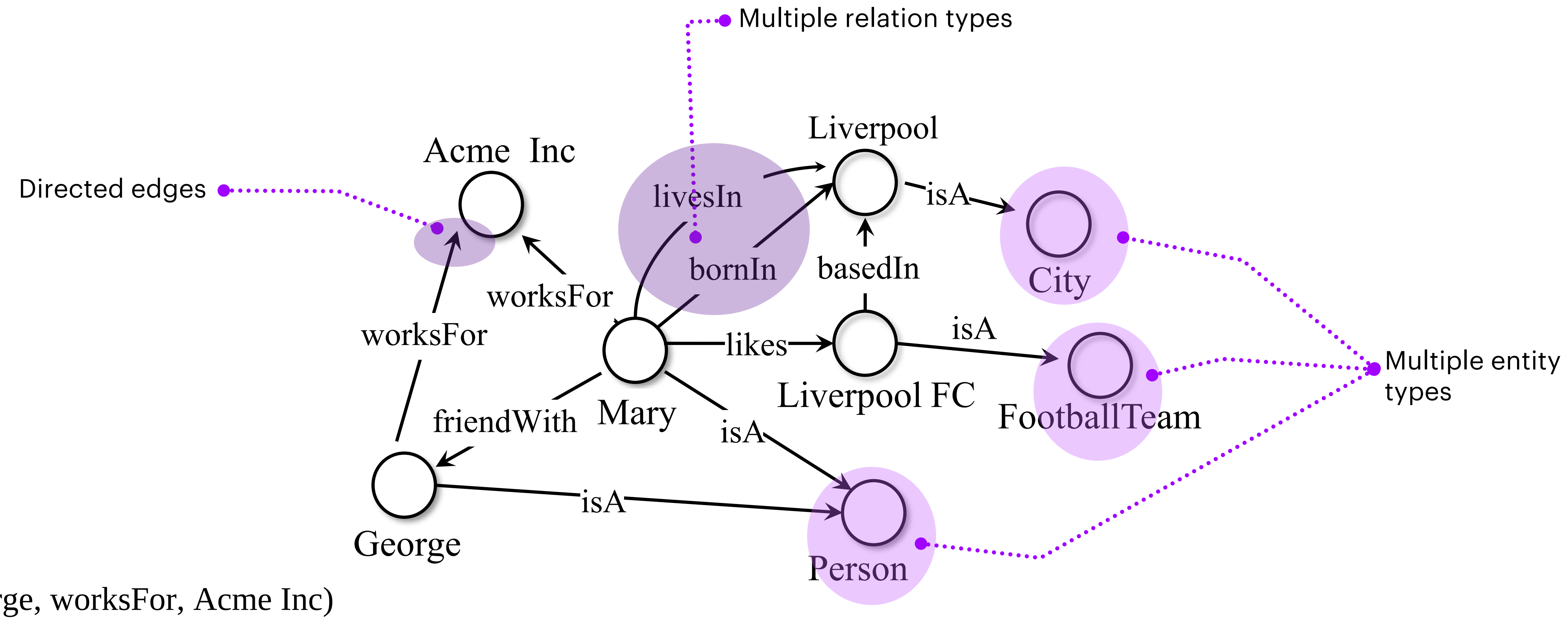
In-depth overview of Knowledge Graphs in [Hogan et al. 2020]

Knowledge Graph



In-depth overview of Knowledge Graphs in [Hogan et al. 2020]

Knowledge Graph

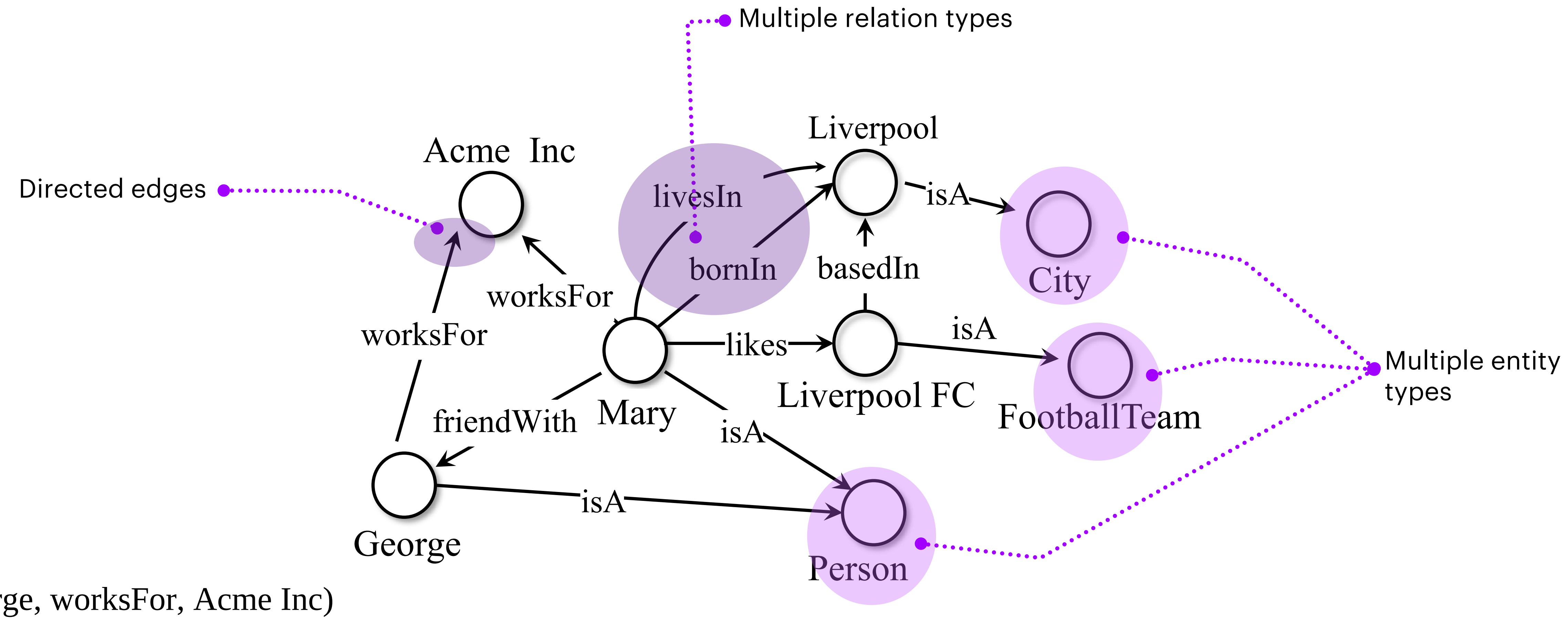


(George, worksFor, Acme Inc)

Triple

In-depth overview of Knowledge Graphs in [Hogan et al. 2020]

Knowledge Graph



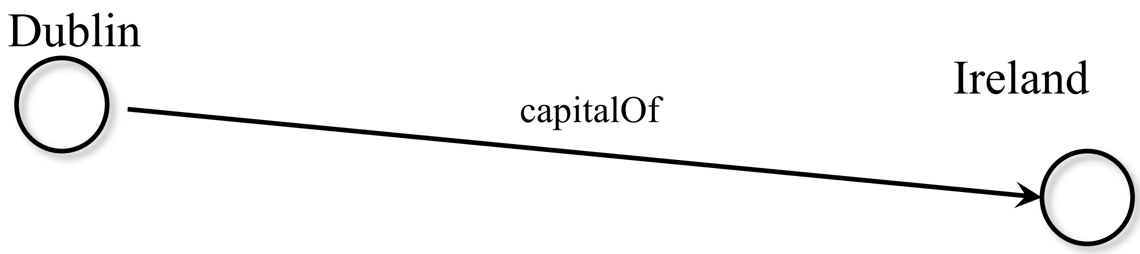
(George, worksFor, Acme Inc)

Triple $\mathcal{G} = \{(s, p, o)\} \subseteq \mathcal{E} \times \mathcal{R} \times \mathcal{E}$

$\mathcal{E}$: set of entities of $\mathcal{G}$ $\mathcal{R}$: set of relations of $\mathcal{G}$

In-depth overview of Knowledge Graphs in [Hogan et al. 2020]

Knowledge Graph: Relational vs Attributive Triples



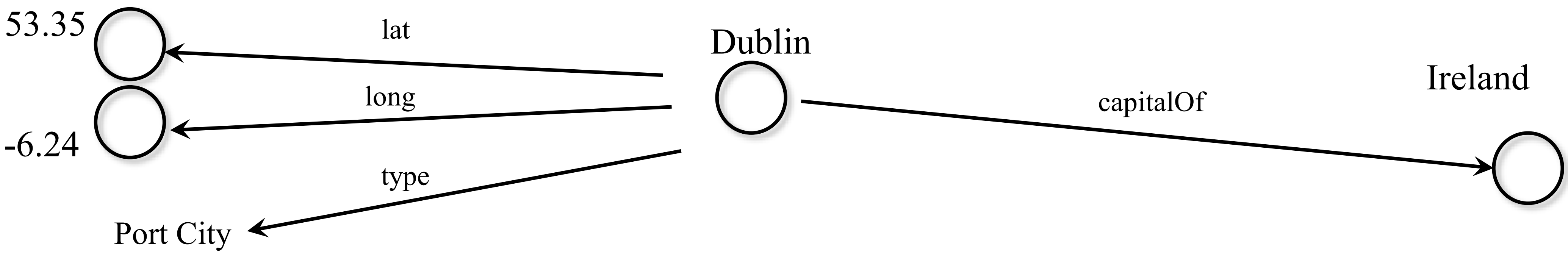
Object properties:

connect an entity to another entity

Relational Triples

(subject node, predicate, object node)

Knowledge Graph: Relational vs Attributive Triples



Data type properties:

connect an entity to a literal

Attributive Triples

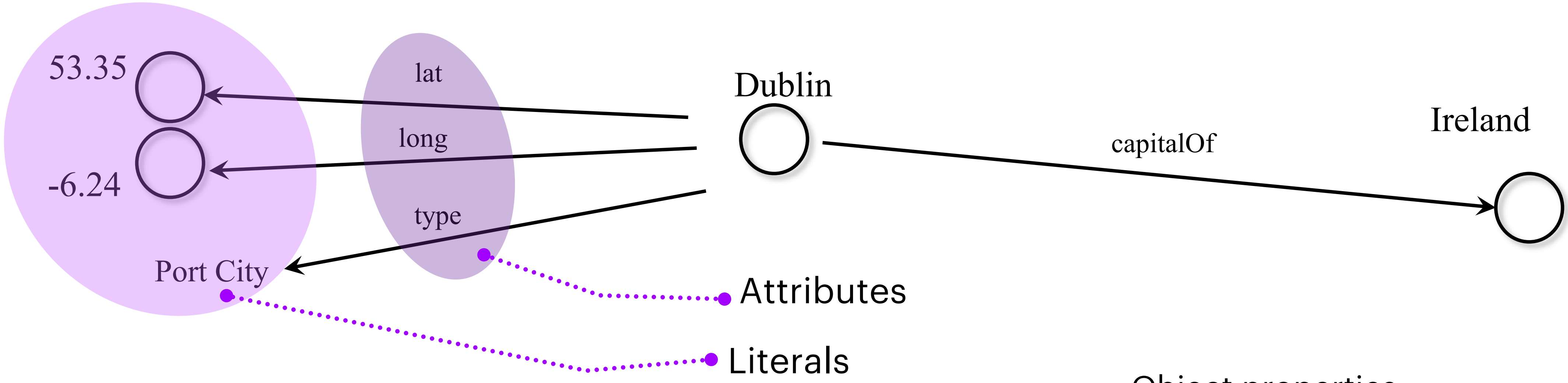
Object properties:

connect an entity to another entity

Relational Triples

(subject node, predicate, object node)

Knowledge Graph: Relational vs Attributive Triples



Data type properties:

connect an entity to a literal

Attributive Triples

(subject node, attribute, literal)

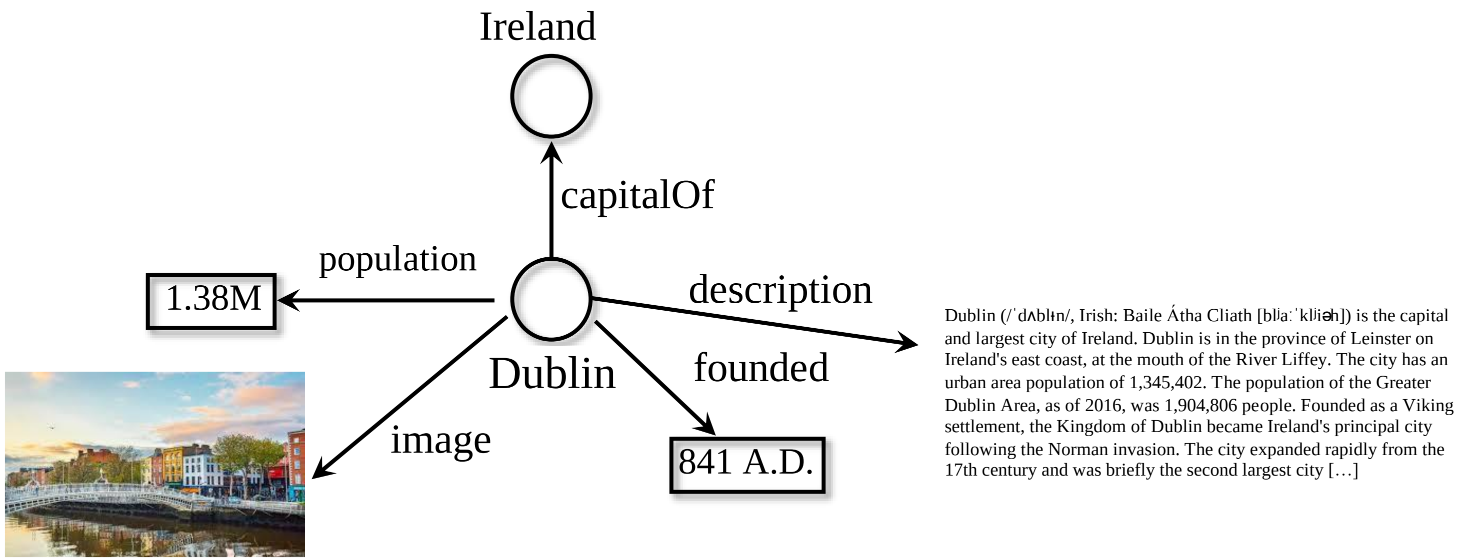
Object properties:

connect an entity to another entity

Relational Triples

(subject node, predicate, object node)

Knowledge Graph: Multi-modal Literals



[Gesese et al. 2019]

Literals can be numerical, categorical, textual, images, sound waves, or other continuous values!

Outline

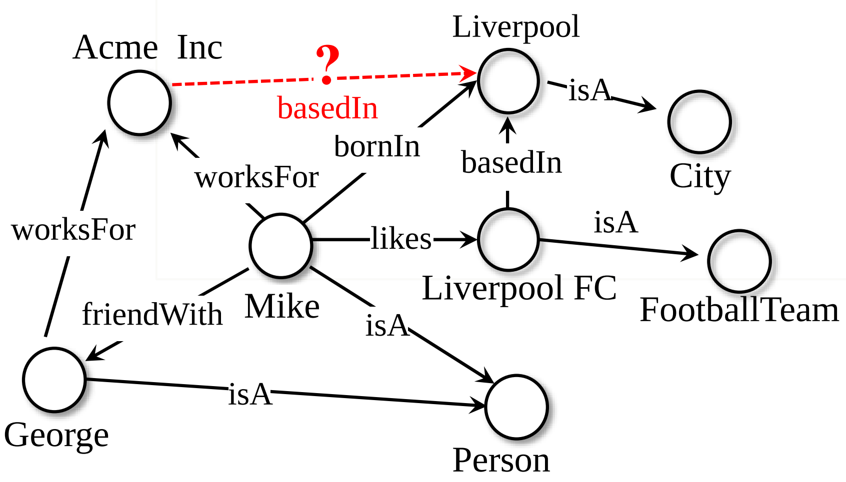
Incorporating Literals for

Knowledge Graph Completion

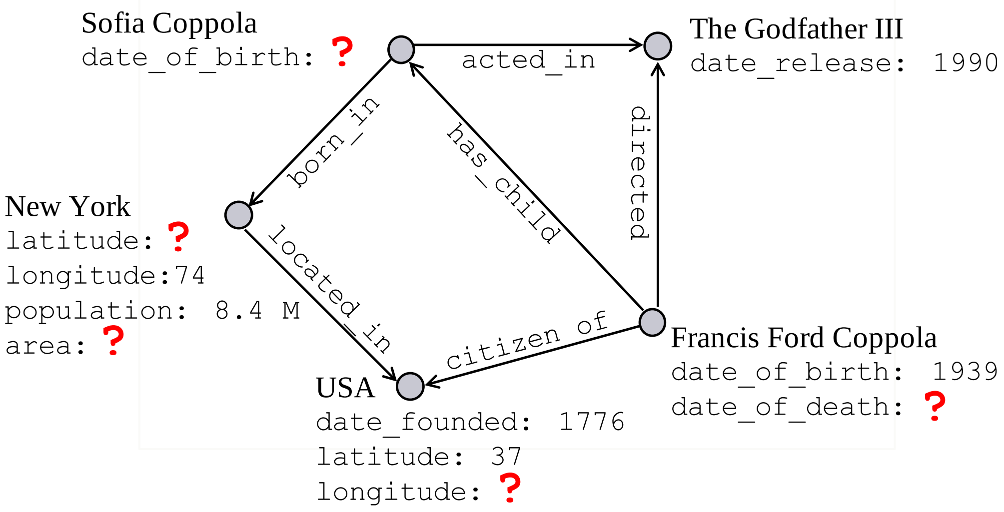
Machine Learning on Knowledge Graphs:

Knowledge graph completion

LINK PREDICTION

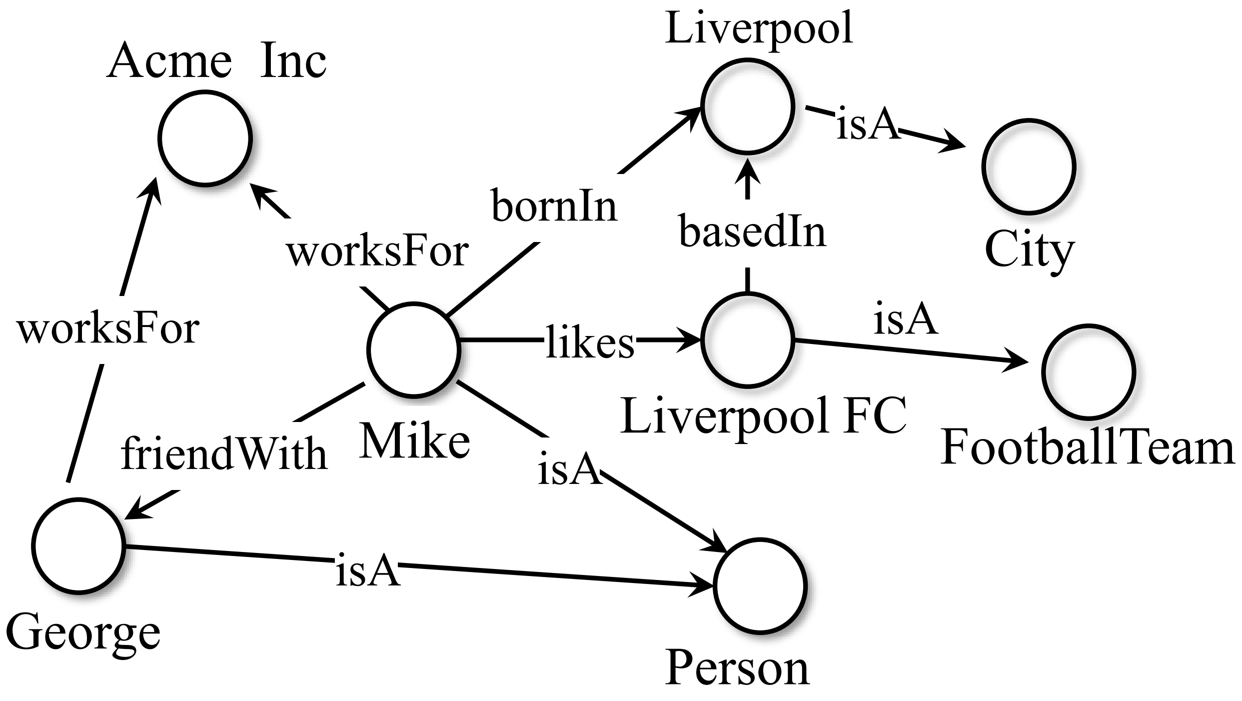


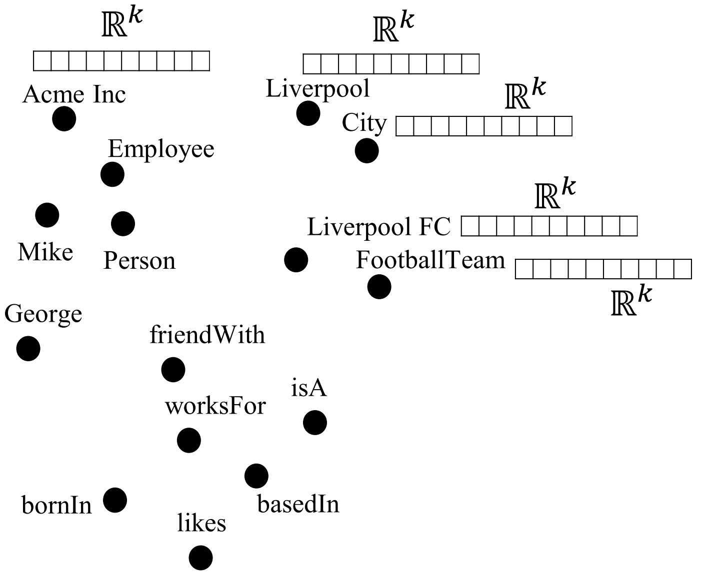
ATTRIBUTE PREDICTION



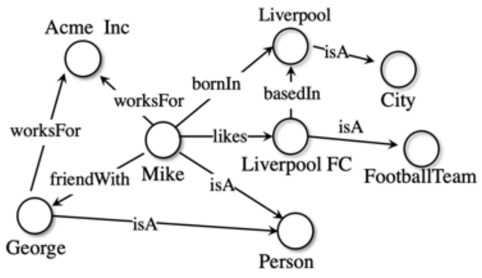
Knowledge Graph Embeddings (KGE)

Automatic, supervised learning of **embeddings**, i.e. projections of entities and relations into a continuous low-dimensional space $\mathbb{R}^k$.





