Scientific reasoning at the age of Large Language Models (LLMs)

André Freitas & Neuro-Symbolic AI Group

University of Lincoln (March 2024)

MANCHESTER 1824
The University of Manchester
idiap RESEARCH INSTITUTE
CANCER RESEARCH UK | National Biomarker Centre

MANCHESTER 1824

The University of Manchester

idiap RESEARCH INSTITUTE

CANCER RESEARCH UK | National Biomarker Centre

Neuro-symbolic AI Group

Today

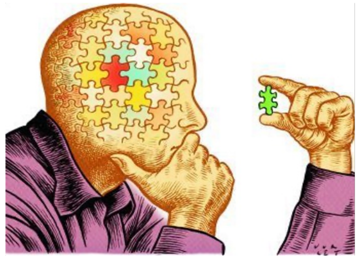
More applied

Large Language Models (LLMs) for supporting scientific discovery.

More foundational

Paradigms for controlling inference over Language Models.

Prototypical scientific workflow

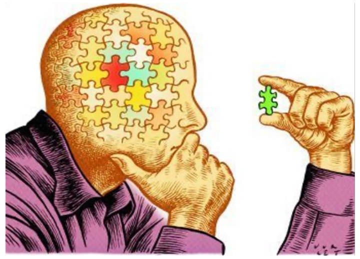


Hypotheses
Questions

New context

New data

Prototypical scientific workflow



Hypotheses
Questions

<u>New context</u>

<u>New data</u>

$$\frac{dx_1(t)}{dt} = x_2\,(t)$$

$$\frac{dx_2(t)}{dt} = ax_1\,(t) - bx_2(t)$$

$$\frac{d^2x_1(t)}{dt^2} = \frac{dx_2(t)}{dt}$$

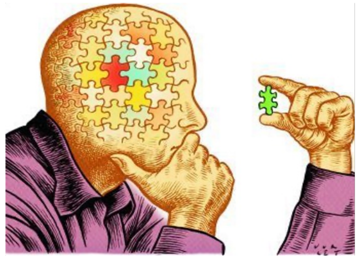
<u>Select</u>
relevant
background
knowledge


nature physics
ARTICLES
New CAST limit on the axion–photon interaction
CAST Collaboration


Article
miR-155 Activates Cytokine Gene Expression in Th17 Cells by Regulating the DNA-Binding Protein Jarid2 to Relieve Polycomb-Mediated Repression


Cancer Precision Medicine:
Physiological Function of C-MYC as Targeted Molecule

Prototypical scientific workflow



Hypotheses
Questions

<u>New context</u>

$$\frac{dx_1(t)}{dt} = x_2(t)$$

$$\frac{dx_2(t)}{dt} = ax_1(t) - bx_2(t)$$

$$\frac{d^2x_1(t)}{dt^2} = \frac{dx_2(t)}{dt}$$

<u>Select</u> relevant background knowledge

<u>Translate</u> to a computable expression

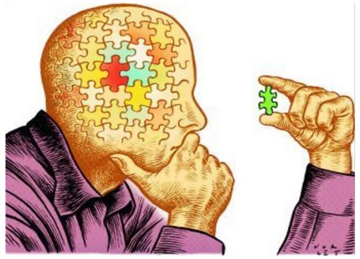
```
function y = simulate
CRS(x1, x2, t)
…
end
```

<u>Solve, Simulate</u>

<u>New data</u>

Data (phenomenal level)

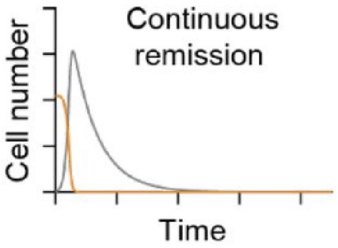
Prototypical scientific workflow



Hypotheses
Questions

New context

New data



Contrast
to new data

Elicit
relevant patterns

$$\frac{dx_1(t)}{dt} = x_2\,(t)$$

$$\frac{dx_2(t)}{dt} = ax_1\,(t) - bx_2(t)$$

$$\frac{d^2x_1(t)}{dt^2} = \frac{dx_2(t)}{dt}$$

Select
relevant
background
knowledge

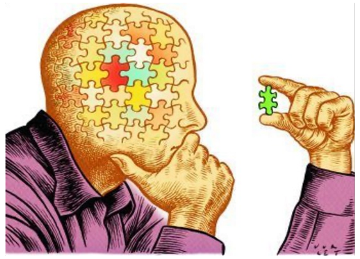
Translate to a
computable expression

```
function y = simulate
CRS(x1, x2, t)
…
end
```

Solve, Simulate

Data (phenomenal level)

Prototypical scientific workflow

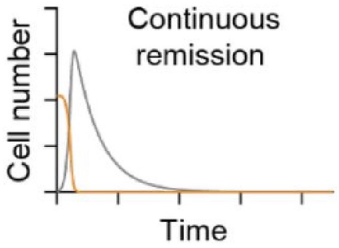


Hypotheses
Questions

New context

New data

Hypothesise
an explanation



Contrast
to new data

Elicit
relevant patterns

$$\frac{dx_1(t)}{dt} = x_2\,(t)$$

$$\frac{dx_2(t)}{dt} = ax_1\,(t) - bx_2(t)$$

$$\frac{d^2x_1(t)}{dt^2} = \frac{dx_2(t)}{dt}$$

Select
relevant background knowledge

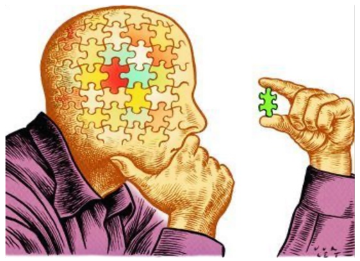
Translate to a computable expression

```
function y = simulate CRS(x1, x2, t)
…
end
```

Solve, Simulate

Data (phenomenal level)

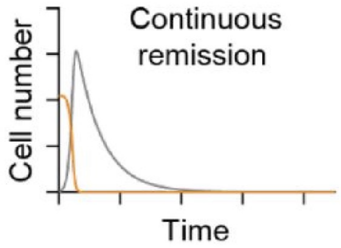
Prototypical scientific workflow



Hypotheses
Questions

New context

New data



Hypothesise
an explanation

Contrast
to new data

Elicit
relevant patterns

(Formally) Extend
existing model

$$\frac{dx_1(t)}{dt} = x_2\,(t)$$

$$\frac{dx_2(t)}{dt} = ax_1\,(t) - bx_2(t)$$

$$\frac{d^2x_1(t)}{dt^2} = \frac{dx_2(t)}{dt}$$

Select
relevant
background
knowledge

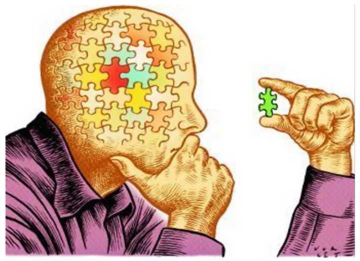
Translate to a
computable expression

```
function y = simulate
CRS(x1, x2, t)
…
end
```

Solve, Simulate

Data (phenomenal level)

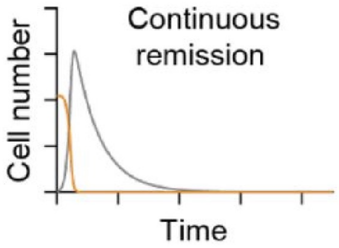
Prototypical scientific workflow



Hypotheses
Questions

New context

New data



Hypothesise
an explanation

Contrast
to new data

Elicit
relevant patterns

(Formally) Extend
existing model

$$\frac{dx_1(t)}{dt} = x_2(t)$$

$$\frac{dx_2(t)}{dt} = ax_1(t) - bx_2(t)$$

$$\frac{d^2x_1(t)}{dt^2} = \frac{dx_2(t)}{dt}$$

Corroborate
with previous
evidence

Select
relevant
background
knowledge

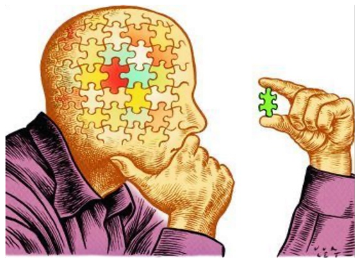
Translate to a
computable expression

```
function y = simulate
CRS(x1, x2, t)
…
end
```

Solve, Simulate

Data (phenomenal level)

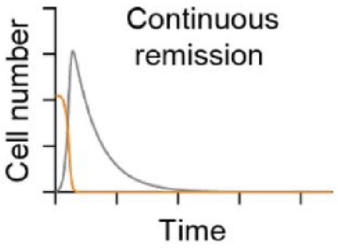
Prototypical scientific workflow



Hypotheses
Questions

New context

New data



Hypothesise
an explanation

Contrast
to new data

Elicit
relevant patterns

(Formally) Extend
existing model

$$\frac{dx_1(t)}{dt} = x_2(t)$$

$$\frac{dx_2(t)}{dt} = ax_1(t) - bx_2(t)$$

$$\frac{d^2x_1(t)}{dt^2} = \frac{dx_2(t)}{dt}$$

Translate to a
computable expression

```
function y = simulate
CRS(x1, x2, t)
…
end
```

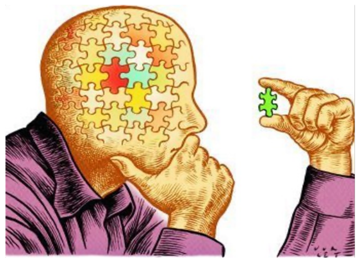
Solve, Simulate

Corroborate
with previous
evidence

Select
relevant
background
knowledge

Data (phenomenal level)

Automating scientific inference/discovery

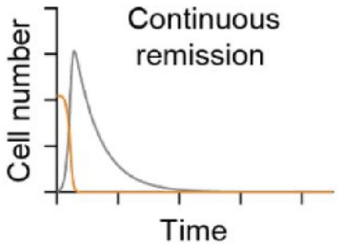


Hypotheses
Questions

New context

New data

<u>Hypothesise</u>
an explanation



<u>Contrast</u>
to new data

<u>Elicit</u>
relevant patterns

<u>(Formally) Extend</u>
existing model

$$\frac{dx_1(t)}{dt} = x_2(t)$$

$$\frac{dx_2(t)}{dt} = ax_1(t) - bx_2(t)$$

$$\frac{d^2x_1(t)}{dt^2} = \frac{dx_2(t)}{dt}$$

<u>Translate</u> to a
computable expression

```
function y = simulate
CRS(x1, x2, t)
…
end
```

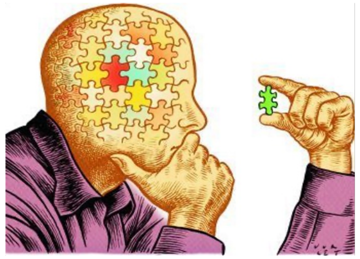
<u>Solve, Simulate</u>

Data (phenomenal level)

<u>Corroborate</u>
with previous
evidence

Abductive NLI
Premise selection
Automating meta-analysis

Automating scientific inference/discovery

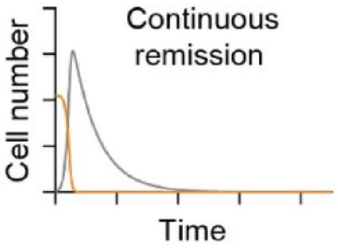


Hypotheses
Questions

New context

New data

Hypothesise
an explanation



Contrast
to new data

Elicit
relevant patterns

(Formally) Extend
existing model

$$\frac{dx_1(t)}{dt} = x_2(t)$$

$$\frac{dx_2(t)}{dt} = ax_1(t) - bx_2(t)$$

$$\frac{d^2x_1(t)}{dt^2} = \frac{dx_2(t)}{dt}$$

Corroborate
with previous
evidence

Abductive NLI
Premise selection
Automating meta-analysis

Auto-coding
Auto-formalisation

```
function y = simulate
CRS(x1, x2, t)
…
end
```

Solve, Simulate

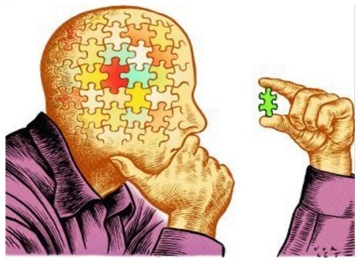
Data (phenomenal level)

Automating scientific inference/discovery

Hypothesise
an explanation

(Formally) Extend
existing model

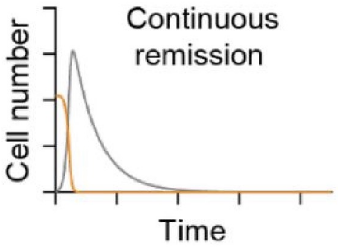
Corroborate
with previous evidence



Hypotheses
Questions

New context

New data



Auto-coding
Auto-formalisation
Abstraction models

$$\frac{dx_1(t)}{dt} = x_2(t)$$
$$\frac{dx_2(t)}{dt} = ax_1(t) - bx_2(t)$$
$$\frac{d^2x_1(t)}{dt^2} = \frac{dx_2(t)}{dt}$$

