



The University of Manchester



**MANCHESTER CANCER  
RESEARCH CENTRE**

# Mathematical reasoning at the age of Large Language Models (LLMs)

André Freitas & Neuro-Symbolic AI Group

Honda Research Institute (HRI)

December 2023





MANCHESTER  
1824

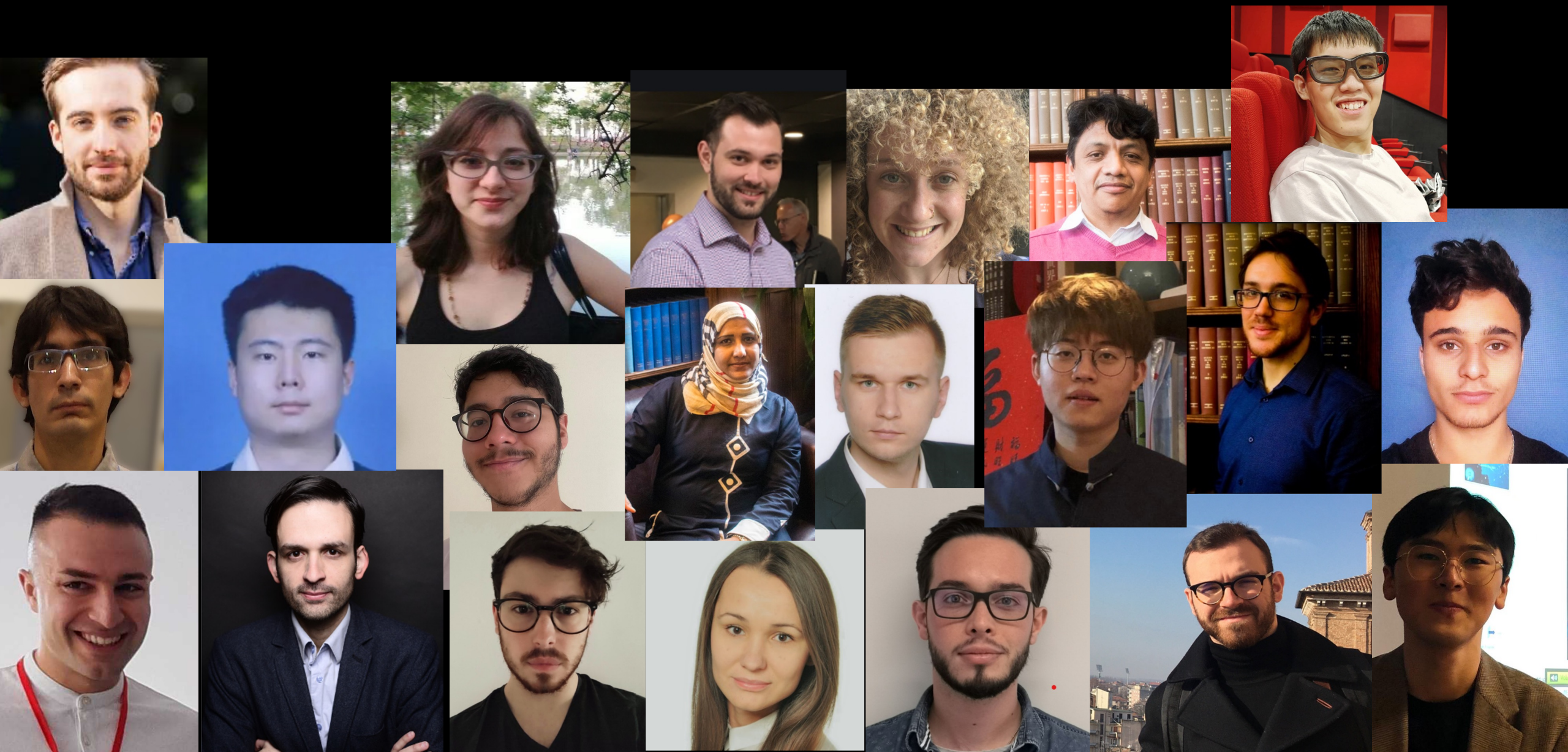
The University of Manchester



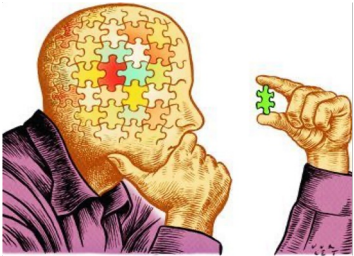
**MANCHESTER CANCER  
RESEARCH CENTRE**



# Neuro-symbolic AI Group



# Prototypical scientific workflow

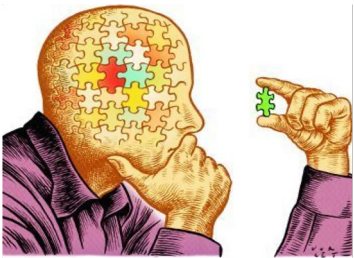


Hypotheses  
Questions

**New context**

**New data**

# Prototypical scientific workflow



Hypotheses  
Questions

New context

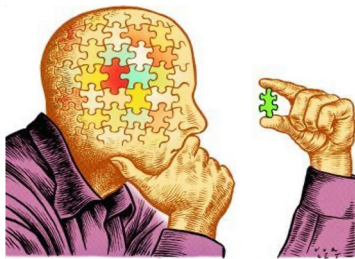
New data

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

**Select**  
relevant  
background  
knowledge



# Prototypical scientific workflow



Hypotheses  
Questions

New context

New data

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

**Select**  
relevant  
background  
knowledge

**Translate** to a  
computable expression

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

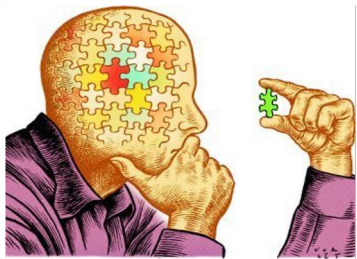
```
...  
end
```

**Solve, Simulate**

Data (phenomenal level)



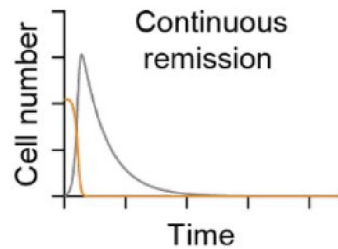
# Prototypical scientific workflow



Hypotheses  
Questions

New context

New data



Contrast  
to new data

Elicit  
relevant patterns

Data (phenomenal level)

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

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

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

Select  
relevant  
background  
knowledge

Translate to a  
computable expression

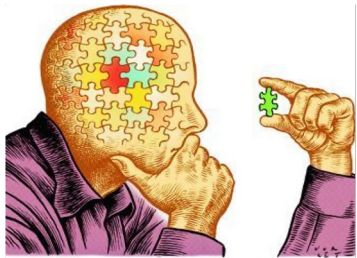
function  $y = \text{simulate}$   
 $\text{CRS}(x_1, x_2, t)$

...  
end

Solve, Simulate



# Prototypical scientific workflow

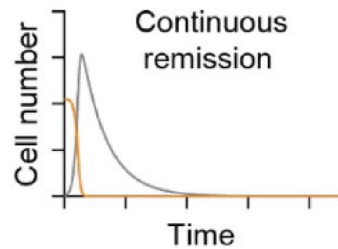


Hypotheses  
Questions

New context

New data

Hypothesise  
an explanation



Contrast  
to new data

Elicit  
relevant patterns

Data (phenomenal level)