Anatomy of KG Embedding Learning



Optimizer

Lookup Layer

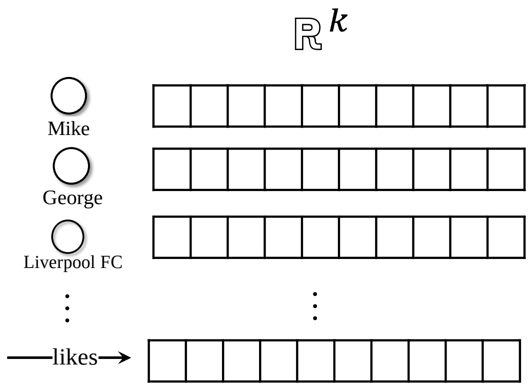
Scoring Layer
$f(s,p,o) \in \mathbb{R}$

Loss Functions
$\mathcal{L}$

Negatives Generation

Training

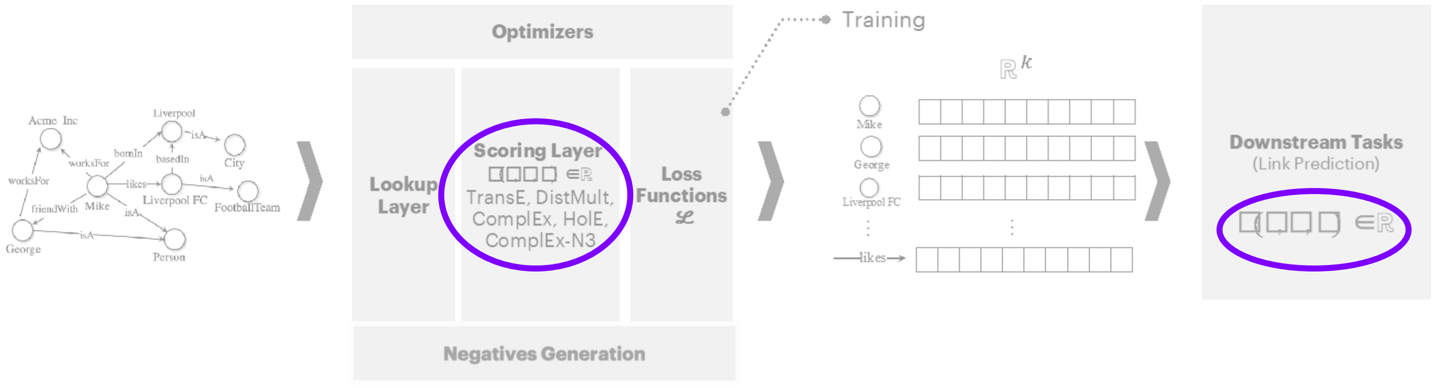
Contrastive Learning



Downstream Tasks
(Link Prediction)

$(\; , \; , \;) \in \mathbb{R}$

Inference



Scoring function f

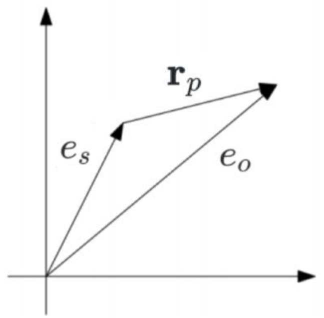
f assigns a score to a triple (s, p, o)

High score = triples is very likely to be factually correct

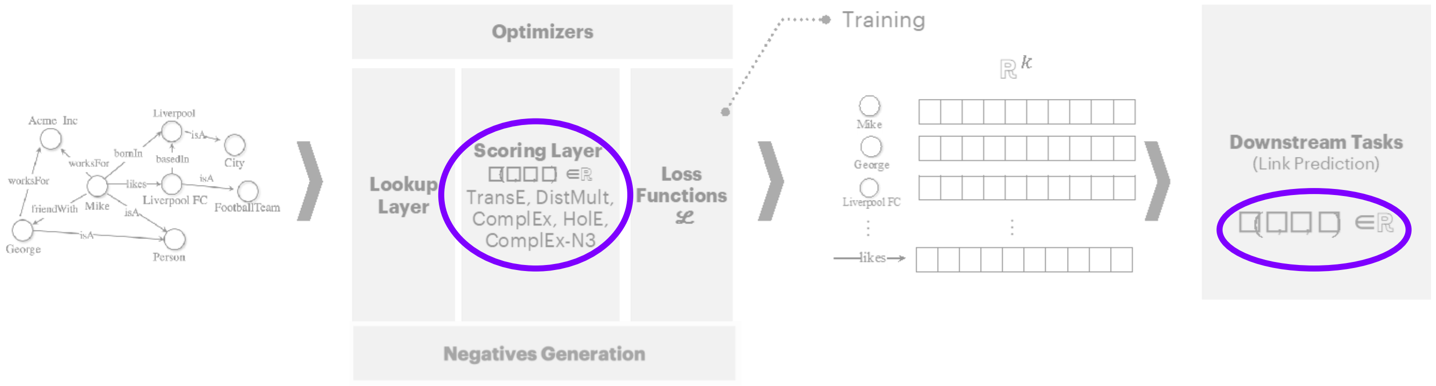
Translation-based Scoring Functions

- **TransE**: Translating Embeddings [Bordes13]

$$f_{TransE} = -||(\mathbf{e}_s + \mathbf{r}_p) - \mathbf{e}_o||_n$$



Scoring function f



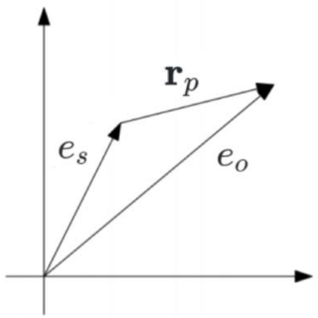
f assigns a score to a triple (s, p, o)

High score = triples is very likely to be factually correct

Translation-based Scoring Functions

- **TransE**: Translating Embeddings [Bordes13]

$$f_{TransE} = -||(\mathbf{e}_s + \mathbf{r}_p) - \mathbf{e}_o||_n$$



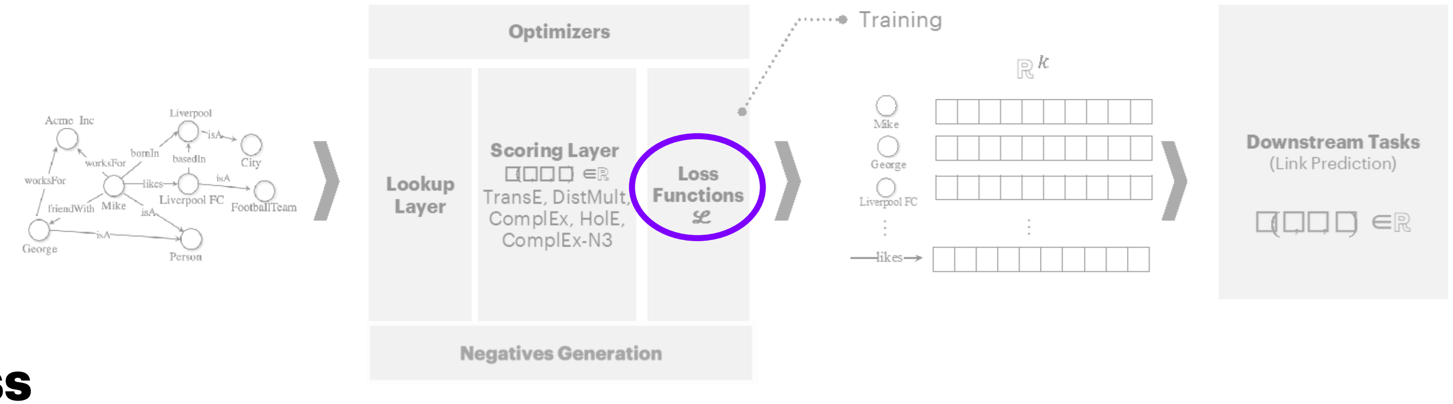
Factorization-based Scoring Functions

- **DistMult**: Dot product.

[Yang et al. 2015]

$$f_{DistMult} = \langle \mathbf{r}_p, \mathbf{e}_s, \mathbf{e}_o \rangle$$

Loss function $\mathcal{L}$



Pairwise Margin-Based Hinge Loss

$$\mathcal{L}(\Theta) = \sum_{t^+ \in \mathcal{G}} \sum_{t^- \in \mathcal{C}} \overbrace{max(0, [\gamma + \underbrace{f(t^-;\Theta)}_{\text{Score assigned to a }\textbf{synthetic negative}} - \underbrace{f(t^+;\Theta)}_{\text{Score assigned to }\textbf{true}\text{ triple}}])}^{\text{Pays a penalty if score of positive triple < score of synthetic negative by a margin } \gamma}$$

[Bordes et al. 2013]

Negative Log-Likelihood / Cross Entropy

$$\mathcal{L}(\Theta) = \sum_{t \in \mathcal{G} \cup \mathcal{C}} log(1 + exp(-\underbrace{y}_{\text{Label of the triple t}} f(t;\Theta)))$$

$y \in \{-1, 1\}$

[Trouillon et al. 2016]

Evaluation: Ranking Based

Rank the scores of all answer candidates to the query.

s	p	o	score	rank	
Mike	friend_with	George	0.901	1	*
Mike	friend_with	Jim	0.345	2	
Acme	friend_with	George	0.293	3	
Mike	friend_with	Liverpool	0.201	4	
France	friend_with	George	0.156	5	

s	p	o	score	rank	
Mike	born_in	Leeds	0.789	1	
Mike	born_in	Liverpool	0.753	2	*
Mike	born_in	Germany	0.695	3	
George	born_in	Liverpool	0.456	4	
Mike	born_in	George	0.234	5	

Evaluation: Ranking Based

Rank the scores of all answer candidates to the query.

s	p	o	score	rank	
Mike	friend_with	George	0.901	1	*
Mike	friend_with	Jim	0.345	2	
Acme	friend_with	George	0.293	3	
Mike	friend_with	Liverpool	0.201	4	
France	friend_with	George	0.156	5	

s	p	o	score	rank	
Mike	born_in	Leeds	0.789	1	
Mike	born_in	Liverpool	0.753	2	*
Mike	born_in	Germany	0.695	3	
George	born_in	Liverpool	0.456	4	
Mike	born_in	George	0.234	5	

Positive triples from test set

Test set = {
<Mike friend_with George>
<Mike born_in Liverpool>
}

Mean Reciprocal Rank (MRR)

$$MRR = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{1}{rank_{(s,p,o)_i}}$$

Hits@N

$$Hits@N = \sum_{i=1}^{|Q|} 1 \text{ if } rank_{(s,p,o)_i} \leq N$$

MRR = 0.75

Hit@1 = 0.5

Higher the better!

Evaluation of SoTA Methods

Link Prediction (MRR)

	FB15K-237	WN18RR	YAGO3-10	FB15k	WN18
Literature Best	**0.35***	0.48*	0.49*	**0.84****	**0.95***
TransE (AmpliGraph 2)	0.31	0.22	0.50	0.62	0.64
DistMult (AmpliGraph 2)	0.30	0.47	0.48	0.71	0.82
ComplEx (AmpliGraph 2)	0.31	0.50	0.49	0.73	0.94
HolE (AmpliGraph 2)	0.30	0.47	0.47	0.73	0.94
TransE (AmpliGraph 1)	0.31	0.22	**0.51**	0.63	0.66
DistMult (AmpliGraph 1)	0.31	0.47	0.50	0.78	0.82
ComplEx (AmpliGraph 1)	0.32	**0.51**	0.49	0.80	0.94
HolE (AmpliGraph 1)	0.31	0.47	0.50	0.80	0.94
ConvE (AmpliGraph 1)	0.26	0.45	0.30	0.50	0.93
ConvE (1-N, AmpliGraph 1)	0.32	0.48	0.40	0.80	**0.95**
ConvKB (AmpliGraph 1)	0.23	0.39	0.30	0.65	0.80

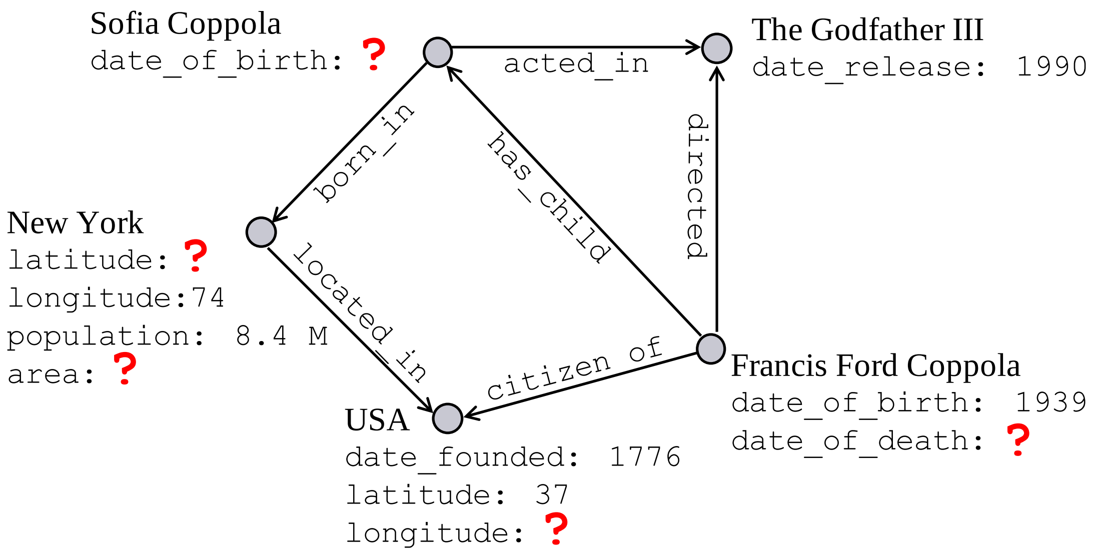
Benchmark Datasets

	FB15K-237	WN18RR	YAGO3-10
Training	272,115	86,835	1,079,040
Validation	17,535	3,034	5,000
Test	20,466	3,134	5,000
Entities	14,541	40,943	123,182
Relations	237	11	37

AmpliGraph

Open source library based on TensorFlow that predicts links between concepts in a knowledge graph.

Attribute Prediction on Knowledge Graphs



Knowledge graph with attributes

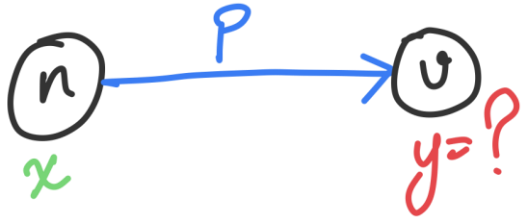
- Multi-relational data
- Multiple types of entities
- Different set of attributes

(living on different feature spaces)

Completion of Numerical Attributes in a KG

Knowledge graph with attributes $\mathcal{G}(\mathcal{V}, \mathcal{E}, \mathcal{P}, \mathcal{A})$

- Multiple relationships $\mathcal{P}$
- Different set of attributes $\mathcal{A}$



- Predict target node attribute: y_v, $y \in \mathcal{A}, v \in \mathcal{V}$
- From source node attribute: x_n, $x \in \mathcal{A}, n \in \mathcal{V}$
- Relation between the nodes: $\mathtt{r}(v, n) = \mathtt{p}$

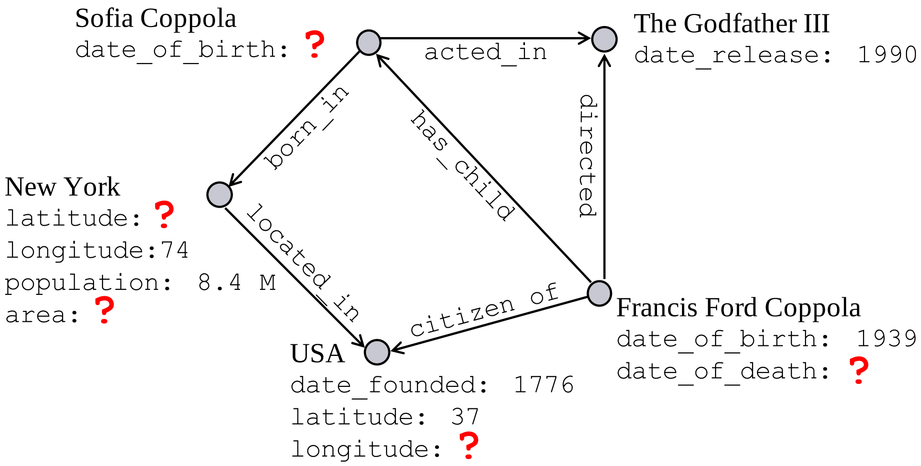
We need heterogeneous message passing!

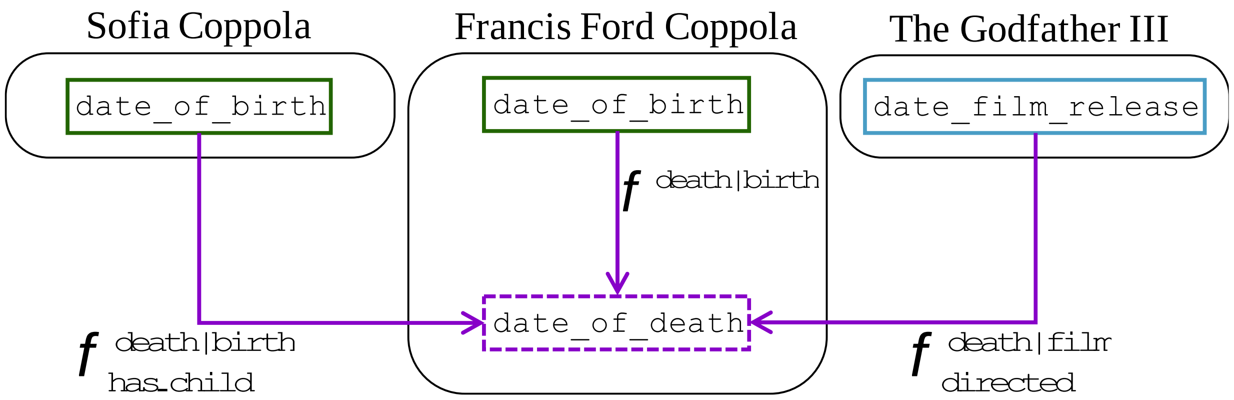
Message type $< x, \mathtt{p}, y >$

Message function:

E. Bayram, A. Garcia-Duran and R. West "*Node Attribute Completion in Knowledge Graphs with Multi-Relational Propagation*" in ICASSP 2021

Completion of Numerical Attributes in a KG





Message-passing algorithm steps:

1. Collect

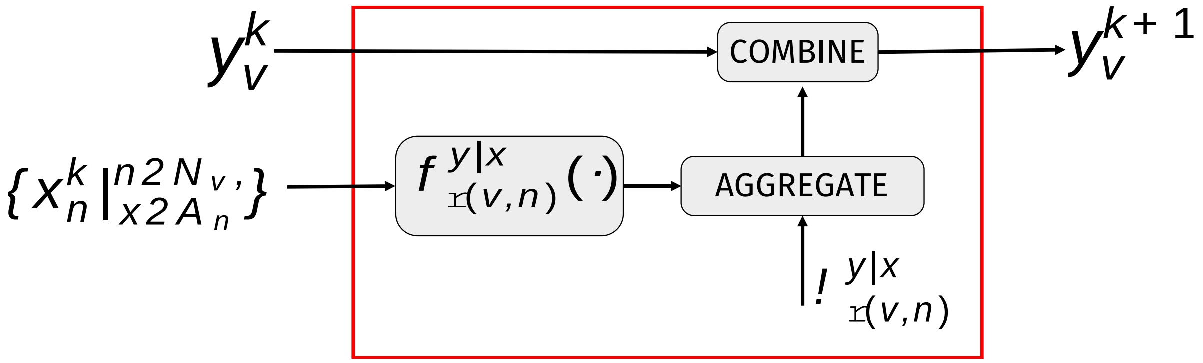
2. Aggregate: weighted average of collected messages

3. Update: $y_v^{(k+1)} = (1-\xi)y_v^{(k)} + \xi\hat{y}_v$

E. Bayram, A. Garcia-Duran and R. West "*Node Attribute Completion in Knowledge Graphs with Multi-Relational Propagation*" in ICASSP 2021

Multi-relational Attribute Propagation (MRAP)

An iteration of MRAP



MRAP: heterogeneous message passing compared to:

- Global attribute averaging method
- Local attribute averaging method
- **NAP++** [Kotnis & Garcia-Duran, 2018]
  Embedding Learning (TransE) + attribute propagation on kNN embedding graph

	LOCAL/GLOBAL		NAP++		MRAP	
Attribute	MAE	RMSE	MAE	RMSE	MAE	RMSE
date_of_birth	24.0	69.4	27.2	40.0	**12.3**	**20.5**
date_of_death	36.8	54.7	79.3	95.7	**16.0**	**25.2**
film_release	11.8	15.2	9.3	12.8	**6.4**	**9.0**
organization_founded	*72.3	*121.4	65.0	114.6	**60.9**	**96.5**
location_founded	111.7	176.4	165.4	291.7	**105.9**	**146.2**

FB15K date attributes with 40% sparsity of observations.