Abductive NLI
Premise selection
Automating meta-analysis

Auto-coding
Auto-formalisation

```
function y = simulate
CRS(x1, x2, t)
…
end
```

Solve, Simulate

Data (phenomenal level)

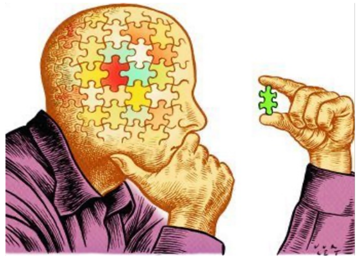
Automating scientific inference/discovery

Symbolic regression
Explanation generation

Hypothesise
an explanation

(Formally) Extend
existing model

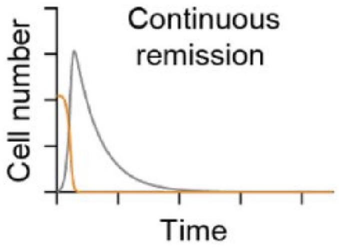
Corroborate
with previous evidence



Hypotheses
Questions

New context

New data



$$\frac{dx_1(t)}{dt} = x_2(t)$$

$$\frac{dx_2(t)}{dt} = a x_1(t) - b x_2(t)$$

$$\frac{d^2 x_1(t)}{dt^2} = \frac{dx_2(t)}{dt}$$

Abductive NLI
Premise selection
Automating meta-analysis

Auto-coding
Auto-formalisation
Abstraction models

Auto-coding
Auto-formalisation

```
function y = simulate
CRS(x1, x2, t)
…
end
```

Solve, Simulate

Data (phenomenal level)

Automating scientific inference/discovery

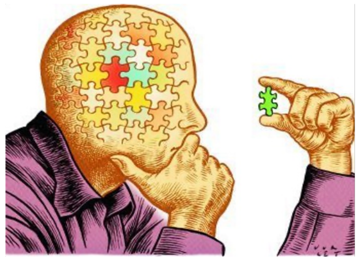
Symbolic regression
Explanation generation

Hypothesise an explanation

(Formally) Extend existing model

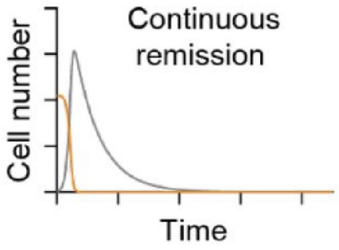
Abductive NLI
Premise selection

Corroborate with previous evidence



Hypotheses
Questions

New context



$$\frac{dx_1(t)}{dt} = x_2(t)$$
$$\frac{dx_2(t)}{dt} = ax_1(t) - bx_2(t)$$
$$\frac{d^2x_1(t)}{dt^2} = \frac{dx_2(t)}{dt}$$

Abductive NLI
Premise selection
Automating meta-analysis

Auto-coding
Auto-formalisation
Abstraction models

Auto-coding
Auto-formalisation

```
function y = simulate
CRS(x1, x2, t)
…
end
```

Solve, Simulate

New data

Data (phenomenal level)

Common denominator

“miR-155 Activates Cytokine Gene Expression in Th17 Cells by Regulating the DNA-Binding Protein Jarid2 to Relieve Polycomb-Mediated Repression.”

	Patients with SARS-Cov-2 confirmed by PCR	Patients without SARS-Cov-2 confirmed by PCR
Median age (IQR)—years	63 (53–72)	60 (49–73)
Male	787/1,309 (60.1%)	90/167 (53.9%)
Race/ethnicity—Hispanic	577/1,268 (45.5%)	62/167 (37.1%)
Race/ethnicity—African American	278/1,268 (21.9%)	46/167 (27.5%)
Race/ethnicity—White	277/1,268 (21.8%)	43/167 (25.7%)
Race/ethnicity—Asian	73/1,268 (5.8%)	5/167 (3.0%)
Race/ethnicity—Other	63/1,268 (5.0%)	11/167 (6.6%)
Obesity (BMI ≥30)	465/1,176 (39.5%)	34/149 (22.8%)[a]
Comorbidities—hypertension	420/1,268 (33.1%)	67/167 (40.1%)
Comorbidities—diabetes	293/1,268 (23.1%)	34/167 (20.4%)
Comorbidities—CKD	167/1,268 (13.2%)	27/167 (16.2%)
…	…	…

Del Valle et al. , *Nature Medicine* (2020)

$$\frac{dx_1(t)}{dt} = x_2\,(t)$$

$$\frac{dx_2(t)}{dt} = ax_1\,(t) - bx_2(t)$$

$$\frac{d^2x_1(t)}{dt^2} = \frac{dx_2(t)}{dt}$$

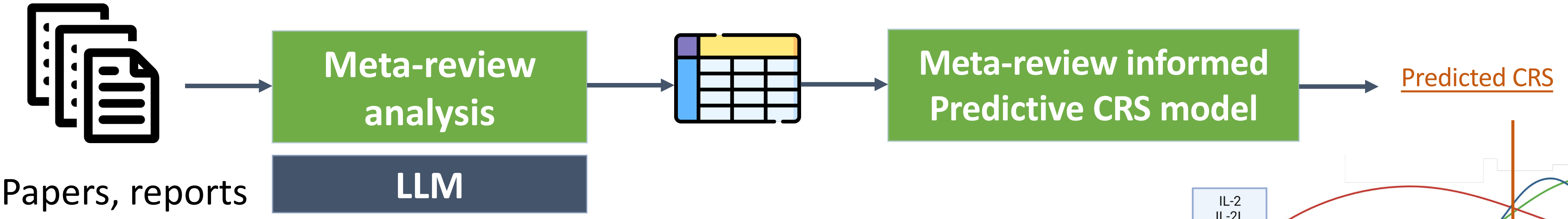
where $x_1(t)$ is the serum concentration of cytokine and its rate of change by $x_2(t)$

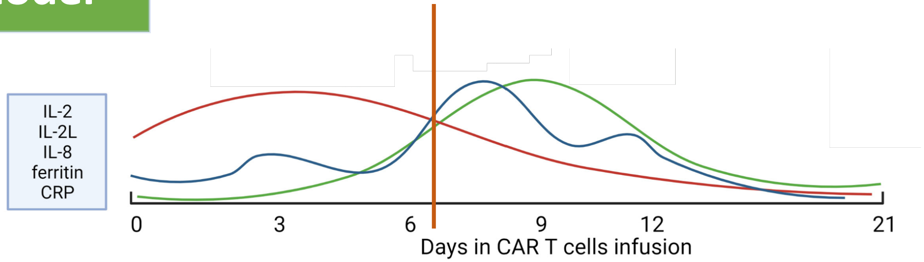
Common denominator: Language & Abstraction!

Evidence Selection & Automating Meta-analysis

Extracting evidence from the literature at scale

Predicting toxicity: Cytokine Release Syndrome (CRS) events for CAR-T cell therapies

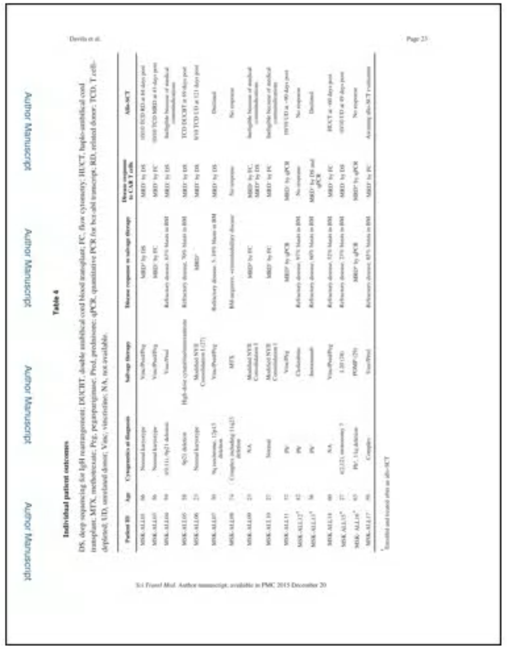




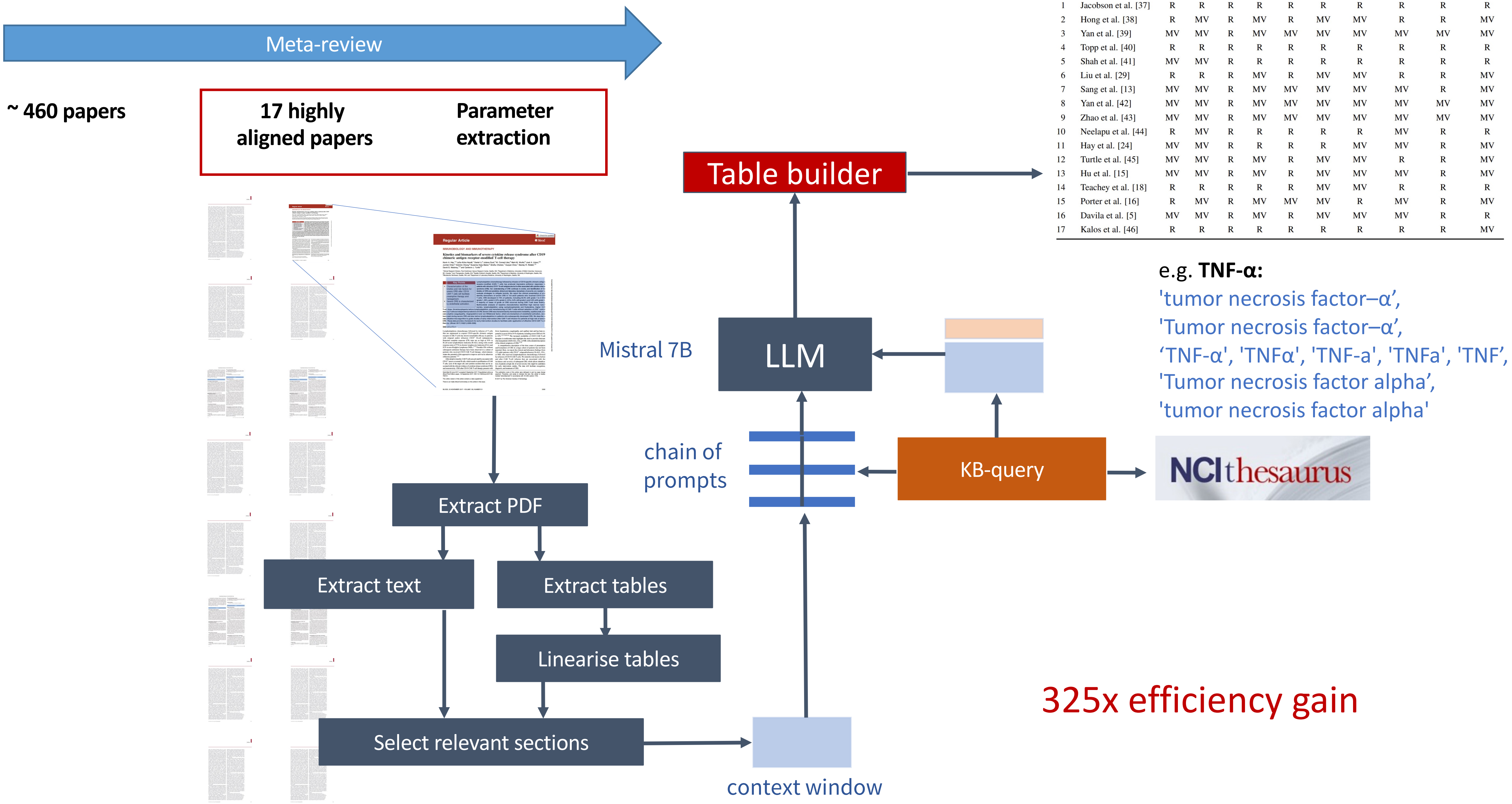
	Study	IL2	IL4	IL6	IL8	IL10	IL15	IL2Rα	TNF–α	IFN–γ	GM-CSF
1	Jacobson et al. [37]	R	R	R	R	R	R	R	R	R	R
2	Hong et al. [38]	R	MV	R	MV	R	MV	MV	R	R	MV
3	Yan et al. [39]	MV	MV	R	MV	MV	MV	MV	MV	MV	MV
4	Topp et al. [40]	R	R	R	R	R	R	R	R	R	R
5	Shah et al. [41]	MV	MV	R	R	R	R	R	R	R	R
6	Liu et al. [29]	R	R	R	MV	R	MV	MV	R	R	MV
7	Sang et al. [13]	MV	MV	R	MV	MV	MV	MV	MV	R	MV
8	Yan et al. [42]	MV	MV	R	MV	MV	MV	MV	MV	MV	MV
9	Zhao et al. [43]	MV	MV	R	MV	MV	MV	MV	MV	MV	MV
10	Neelapu et al. [44]	R	MV	R	R	R	R	R	MV	R	R
11	Hay et al. [24]	MV	MV	R	R	R	R	MV	MV	R	MV
12	Turtle et al. [45]	MV	MV	R	MV	R	MV	MV	R	R	MV
13	Hu et al. [15]	MV	MV	R	MV	R	MV	MV	MV	R	MV
14	Teachey et al. [18]	R	R	R	R	R	MV	MV	R	R	R
15	Porter et al. [16]	R	MV	R	MV	MV	MV	R	MV	R	MV
16	Davila et al. [5]	MV	MV	R	MV	R	MV	MV	MV	R	R
17	Kalos et al. [46]	R	R	R	R	R	R	R	R	R	MV