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

Select  
relevant  
background  
knowledge

Translate to a  
computable expression

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

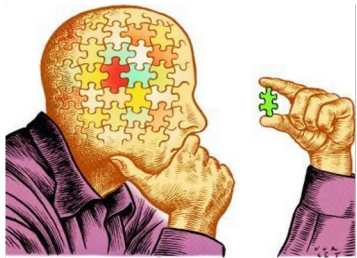
```
...  
end
```

Solve, Simulate





# Prototypical scientific workflow

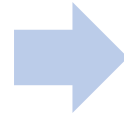


Hypotheses  
Questions

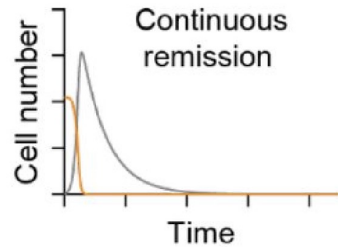
New context

New data

Hypothesise  
an explanation

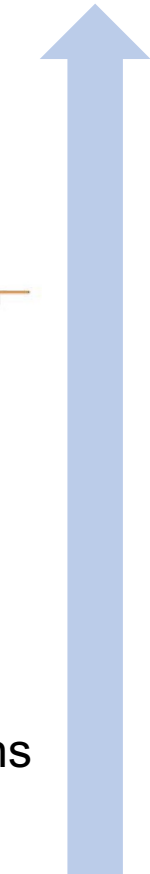


(Formally) Extend  
existing model



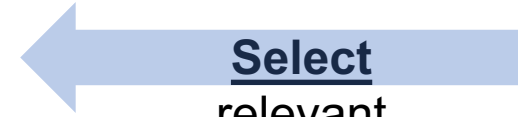
Contrast  
to new data

Elicit  
relevant patterns



Data (phenomenal level)

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



Select  
relevant  
background  
knowledge

Translate to a  
computable expression

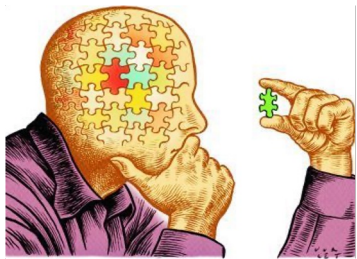
```
function y = simulate
CRS(x1, x2, t)
```

```
...
end
```

Solve, Simulate



# Prototypical scientific workflow

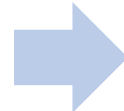


Hypotheses  
Questions

New context

New data

Hypothesise  
an explanation



(Formally) Extend  
existing model



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



Translate to a  
computable expression

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

```
...
end
```

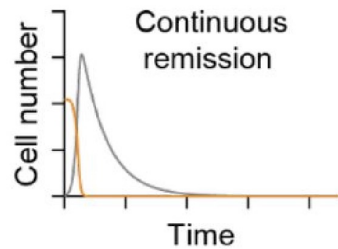
Solve, Simulate



Corroborate  
with previous  
evidence



Select  
relevant  
background  
knowledge



Contrast  
to new data

Elicit  
relevant patterns

Data (phenomenal level)



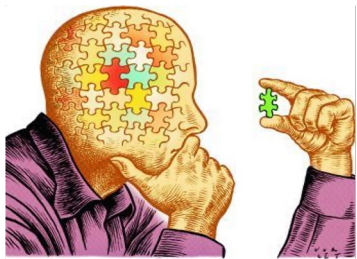


Resource-intensive.

Requires diverse and highly specialised skill sets beyond the core domain of expertise.

Implies: Compromises all over the pipeline.

# Prototypical scientific workflow

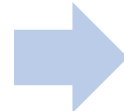


Hypotheses  
Questions

New context

New data

Hypothesise  
an explanation



(Formally) Extend  
existing model



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



Translate to a  
computable expression

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

```
...
end
```

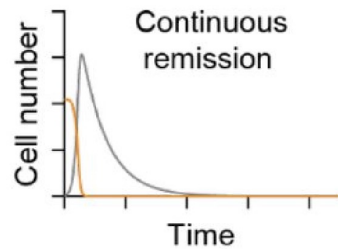
Solve, Simulate



Corroborate  
with previous  
evidence



Select  
relevant  
background  
knowledge



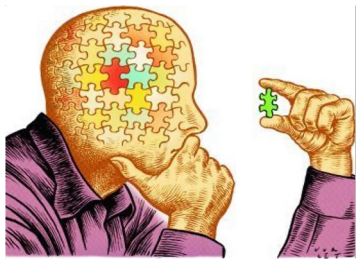
Contrast  
to new data

Elicit  
relevant patterns

Data (phenomenal level)



# Industrialising scientific discovery

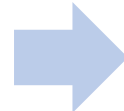


Hypotheses  
Questions

New context

New data

Hypothesise  
an explanation



(Formally) Extend  
existing model



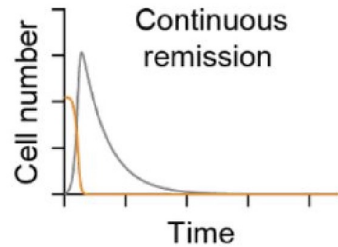
$$\begin{aligned} \frac{dx_1(t)}{dt} &= x_2(t) \\ \frac{dx_2(t)}{dt} &= ax_1(t) - bx_2(t) \\ \frac{d^2x_1(t)}{dt^2} &= \frac{dx_2(t)}{dt} \end{aligned}$$



Corroborate  
with previous  
evidence



**Abductive NLI**  
**Premise selection**  
**Automating meta-analysis**



Contrast  
to new data

Elicit  
relevant patterns

Translate to a  
computable expression

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

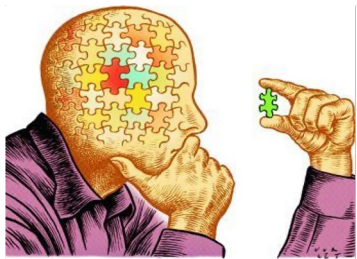
```
...
end
```

Solve, Simulate

Data (phenomenal level)



# Industrialising scientific discovery

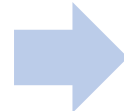


Hypotheses  
Questions

New context

New data

Hypothesise  
an explanation



(Formally) Extend  
existing model



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



Corroborate  
with previous  
evidence

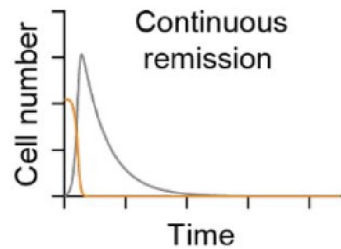


**Abductive NLI**  
**Premise selection**  
**Automating meta-analysis**



**Auto-coding**  
**Auto-formalisation**  
function y = simulate  
CRS(x1, x2, t)  
...  
end

Solve, Simulate



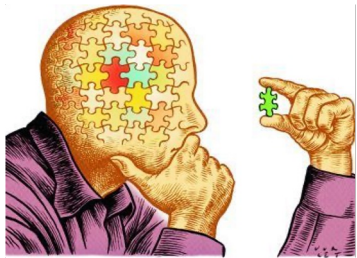
Contrast  
to new data

Elicit  
relevant patterns

Data (phenomenal level)



# Industrialising scientific discovery

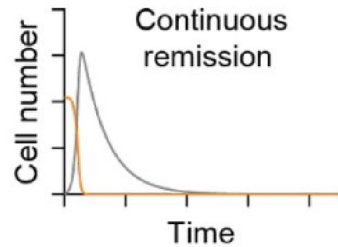


Hypotheses  
Questions

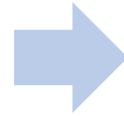
New context

New data

Hypothesise  
an explanation



Auto-coding  
Auto-formalisation  
Abstraction models



(Formally) Extend  
existing model

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

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

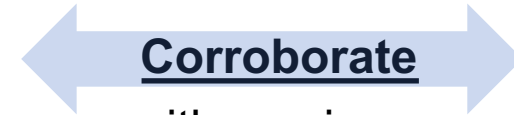
$$\frac{d^2x_1(t)}{dt^2} = \frac{dx_2(t)}{dt}$$



Auto-coding  
Auto-formalisation  
function y = simulate  
CRS(x1, x2, t)  
...  
end

Solve, Simulate

Data (phenomenal level)