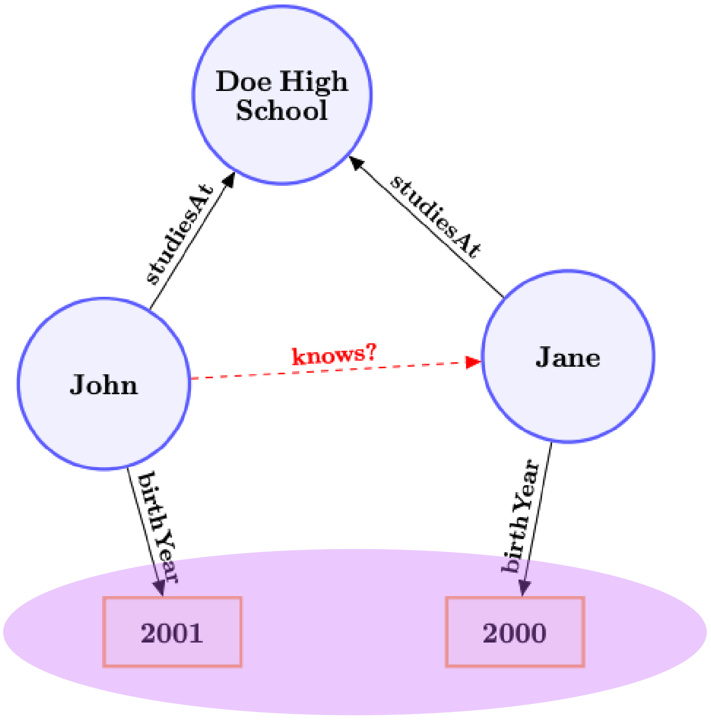
E. Bayram, A. Garcia-Duran and R. West “*Node Attribute Completion in Knowledge Graphs with Multi-Relational Propagation*” in ICASSP 2021

Outline

Incorporating Literals for

Knowledge Graph Completion

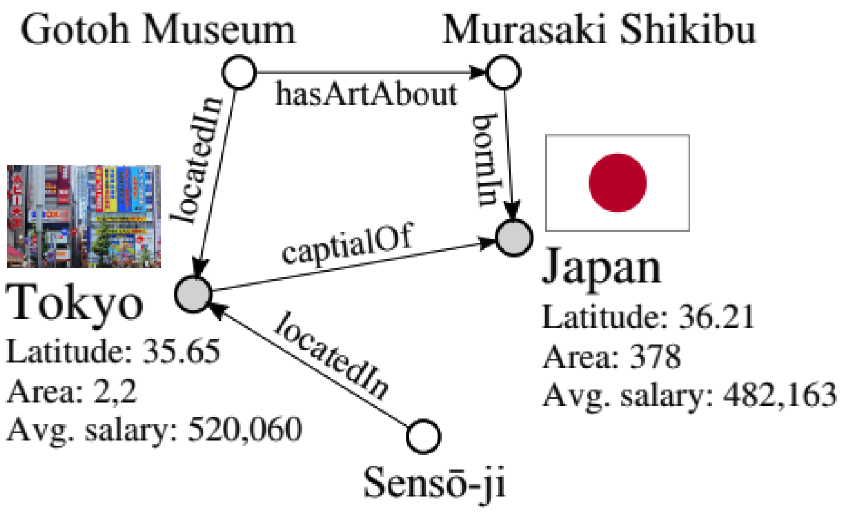
Incorporating Literals for Link Prediction



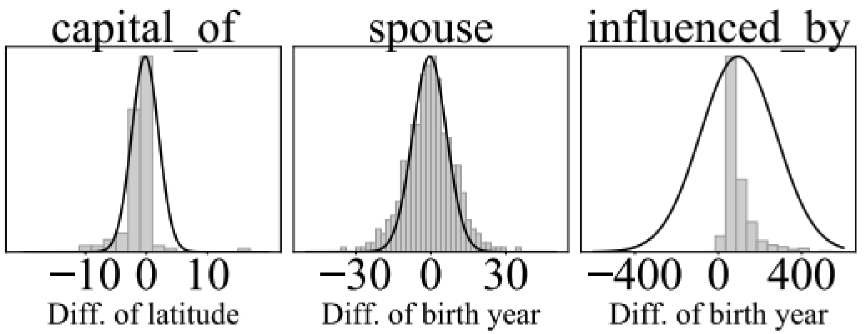
Literals (attributive triples) can be useful while reasoning about the relational triples!

Incorporating Literals for Link Prediction

KBLRN [Garcia-Duran et al., 2018]



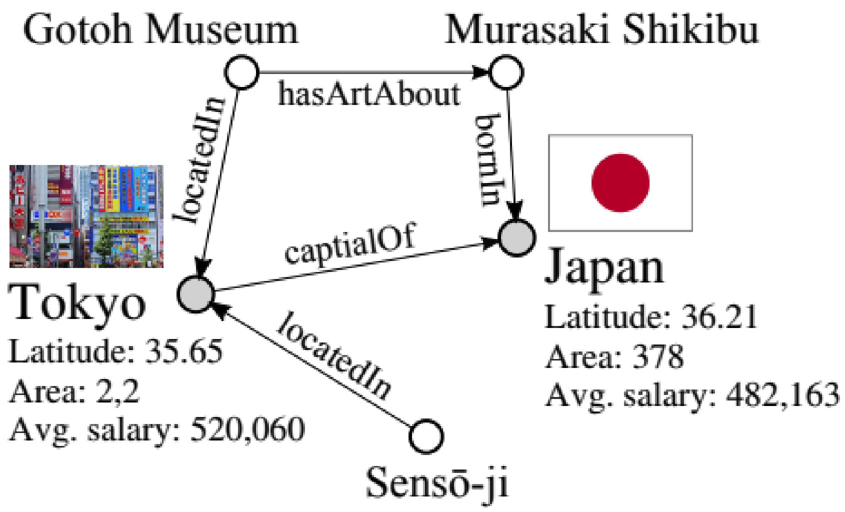
1. Construct RBFs with literal distributions over triples (difference on subject and object attributes)



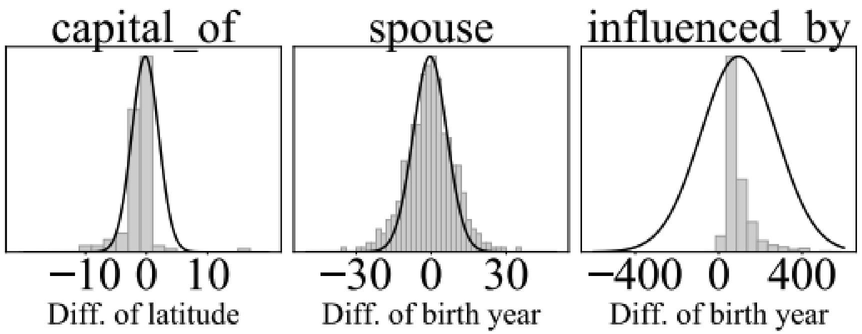
$$\underbrace{\text{f\_att(Tokyo's lat, capitalOf, Japan's lat)}}_{\text{RBF scoring the plausibility of literal difference}}$$

Incorporating Literals for Link Prediction

KBLRN [Garcia-Duran et al., 2018]



1. Construct RBFs with literal distributions over triples (difference on subject and object attributes)



2. Incorporate the attribute prediction score into the loss of the link prediction model

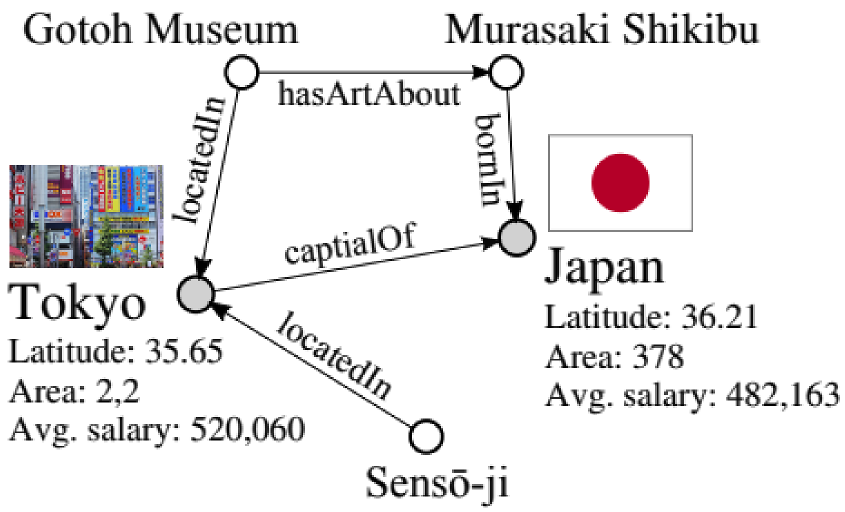
f_rel(Tokyo, capitalOf, Japan)

f_att(Tokyo's lat, capitalOf, Japan's lat)

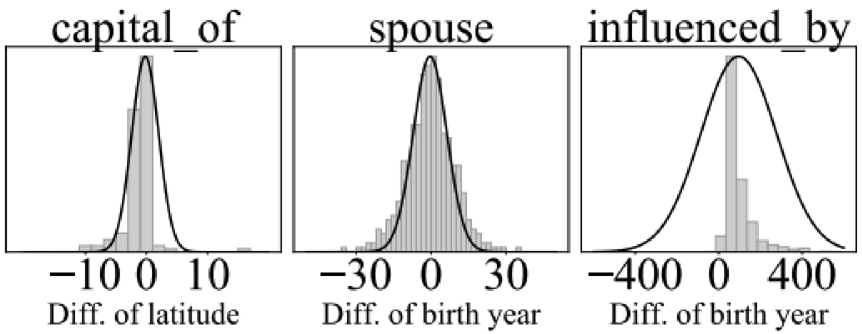
RBF scoring the plausibility of literal difference

Incorporating Literals for Link Prediction

KBLRN [Garcia-Duran et al., 2018]



1. Construct RBFs with literal distributions over triples (difference on subject and object attributes)



2. Incorporate the attribute prediction score into the loss of the link prediction model

f_rel(Tokyo, capitalOf, Japan)

f_att(Tokyo's lat, capitalOf, Japan's lat)

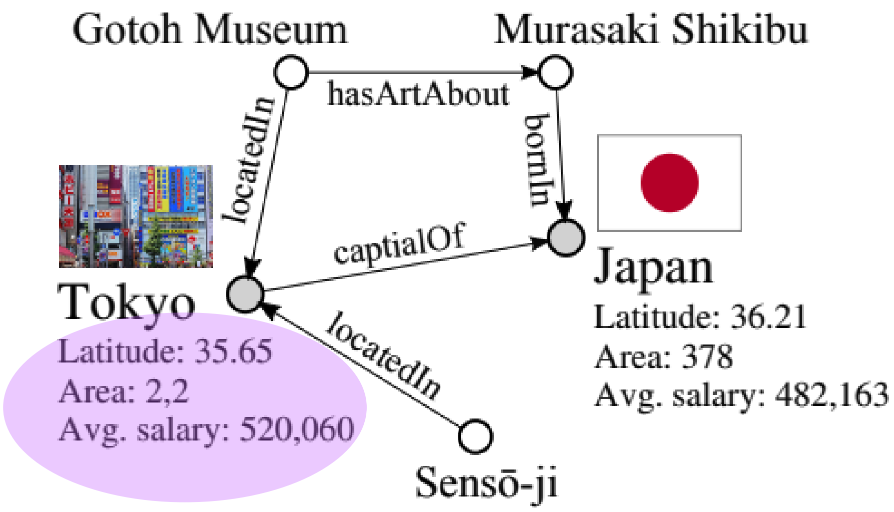
RBF scoring the plausibility of literal difference

✓ Intepretable

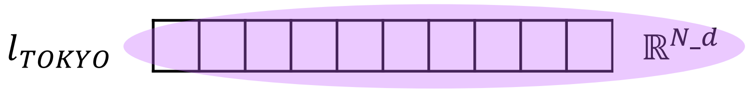
✗ Same type of attribute appear in both its subject and object are quite sparse

Incorporating Literals for Link Prediction

LiteralE [Garcia-Duran et al., 2018]

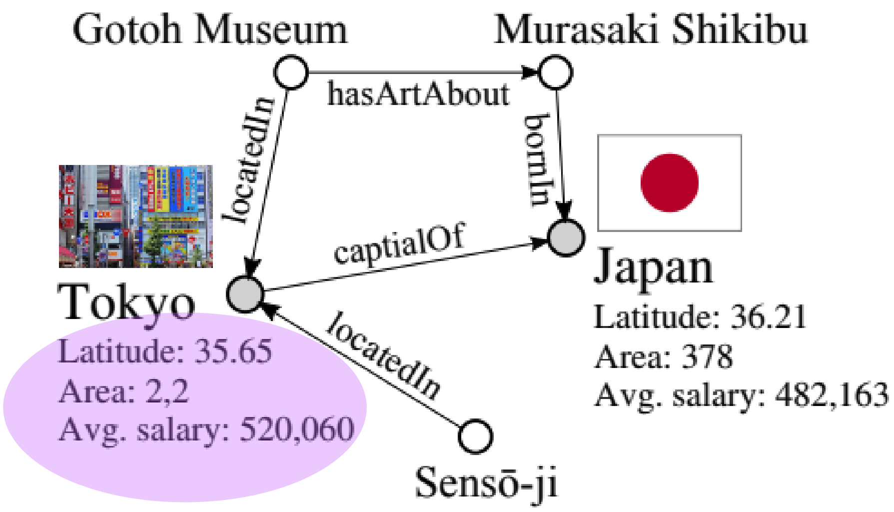


1. Construct a literal vector for each entity

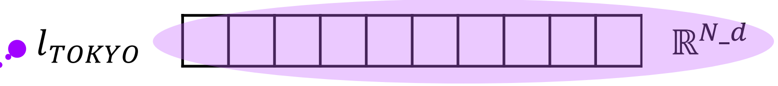


Incorporating Literals for Link Prediction

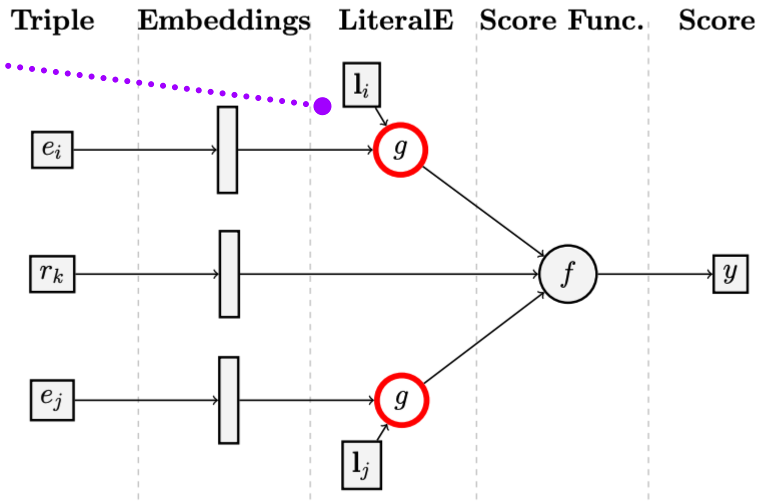
LiteralE [Garcia-Duran et al., 2018]



1. Construct a literal vector for each entity

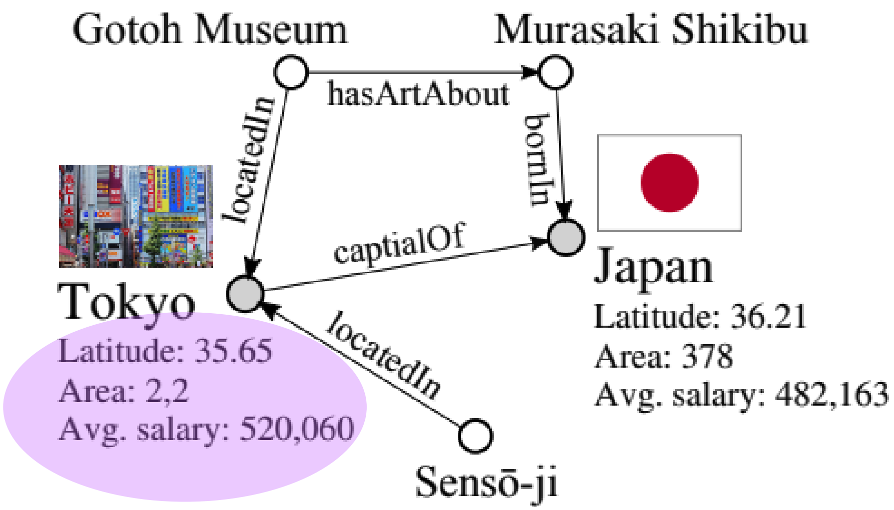


2. Enrich the embedding of the entity using its literal vector through a learnable function

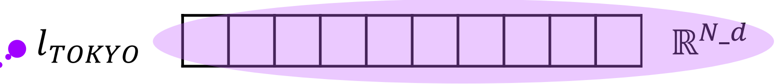


Incorporating Literals for Link Prediction

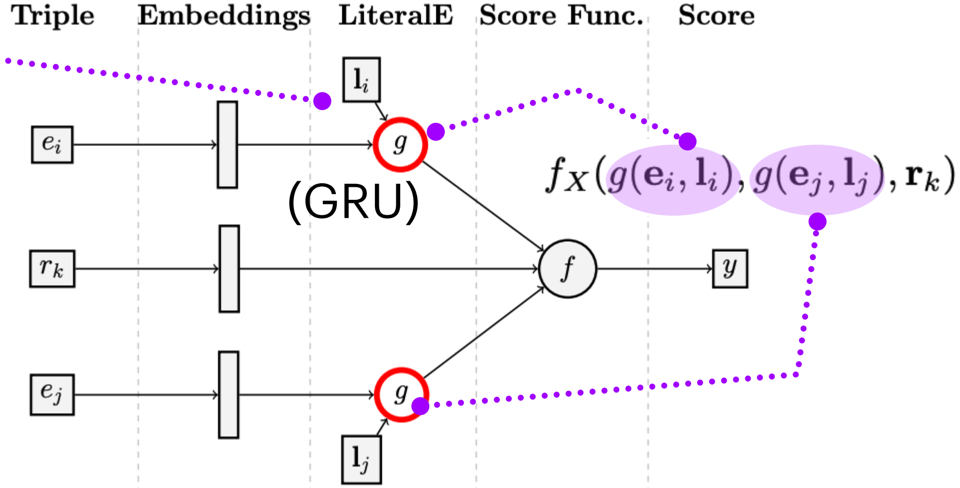
LiteralE [Garcia-Duran et al., 2018]



1. Construct a literal vector for each entity

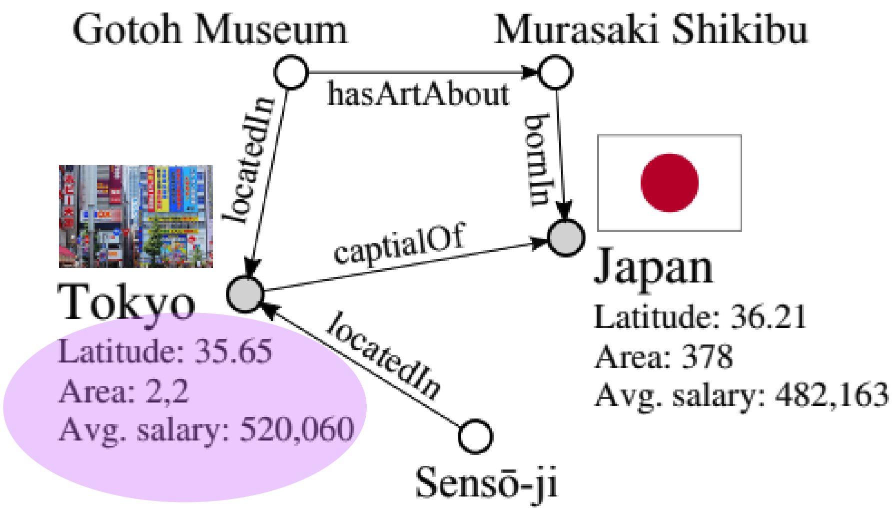


2. Enrich the embedding of the entity using its literal vector through a learnable function



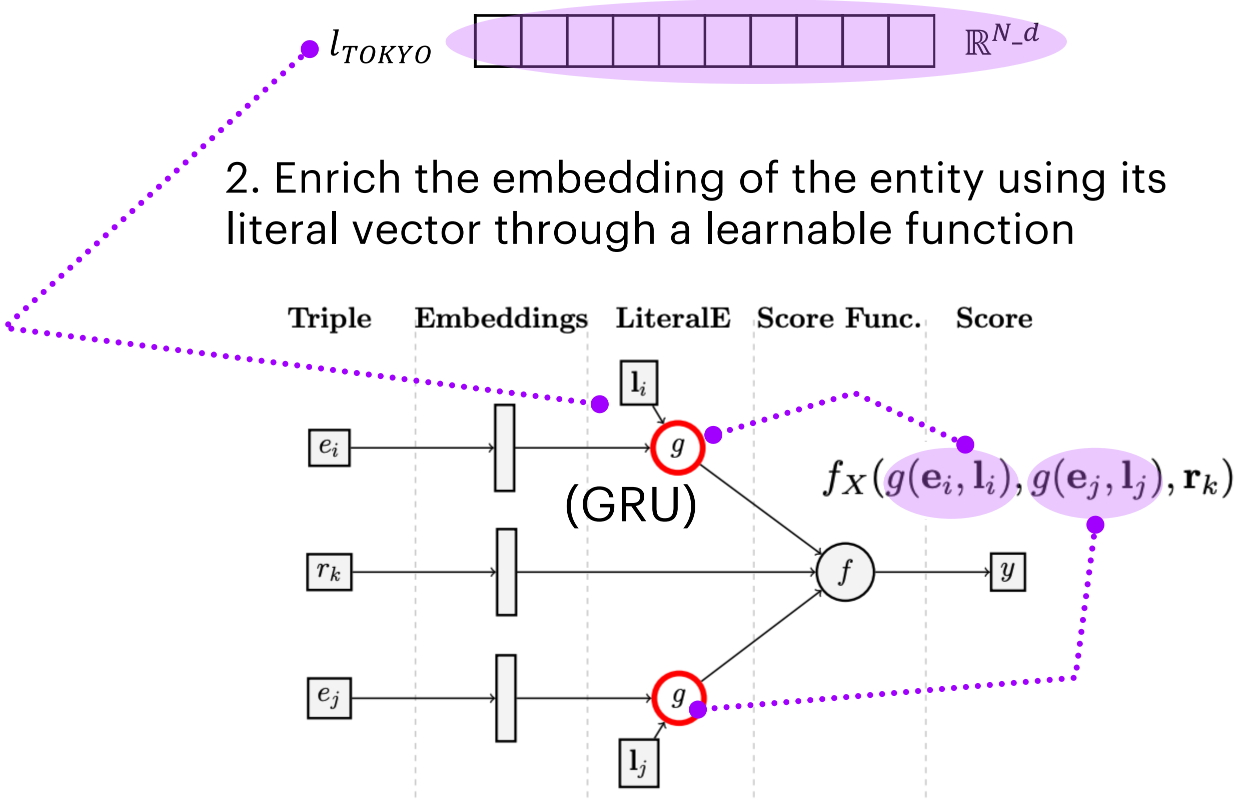
Incorporating Literals for Link Prediction