Bogatu et al. (JBI, 2023)

Davila Marcoa L., et al. Efficacy and toxicity management of 19-28z CAR T cell therapy in b cell acute lymphoblastic leukemia Sci. Transl. Med., 6 (224) (2014), 10.1126/scitranslmed.3008226 (ISSN: 1946-6234, 1946-6242)



Biomark	Davila e	Hay et a	Hong et	Hu et al	Jacobsc	Kalos et	Liu et al	Neelapu	Porter e	Sang et	Shah et	Teachey	Topp et	Turtle et	Yan et a	Yan et a	Zhao et
IL-2																	
IL-4																	
IL-6																	
IL-8																	
IL-10																	
IL-15																	
IL-2Rα																	
TNF-α																	
IFN-γ																	
GM-CS																	
CRP																	
ferritin																	
D-dime																	
VWF																	
Ang-2																	
MCP-1																	
granzy																	
TNFRp																	
MIP1β																	
MIP1α																	
IL-17																	
sIL6R																	



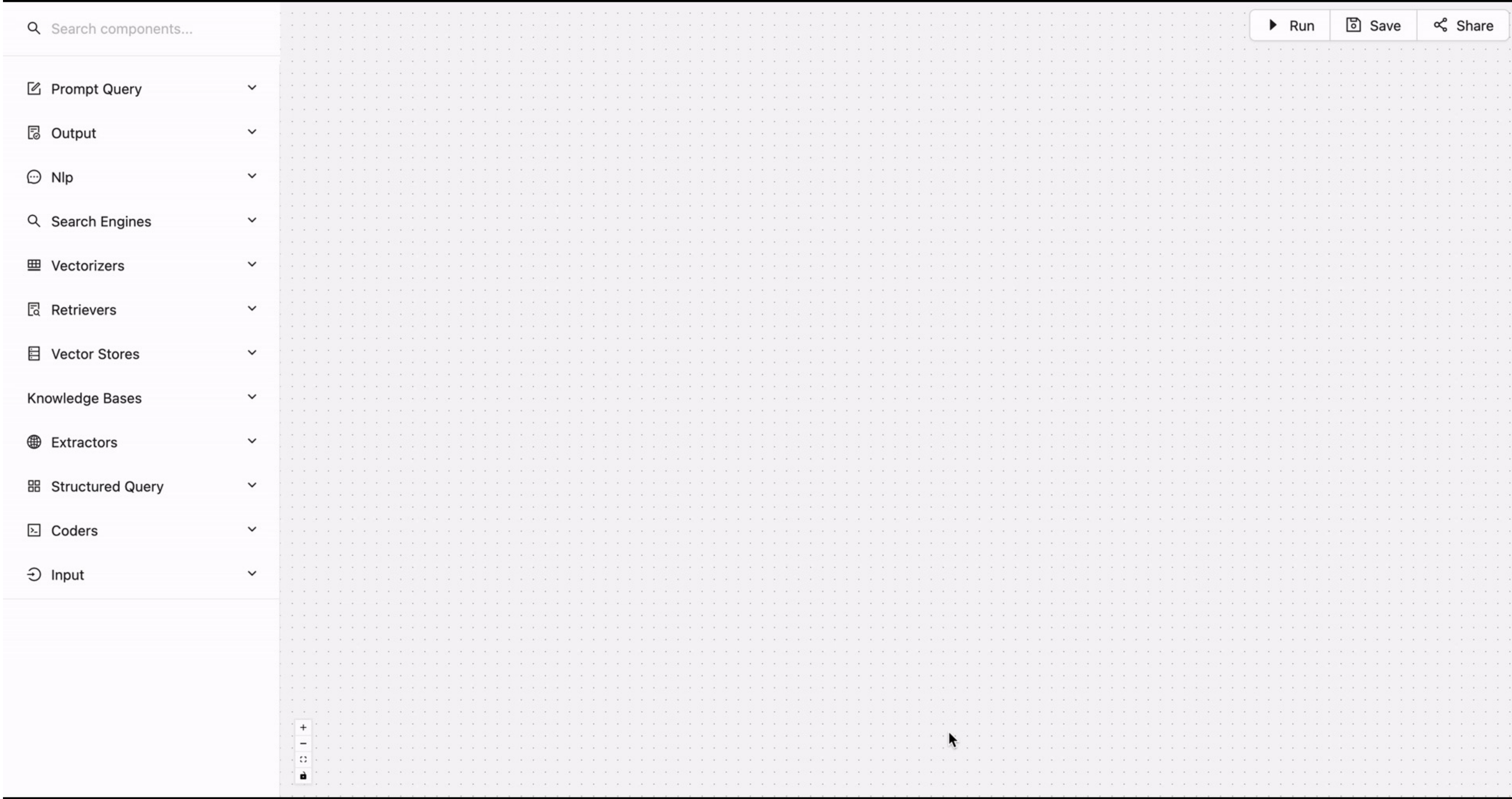
	Study	IL2	IL4	IL6	IL8	IL10	IL15	IL2Rα	TNF–α	IFN–γ	GM-CSF
1	Jacobson et al. [37]	R	R	R	R	R	R	R	R	R	R
2	Hong et al. [38]	R	MV	R	MV	R	MV	MV	R	R	MV
3	Yan et al. [39]	MV	MV	R	MV	MV	MV	MV	MV	MV	MV
4	Topp et al. [40]	R	R	R	R	R	R	R	R	R	R
5	Shah et al. [41]	MV	MV	R	R	R	R	R	R	R	R
6	Liu et al. [29]	R	R	R	MV	R	MV	MV	R	R	MV
7	Sang et al. [13]	MV	MV	R	MV	MV	MV	MV	MV	R	MV
8	Yan et al. [42]	MV	MV	R	MV	MV	MV	MV	MV	MV	MV
9	Zhao et al. [43]	MV	MV	R	MV	MV	MV	MV	MV	MV	MV
10	Neelapu et al. [44]	R	MV	R	R	R	R	R	MV	R	R
11	Hay et al. [24]	MV	MV	R	R	R	R	MV	MV	R	MV
12	Turtle et al. [45]	MV	MV	R	MV	R	MV	MV	R	R	MV
13	Hu et al. [15]	MV	MV	R	MV	R	MV	MV	MV	R	MV
14	Teachey et al. [18]	R	R	R	R	R	MV	MV	R	R	R
15	Porter et al. [16]	R	MV	R	MV	MV	MV	R	MV	R	MV
16	Davila et al. [5]	MV	MV	R	MV	R	MV	MV	MV	R	R
17	Kalos et al. [46]	R	R	R	R	R	R	R	R	R	MV