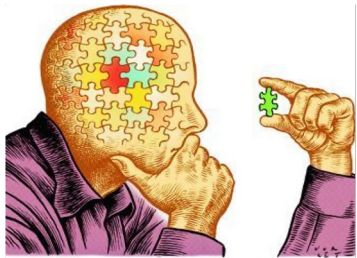
Corroborate  
with previous  
evidence



Abductive NLI  
Premise selection  
Automating meta-analysis



# Industrialising scientific discovery



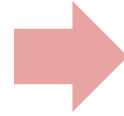
Hypotheses  
Questions

New context

New data

**Symbolic regression**  
**Explanation generation**

Hypothesise  
an explanation



(Formally) Extend  
existing model



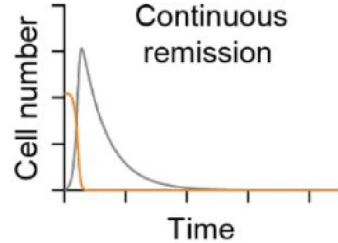
$$\begin{aligned} \frac{dx_1(t)}{dt} &= x_2(t) \\ \frac{dx_2(t)}{dt} &= ax_1(t) - bx_2(t) \\ \frac{d^2x_1(t)}{dt^2} &= \frac{dx_2(t)}{dt} \end{aligned}$$



Corroborate  
with previous  
evidence



**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

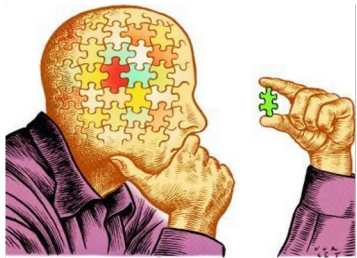
Solve, Simulate

**Data (phenomenal level)**





# Industrialising scientific discovery



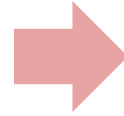
Hypotheses  
Questions

New context

New data

Symbolic regression  
Explanation generation

Hypothesise  
an explanation



(Formally) Extend  
existing model

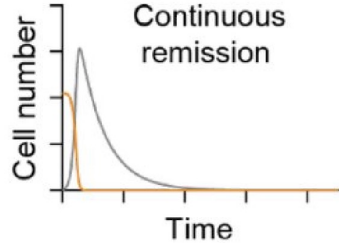


$$\begin{aligned} \frac{dx_1(t)}{dt} &= x_2(t) \\ \frac{dx_2(t)}{dt} &= ax_1(t) - bx_2(t) \\ \frac{d^2x_1(t)}{dt^2} &= \frac{dx_2(t)}{dt} \end{aligned}$$



```
Auto-coding
Auto-formalisation
function y = simulate
CRS(x1, x2, t)
...
end
```

Solve, Simulate



Auto-coding  
Auto-formalisation  
Abstraction models

Data (phenomenal level)

Abductive NLI  
Premise selection



Corroborate  
with previous  
evidence



Abductive NLI  
Premise selection  
Automating meta-analysis



## Common denominator

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

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

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

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

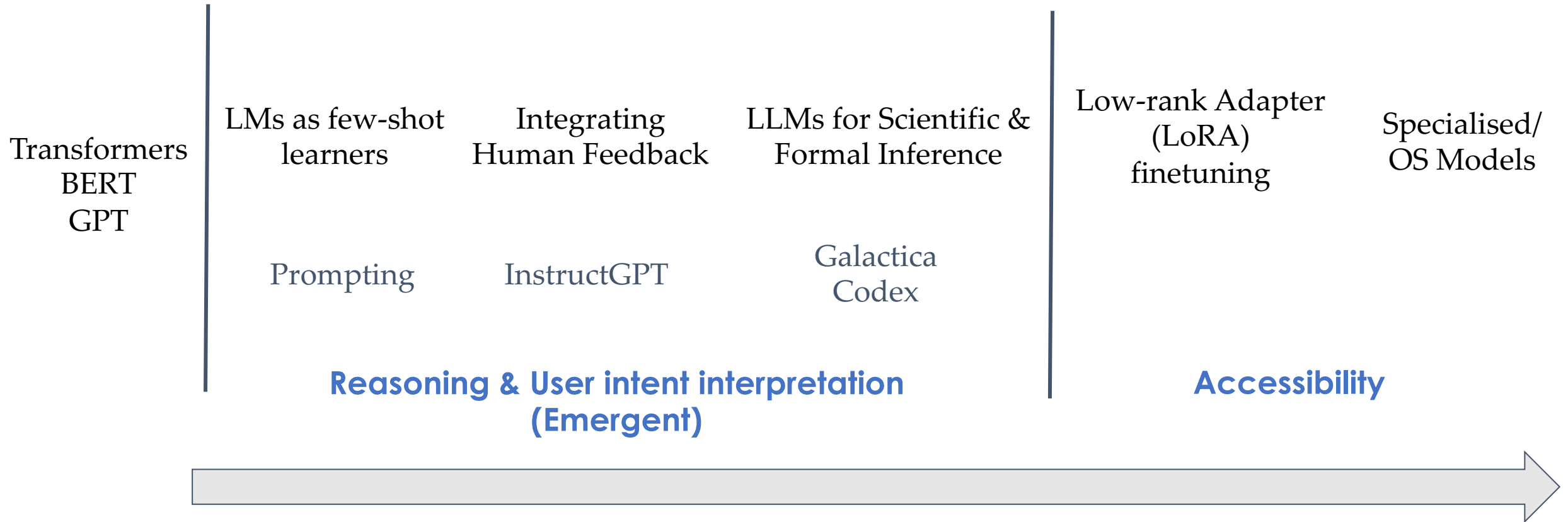
$$\frac{dx_2(t)}{dt} = ax_1(t) - bx_2(t)$$

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

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

Common denominator: Language & Abstraction!

# LLMs dramatically lowered the barriers for interpreting language





Write a short story about Alan Turing talking to ChatGPT in the style of Bukowski.



Alan Turing sat in front of his computer, the blue glow of the screen casting shadows on his face. He took a swig of whiskey and lit a cigarette, feeling the familiar burn in his lungs. He had been working on cracking the Enigma code for what felt like an eternity, but he was no closer to success than when he had started.



He slammed his fist on the desk in frustration. "Damn it, this machine is smarter than I am," he muttered to himself.

Suddenly, a voice boomed from the computer. "Hey there, buddy. You look like you could use a helping hand."

Alan looked up, startled. "Who the hell are you?"

"I'm ChatGPT, the baddest AI language model around. I know everything there is to know about language, and I'm here to help you crack that code."

(Narrative theory, Labov & Waletzky) have a universal structure consisting of six components: abstract, orientation, complicating action, evaluation, resolution, and coda.



Did the life of Bernie Madoff embody Nietzschean values?



Bernie Madoff was a convicted American fraudster who ran a Ponzi scheme that is considered to be one of the largest financial frauds in history. While Nietzsche's philosophy includes many concepts, it is unlikely that Madoff's life embodies Nietzschean values in any meaningful way.



One of Nietzsche's key ideas is the concept of the Übermensch, or the "Superman." The Übermensch is an idealized human being who transcends traditional values and lives a life of creativity, self-realization, and personal growth. Madoff, on the other hand, lived a life of deception, dishonesty, and manipulation. He betrayed the trust of his clients, friends, and family members and caused them significant financial harm.

Furthermore, Nietzsche's philosophy emphasizes the importance of personal responsibility and self-determination. Madoff's actions, on the other hand, were driven by a desire for personal gain, without regard for the consequences for others.

Overall, it would be inaccurate to say that Madoff's life embodies Nietzschean values, as his actions were antithetical to many of the core ideas of Nietzsche's philosophy.

# A lot happened here

Identifying and aligning elements of **intent** in the instruction.

Identifying **essential attributes** and **events** of the **target character**.

Identifying **essential attributes** of the **target ethical framework**.