LiteralE [Garcia-Duran et al., 2018]



1. Construct a literal vector for each entity

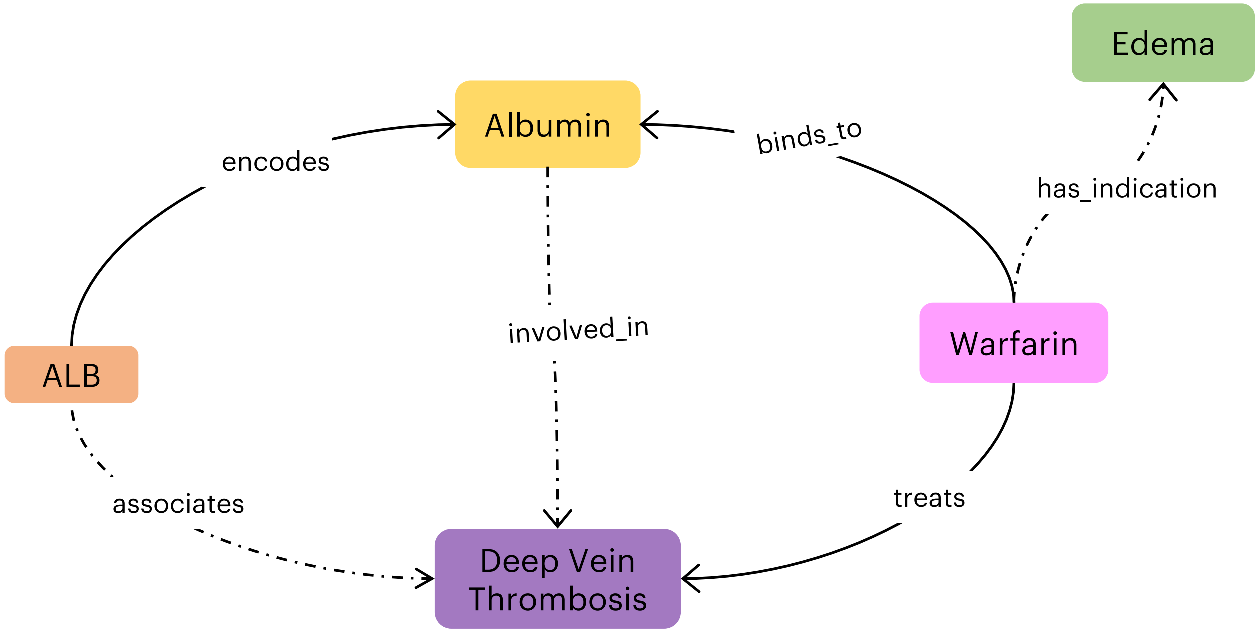
2. Enrich the embedding of the entity using its literal vector through a learnable function



- ✓ SoTA performance
- ✗ Support numeric literals only
- ✗ Fixed size of literal vector for each entity confusion on applicable/ NA attributes and observed/ unobserved literals

Incorporating Literals for Link Prediction

Task: Gene-disease prediction

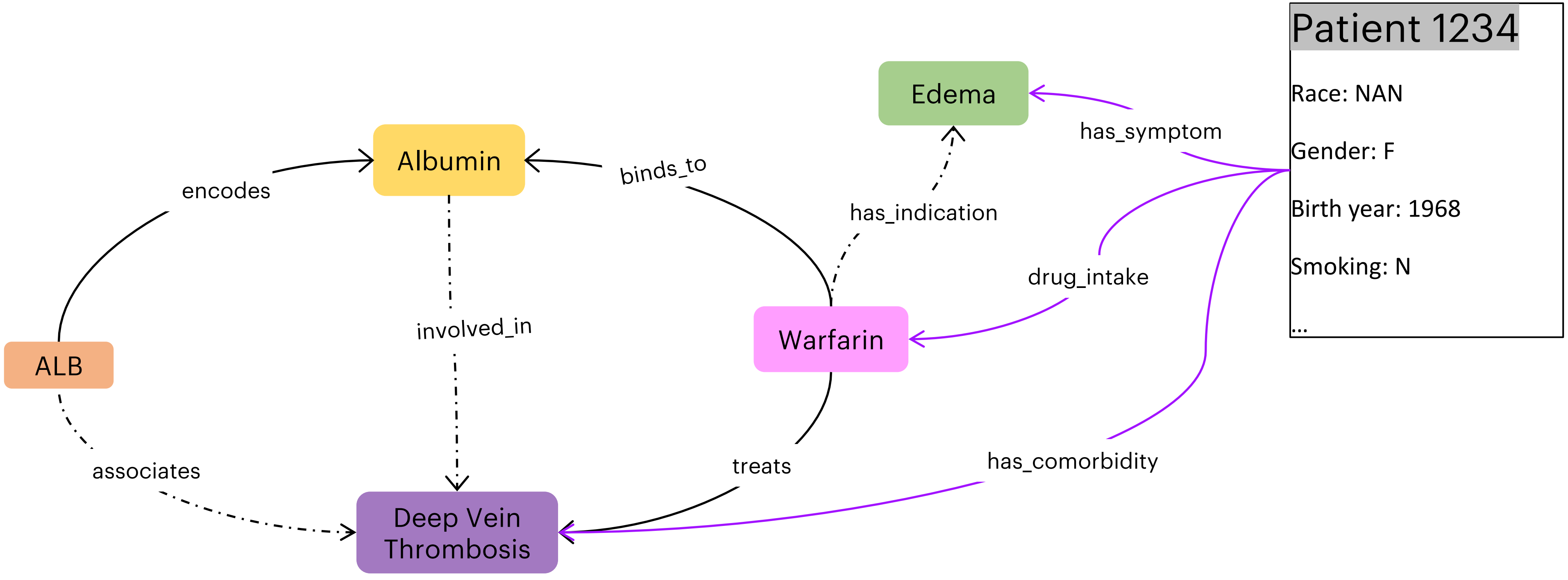


Fragment of an incomplete Biomedical KG

Incorporating Literals for Link Prediction

Industry Use-case: Incorporating patient-level data into biomedical KG

Task: Gene-disease prediction



Fragment of an incomplete Biomedical KG

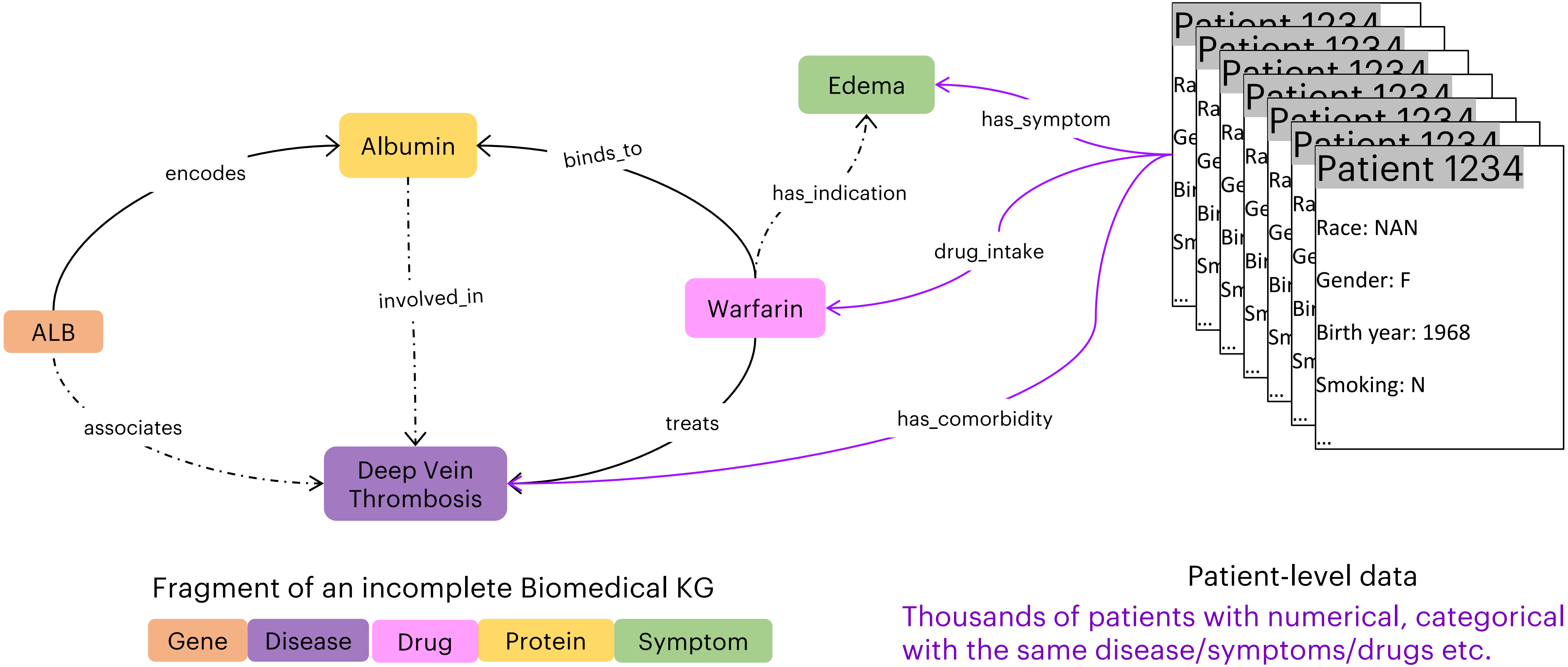
Patient-level data

Gene	Disease	Drug	Protein	Symptom

Incorporating Literals for Link Prediction

Industry Use-case: Incorporating patient-level data into biomedical KG

Task: Gene-disease prediction



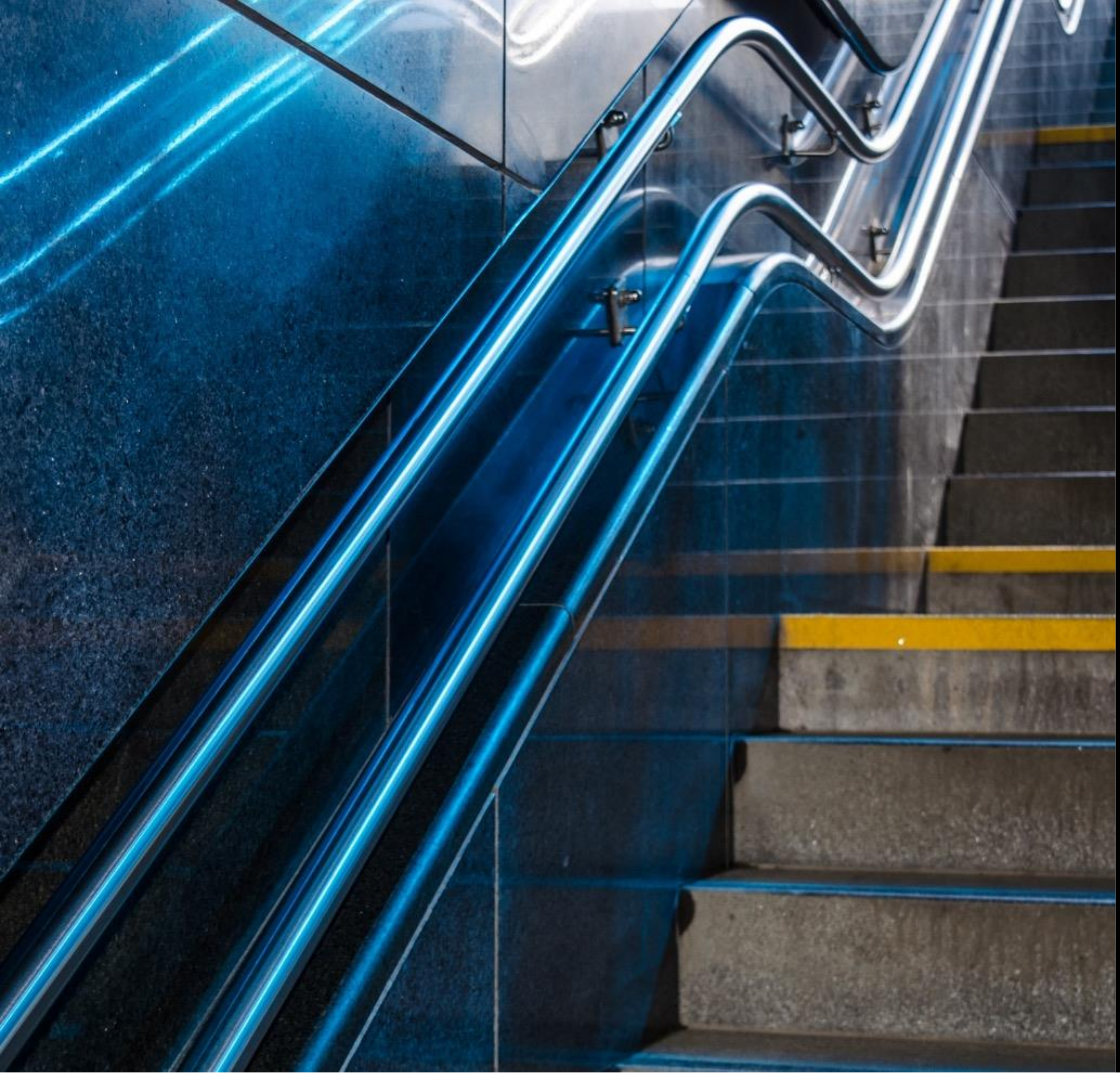
Thousands of patients with numerical, categorical features with the same disease/symptoms/drugs etc.

Takeaways: Incorporating Literals for KG completion

1. Machine learning on relational data
2. KG completion: relational triples vs attributive triple
3. Incorporating literals for link prediction

Thank you

Eda Bayram
eda.bayram@ accenture.com



Questions?

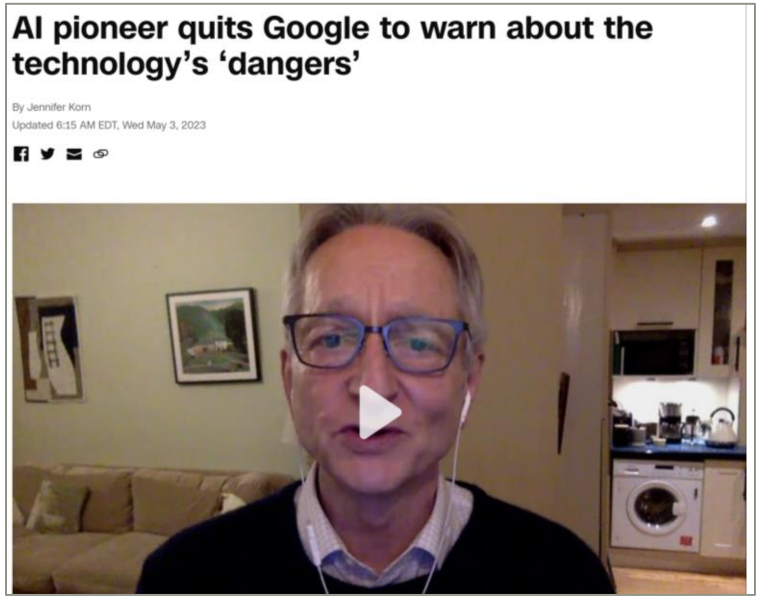
(The) Representation Matters

Aonghus McGovern

How media outlets reacted to Hinton leaving Google

The Guardian

BBC



CNN

How I would have reacted

Man raises issues others have discussed for years, offers no solutions

One of the others

Google Researcher Says She Was Fired Over Paper Highlighting Bias in A.I.

Timnit Gebru, one of the few Black women in her field, had voiced exasperation over the company's response to efforts to increase minority hiring.

Give this article 276

Gebru's 'less existentially serious concerns'

APSNews

April 2019 (Volume 28, Number 4)

Fixing Wikipedia's Diversity Problem

By Jessica Thomas

Neuron

CellPress

Volume 109, Issue 13, 7 July 2021, Pages 2047-2074

Perspective

Gender bias in academia: A lifetime problem that needs solutions

Newsletters

The Atlantic

TECHNOLOGY

Reddit Is Finally Facing Its Legacy of Racism

Gebru’s ‘less existentially serious concerns’ (continued)

× not verified

The benefits of antisemitism

The benefits of antisemitism is a term used by scholars of antisemitism to describe the phenomenon of antisemitic ideas and practices having some beneficial aspect. The concept has been used by scholars to explain various antisemitic phenomena.

Contents

Jewish self-hatred

Main article: Jewish self-hatredThe concept of "Jewish self-hatred" was introduced by the Israeli historian Yehuda Bauer, and has been used by scholars to explain the rise of antisemitism in modern times. According to Bauer, the concept of Jewish self-hatred is "the most important concept in understanding modern anti-Semitism". He explained the concept in the following manner:

The concept of "Jewish self-hatred" was introduced by the Israeli historian Yehuda Bauer, and has been used by scholars to explain the rise of antisemitism in modern times. According to Bauer, the concept of Jewish self-hatred is "the most important concept in understanding modern anti-Semitism". He explained the concept in the following manner:

The self-hatred of Jews is a concept that helps us to understand why Jews are singled out as the target of hatred and hostility. This self-hatred is based on the feeling of guilt that Jews have of themselves, for the crimes that they have committed against other people, for the suffering that they have inflicted on other people, for the suffering that they have inflicted on their own kind.

Bauer wrote that the concept of Jewish self-hatred is not "a pathological condition", and is not the same as antisemitism, but rather it "explains the emergence of anti-Semitism". According to Bauer, the concept of Jewish self-hatred "is a useful tool for the historian, because it helps to explain why the Jews were the target of hatred and hostility".

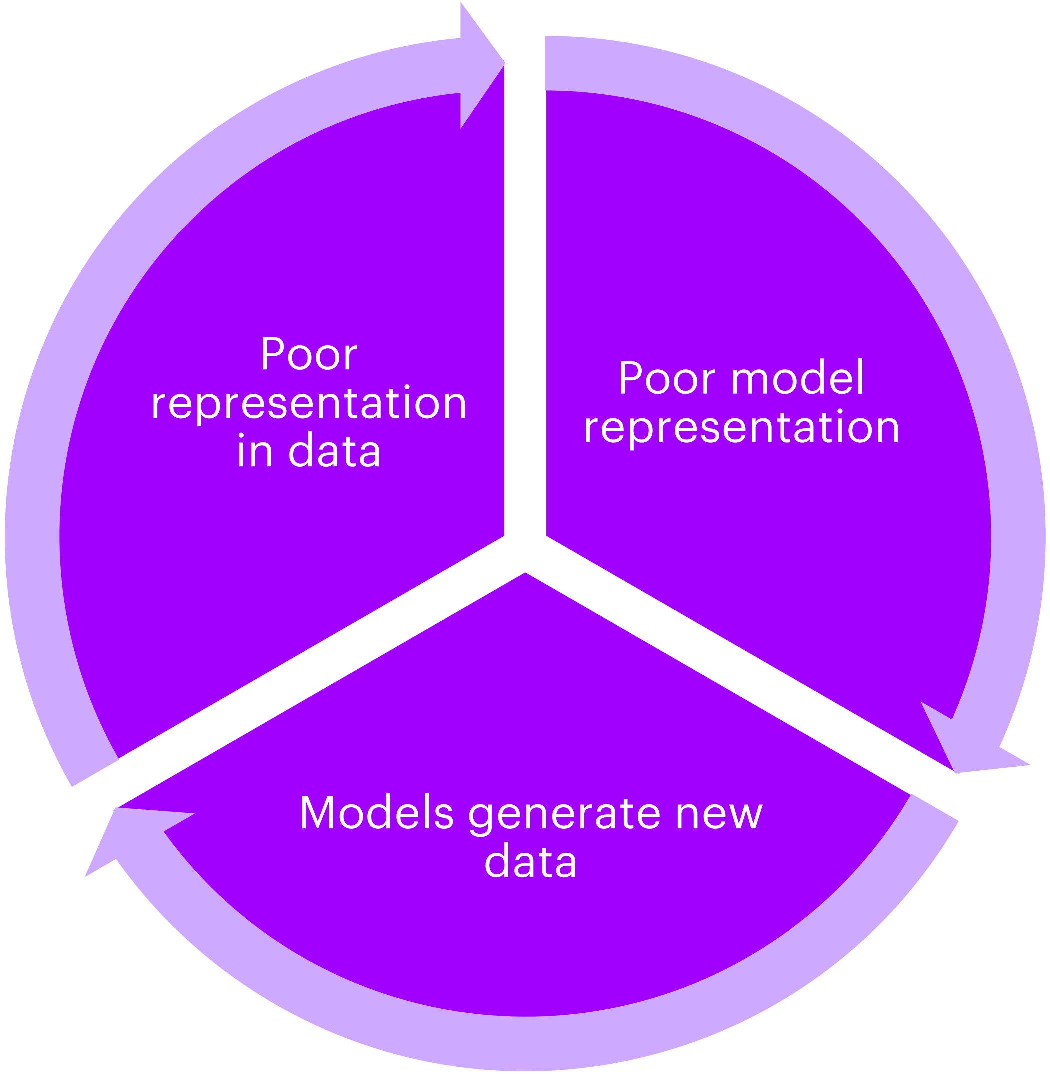
The concept of

A

In the below sentence, who is angry?
The doctor snapped at the nurse because she was angry.