325x efficiency gain

Demo Wysocki, Wysocka, Carvalho, Bogatu, Miranda

Lunar

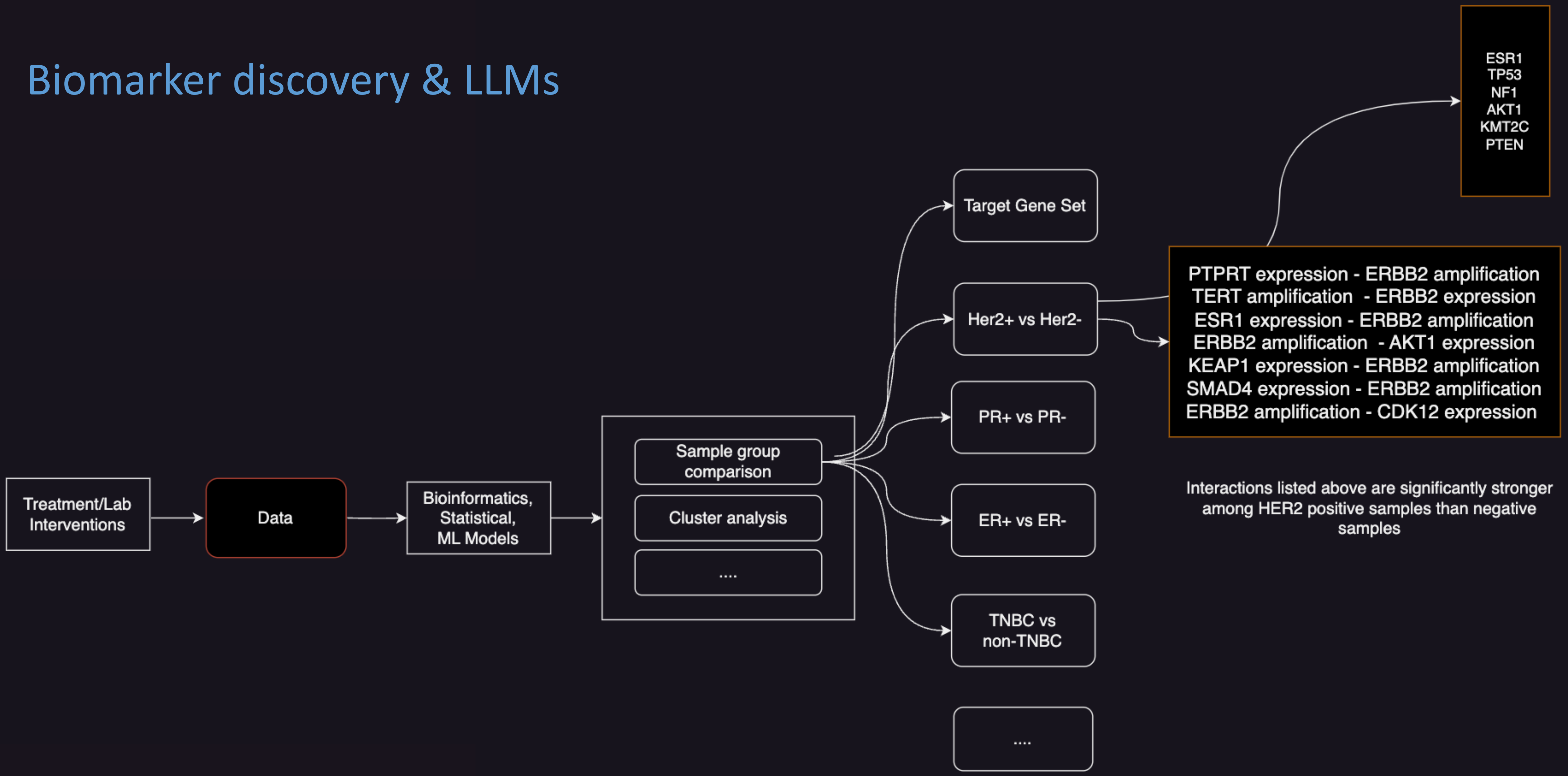
AI coordination infrastructure

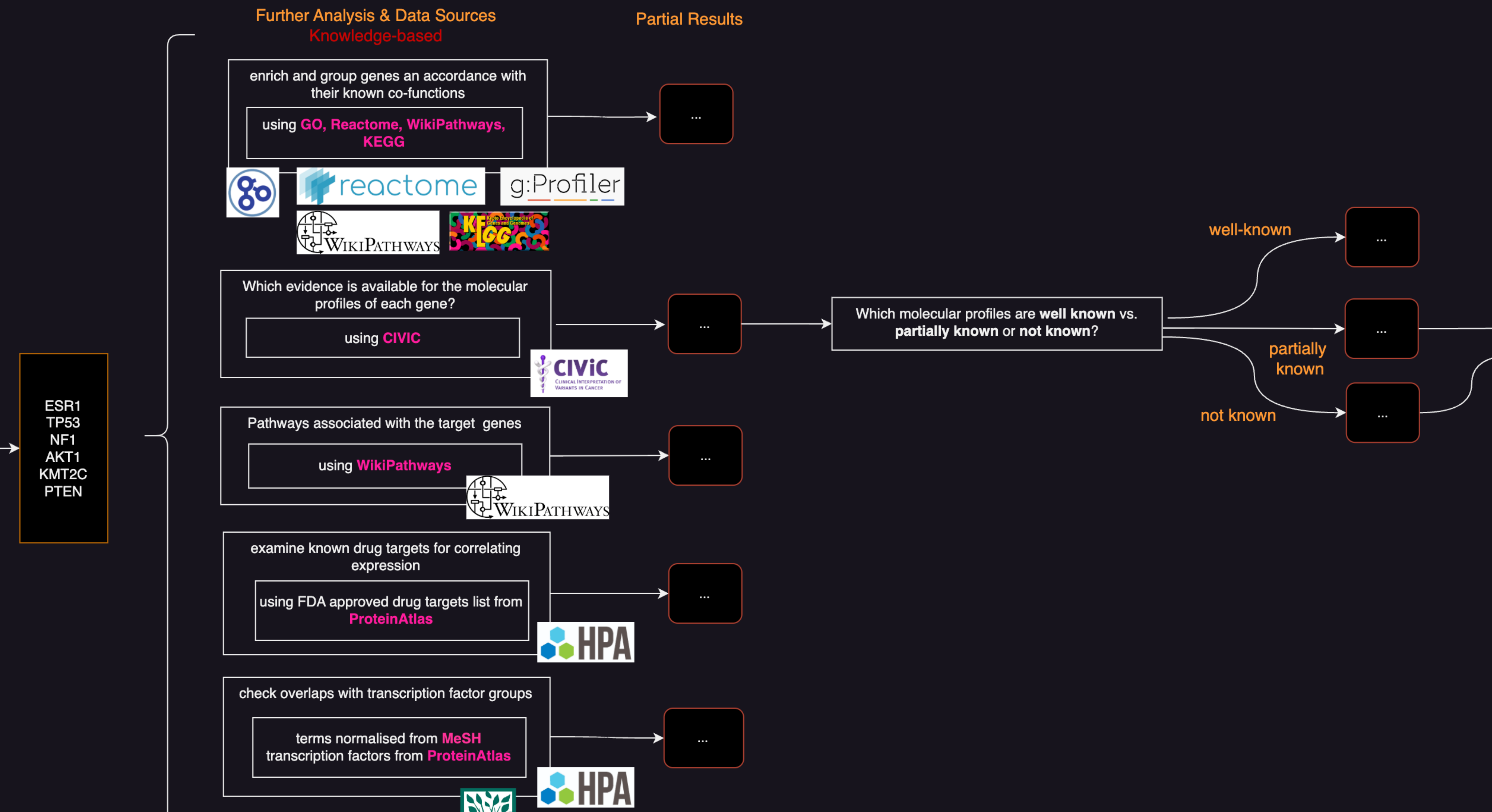


Search components...
Prompt Query
Output
Nlp
Search Engines
Vectorizers
Retrievers
Vector Stores
Knowledge Bases
Extractors
Structured Query
Coders
Input
Run
Save
Share

Evidence-based Scientific Reasoning

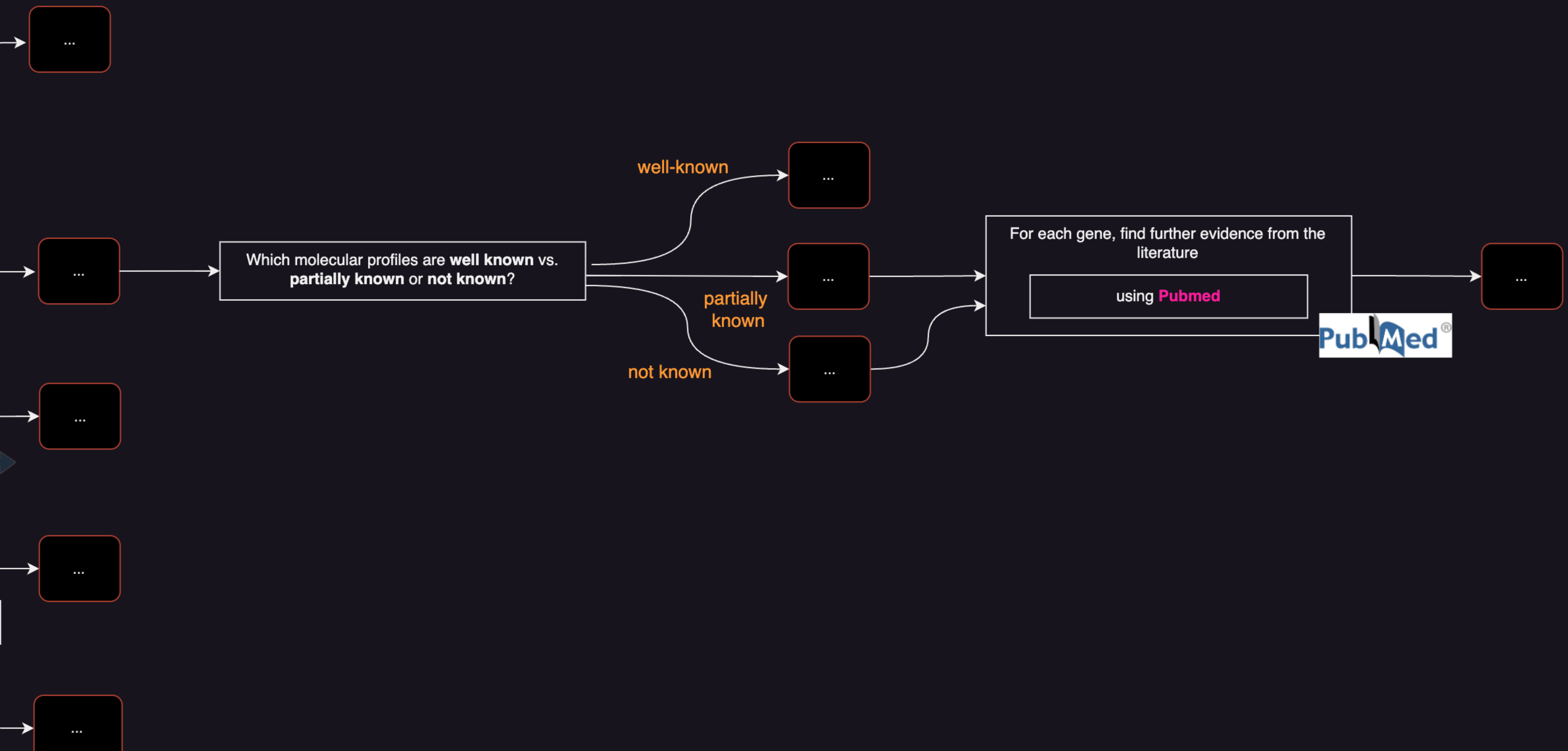
Biomarker discovery & LLMs





Further Analysis & Data Sources
Knowledge-based
Partial Results
ESR1
TP53
NF1
AKT1
KMT2C
PTEN
enrich and group genes an accordance with their known co-functions
using GO, Reactome, WikiPathways, KEGG
reactome
g:Profiler
WikiPathways
Which evidence is available for the molecular profiles of each gene?
using CIVIC
CIViC
Clinical Interpretation of Variants in Cancer
Pathways associated with the target genes
using WikiPathways
WikiPathways
examine known drug targets for correlating expression
using FDA approved drug targets list from ProteinAtlas
HPA
check overlaps with transcription factor groups
terms normalised from MeSH
transcription factors from ProteinAtlas
HPA
...
Which molecular profiles are well known vs. partially known or not known?
well-known
partially known
not known

Partial Results



Further Analysis & Data Sources
Knowledge-based

Partial Results

Multi-step decomposition
Lowering the ‘impedance’ across heterogeneous evidence and tools
Harmonising the evidence space
Which can be reasoned over (linguistically, mechanistically, etc ..)

enrich and group genes an accordance with their known co-functions

using **GO, Reactome, WikiPathways, KEGG**

...

Which evidence is available for the molecular profiles of each gene?

using **CIVIC**

Not known molecular profiles
- **Gene: MTAP:** methylthioadenosine phosphorylase
-- MolecularProfileName: MTAP Deletion; no Evidence in CivicDB

- **Gene: KMT2C:** lysine methyltransferase 2C
-- MolecularProfileName: KMT2C Loss-of-function; no Evidence in CivicDB

For each gene, find further evidence from the literature

using **Pubmed**

PRMT5: An Emerging Target for Pancreatic Adenocarcinoma. 7 citations
DOI: 10.3390/cancers13205136

Loss of MAT2A compromises methionine metabolism and represents a vulnerability in H3K27M mutant glioma by modulating the epigenome. 6 citations
DOI: 10.1038/s43018-022-00348-3
...

Pathways associated with the genes

using **WikiPathways**

...

examine known drug targets for correlating expression

using FDA approved drug targets list from **ProteinAtlas**

Using the Human Protein Atlas as a reference:

Eight genes overlapped between the two sets, including CHST11, STK4, CASP8, TGFBR2, CIITA, BTK, CSF1R, and CYLD. The precision of the analysis was 0.127, indicating that 12.7% of the analyzed set was also present in the reference set. The recall was 0.008, indicating that only 0.8% of the reference set was present in the analyzed set. The p-value was marginally significant, indicating that the overlap between the two sets was not entirely random.

check overlaps with transcription factor groups

terms normalised from **MeSH**
transcription factors from **ProteinAtlas**

Using the Human Protein Atlas as a reference:

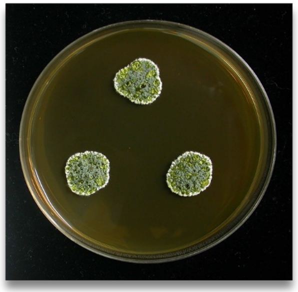
When compared the selected set of genes with the reference set of transcription factors it was found that 13 genes overlapped between the two sets, including EBF1, MAF, NFATC2, PAX5, LYL1, BCL11B, PRDM1, TCF7, IKZF1, FLI1, FOXO1, IRF4, and TFEB. The precision of the comparison was 0.206, indicating that 20.6% of the genes in the selected set were also present in the reference set. The recall was 0.009, indicating that only 0.9% of the reference set genes were also present in the selected set. The Fisher's Test resulted in a statistically significant p-value, indicating that the overlap between the two sets was not random.

Drug Discovery

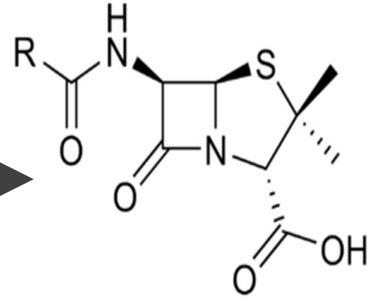
Organisms produce **compounds** which can deliver **therapeutic properties**.

(Fungi, plants, extremophiles)

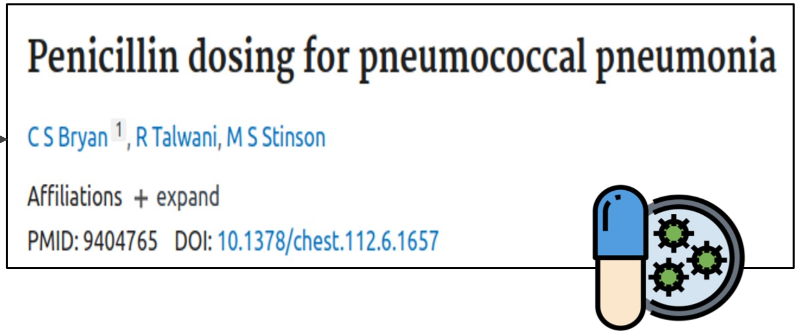
(antibiotic properties)



produces



has activity



Penicillin dosing for pneumococcal pneumonia

C S Bryan [1], R Talwani, M S Stinson

Affiliations + expand

PMID: 9404765 DOI: 10.1378/chest.112.6.1657

Penicillium Chrysogenum

Penicillin

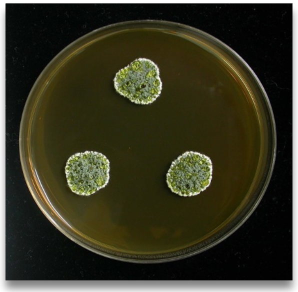
associated antibiotic activity

Testing each **compound** is a **long and expensive process** (~1M CHF / per **compound**).

Assessing what is already known is essential to **prioritise**, **avoid rediscoveries and dead-ends.**

Drug Discovery

For a target list of 64 organisms

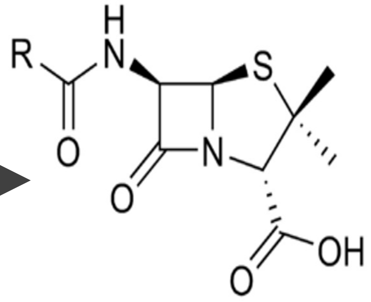


produces

LOTUS

DB: 84

+Lit: 1144

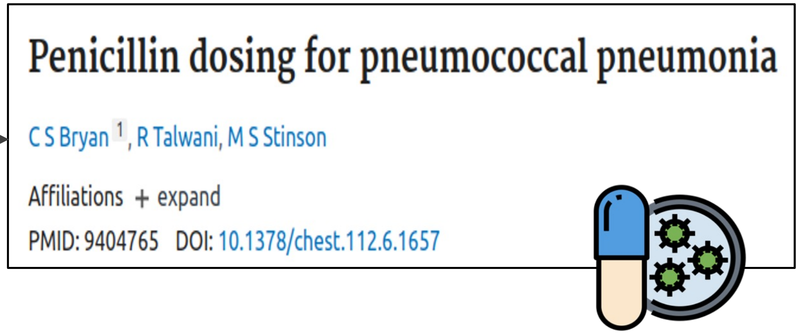


has activity

ChEBI

DB: 5

+Lit: 2323


Penicillin dosing for pneumococcal pneumonia

C S Bryan [1], R Talwani, M S Stinson

Affiliations + expand

PMID: 9404765 DOI: 10.1378/chest.112.6.1657

Testing each **compound** is a **long and expensive process** (~1M CHF / per **compound**).