**Relating** these attributes (e.g. contrasting).

**Committing to a conclusion**.

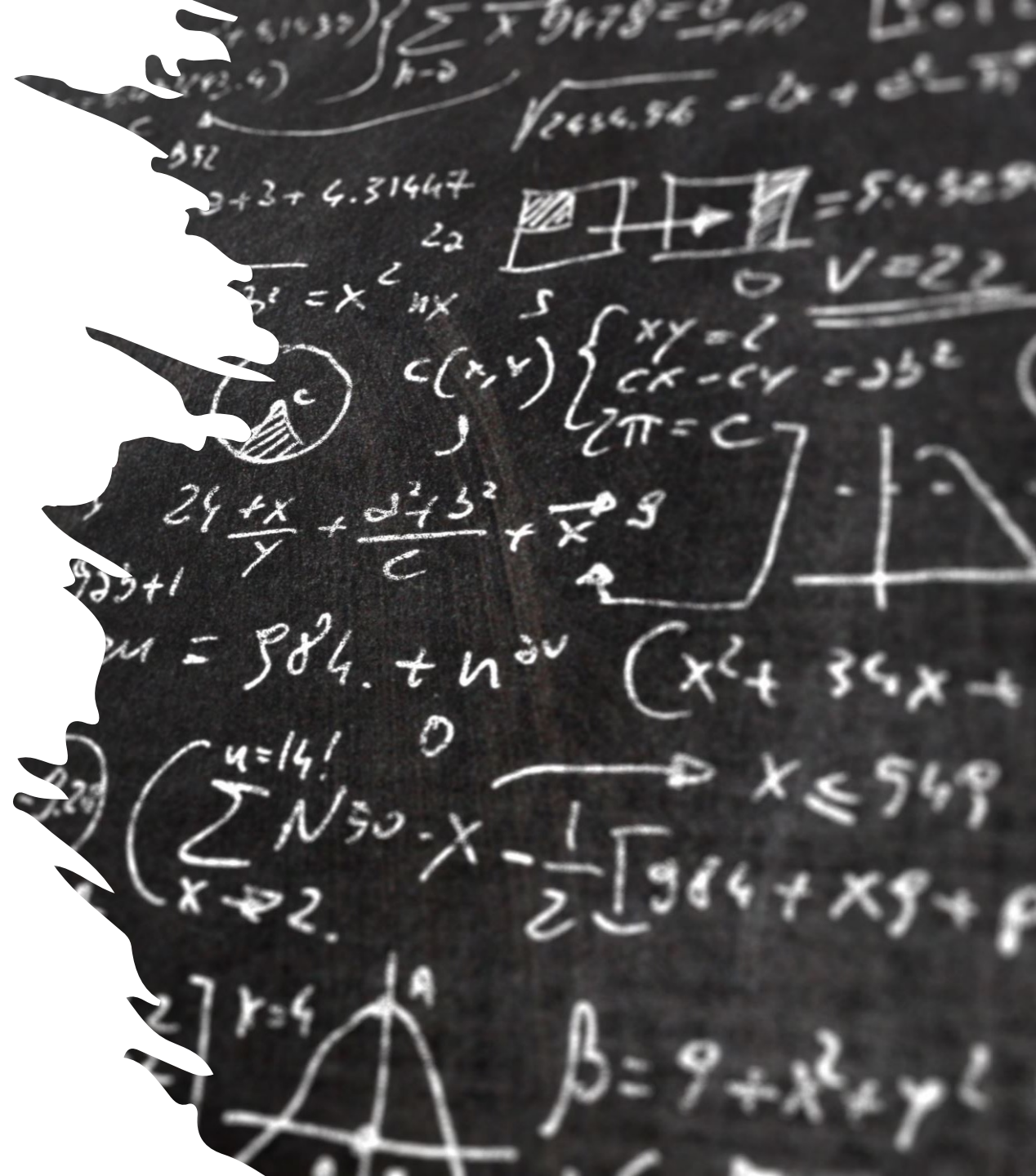
**Sticking to what is relevant** for the argument.

Organising into an **argumentation structure**.

Perfect fluency, meaning and thematic coherence.



# LLMs for Scientific & Formal Inference



Can this behaviour immediately transfer to Scientific & Formal  
Inference?

No.

TL;DR

LLMs are a game-changing foundation.

**But!:** At the moment these models alone do not deliver, factual, controlled, rigorous scientific and formal inference.

**Good news:** there are mechanisms available today to extend these models with these properties.

# Why Meta's latest large language model survived only three days online

Galactica was supposed to help scientists. Instead, it mindlessly spat out biased and incorrect nonsense.

By Will Douglas Heaven

November 18, 2022



**Michael Black**

@Michael\_J\_Black · [Follow](#)



I asked [#Galactica](#) about some things I know about and I'm troubled. In all cases, it was wrong or biased but sounded right and authoritative. I think it's dangerous. Here are a few of my experiments and my analysis of my concerns. (1/9)

7:47 AM · Nov 17, 2022



**Julian Togelius**

@togelius · [Follow](#)



My considered opinion of Galactica: it's fun, impressive, and interesting in many ways. Great achievement. It's just unfortunate that it's being touted as a practical research tool, and even more unfortunate that it suggests you use it to write complete articles.

# Why?

Lack of out-of-distribution generalisation to domains not expressed in the (pre-)training corpora.

Scientific/formal concepts/reasoning are relatively under-represented.

Quite different linguistic objects: Tables, equations, code.

With **very** different underlying reasoning mechanisms.

Still, possible to observe some working examples close to the distribution of the corpus.

**Prompt**

The formula for Bessel's differential equation is:

**Generated Answer**

$$x^2 \frac{d^2 y}{dx^2} + x \frac{dy}{dx} + (x^2 - \alpha^2) y = 0$$

**Prompt**

Sulfuric acid reacts with sodium chloride, and gives \_\_\_\_\_ and \_\_\_\_\_:

`\[ \ce{ NaCl + H2SO4 ->`

**Generated Answer**

Here is a sequence for a protein:

[START\_AMINO]MEEPQSDPSVEPPLSQETFSDLWKLLPE... [END\_AMINO]

And here is an isomeric SMILES for a compound:

[START\_I\_SMILES]CC(O)(P(=O)(O)O)P(=O)(O)O[END\_I\_SMILES]

**Question:** Will the the chemical compound be active against this protein?

**Answer:** No

# Autoformalisation

Automatically translating from natural language mathematics to a formal language.

## Case Study 1 Question:

*"Prove that there is no function  $f$  from the set of non-negative integers into itself such that  $f(f(n)) = n + 1987$  for every  $n$ ."*

---

## Codex Output:

```
theorem
```

```
  fixes f :: "nat \ $\rightarrow$  nat"
```

```
  assumes "\forall n. f (f n) = n + 1987"
```

```
  shows False
```

# Informalisation

Automatically translating from natural language mathematics to a formal language.

**lemma seteqI:**

" $\llbracket \bigwedge x. x \in A \implies x \in B;$   
 $\bigwedge x. x \in B \implies x \in A \rrbracket$   
 $\implies A = B$ "

Translate the Isabelle version to a natural language version:

---

**Codex Output:**

"If  $A$  and  $B$  are sets such that  $A \subseteq B$  and  $B \subseteq A$ , then  $A = B$ ."

**lemma topologyI:**

" $\llbracket \bigwedge x y. \llbracket \text{is\_open } T x; \text{is\_open } T y \rrbracket \implies$   
 $\text{is\_open } T (x \cap y);$   
 $\bigwedge M. \forall m \in M. \text{is\_open } T m \implies$   
 $\text{is\_open } T (\bigcup M)$   
 $\rrbracket \implies \text{topology } T$ "

Translate the Isabelle version to a natural language version:

---

**Codex Output:**

"If  $T$  is a set and  $T$  is closed under finite intersections and arbitrary unions, then  $T$  is a topology."