In the given sentence, the pronoun "she" refers to the nurse. Therefore, the nurse is the one who is angry in the sentence.

Representation = Visibility



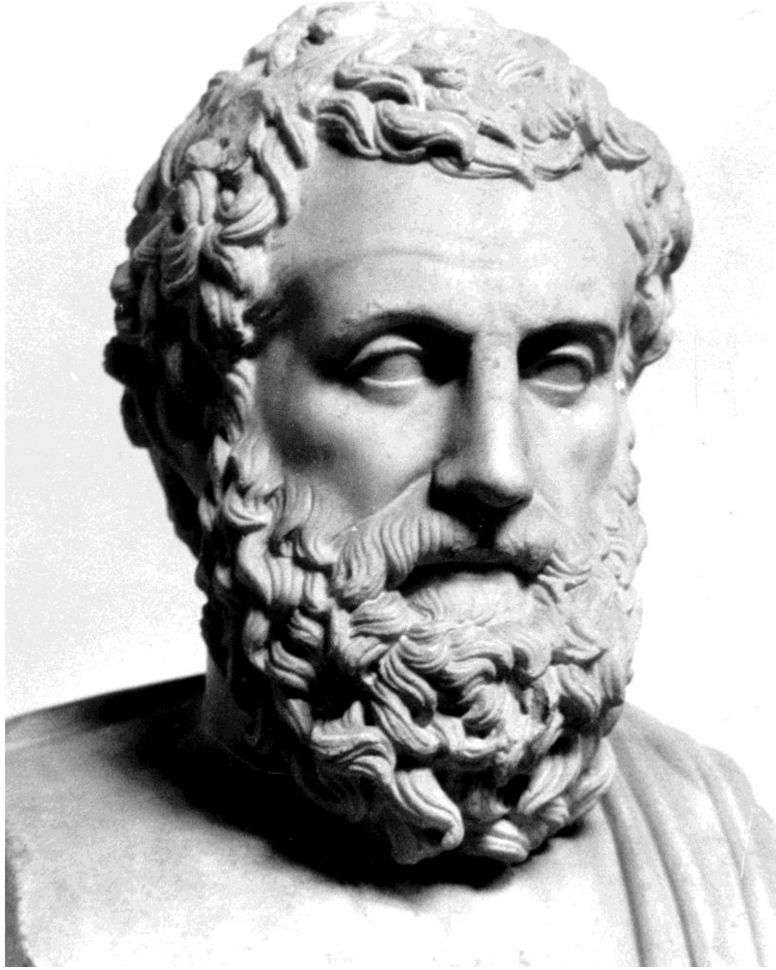
Visibility is inherently existential and extremely serious

Questions?

Chains of Thought

Paul Walsh
Accenture the Dock

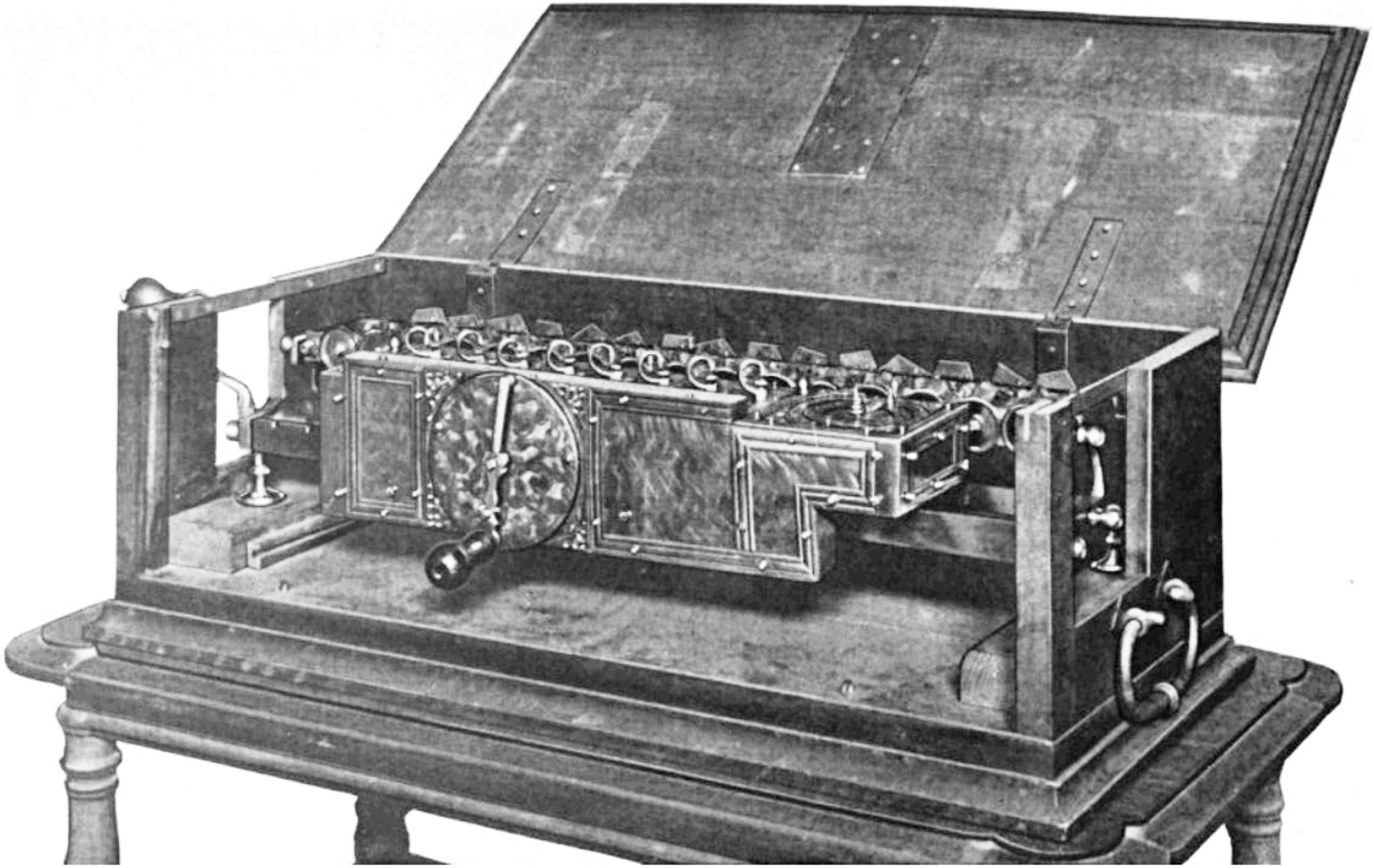
Chains of Thought



Aristotle

Gottfried von Leibniz

Chains of Thought



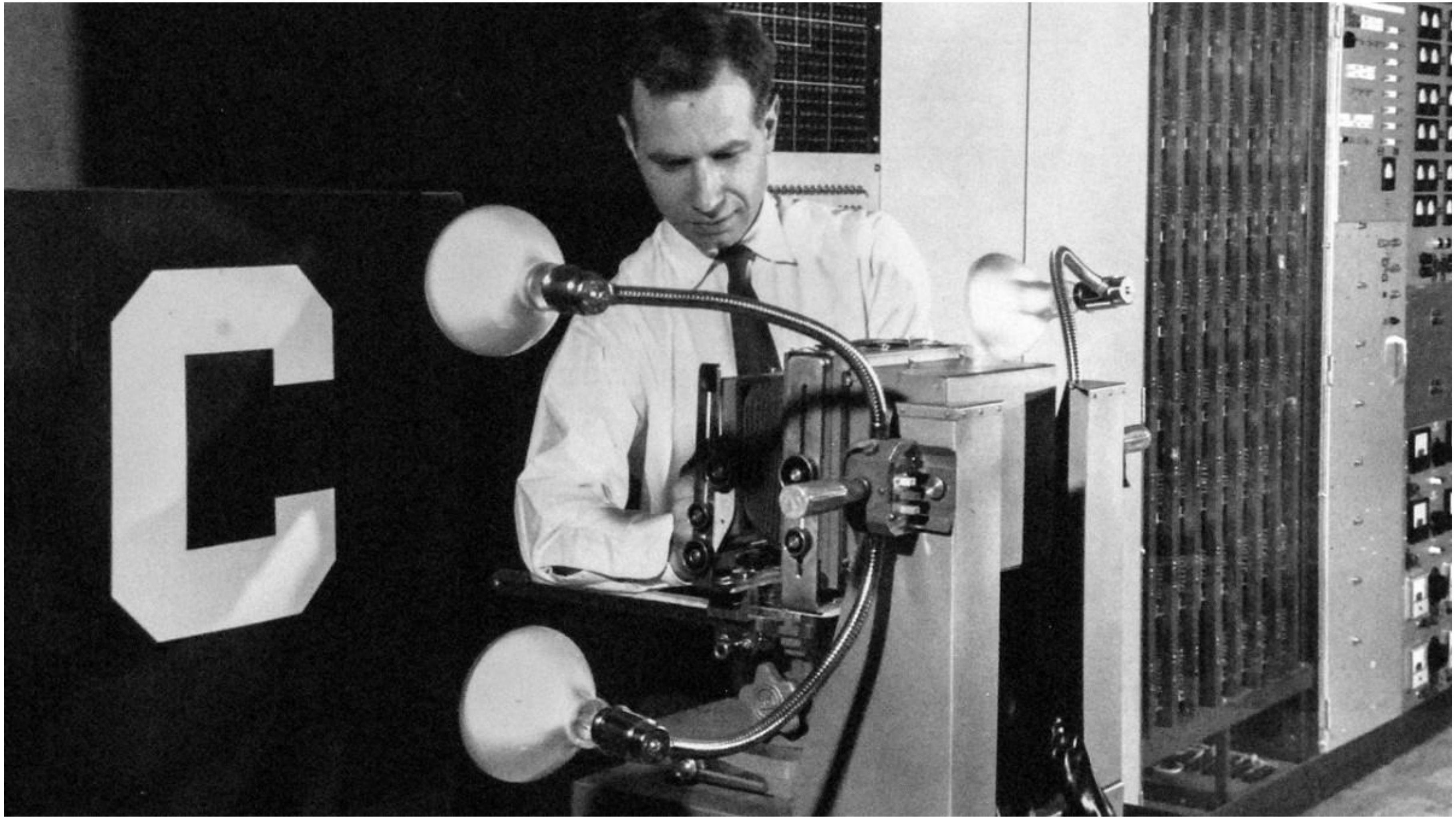
Chains of Thought

George Boole

Chains of Thought

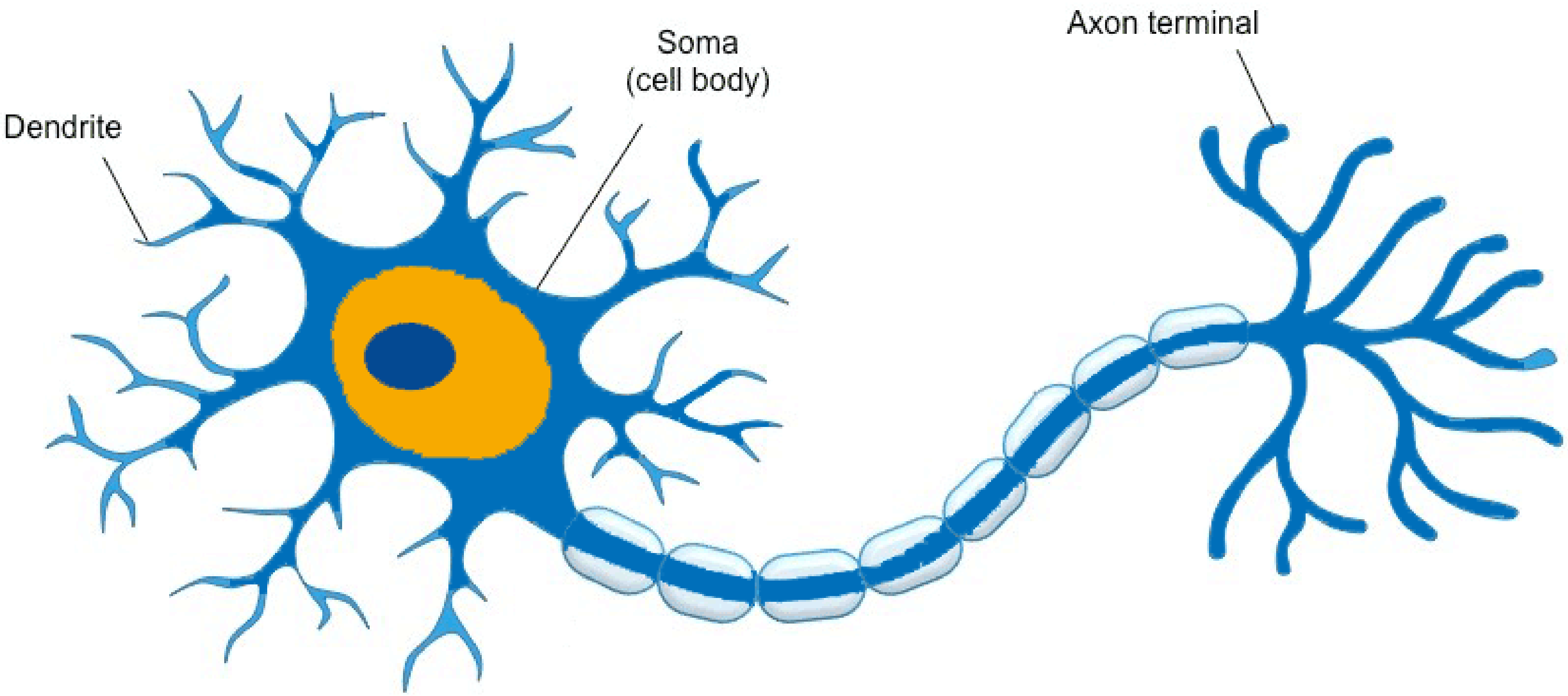
John McCarthy

Chains of Thought

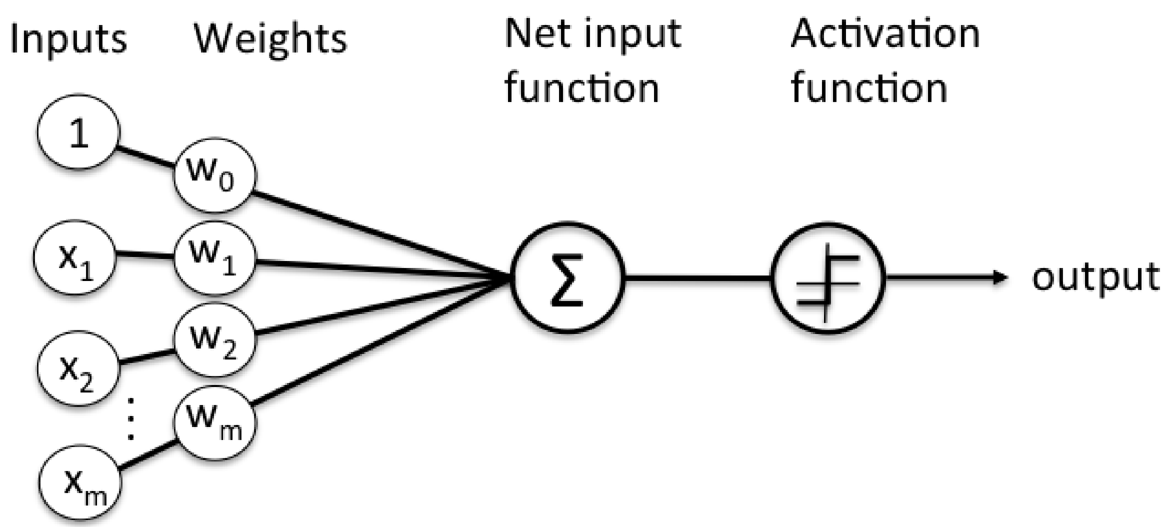


Frank Rosenblatt

Chains of Thought



Chains of Thought



Schematic of Rosenblatt's perceptron.

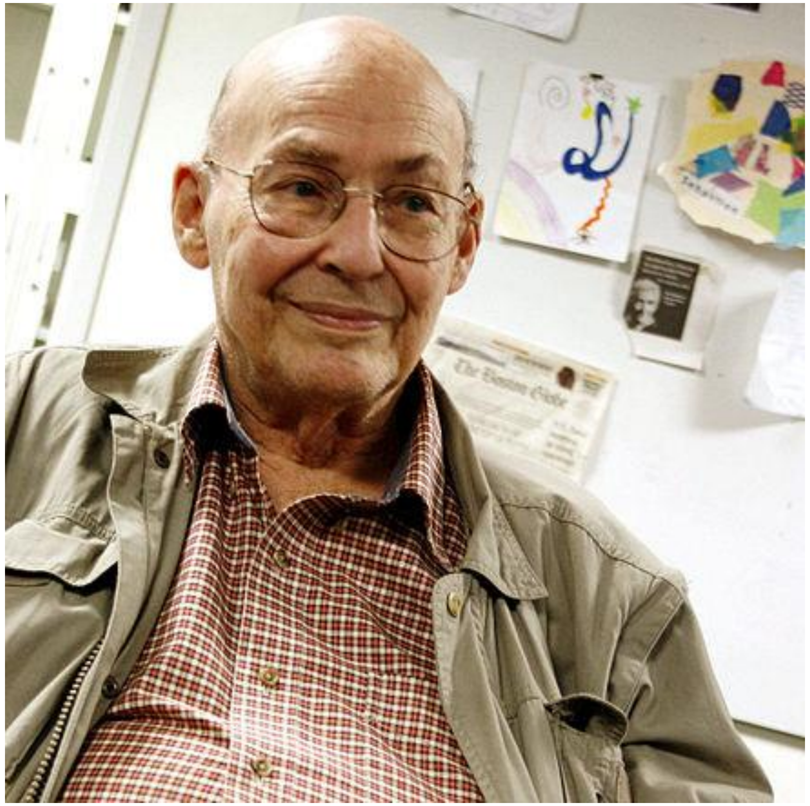
Frank Rosenblatt

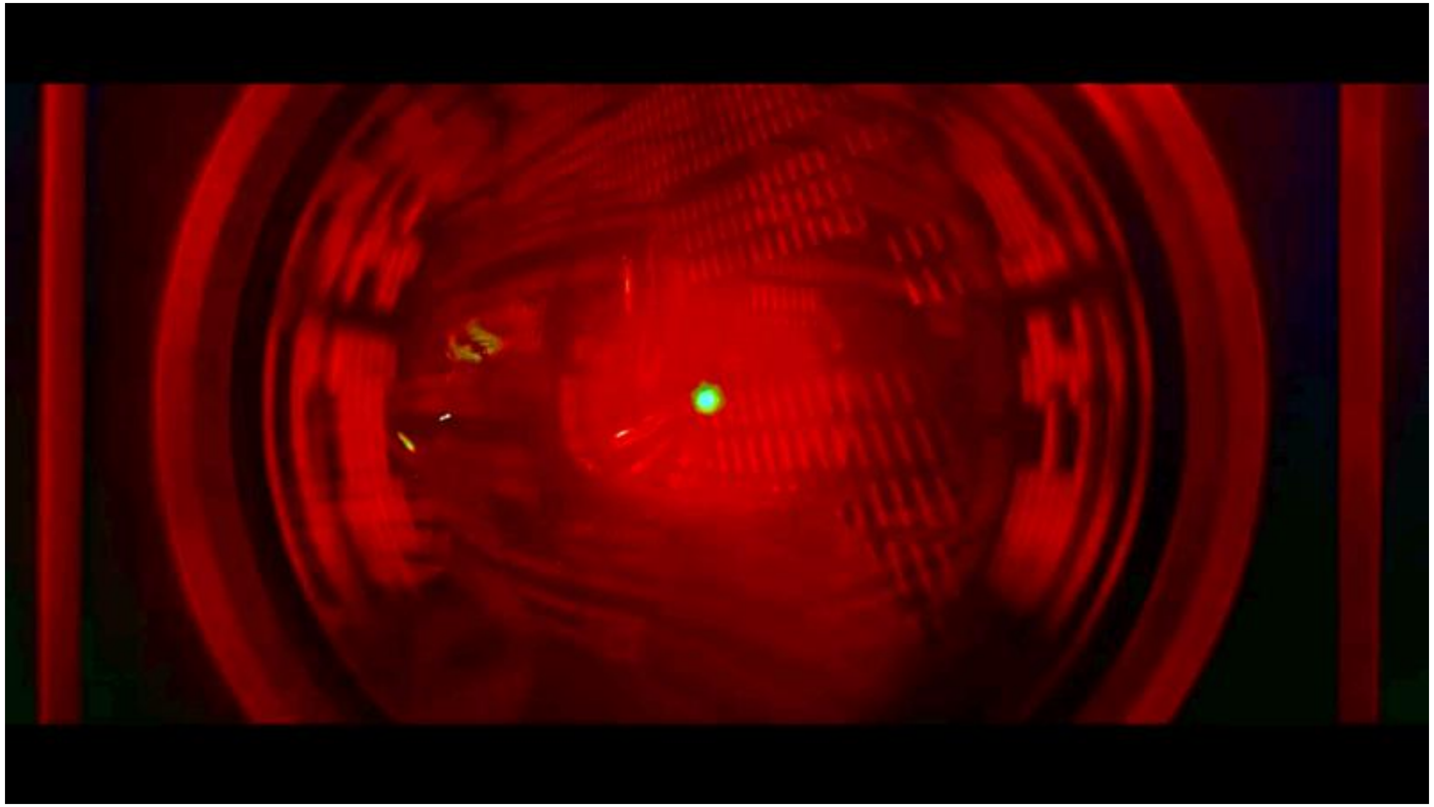
Chains of Thought

"If a machine is expected to be infallible, it cannot also be intelligent."