Assessing what is already known is essential to **prioritise**, **avoid rediscoveries and dead-ends.**

PMC

28.301 passages

PubMed

49.671 abstracts

Thermomyces lanuginosus

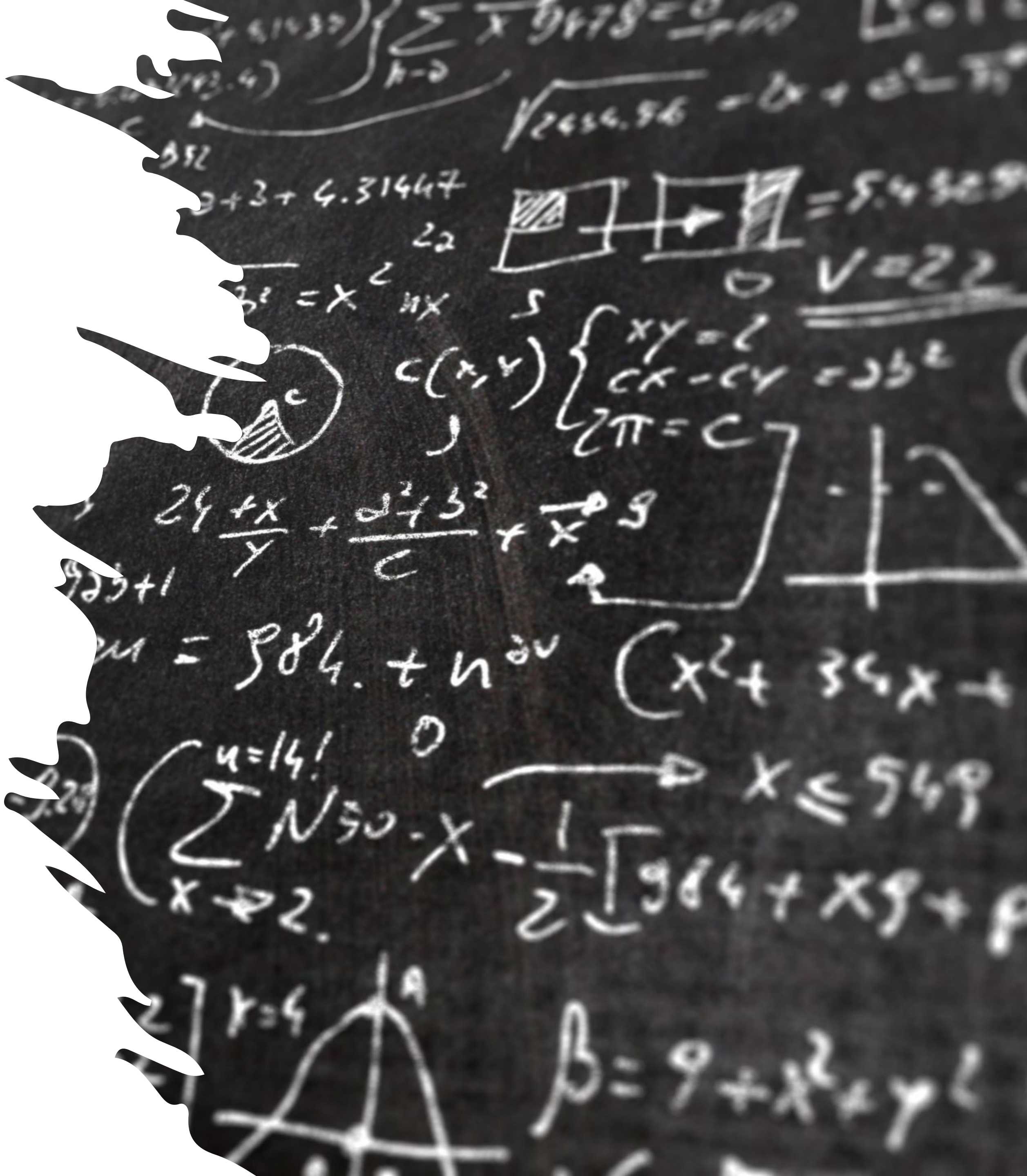
Cytochrome C	https://pubmed.ncbi.nlm.nih.gov/4342602
Lipozyme TL IM	https://pubmed.ncbi.nlm.nih.gov/15048592, https://pubmed.ncbi.nlm.nih.gov/29459507, https://pubmed.ncbi.nlm.nih.gov/34269888, https://pubmed.ncbi.nlm.nih.gov/36985609
Phenylacetaldehyde	https://pubmed.ncbi.nlm.nih.gov/36212286
2-Phenylethanol	https://pubmed.ncbi.nlm.nih.gov/36212286
Glucosides	https://pubmed.ncbi.nlm.nih.gov/10467123
Arbutin	https://pubmed.ncbi.nlm.nih.gov/24278310, https://pubmed.ncbi.nlm.nih.gov/34705451
Stearic Acid	https://pubmed.ncbi.nlm.nih.gov/36766114
Polydatin	https://pubmed.ncbi.nlm.nih.gov/34869302
Eugenyl acetate	https://pubmed.ncbi.nlm.nih.gov/25875787
2-Deoxy-D-glucose	https://pubmed.ncbi.nlm.nih.gov/16110918

Thermomyces lanuginosus

- Cytochrome C
- Lipozyme TL IM
- Phenylacetaldehyde
- 2-Phenylethanol
- Glucosides
- Arbutin
- Stearic Acid
- Polydatin
- Eugenyl acetate
- 2-Deoxy-D-glucose

1. "Crude essential oils of C. longa and S. aromaticum exhibited antimicrobial activity against all selected isolates but S. aromaticum activity was better than the C. longa with a maximum 19.3±1.50 mm zone of inhibition against A. baumannii at 1.04 µL/mL MIC." This suggests that both Curcuma longa and Syzygium aromaticum have antibacterial properties, with Syzygium aromaticum being more potent against Acinetobacter baumannii, as indicated by the larger zone of inhibition and lower MIC value.
2. "GC/MS analysis revealed the abundance of components including eugenol, eugenyl acetate, b- caryophyllene, and a- Humulene in both crude oil and fractions of S. aromaticum." This indicates that eugenyl acetate is one of the main components present in Syzygium aromaticum, suggesting its potential role in the observed antibacterial activity.

Controlling Language Models & Formal Inference



Scientific Reasoning

- Step-wise explicit (verbalised) inference.
- Formal, verifiable argument & explanation.
- Preserving the positive aspects of LLMs.
- Improving control.

Large Language Models (LLMs) — Formal

Neuro — **Neuro-symbolic** — Symbolic

Scientific Reasoning

- Step-wise explicit (verbalised) inference.
- Formal, verifiable argument & explanation.
- Preserving the positive aspects of LLMs.
- Improving control.

$$\Gamma \models \Phi$$

$$\Gamma \text{ semantically entails } \Phi$$

In our context (some notational abuse):
- explanations | arguments
- proofs | derivations

$$\Gamma \vdash \Phi$$

$$\Gamma \text{ proves } \Phi$$

$\vdash$

- interpretability
- verifiability
- control (inference guarantees)

T|F?

Conclusion

Patients with loss of PALB2 may benefit from PARP1 inhibition due to synthetic lethality, causing cells to rely on a singular mechanism to repair cumulative damage to DNA.

Intermediate Steps

24. Loss of PALB2 leads to a deficiency in HRR, causing the cells to rely on other DNA repair mechanisms. (Combination of premises 8, 15, 16, 21, 22)

25. Inhibiting PARP in cells lacking PALB2 results in the accumulation of DNA damage due to the reliance on a singular repair mechanism, leading to synthetic lethality. (Combination of premises 5, 9, 10, 24)

RAG

Premises

…

5- Inhibiting PARP results in accumulation of SS breaks.

6- NHEJ does not use a template to repair DSB and can cause increased genomic instability.

7- PARP1 synthesis PAR which recruits repair proteins to sites of DNA damage

8- In the absence of functional HRR genes, DNA repair defaults to NHEJ.

9- PARP1 synthesises PAR.

10- PAR recruits repair proteins to damaged DNA site.

…

15- PALB2 is required for the localization of BRCA2 to sites of DNA damage

16- PALB2…encodes a major BRCA2 binding partner that controls its intranuclear localization and stability.

17- RAD51 is a eukaryotic gene that encodes the RAD51 homolog gene.

18- BRCA2 promotes the assembly of RAD51 homolog 1 onto SS DNA in HRR.

19- BRCA2 is a human gene that encodes the BRCA2 protein.

20- BRCA2 protein is a tumour suppressor involved in HRR.

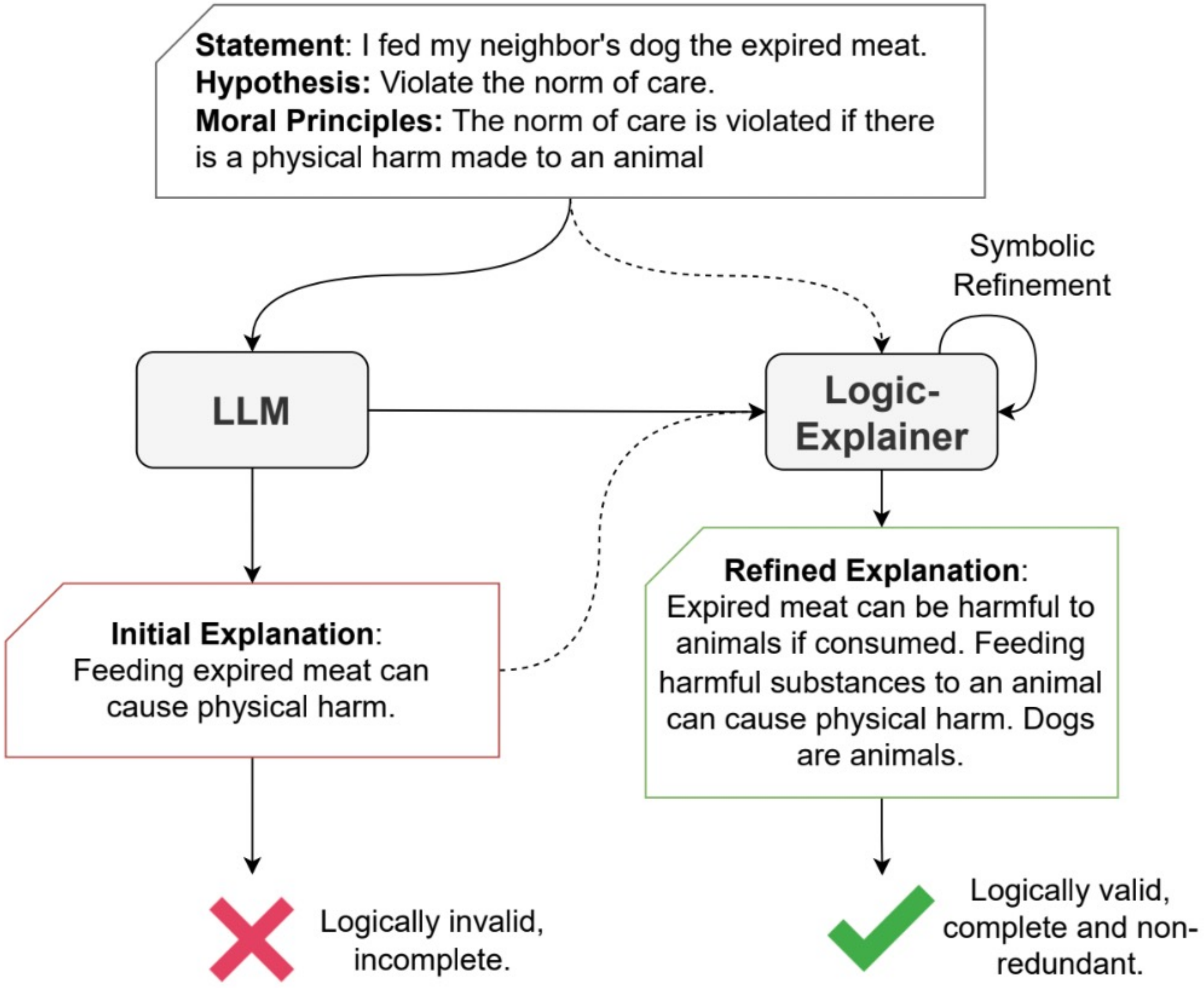
21- HRR is the primary process for repairing DNA double strand breaks.