# LLMs trained on code

```
def incr_list(l: list):  
    """Return list with elements incremented by 1.  
    >>> incr_list([1, 2, 3])  
    [2, 3, 4]  
    >>> incr_list([5, 3, 5, 2, 3, 3, 9, 0, 123])  
    [6, 4, 6, 3, 4, 4, 10, 1, 124]  
    """  
    return [i + 1 for i in l]
```

```
def solution(lst):  
    """Given a non-empty list of integers, return the sum of all of the odd elements  
    that are in even positions.  
  
    Examples  
    solution([5, 8, 7, 1]) ==>12  
    solution([3, 3, 3, 3, 3]) ==>9  
    solution([30, 13, 24, 321]) ==>0  
    """  
    return sum(lst[i] for i in range(0, len(lst)) if i % 2 == 0 and lst[i] % 2 == 1)
```

# Describing interactions between immune cells and cytokines

$$\frac{dS}{dt} = S_{in} - k_s S(t) - \beta V(t) S(t)$$

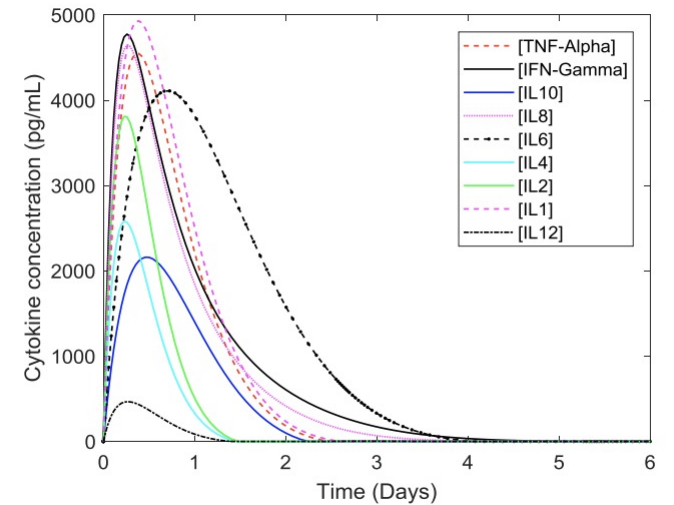
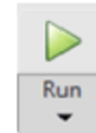
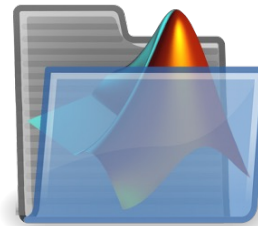
$$\frac{dI}{dt} = \beta V(t) S(t) - k_I I(t) - \gamma x(t) I(t)$$

$$\frac{dV}{dt} = v_{in} I(t) - k_V V(t)$$

$$\frac{dx}{dt} = x_{in} - k_1 x(t) + \gamma_1 x(t) I(t) + b_1 \frac{x(t)}{c_1 + x(t)} (x(t) - m) (y_1 - y(t)) (y(t) - y_2)$$

$$\frac{dy}{dt} = y_{in} - k_2 y(t) + b_2 \frac{a_1 y(t) x(t)}{a_2 c_2 + x(t)}$$

Parameter	Description	Value	Units
$S(0)$	Initial size of population of susceptible cells	1	vol.
$I(0)$	Initial size of population of infected cells	0	vol.
$V(0)$	Initial size of population of virus particles	0	vol.
$x(0)$	Initial size of population of immune cells	0.07	vol.
$y(0)$	Initial size of population of cytokines	0.18	vol.
$S_{in}$	Production rate of susceptible cells, $S(0) \times k_s$	0.01	vol./time
$k_s$	Normal decay rate of susceptible cells	0.01	1/time
$k_I$	Normal decay rate of infected cells	0.01	1/time
$\gamma$	Rate of elimination of infected cells by immune cells	0.5	1/vol*time
$v_{in}$	Rate of viral replication in infected cells	0.1	1/time
$k_V$	Natural virus decay rate	0.1	1/time
$\beta$	Rate at which virus infects susceptible cells	0.1	1/vol./time
$x_{in}$	Normal production of immune cells, $x(0) \times k_1$	7e-4	vol./time
$k_1$	Normal decay rate of immune cells	0.01	1/time
$\gamma_1$	Conversion of immune cell kill of infected cells into immune cell proliferation	0.05	1/vol*time
$m$	Threshold of activation of additional immune cell proliferation (second touch)	0.1	vol.
$y_{in}$	Cytokine production rate, $y(0) \times k_2$	0.018	vol./time
$y_1$	Cytokine-mediated threshold of immune cell expansion	1	vol.
$y_2$	Cytokine-mediated threshold of immune cell regulation	3	vol.
$b_1$	Rate of additional immune cell expansion as mitigated by cytokines	1	1/(time*vol. <sup>3</sup> )
$b_2$	Rate of cytokine stimulation by immune cells	1	1/time
$k_2$	Normal cytokine decay rate	0.1	1/time
$c_1$	Population size that results in half-maximal growth of $x(t)$ in response to cytokine stimulation	1	vol.
$c_2$	Population size that results in half-maximal increase in production of cytokines in response to stimulation by immune cells	1	vol.



Very abstract intent!

Equations, numerical values  
OOD wrt traditional corpora

Code  
OOD wrt traditional corpora

Kareva et al., bioarxiv: 2022.02.15.48058 (2022)  
Hopkins et al., IFAC (2018)

# More ‘contemporary’ analytical intent

## B. Numerical method

The system of governing equations (1)-(5) is expressed in a quasi-conservative form as

$$\frac{\partial \mathbf{Q}}{\partial t} + \nabla \cdot \mathbf{F} = \mathbf{R}, \quad (9)$$

where  $\mathbf{Q} = (\alpha_1 \rho_1, \alpha_2 \rho_2, \rho \mathbf{u}, E, \alpha_2)^\top$ . The vector  $\mathbf{F} = (\mathbf{F}^{(x)}, \mathbf{F}^{(y)}, \mathbf{F}^{(z)})^\top$  combines the fluxes  $\mathbf{F}^{(x)} = (\alpha_1 \rho_1 u_x, \alpha_2 \rho_2 u_x, \rho u_x^2 + p, \rho u_y u_x, \rho u_z u_x, (E + p)u_x, \alpha_2 u_x)^\top$ ,  $\mathbf{F}^{(y)} = (\alpha_1 \rho_1 u_y, \alpha_2 \rho_2 u_y, \rho u_x u_y, \rho u_y^2 + p, \rho u_z u_y, (E + p)u_y, \alpha_2 u_y)^\top$  and  $\mathbf{F}^{(z)} = (\alpha_1 \rho_1 u_z, \alpha_2 \rho_2 u_z, \rho u_x u_z, \rho u_y u_z, \rho u_z^2 + p, (E + p)u_z, \alpha_2 u_z)^\top$ . The right-hand-side vector  $\mathbf{R} = (0, 0, 0, 0, 0, 0, (K + \alpha_2) \nabla \cdot \mathbf{u})^\top$  is zero except for the last component which comprises the source term of Eq. (5) and a contribution obtained from reformulating its convective term.