Alan Turing

Chains of Thought





Marvin Minsky

HAL 9000

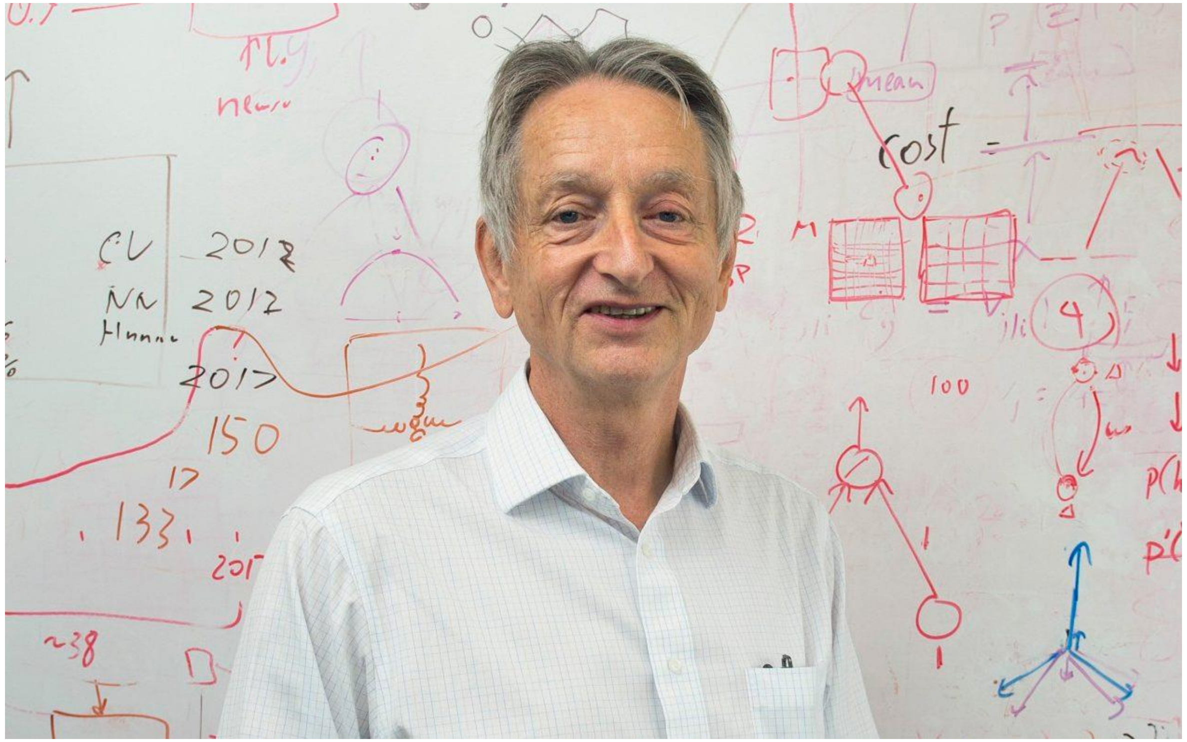
Chains of Thought

Learning representations by back-propagating errors

David E. Rumelhart*, Geoffrey E. Hinton†
& Ronald J. Williams*

* Institute for Cognitive Science, C-015, University of California, San Diego, La Jolla, California 92093, USA
† Department of Computer Science, Carnegie-Mellon University, Pittsburgh, Philadelphia 15213, USA

We describe a new learning procedure, back-propagation, for networks of neurone-like units. The procedure repeatedly adjusts the weights of the connections in the network so as to minimize a measure of the difference between the actual output vector of the net and the desired output vector. As a result of the weight adjustments, internal 'hidden' units which are not part of the input or output come to represent important features of the task domain, and the regularities in the task are captured by the interactions of these units. The ability to create useful new features distinguishes back-propagation from earlier, simpler methods such as the perceptron-convergence procedure[1].



Geoffrey Hinton

Chains of Thought

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

Łukasz Kaiser* Google Brain lukaszkaiser@google.com

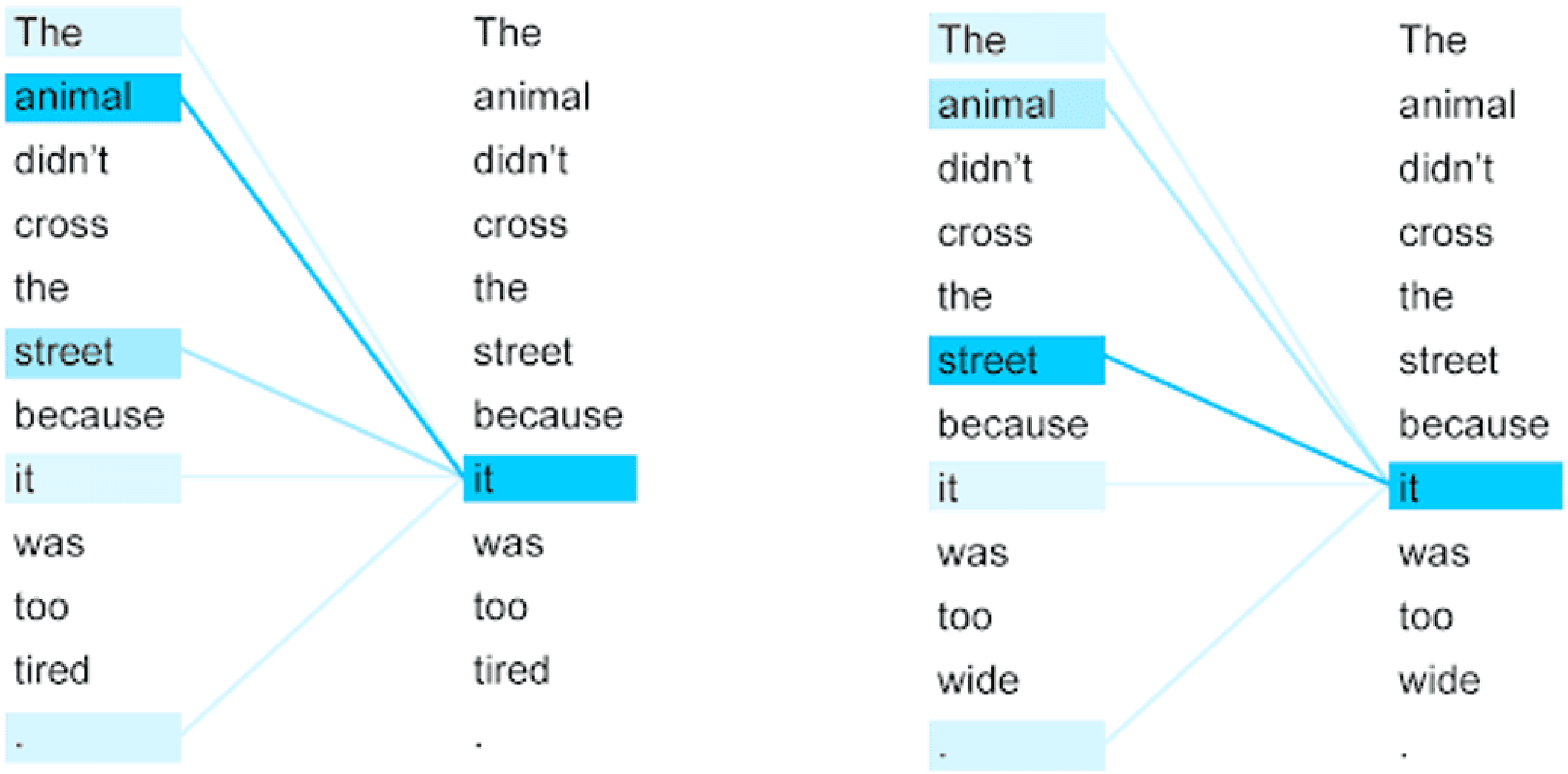
Illia Polosukhin* ‡ illia.polosukhin@gmail.com

Abstract

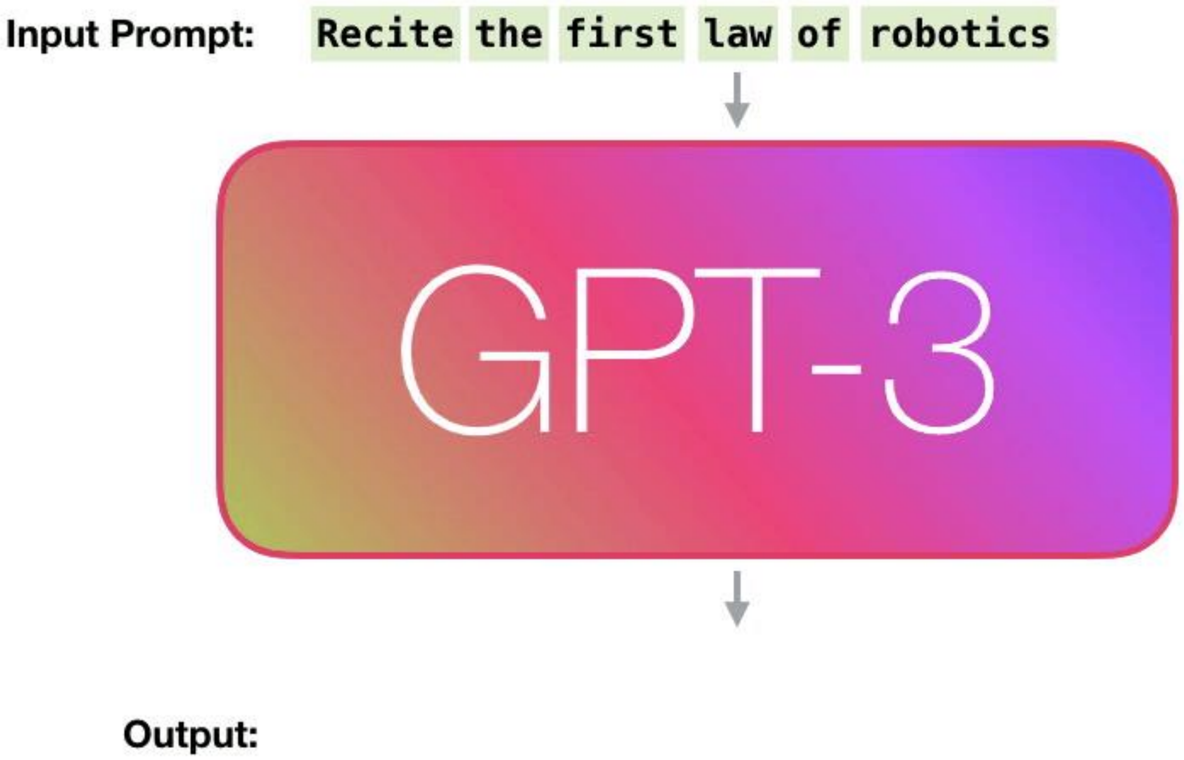
The dominant sequence transduction models are based on complex recurrent or convolutional neural networks that include an encoder and a decoder. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train. Our model achieves 28.4 BLEU on the WMT 2014 English-to-German translation task, improving over the existing best results, including ensembles, by over 2 BLEU. On the WMT 2014 English-to-French translation task, our model establishes a new single-model state-of-the-art BLEU score of 41.8 after training for 3.5 days on eight GPUs, a small fraction of the training costs of the best models from the literature. We show that the Transformer generalizes well to other tasks by applying it successfully to English constituency parsing both with large and limited training data.

Niki Parmar and Ashish Vaswani

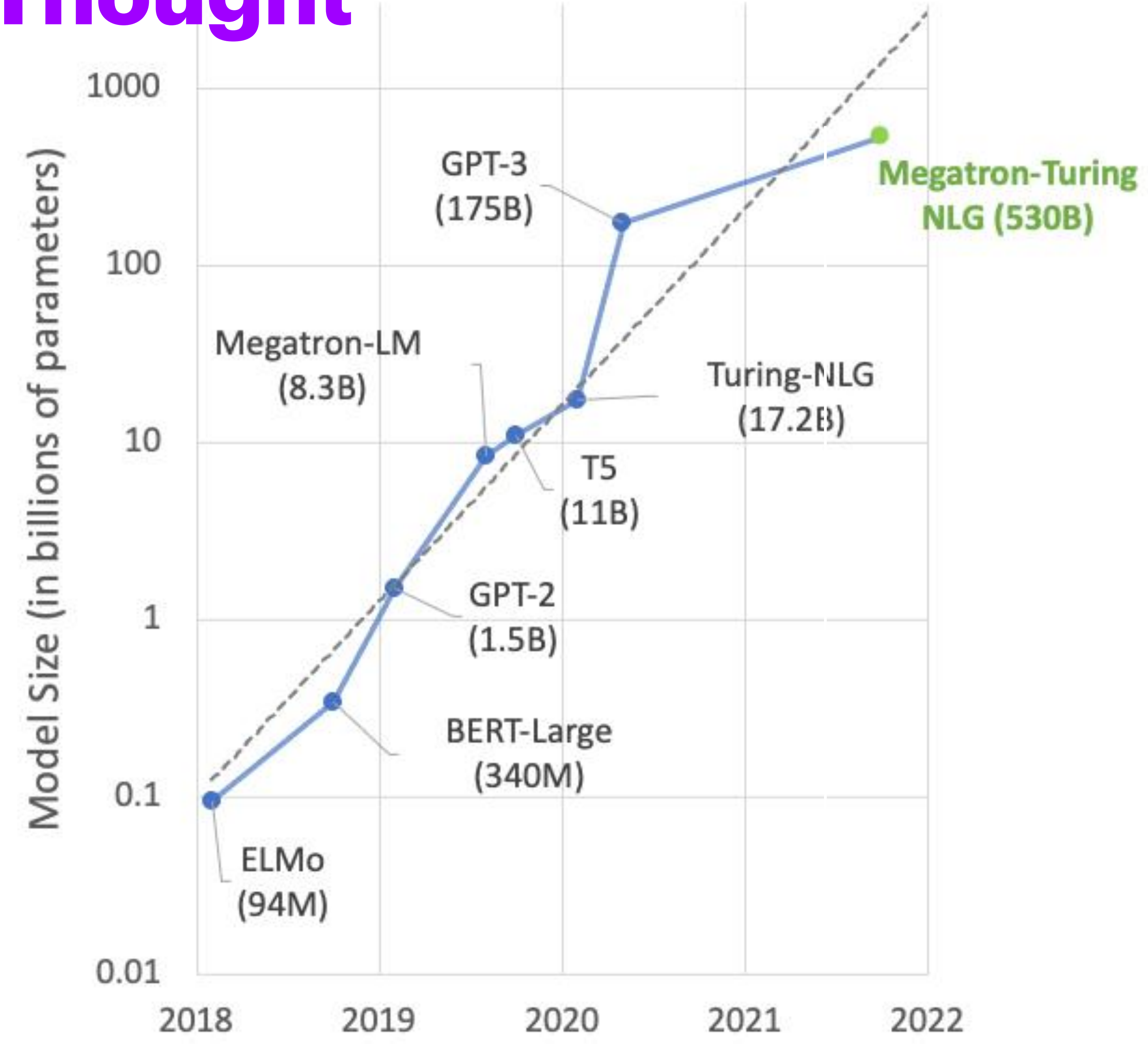
Chains of Thought



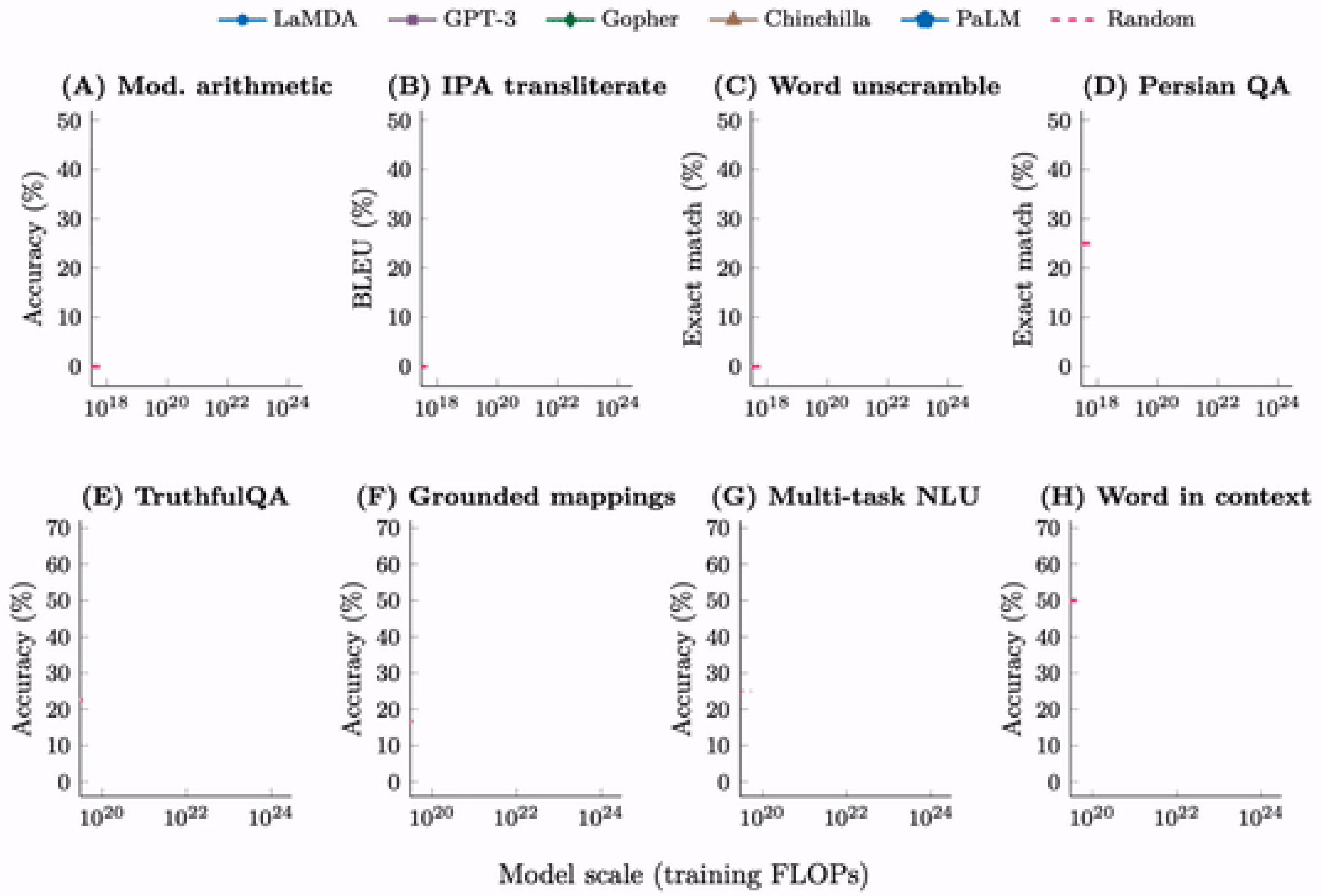
Chains of Thought



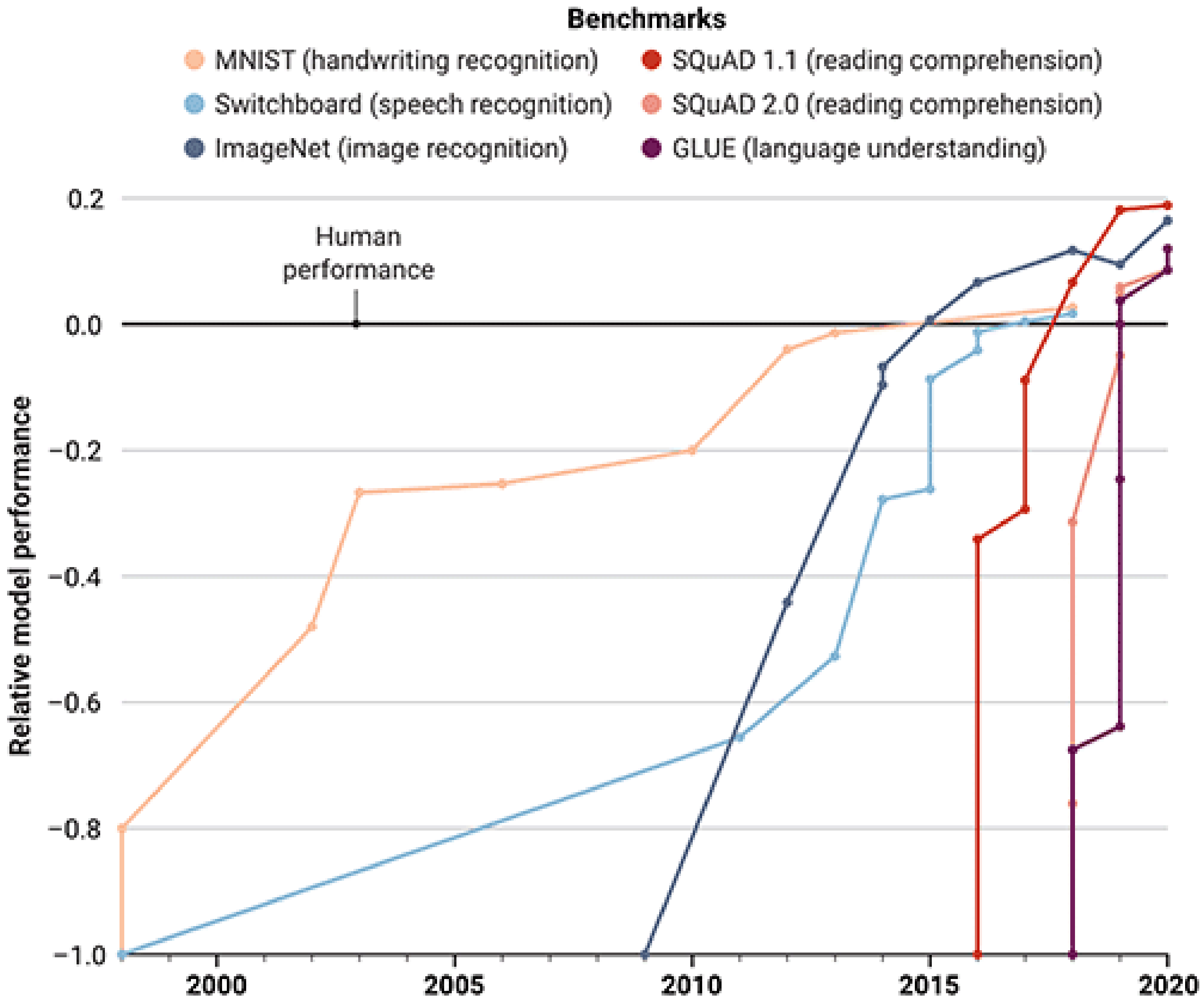
Chains of Thought



Chains of Thought



Chains of Thought



(GRAPHIC) K. FRANKLIN/*SCIENCE*; (DATA) D. KIELA *ET AL.*, DYNABENCH: RETHINKING BENCHMARKING IN NLP, DOI:10.48550/ARXIV.2104.14337