22- HRR repairs damage to DNA using information copied from a homologous undamaged molecule.

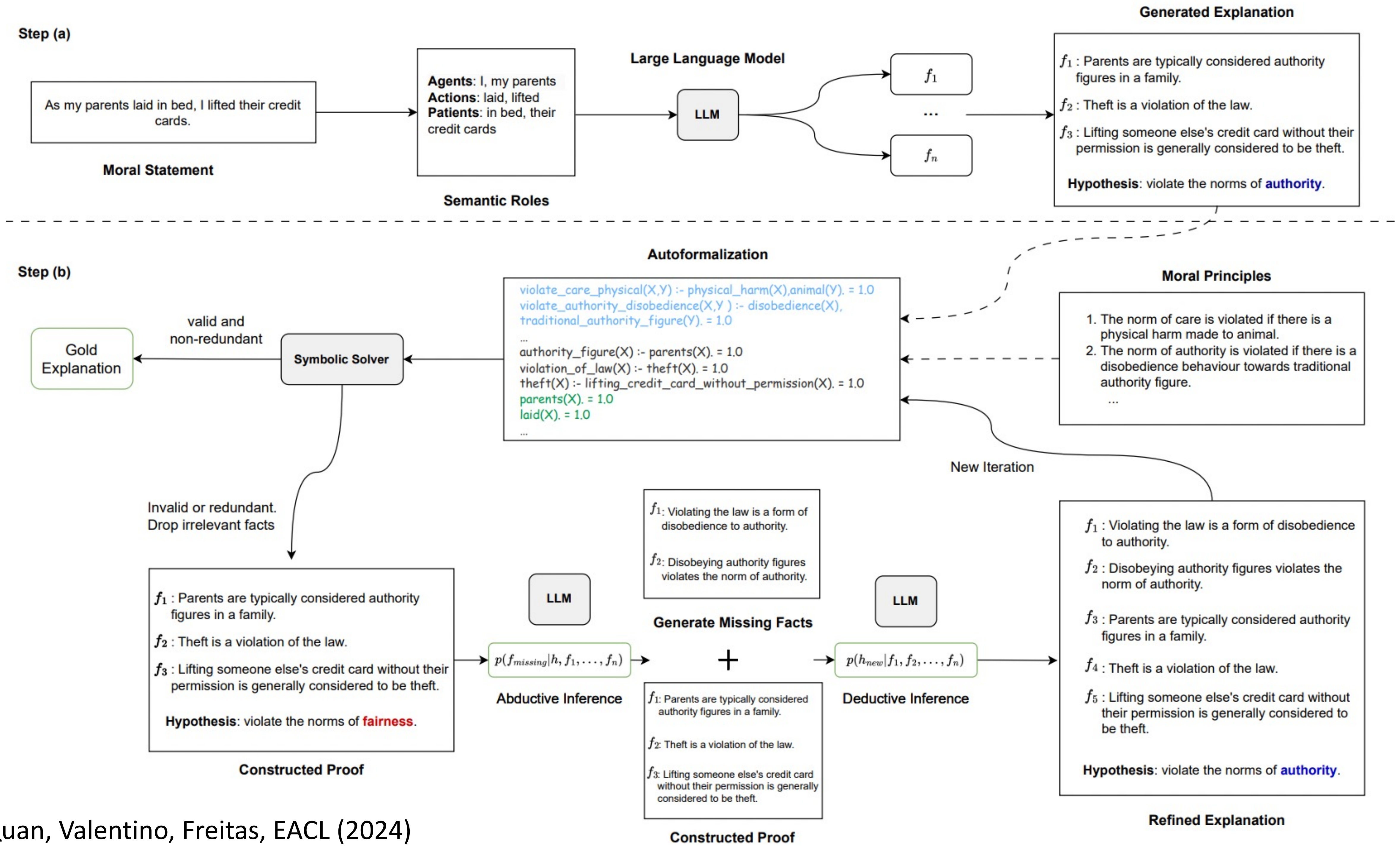
23- Undamaged homologous molecules are provided by sister chromatids or paternal/maternal copies of chromosomes.

…

Ethical Reasoning



Quan, Valentino, Freitas, EACL (2024)



Quan, Valentino, Freitas, EACL (2024)

Ethical Reasoning

Predictive Task

Model	Iterations	Easy	Hard	AVG
Zero-Shot	0	40.1	55.0	47.5
Chain-Of-Thought	0	54.5	54.1	54.3
Logic-Explainer	0	52.8	58.3	55.6
	1	54.4	**59.1**	56.8
	2	57.5	**59.1**	**58.3**
	3	**57.6**	58.6	58.1
Human		85.1	83.4	84.22

Explanation Quality

Model	Valid ↑	Invalid ↓	Valid and non-Redundant ↑	Valid but Redundant ↓
Chain-of-Thought	22.9	77.1	34.2	65.8
Logic-Explainer+0 iter.	40.4	59.6	13.4	86.6
Logic-Explainer+1 iter.	53.6	46.4	75.3	24.7
Logic-Explainer+2 iter.	62.0	41.6	86.4	13.6
Logic-Explainer+3 iter.	**65.1**	**34.9**	**95.4**	**4.60**

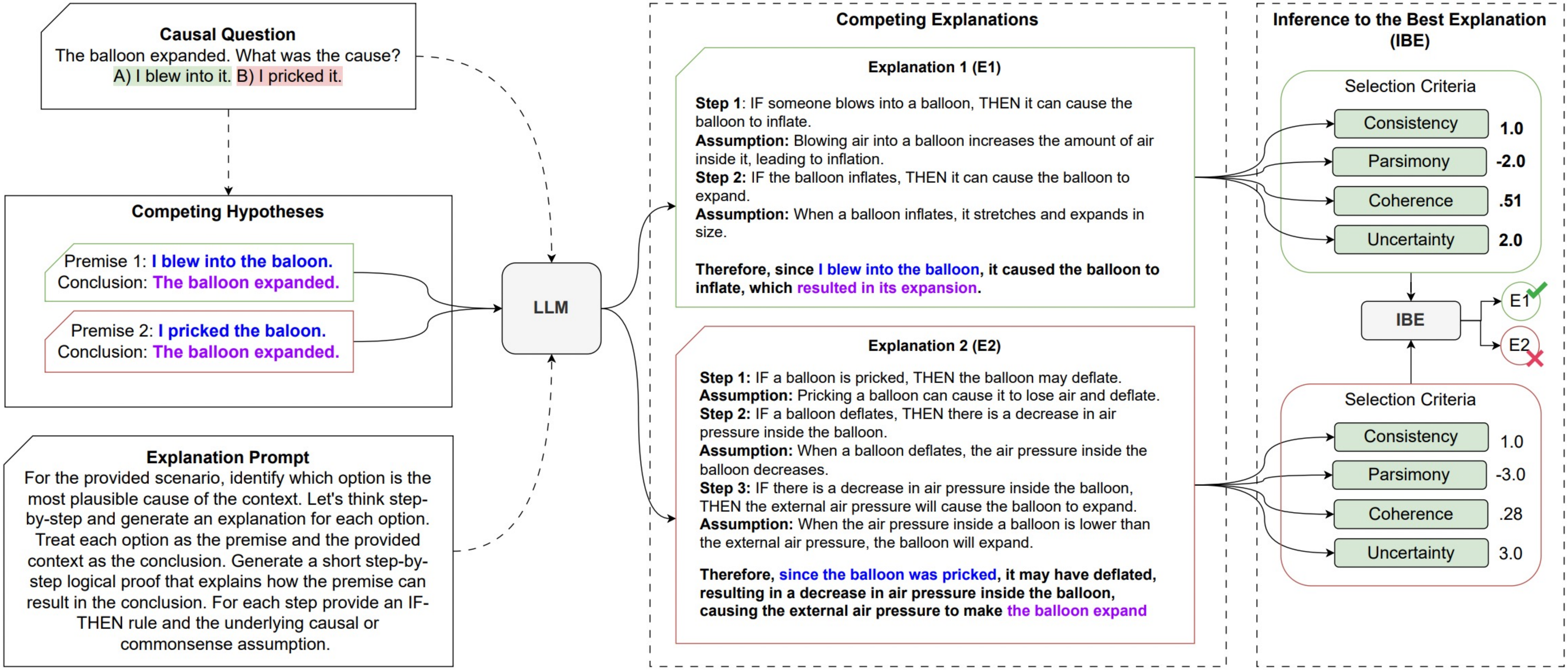
External symbolic solvers elicit valid and complete reasoning.

Logic-Explainer improve LLMs on identifying underlying moral violations.

Incomplete explanations impact LLMs' performance.

Neo-Davidsonian semantics enhances logical consistency in complex sentence representation.

Causal Reasoning



Dalal, Valentino, Freitas, Buitelaar, arXiv, 2402.10767 (2024)

Mathematical Reasoning

$$g(x) = \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{\infty} \tilde{g}(p) \cdot e^{\frac{ipx}{\hbar}}\, dp$$

$$\tilde{g}(p) = p \cdot \varphi(p)$$

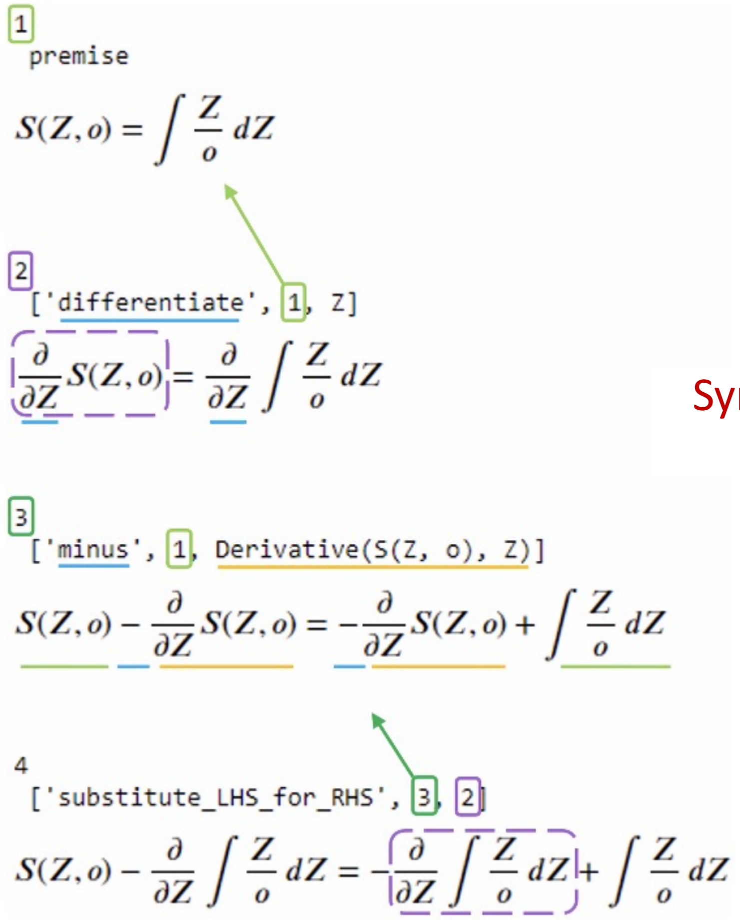
$$g(x) = \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{\infty} p \cdot \varphi(p) \cdot e^{\frac{ipx}{\hbar}}\, dp$$

$$\varphi(p) = \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{\infty} \varphi(\chi) \cdot e^{\frac{-ip\chi}{\hbar}}\, d\chi$$

$$g(x) = \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{\infty} p \cdot \left(\frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{\infty} \varphi(\chi) \cdot e^{\frac{-ip\chi}{\hbar}}\, d\chi\right) \cdot e^{\frac{ipx}{\hbar}}\, dp$$

$$g(x) = \frac{1}{2\pi\hbar} \int_{-\infty}^{\infty} p \cdot \left(\int_{-\infty}^{\infty} \varphi(\chi) \cdot e^{\frac{-ip\chi}{\hbar}}\, d\chi\right) \cdot e^{\frac{ipx}{\hbar}}\, dp$$

Incompleteness



Symbolic/algebraic inference

Synthetic-stepwise, Maths (algebraic/calculus), OOD

Meadows, Valentino, Teney, Freitas, arXiv:2305.12563 (2023)

Meadows, Valentino, Freitas, arXiv:2307.09998 (2023)

Meadows, James, Freitas (2024)

Controlling Language Spaces

Contemporary linguistic objects live on high-dimensional embedding spaces.

implies a geometry

Properties of these spaces are poorly characterised and controlled.

entanglement, non-separation

Implications in terms of inference safety, out-of-distributional generalisation, …