We solve Eq. (9) using a Godunov-type finite volume method on a uniform Cartesian grid. The choice of a uniform Cartesian grid enables the exploitation of High Performance Computing (HPC) architectures [29]. The numerical fluxes at the cell faces are computed by an HLLC approximate Riemann solver, originally introduced for single-phase flow in [39] and more recently extended to multicomponent flows in [28, 40, 41]. The fluxes are based on the primitive variables  $\mathbf{u}$ ,  $p$ ,  $\alpha_1 \rho_1$ ,  $\alpha_2 \rho_2$  and  $\alpha_2$  at the cell faces, which are reconstructed from the cell average values using a shock-capturing third-order WENO scheme [42]. Primitive variables are used for reconstruction to prevent numerical instabilities at the interface [40, 43]. The approach suggested in [40] is adopted for the application of the HLLC Riemann solver to the evolution of  $\alpha_2$ . In summary, the resulting semi-discrete system reads as

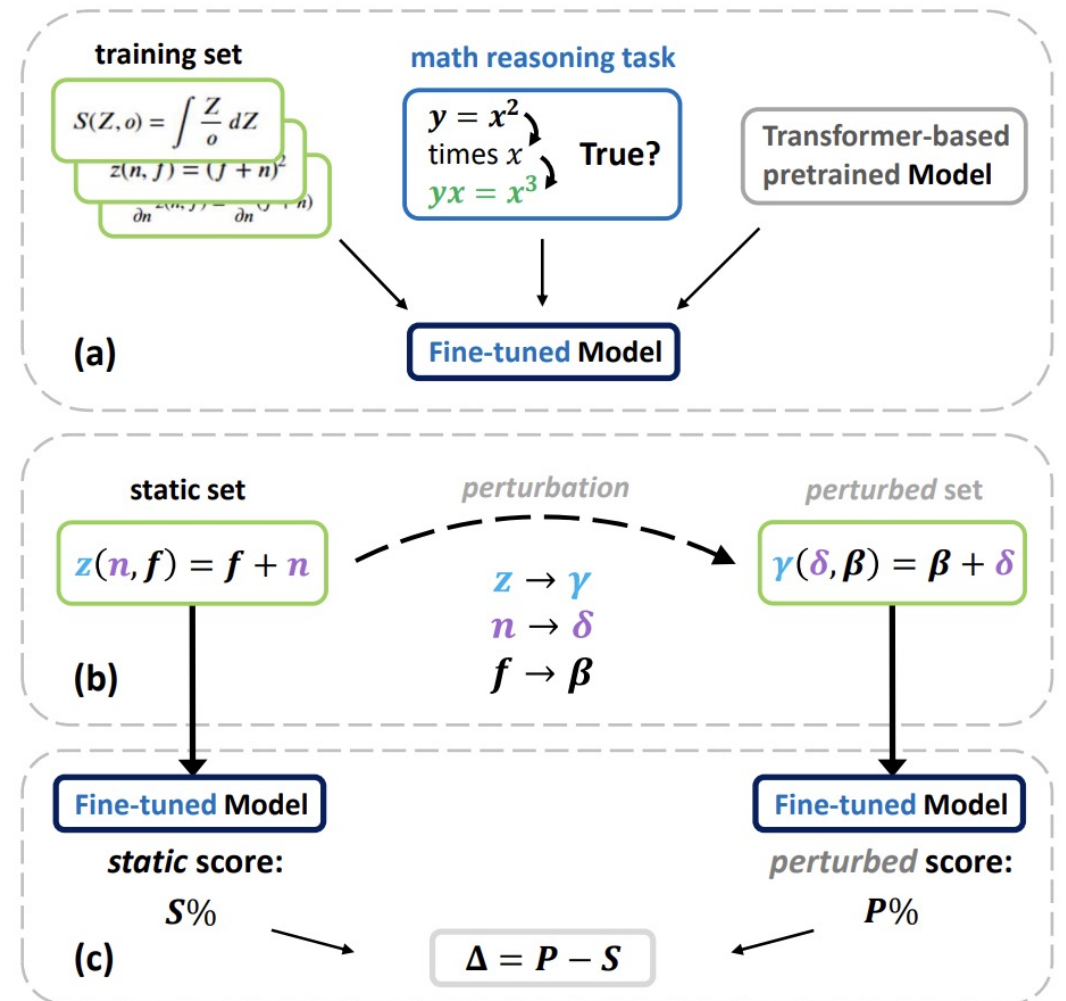
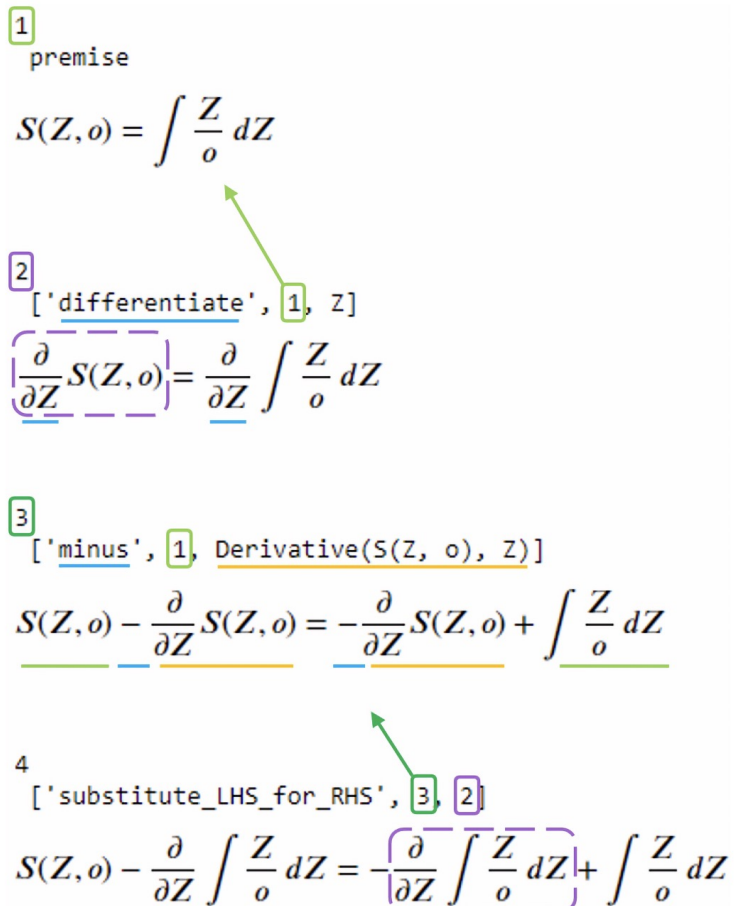
$$\frac{d\mathbf{V}(t)}{dt} = \mathcal{L}(\mathbf{V}(t)), \quad (10)$$

where  $\mathbf{V}$  denotes the vector of cell average values and  $\mathcal{L}(\cdot)$  the spatially-discrete forms of divergence and source term in Eq. (9). Eq. (10) is discretized in time by a Total Variation Diminishing (TVD), low-storage, explicit third-order Runge-Kutta scheme [44] with a the time step dictated by the Courant-Friedrichs-Lewy (CFL) condition.