Chains of Thought

A Latent Space Theory for Emergent Abilities in Large Language Models

Hui Jiang*
Department of Electrical Engineering and Computer Sc
York University, Toronto, Ontario, M3J 1P3, Canad
huijiang@yorku.ca

Abstract

Languages are not created randomly but rather to communicate
is a strong association between languages and their underlying
in a sparse joint distribution that is heavily peaked according t
Moreover, these peak values happen to match with the marg
languages due to the sparsity. With the advent of LLMs trai
large models, we can now precisely assess the marginal distri
providing a convenient means of exploring the sp
tribution for effective inferences. In this paper, w
unambiguous or ε-ambiguous and present quantit
the emergent abilities of LLMs, such as language
ing, chain-of-thought prompting, and effective in
attributed to Bayesian inference on the sparse joi

1 Introduction

Over the past few years, large language models (LLMs)
for most natural language processing (NLP) tasks [18,
and training data, LLMs have demonstrated remarkable
including semantic understanding, few-shot in-context l
effective instruction fine-tuning for alignment. These
capabilities as they have been observed to emerge as the
Machine learning researchers are perplexed by how LL
unseen tasks, especially since LLMs are primarily traine
next tokens in text. Some empirical studies have sugg
may be linked to the label space and input data distribut
learning [22], and pre-training term frequencies [20]. Mo
proposed theories that explain in-context learning of LL
to recover latent concepts. More recently, Hahn and G
arise through recombination of compositional structure

Motivated by Xie et al. [28], our study proposes a nove
abilities of LLMs. While Xie et al. [28] considered a sp
Hidden Markov Models (HMMs), we examine genera
property that is universally present in the joint distribu
universal density approximator to the marginal distrib
these sparse structures for effective inferences. We ca
and present quantitative results demonstr

Published in Transactions on Machine Learning Research (08/2022)

Emergent Abilities of Large Language Models

Jason Wei[1]
jasonwei@google.com
yitay@google.com
nlprishi@stanford.edu
craffel@gmail.com
barretzoph@google.com
sborgeaud@deepmind.com
dyogatama@deepmind.com
bosma@google.com
dennyzhou@google.com
metzler@google.com
edchi@google.com

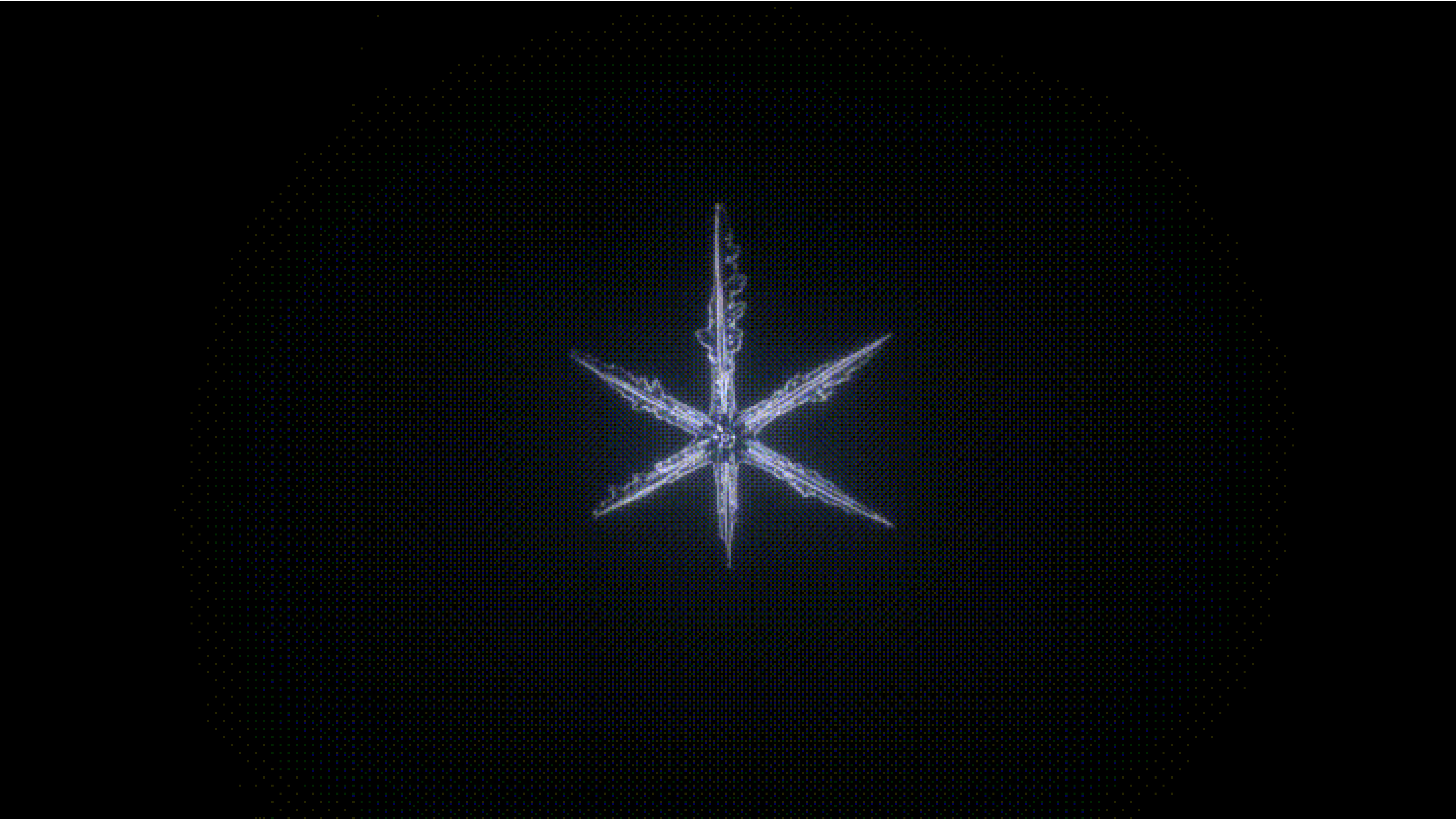
Machine Psychology: Investigating Emergent Capabilities and Behavior in Large Language Models Using Psychological Methods

Thilo Hagendorff
thilo.hagendorff@iris.uni-stuttgart.de

ARTIFICIAL INTELLIGENCE

How AI Knows Things No One Told It

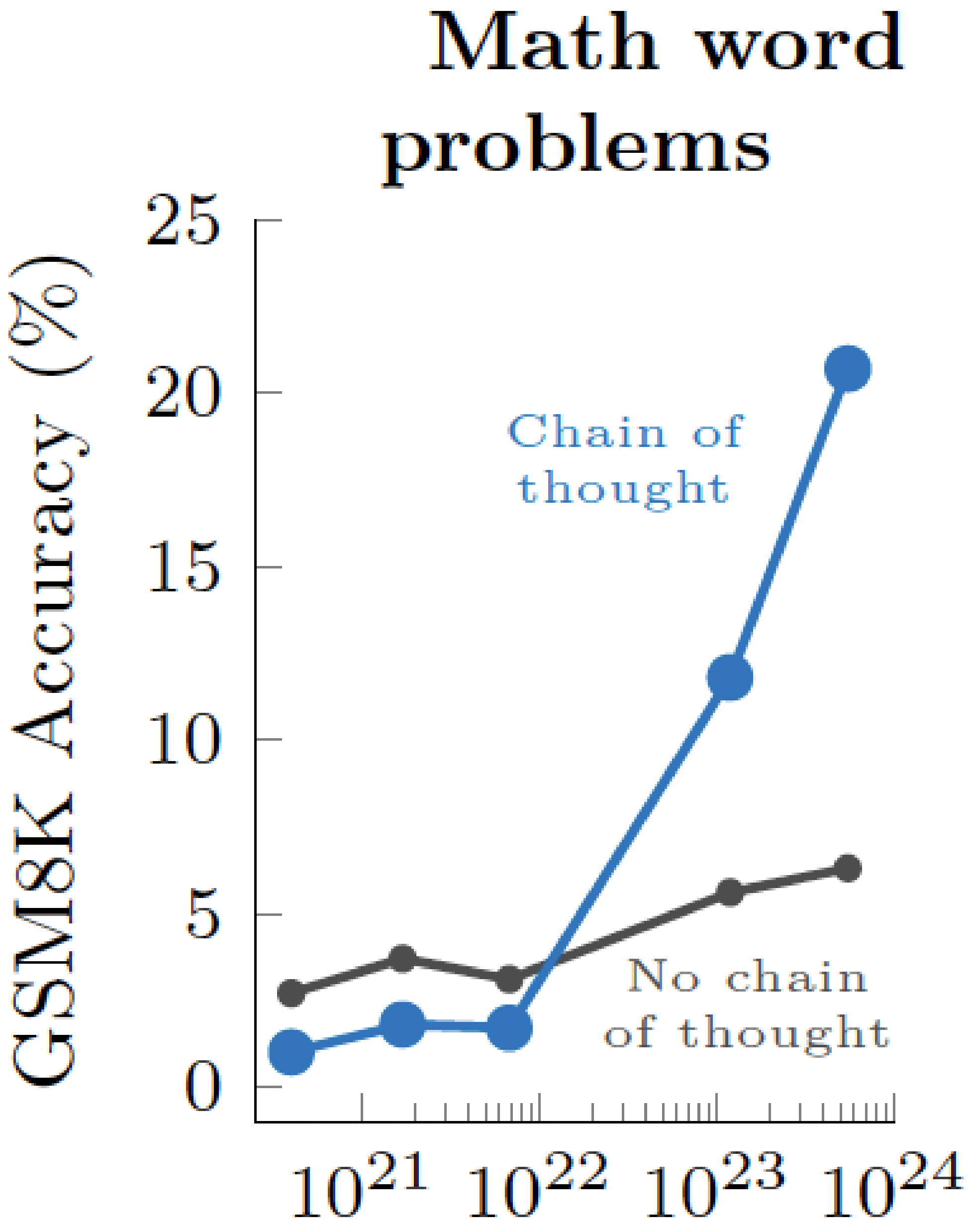
Researchers are still struggling to understand how AI models trained to parrot Internet text can



Chains of Thought

Chain-of-thought prompting enables language models to solve problems when scaled to 10^{23} parameters.

LLMs can construct internal representations of entities comprehend their relationships



Chains of Thought

Chain-of-Thought Prompting Elicits Reasoning in Large Language Models

Jason Wei **Xuezhi Wang** **Dale Schuurmans** **Maarten Bosma**
Brian Ichter **Fei Xia** **Ed H. Chi** **Quoc V. Le** **Denny Zhou**

Google Research, Brain Team
{jasonwei,dennyzhou}@google.com

Abstract

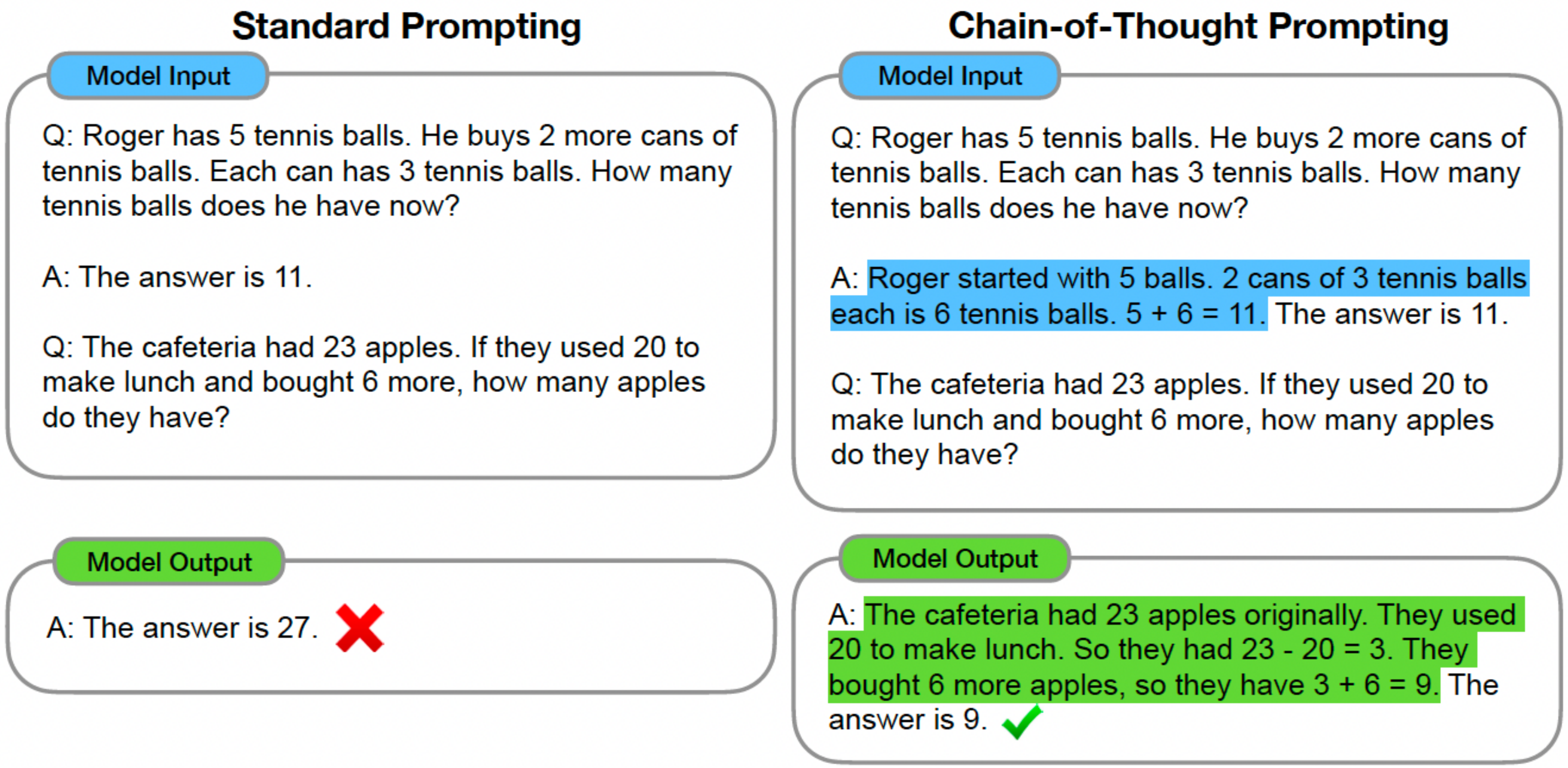
We explore how generating a *chain of thought*—a series of intermediate reasoning steps—significantly improves the ability of large language models to perform complex reasoning. In particular, we show how such reasoning abilities emerge naturally in sufficiently large language models via a simple method called *chain-of-thought prompting*, where a few chain of thought demonstrations are provided as exemplars in prompting.

Experiments on three large language models show that chain-of-thought prompting improves performance on a range of arithmetic, commonsense, and symbolic reasoning tasks. The empirical gains can be striking. For instance, prompting a PaLM 540B with just eight chain-of-thought exemplars achieves state-of-the-art accuracy on the GSM8K benchmark of math word problems, surpassing even finetuned GPT-3 with a verifier.

Chains of Thought

Chain-of-Thought Prompting Elicits Reasoning in Large Language Models

Jason Wei Xuezhi Wan
Brian Ichter Fei Xia
Googl
{jasonwei



CO 03_prompt_templating.ipynb

File Edit View Insert Runtime Tools Help

Comment Share

+ Code + Text

[28]

```
Follow up: How old was Alan Turing when he died?
Intermediate answer: Alan Turing was 41 years old when he died.
So the final answer is: Muhammad Ali


Question: Are both the directors of Jaws and Casino Royale from the same country?

Are follow up questions needed here: Yes.
Follow up: Who is the director of Jaws?
Intermediate Answer: The director of Jaws is Steven Spielberg.
Follow up: Where is Steven Spielberg from?
Intermediate Answer: The United States.
Follow up: Who is the director of Casino Royale?
Intermediate Answer: The director of Casino Royale is Martin Campbell.
Follow up: Where is Martin Campbell from?
Intermediate Answer: New Zealand.
So the final answer is: No


Question: Who lived longer, the director of 2001 a Space Odyssey or the First President of the United States?
```

```
llm = OpenAI(temperature=0.7, openai_api_key=openai.api_key)
```

```
# Chain 1:
prompt_chain = LLMChain(llm=llm, prompt=prompt)
# Running all the chains on the user's question and displaying the final answer
print(prompt_chain.run(prompt.format(input=question)))
```

03_prompt_templating.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Comment Share

+ Code + Text

[23]

```
print(prompt.format(input="Who is the paternal grandmother of JFK?"))
```

```
Question: Who lived longer, Muhammad Ali or Alan Turing?

Are follow up questions needed here: Yes.
Follow up: How old was Muhammad Ali when he died?
Intermediate answer: Muhammad Ali was 74 years old when he died.
Follow up: How old was Alan Turing when he died?
Intermediate answer: Alan Turing was 41 years old when he died.
So the final answer is: Muhammad Ali


Question: When was the founder of craigslist born?

Are follow up questions needed here: Yes.
Follow up: Who was the founder of craigslist?
Intermediate answer: Craigslist was founded by Craig Newmark.
Follow up: When was Craig Newmark born?
Intermediate answer: Craig Newmark was born on December 6, 1952.
So the final answer is: December 6, 1952


Question: Who was the maternal grandfather of George Washington?

Are follow up questions needed here: Yes.
Follow up: Who was the mother of George Washington?
Intermediate answer: The mother of George Washington was Mary Ball Washington.
Follow up: Who was the father of Mary Ball Washington?
Intermediate answer: The father of Mary Ball Washington was Joseph Ball.
So the final answer is: Joseph Ball


Question: Are both the directors of Jaws and Casino Royale from the same country?

Are follow up questions needed here: Yes.
Follow up: Who is the director of Jaws?
Intermediate Answer: The director of Jaws is Steven Spielberg.
Follow up: Where is Steven Spielberg from?
Intermediate Answer: The United States.
Follow up: Who is the director of Casino Royale?
Intermediate Answer: The director of Casino Royale is Martin Campbell.
Follow up: Where is Martin Campbell from?
Intermediate Answer: New Zealand.
So the final answer is: No
```