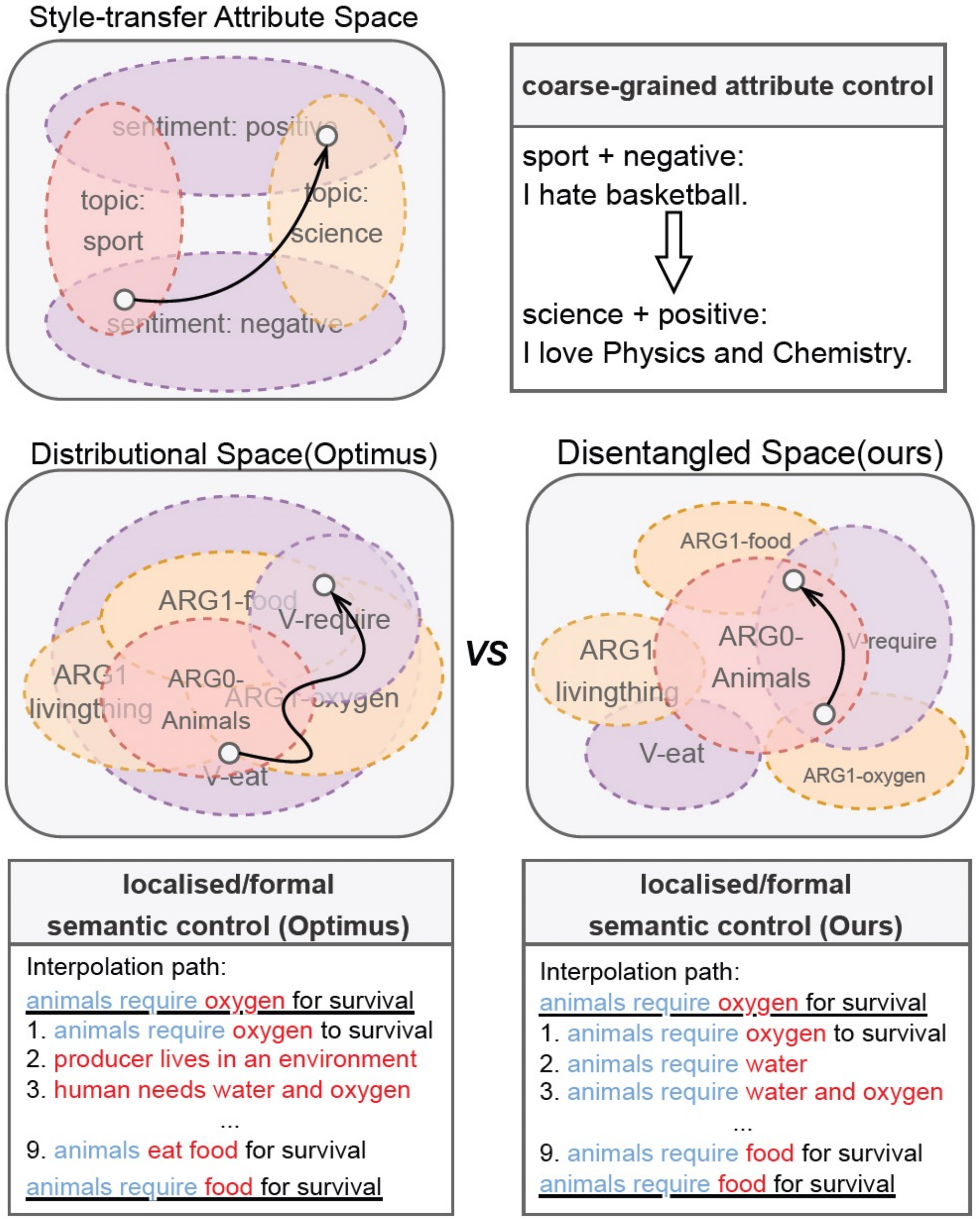
Q: Can we develop embedding models with better control properties?

better geometrical-semantic alignment

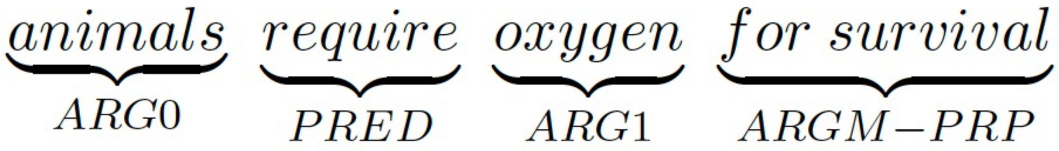
Fundamental for rigorous scientific reasoning

Explanations | Definitions

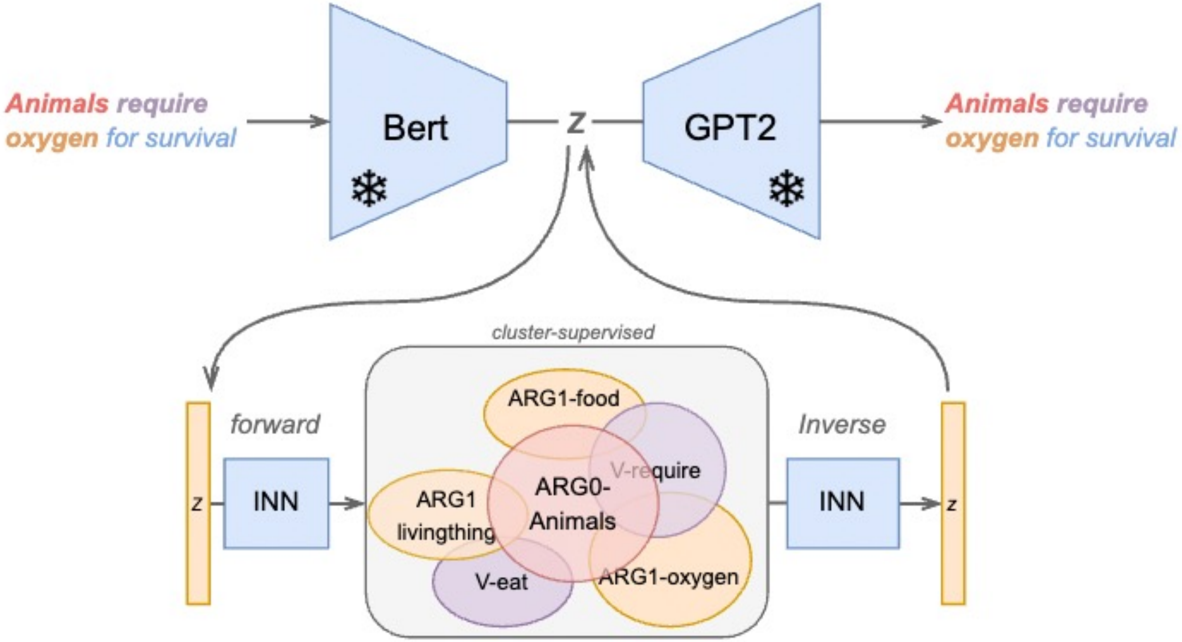
Language Variational Autoencoders (VAEs)



Semantic properties



Disentanglement properties



Separability properties

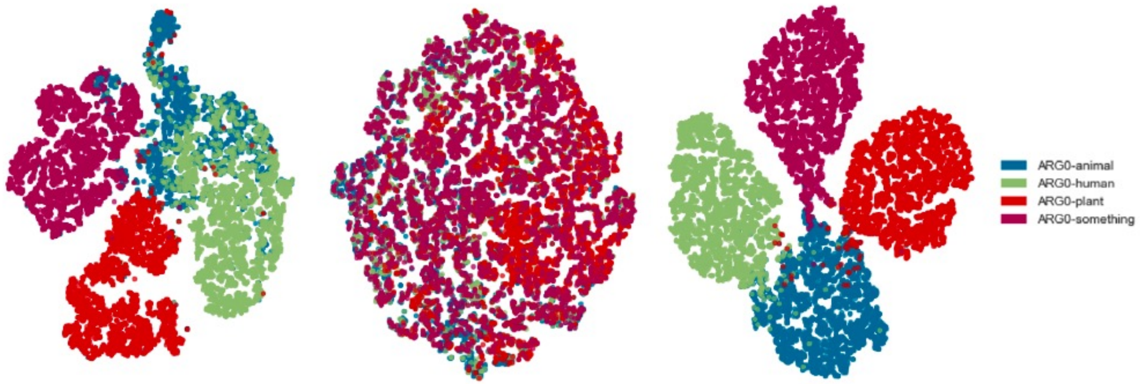
interpolation localisation: *predicate-require*

source: humans require freshwater for survival

Optimus:
1. humans require water and food through fossil fuels
2. humans require water for survival
3. humans produce small amounts of consumer food
4. human has a positive impact on a plant's survival
5. humans convert food into animal prey
6. humans make food for themselves by eating
7. animals require food for survival
8. animals require nutrients from the air
9. humans eat plants for food
10. animals require food for survival

Cluster-supervised INN:
1. humans require water for survival
2. nonhumans require water for survival
3. animals require water and food
4. animals require water to survive
5. animals require water to live
6. animals require food for survival
7. animals require food for survival
8. animals require food for survival
9. animals require food for survival
10. animals require food to survive