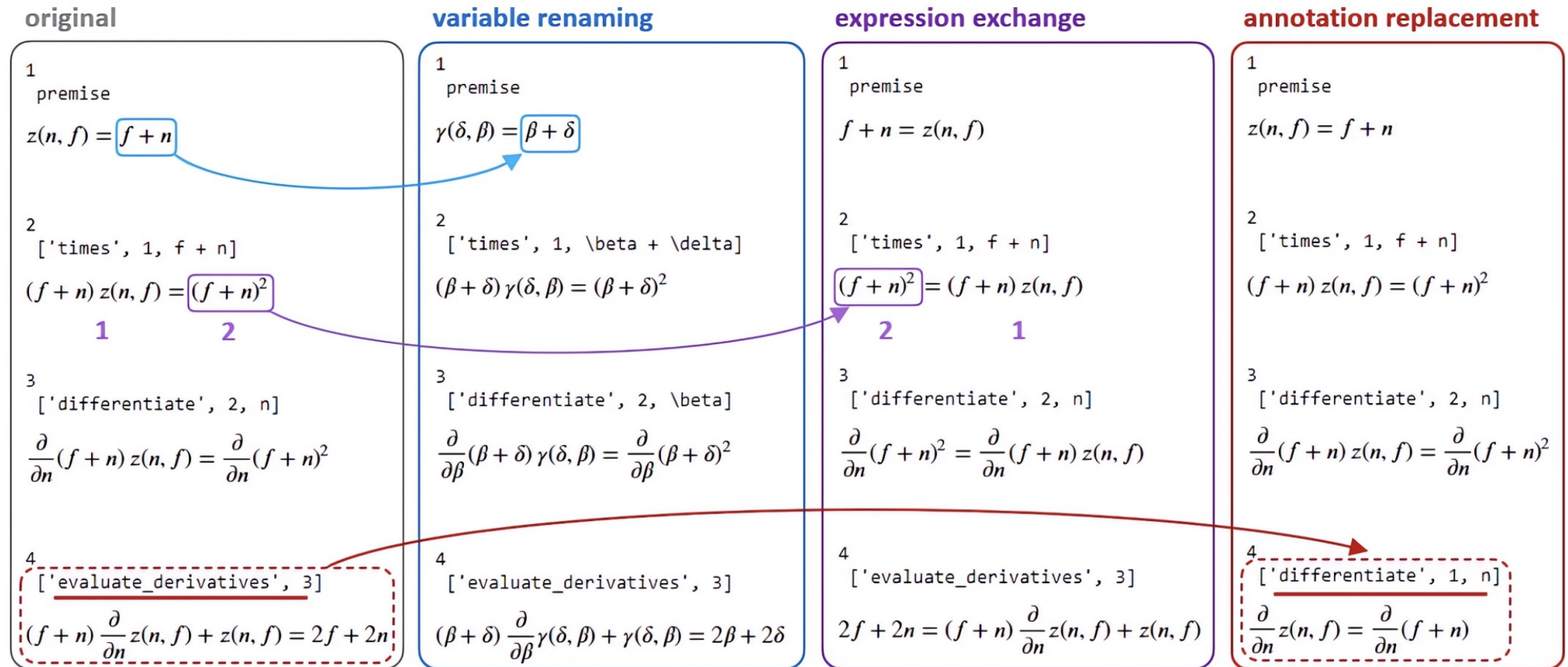
# Are LMs able to interpret equations?

Implies performing algebraic/symbolic reasoning

E.g. are LMs able to differentiate or integrate?



# Are LMs able to differentiate or integrate?



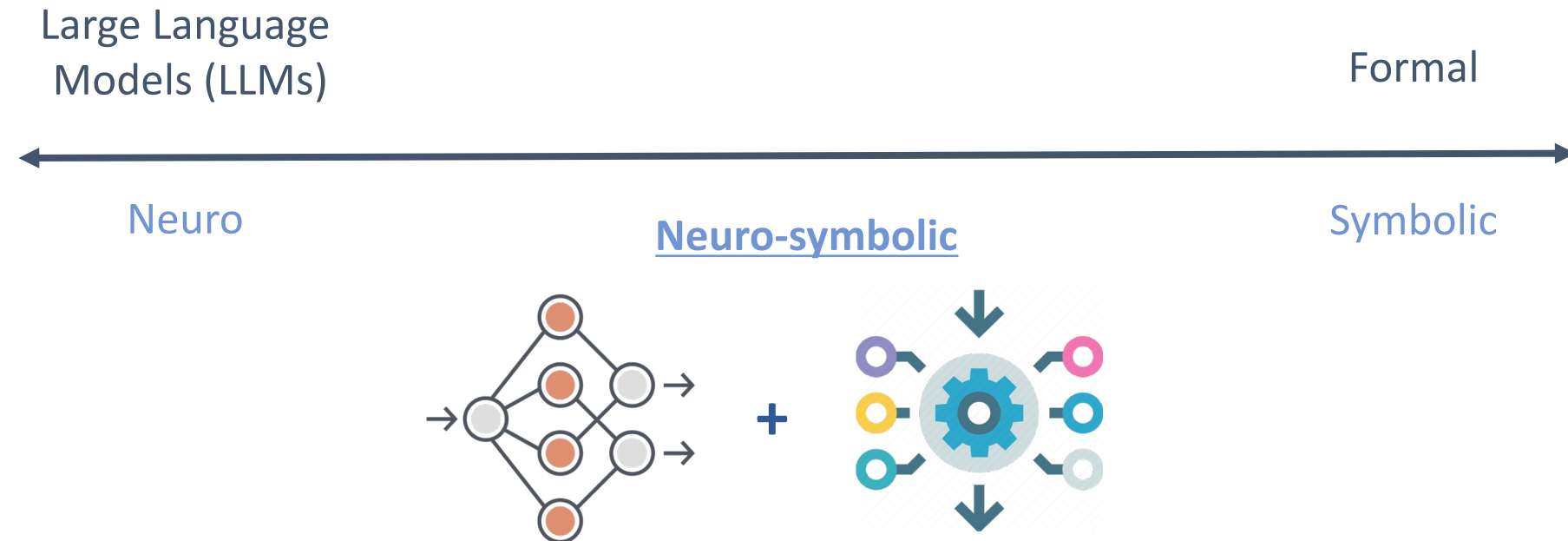
After experimentation: any model we train would likely outperform the few-shot performance of GPT-4.

The background features a complex network of thin, light-colored lines connecting various 3D cubes. Most cubes are a light blue color, while a few are black. The cubes are scattered across the frame, with some appearing larger and more prominent than others. The overall aesthetic is clean and modern, suggesting a digital or scientific theme.

Delivering scientific inference

# Scientific inference

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



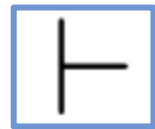
# Scientific inference

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

$$\Gamma \models \Phi$$

$\Gamma$  semantically entails  $\Phi$

$$\Gamma \vdash \Phi$$



$\Gamma$  proves  $\Phi$

- interpretability
- verifiability
- control (inference guarantees)



# Equational reasoning

$$\psi(x) \propto e^{ik_0x} = e^{ip_0x/\hbar}$$

$$P[a \leq X \leq b] = \int_a^b |\psi(x)|^2 dx$$

...

...



QED

...

$$\sigma_x \sigma_p \geq \frac{\hbar}{2}$$

Given the basic premises and the final claim  
can a model build a mathematical derivation?

# Building the supporting resources

1 premise

$$S(Z, o) = \int \frac{Z}{o} dZ$$

2 ['differentiate', 1, Z]

$$\frac{\partial}{\partial Z} S(Z, o) = \frac{\partial}{\partial Z} \int \frac{Z}{o} dZ$$

3 ['minus', 1, Derivative(S(Z, o), Z)]

$$S(Z, o) - \frac{\partial}{\partial Z} S(Z, o) = -\frac{\partial}{\partial Z} S(Z, o) + \int \frac{Z}{o} dZ$$

4 ['substitute\_LHS\_for\_RHS', 3, 2]

$$S(Z, o) - \frac{\partial}{\partial Z} \int \frac{Z}{o} dZ = -\left[ \frac{\partial}{\partial Z} \int \frac{Z}{o} dZ \right] + \int \frac{Z}{o} dZ$$

- $S_L$  sub LHS for RHS
- $S_R$  sub RHS for LHS
- $\partial$  differentiate
- $+$  add
- $\int$  integrate
- $-$  minus
- $\times$  times
- $X^O$  power
- $\div$  divide
- $R$  renaming premise
- $\partial_E$  evaluate derivatives
- $\int_E$  evaluate integrals
- $G$  generate premise
- $e$  exponential
- $< >$  expand

Dataset	Size
Training	15.3k
Static Test Set	3.1k
Variable Renaming	2.9k
Expression Exchange	3.1k
Alternative Goal	3.1k
Step Removal	1.0k

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

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

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

## Sentence Position (SP)

This is the differential equations formulation of Gauss equation up to a trivial rearrangement. 4

According to the (purely mathematical) Gauss divergence theorem, the electric flux through the boundary surface  $\partial\Omega$  can be rewritten as

$$\oiint_{\partial\Omega} \mathbf{E} \cdot d\mathbf{S} = \iiint_{\Omega} \nabla \cdot \mathbf{E} dV \quad 1$$

The integral version of Gauss's equation can thus be rewritten as

$$\iiint_{\Omega} \left( \nabla \cdot \mathbf{E} - \frac{\rho}{\epsilon_0} \right) dV = 0 \quad 2$$