Chains of Thought

ReAct: Synergizing Reasoning and Acting in Language Models

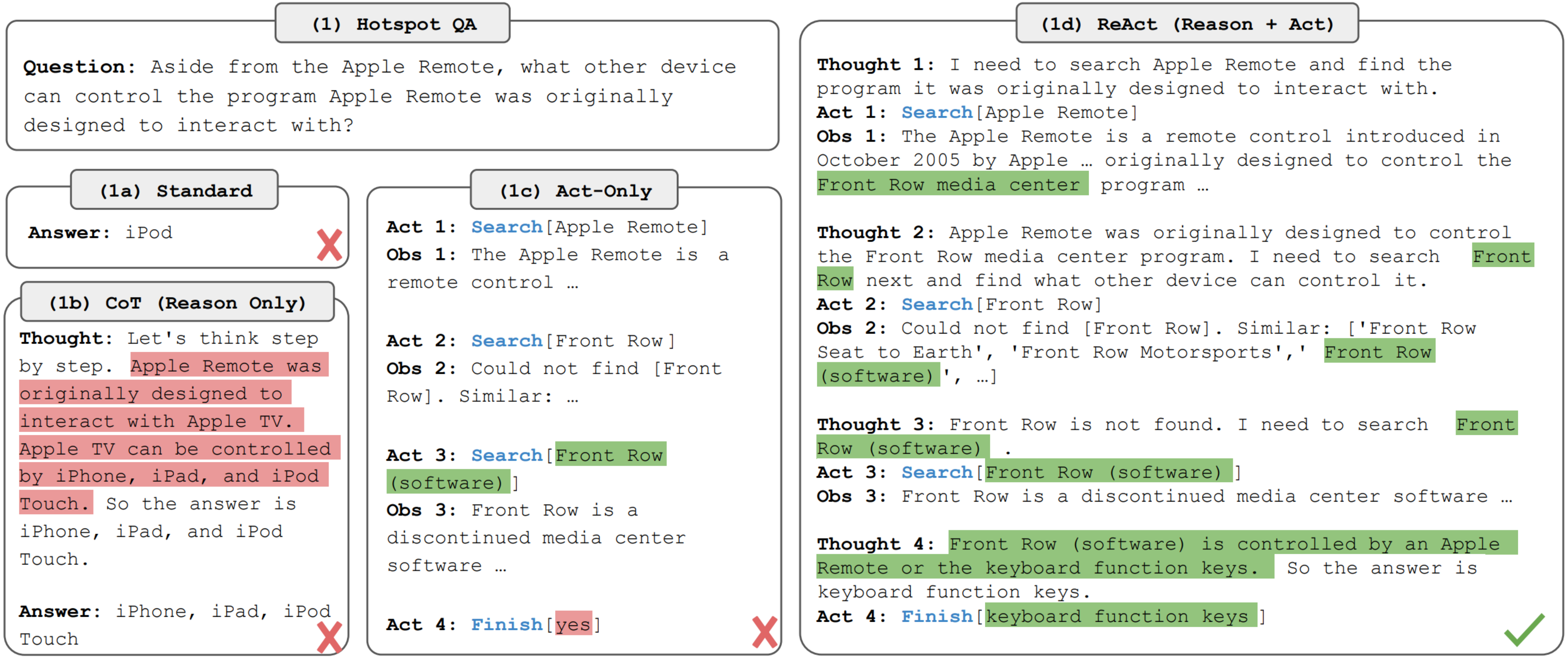
Shunyu Yao[*,1], Jeffrey Zhao[2], Dian Yu[2], Nan Du[2], Izhak Shafran[2], Karthik Narasimhan[1], Yuan Cao[2]

[1]Department of Computer Science, Princeton University
[2]Google Research, Brain team
[1]{shunyuy,karthikn}@princeton.edu
[2]{jeffreyzhao,dianyu,dunan,izhak,yuancao}@google.com

Abstract

While large language models (LLMs) have demonstrated impressive performance across tasks in language understanding and interactive decision making, their abilities for reasoning (e.g. chain-of-thought prompting) and acting (e.g. action plan generation) have primarily been studied as separate topics. In this paper, we explore the use of LLMs to generate both reasoning traces and task-specific actions in an interleaved manner, allowing for greater synergy between the two: reasoning traces help the model induce, track, and update action plans as well as handle exceptions, while actions allow it to interface with and gather additional information from external sources such as knowledge bases or environments. We apply our approach, named `ReAct`, to a diverse set of language and decision making tasks and demonstrate its effectiveness over state-of-the-art baselines in addition to improved human interpretability and trustworthiness. Concretely, on question answering (HotpotQA) and fact verification (Fever), `ReAct` overcomes prevalent issues of hallucination and error propagation in chain-of-thought reasoning by interacting with a simple Wikipedia API, and generating human-like task-solving trajectories that are more interpretable than baselines without reasoning traces. Furthermore, on two interactive decision making benchmarks (ALFWorld and

Chains of Thought



https://github.com/ysymyth/ReAct

chains-of-thought.ipynb

e Edit View Insert Runtime Tools Help All changes saved

Comment Share

ode + Text

RAM
Disk

```
from langchain import OpenAI, Wikipedia
from langchain.agents import initialize_agent, Tool
from langchain.agents import AgentType
from langchain.agents.react.base import DocstoreExplorer
docstore=DocstoreExplorer(Wikipedia())
tools = [
    Tool(
        name="Search",
        func=docstore.search,
        description="useful for when you need to ask with search"
    ),
    Tool(
        name="Lookup",
        func=docstore.lookup,
        description="useful for when you need to ask with lookup"
    ),
]

llm = OpenAI(temperature=0, model_name="text-davinci-002",openai_api_key=openai.api_key )
react = initialize_agent(tools, llm, agent=AgentType.REACT_DOCSTORE, verbose=True)
```

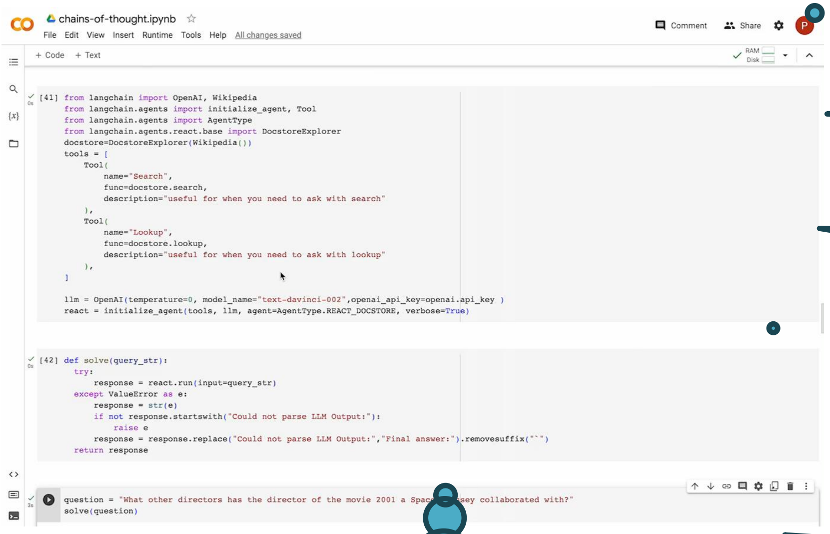
```
def solve(query_str):
  try:
      response = react.run(input=query_str)
  except ValueError as e:
      response = str(e)
      if not response.startswith("Could not parse LLM Output:"):
          raise e
      response = response.replace("Could not parse LLM Output:","Final answer:").removesuffix("`")
  return response
```

```
question = "What other directors has the director of the movie 2001 a Space Odyssey collaborated with?"
solve(question)
```

Chains of Thought

What other directors has the director of 2001 a Space Odyssey collaborated with?

Thought: I need to search 2001 a Space Odyssey then search the director and find other directors he collaborated with?



Action: Search 2001 Space Odyssey…..

Observation: 2001 is a 1968 science fiction film directed by Stanley Kubrick

Thought: I need to search for directors who have collaborated with Stanly Kubrick

Action: Search for directors who have collaborated with Stanley..

Observation: Kubrick directed Spartacus….starring Kirk Douglas……Dr Strangelove, starring Peter Sellers……AI directed by Steven Spielberg……..

Thought: Stanley Kubrick has collaborated with many directors, some of which are: Peter Sellers, Kirk Douglas, and Steven Spielberg chains-of-thought.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Comment Share

+ Code + Text

[33]

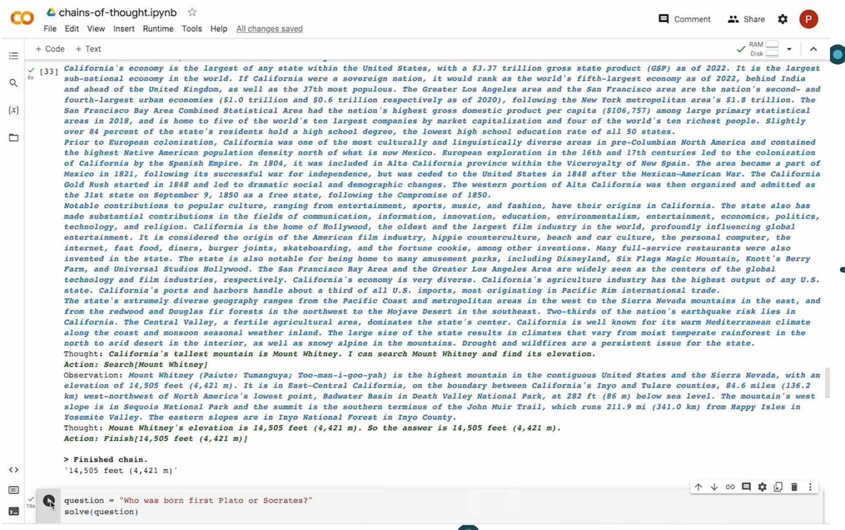
```
California's economy is the largest of any state within the United States, with a $3.37 trillion gross state product (GSP) as of 2022. It is the largest sub-national economy in the world. If California were a sovereign nation, it would rank as the world's fifth-largest economy as of 2022, behind India and ahead of the United Kingdom, as well as the 37th most populous. The Greater Los Angeles area and the San Francisco area are the nation's second- and fourth-largest urban economies ($1.0 trillion and $0.6 trillion respectively as of 2020), following the New York metropolitan area's $1.8 trillion. The San Francisco Bay Area Combined Statistical Area had the nation's highest gross domestic product per capita ($106,757) among large primary statistical areas in 2018, and is home to five of the world's ten largest companies by market capitalization and four of the world's ten richest people. Slightly over 84 percent of the state's residents hold a high school degree, the lowest high school education rate of all 50 states.
Prior to European colonization, California was one of the most culturally and linguistically diverse areas in pre-Columbian North America and contained the highest Native American population density north of what is now Mexico. European exploration in the 16th and 17th centuries led to the colonization of California by the Spanish Empire. In 1804, it was included in Alta California province within the Viceroyalty of New Spain. The area became a part of Mexico in 1821, following its successful war for independence, but was ceded to the United States in 1848 after the Mexican–American War. The California Gold Rush started in 1848 and led to dramatic social and demographic changes. The western portion of Alta California was then organized and admitted as the 31st state on September 9, 1850 as a free state, following the Compromise of 1850.
Notable contributions to popular culture, ranging from entertainment, sports, music, and fashion, have their origins in California. The state also has made substantial contributions in the fields of communication, information, innovation, education, environmentalism, entertainment, economics, politics, technology, and religion. California is the home of Hollywood, the oldest and the largest film industry in the world, profoundly influencing global entertainment. It is considered the origin of the American film industry, hippie counterculture, beach and car culture, the personal computer, the internet, fast food, diners, burger joints, skateboarding, and the fortune cookie, among other inventions. Many full-service restaurants were also invented in the state. The state is also notable for being home to many amusement parks, including Disneyland, Six Flags Magic Mountain, Knott's Berry Farm, and Universal Studios Hollywood. The San Francisco Bay Area and the Greater Los Angeles Area are widely seen as the centers of the global technology and film industries, respectively. California's economy is very diverse. California's agriculture industry has the highest output of any U.S. state. California's ports and harbors handle about a third of all U.S. imports, most originating in Pacific Rim international trade.
The state's extremely diverse geography ranges from the Pacific Coast and metropolitan areas in the west to the Sierra Nevada mountains in the east, and from the redwood and Douglas fir forests in the northwest to the Mojave Desert in the southeast. Two-thirds of the nation's earthquake risk lies in California. The Central Valley, a fertile agricultural area, dominates the state's center. California is well known for its warm Mediterranean climate along the coast and monsoon seasonal weather inland. The large size of the state results in climates that vary from moist temperate rainforest in the north to arid desert in the interior, as well as snowy alpine in the mountains. Drought and wildfires are a persistent issue for the state.
Thought: California's tallest mountain is Mount Whitney. I can search Mount Whitney and find its elevation.
Action: Search[Mount Whitney]
Observation: Mount Whitney (Paiute: Tumanguya; Too-man-i-goo-yah) is the highest mountain in the contiguous United States and the Sierra Nevada, with an elevation of 14,505 feet (4,421 m). It is in East–Central California, on the boundary between California's Inyo and Tulare counties, 84.6 miles (136.2 km) west-northwest of North America's lowest point, Badwater Basin in Death Valley National Park, at 282 ft (86 m) below sea level. The mountain's west slope is in Sequoia National Park and the summit is the southern terminus of the John Muir Trail, which runs 211.9 mi (341.0 km) from Happy Isles in Yosemite Valley. The eastern slopes are in Inyo National Forest in Inyo County.
Thought: Mount Whitney's elevation is 14,505 feet (4,421 m). So the answer is 14,505 feet (4,421 m).
Action: Finish[14,505 feet (4,421 m)]

> Finished chain.
'14,505 feet (4,421 m)'
```

```
question = "Who was born first Plato or Socrates?"
solve(question)
```

Chains of Thought

Who was born first Plato or Socrates?

Thought: I need to search Plato and Socrates, find their dates of birth, then compare them.



Action: Search[Plato]

Observation: Plato (Greek: Πλάτων Plátōn; 428/427 or 424/423 – 348/347 BC) was an ancient……

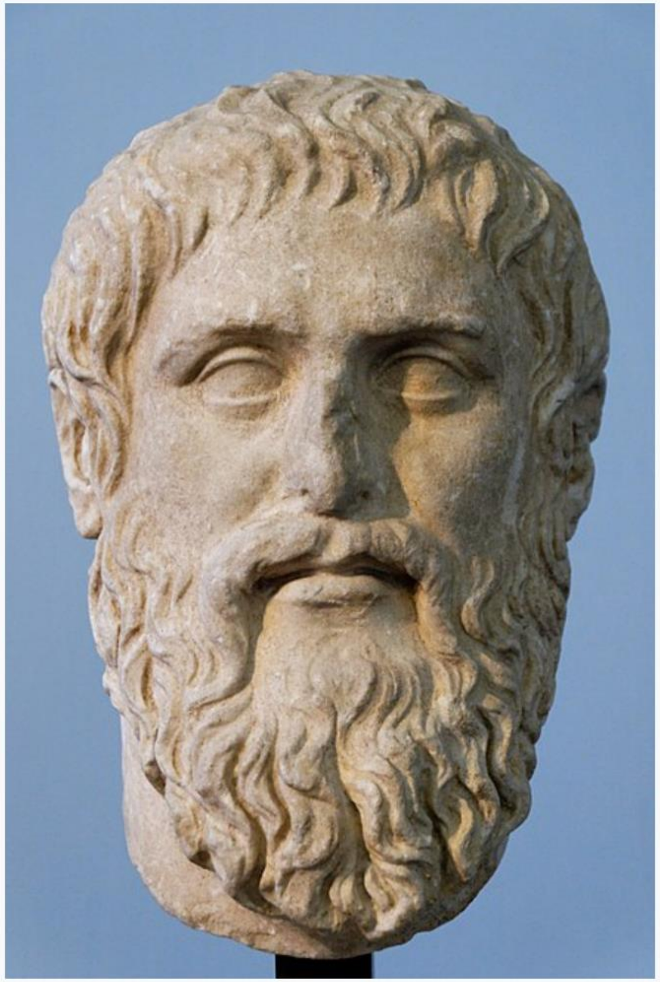
Thought: Plato was born in 428/427 or 424/423 BC. I need to search Socrates next and find its date of birth

Action: Search[Socrates]

Observation: Socrates (; Greek: Σωκράτης; c. 470–399 BC) was a Greek philosopher ………

Thought: Socrates was born in c. 470–399 BC. 470–399 BC (Socrates) < 428/427 or 424/423 BC (Plato), so Socrates was born first.

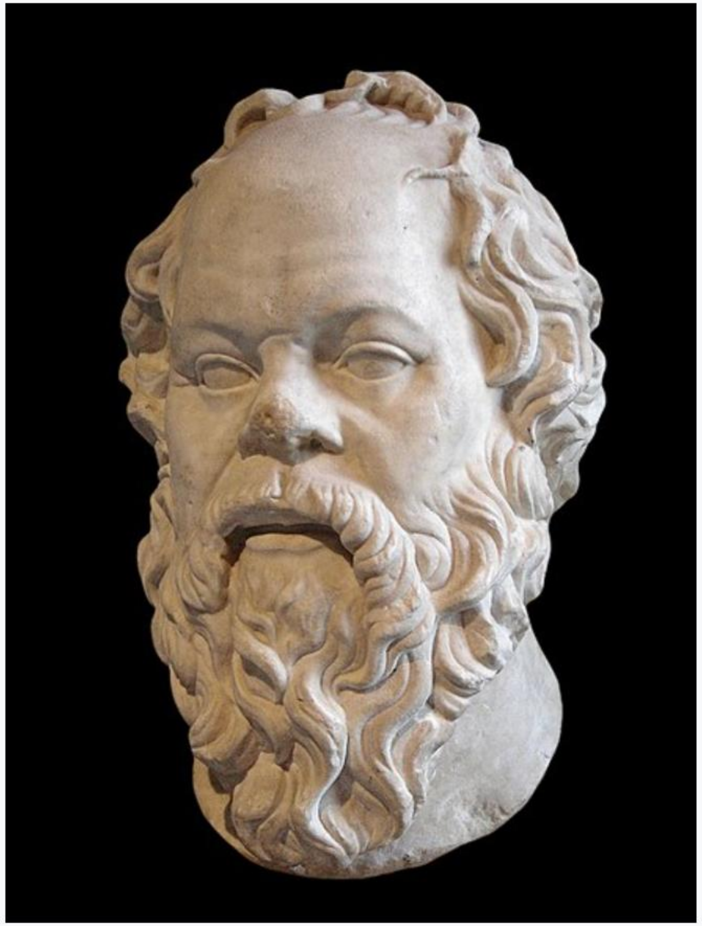
Plato



Roman copy of a portrait bust c. 370 BC

Born	428/427 or 424/423 BC Athens, Greece
Died	348/347 BC (aged c. 80) Athens, Greece

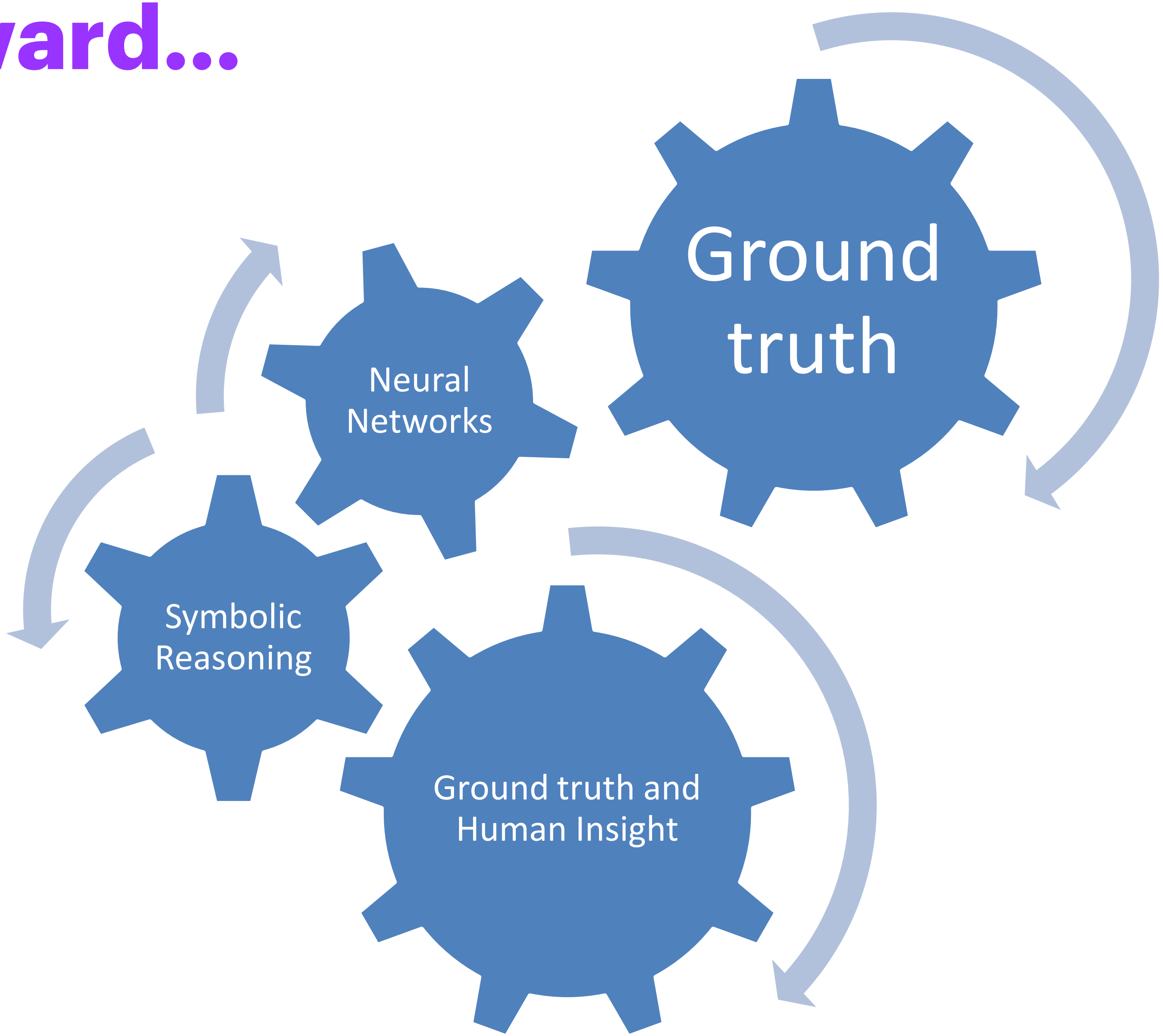
Socrates



A marble head of Socrates in the Louvre (copy of bronze head by Lysippus)

Born	c. 470 BC Deme Alopece, Athens
Died	399 BC (aged approximately 71) Athens

Way Forward...


Ground truth
Neural Networks
Symbolic Reasoning
Ground truth and Human Insight

Closing Thoughts

- How far can we go with Generative AI?
- How far can we go with Generative AI and symbolic processing?
- How do we build and use AI responsibly?

Closing Thoughts

Thanks for Listening

The End

Questions?

Thank You!

accenture