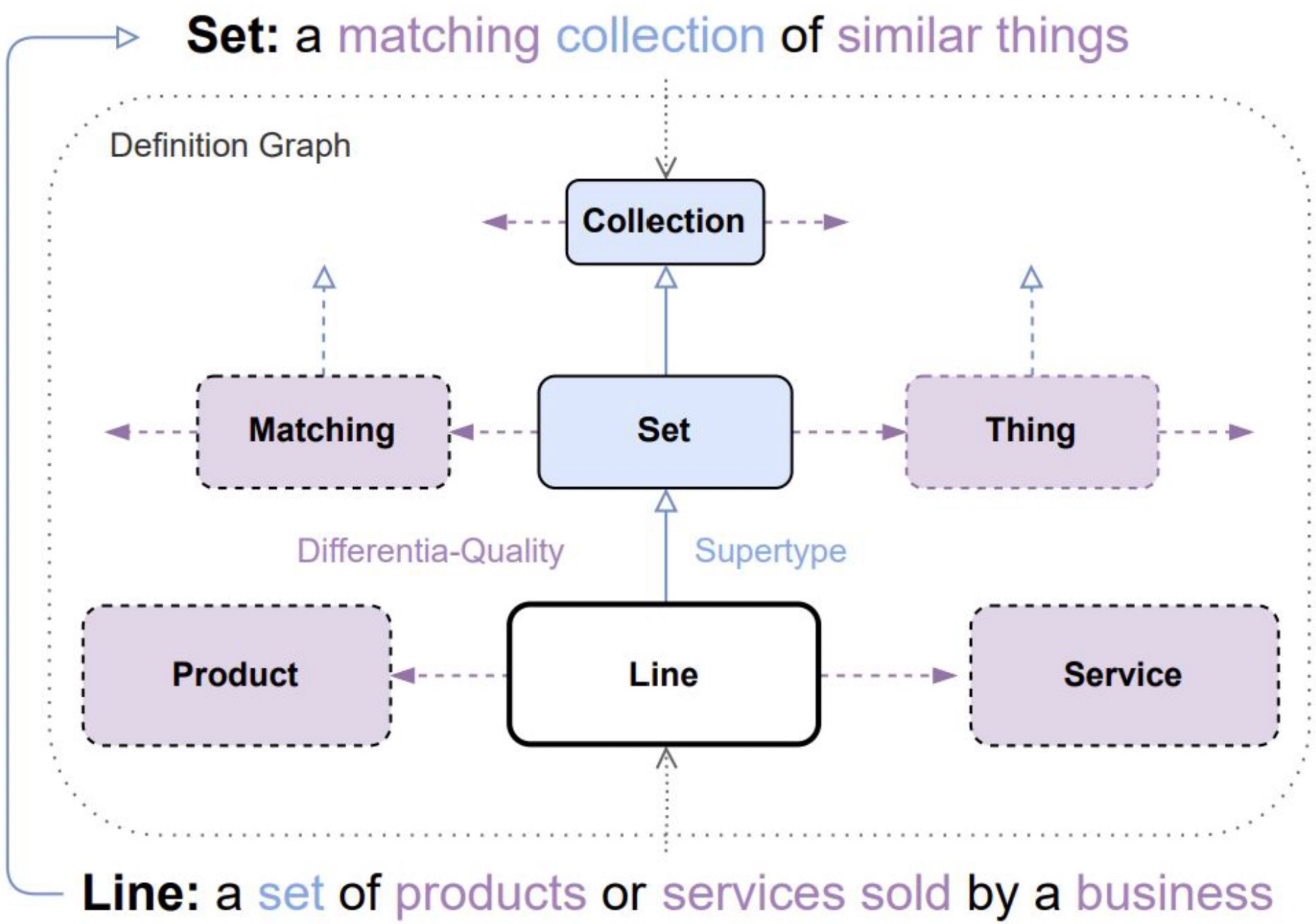
target: animals require food to survive



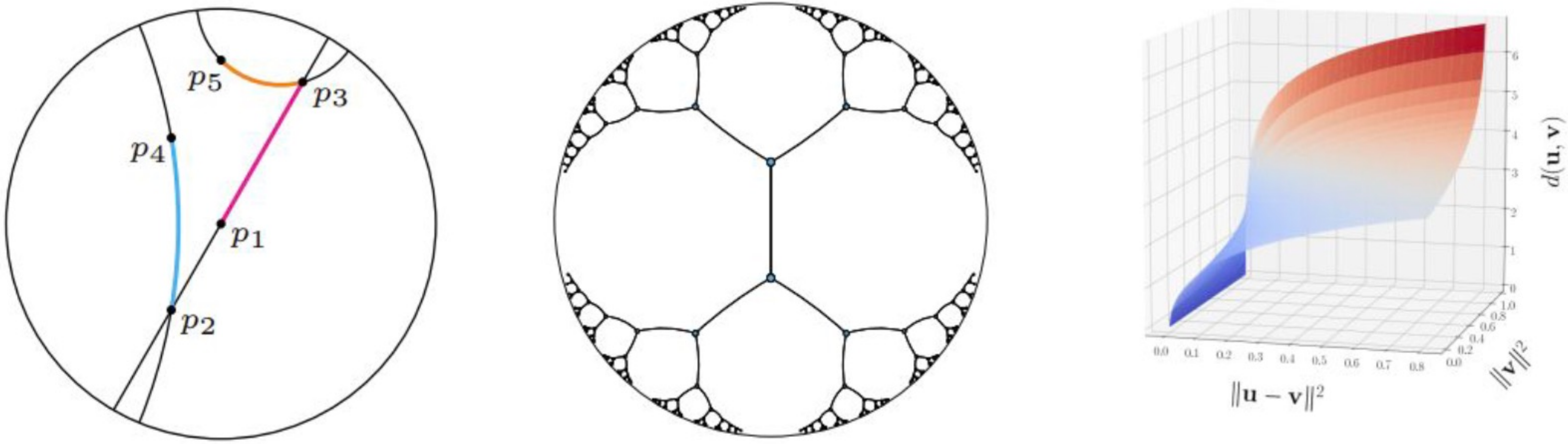
Evaluation Metrics	avg IS↑	max IS↑	min IS↑
DAE (Vincent et al., 2008)	0.144	0.330	0.055
AAE (Makhzani et al., 2015)	0.142	0.284	0.054
LAAE(Rubenstein et al., 2018)	0.172	0.347	0.056
DAAE (Shen et al., 2020)	0.055	0.061	0.023
β-VAE (Higgins et al., 2016)	0.198	0.379	0.041
AdaVAE (Tu et al., 2022)	0.085	0.105	0.050
Della (Hu et al., 2022)	0.253	0.416	0.155
Optimus (Li et al., 2020b)	0.220	0.525	0.130
AutoEncoder (Bert-GPT2)	0.259	0.585	0.165
INN (U) (our)	0.251	0.540	0.159
INN (C) (our)	**0.282**	**0.607**	**0.206**

Zhang, Carvalho, Valentino, Pratt-Hartmann, Freitas, EACL Findings (2024)
Zhang, Carvalho, Pratt-Hartmann, Freitas, arXiv:2305.01713 (2023)
Carvalho, Zhang, Freitas, EACL Findings (2022)

Reasoning over definitions



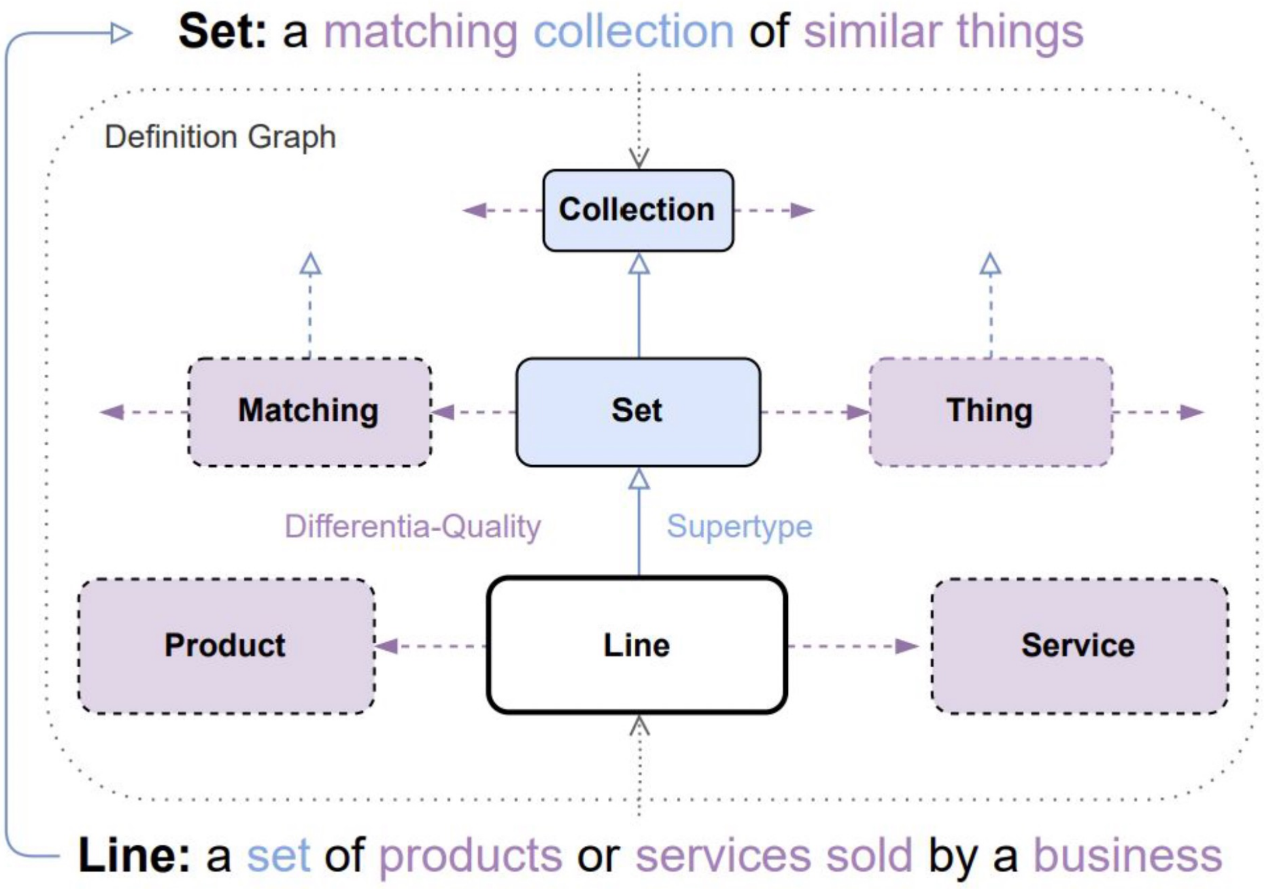
Multi-relational Hyperbolic Embeddings

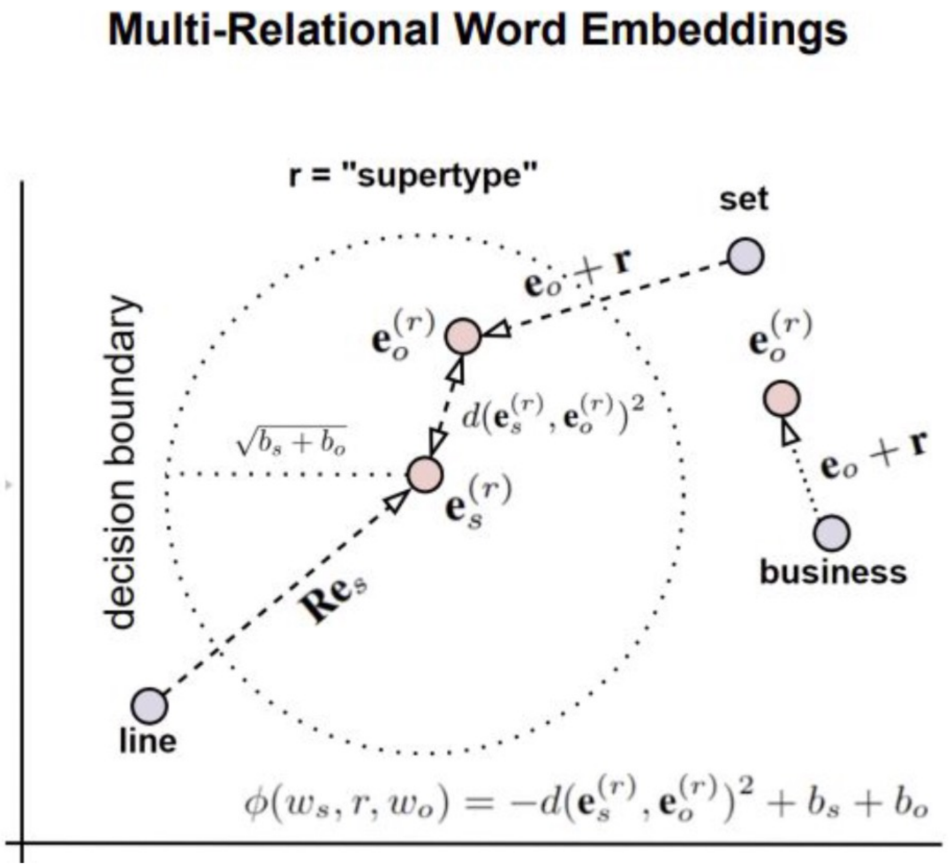


(a) Geodesics of the Poincaré disk (b) Embedding of a tree in $\mathcal{B}^2$ (c) Growth of Poincaré distance

Nickel & Kiela, NIPS (2017)

Multi-relational Hyperbolic Embeddings





Valentino, Carvalho, Freitas, EACL (2024)

Multi-relational Hyperbolic Embeddings

Model	Dim	FT	PT	SV-d	MEN-d	SV-t	MEN-t	SL999	SCWS	353	RG
Glove	300	yes	no	12.0	54.8	7.8	57.0	19.8	46.8	44.4	57.5
Word2Vec	300	yes	no	35.2	62.3	36.4	59.9	34.5	54.5	**61.9**	65.7
AE	300	yes	no	34.9	42.7	32.5	42.2	35.6	50.2	41.4	64.8
CPAE	300	yes	no	42.8	48.5	34.8	49.2	39.5	54.3	48.7	67.1
CPAE-P	300	yes	yes	44.1	**65.1**	42.3	63.8	45.8	**60.4**	61.3	72.0
bert-base	768	no	yes	13.5	27.8	13.3	30.6	15.1	37.8	20.0	68.1
bert-large	1024	no	yes	16.1	23.4	14.4	26.8	13.4	35.7	19.8	60.7
defsent-bert	768	yes	yes	40.0	60.2	40.0	60.0	42.0	56.8	46.6	82.4
defsent-roberta	768	yes	yes	43.0	55.0	44.0	52.6	47.7	54.3	44.9	80.6
distilroberta-v1	768	no	yes	35.8	61.2	36.7	62.2	43.4	57.1	52.0	77.4
mpnet-base-v2	768	no	yes	45.9	64.9	42.5	**67.5**	49.5	58.6	56.5	81.3
sentence-t5-large	768	no	yes	**49.4**	63.1	**50.2**	66.3	**57.3**	56.1	51.8	**85.3**
Multi-Relational											
Euclidean	40	yes	no	39.1	62.9	35.7	65.4	36.3	58.2	52.1	80.9
Euclidean	80	yes	no	44.1	65.6	39.5	66.2	41.2	58.4	55.8	78.0
Euclidean	200	yes	no	47.3	67.0	41.0	67.6	43.4	60.6	55.4	78.1
Euclidean	300	yes	no	47.9	68.3	43.1	69.1	**44.7**	61.0	54.4	79.0
Hyperbolic	40	yes	no	36.7	66.2	34.3	66.4	31.8	57.7	49.9	75.5
Hyperbolic	80	yes	no	42.7	68.2	40.7	68.6	38.3	60.5	57.3	81.0
Hyperbolic	200	yes	no	48.8	71.9	44.7	73.2	40.7	62.5	62.5	**81.6**
Hyperbolic	300	yes	no	**50.6**	**72.6**	**45.4**	**74.2**	42.3	**63.0**	**63.3**	80.5

Valentino, Carvalho, Freitas, EACL (2024)

Take-away

Emerging foundations for **scaling-up scientific inference**

Universal framework for **integrating and reasoning over heterogeneous evidence**

Large Language Models

Are a game-changing foundation.
Transformers are an efficient substrate for modelling language.
Alone they are not fit for purpose for full scientific reasoning.

Controlling reasoning

Decomposition: Scientific reasoning requires coordination infrastructures.
Formal augmentation: Close integration LLMs with symbolic solvers.
Geometrical-semantic alignment: Language VAEs.

Thank you for your attention!

Generously supported by:

Swiss National Science Foundation

UKRI Engineering and Physical Sciences Research Council

the ark

Innosuisse

Horizon Europe

CANCER RESEARCH UK

contact: andre.freitas@manchester.ac.uk

ai-reasoning.net