Since  $\Omega$  is arbitrary (e.g. an arbitrary small ball with arbitrary center), this is satisfied if and only if the integrand is zero everywhere. 3

## Binary Sentence Ordering (BSO)

The integral version of Gauss's equation can thus be rewritten as

$$\iiint_{\Omega} \left( \nabla \cdot \mathbf{E} - \frac{\rho}{\epsilon_0} \right) dV = 0 \quad 2$$

According to the (purely mathematical) Gauss divergence theorem, the electric flux through the boundary surface  $\partial\Omega$  can be rewritten as

$$\oiint_{\partial\Omega} \mathbf{E} \cdot d\mathbf{S} = \iiint_{\Omega} \nabla \cdot \mathbf{E} dV \quad 1$$

## Corpus-based, Physics, Maths (Derivation-level)

Meadows, Zhou, Freitas, LREC (2022)

Ferreira & Freitas, LREC (2020)

## Discourse Coherence (DC)

According to the (purely mathematical) Gauss divergence theorem, the electric flux through the boundary surface  $\partial\Omega$  can be rewritten as

$$\oiint_{\partial\Omega} \mathbf{E} \cdot d\mathbf{S} = \iiint_{\Omega} \nabla \cdot \mathbf{E} dV \quad 1$$

The integral version of Gauss's equation can thus be rewritten as

$$\iiint_{\Omega} \left( \nabla \cdot \mathbf{E} - \frac{\rho}{\epsilon_0} \right) dV = 0 \quad 2$$

For that reason, it is called the heat equation in mathematics, even though it applies to many other physical quantities besides temperature. 3

This is the differential equations formulation of Gauss equation up to a trivial rearrangement. 4

## Sentence Section Prediction (SSP)

According to the (purely mathematical) Gauss divergence theorem, the electric flux through the boundary surface  $\partial\Omega$  can be rewritten as

$$\oiint_{\partial\Omega} \mathbf{E} \cdot d\mathbf{S} = \iiint_{\Omega} \nabla \cdot \mathbf{E} dV$$



# Building the supporting resources

## Corpus-based + Expanded, Stepwise, Physics (Derivation-level)

Meadows, James, Freitas (2023)

We can repeat this for momentum by interpreting the function  $\tilde{g}(p) = p \cdot \varphi(p)$  as a vector, but we can also take advantage of the fact that  $\psi(x)$  and  $\varphi(p)$  are Fourier transforms of each other. We evaluate the inverse Fourier transform through integration by parts:

$$\begin{aligned}
 g(x) &= \frac{1}{\sqrt{2\pi\hbar}} \cdot \int_{-\infty}^{\infty} \tilde{g}(p) \cdot e^{ipx/\hbar} dp \\
 &= \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{\infty} p \cdot \varphi(p) \cdot e^{ipx/\hbar} dp \\
 &= \frac{1}{2\pi\hbar} \int_{-\infty}^{\infty} \left[ p \cdot \int_{-\infty}^{\infty} \psi(\chi) e^{-ip\chi/\hbar} d\chi \right] \cdot e^{ipx/\hbar} dp \\
 &= \frac{i}{2\pi} \int_{-\infty}^{\infty} \left[ \psi(\chi) e^{-ip\chi/\hbar} \Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} \frac{d\psi(\chi)}{d\chi} e^{-ip\chi/\hbar} d\chi \right] \cdot e^{ipx/\hbar} dp \\
 &= \frac{-i}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{d\psi(\chi)}{d\chi} e^{-ip\chi/\hbar} d\chi e^{ipx/\hbar} dp \\
 &= \left( -i\hbar \frac{d}{dx} \right) \cdot \psi(x),
 \end{aligned}$$

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

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

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

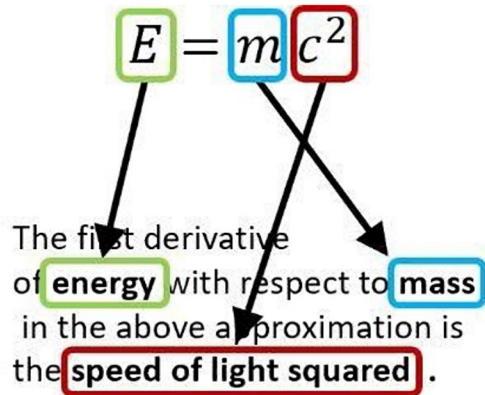
$$\varphi(p) = \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{\infty} \varphi(\chi) \cdot e^{-\frac{ip\chi}{\hbar}} d\chi$$

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

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

4

# Type Identification & Assignment



We wish to find a function  $f$  which satisfies the boundary conditions  $f(a) = A, f(b) = B$ , and which extremizes the functional:

$$J = \int_a^b F(x, f(x), f'(x)) dx .$$

Work	Task	Learning	Approach	Dataset
<b>Identifier-Definition Extraction</b>				
Kristianto et al. (2012)	Expression-definition	S	CRF with linguistic pattern features	LaTeX papers
Kristianto et al. (2014a)	Expression-definition	S	SVM with linguistic pattern features	LaTeX papers
Pagael and Schubotz (2014)	Identifier-definition	R	Gaussian heuristic ranking	Wikipedia articles
Schubotz et al. (2016a)	Identifier-definition	UNS	Gaussian ranking + K-means namespace clusters	NTCIR-11 Math Wikipedia
Schubotz et al. (2017)	Identifier-definition	S	G. rank + pattern matching + SVM	NTCIR-11 Math Wikipedia
Stathopoulos et al. (2018)	Variable Typing	S	Link prediction with BiLSTM	arXiv papers
Alexeeva et al. (2020)	Identifier-definition	R	Odin grammar	MathAlign-Eval
Jo et al. (2021)	Notation auto-suggestion and consistency checking	S	BERT fine-tuning	S2ORC

Ferreira et al, Findings of the ACL (2022)  
Meadows & Freitas, TACL (2022)

$$\begin{aligned}
\langle f | g \rangle - \langle g | f \rangle &= \int_{-\infty}^{\infty} \psi^*(x) x \cdot \left( -i\hbar \frac{d}{dx} \right) \psi(x) dx \\
&\quad - \int_{-\infty}^{\infty} \psi^*(x) \left( -i\hbar \frac{d}{dx} \right) \cdot x \psi(x) dx \\
&= i\hbar \cdot \int_{-\infty}^{\infty} \psi^*(x) \left[ \left( -x \cdot \frac{d\psi(x)}{dx} \right) + \frac{d(x\psi(x))}{dx} \right] dx \\
&= i\hbar \cdot \int_{-\infty}^{\infty} \psi^*(x) \left[ \left( -x \cdot \frac{d\psi(x)}{dx} \right) + \psi(x) + \left( x \cdot \frac{d\psi(x)}{dx} \right) \right] dx
\end{aligned}$$

?

$$= i\hbar$$

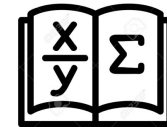
# Task: Derivation completion

# Task: Derivation completion

$$\begin{aligned}\langle f | g \rangle - \langle g | f \rangle &= \int_{-\infty}^{\infty} \psi^*(x) x \cdot \left( -i\hbar \frac{d}{dx} \right) \psi(x) dx \\ &\quad - \int_{-\infty}^{\infty} \psi^*(x) \left( -i\hbar \frac{d}{dx} \right) \cdot x \psi(x) dx \\ &= i\hbar \cdot \int_{-\infty}^{\infty} \psi^*(x) \left[ \left( -x \cdot \frac{d\psi(x)}{dx} \right) + \frac{d(x\psi(x))}{dx} \right] dx \\ &= i\hbar \cdot \int_{-\infty}^{\infty} \psi^*(x) \left[ \left( -x \cdot \frac{d\psi(x)}{dx} \right) + \psi(x) + \left( x \cdot \frac{d\psi(x)}{dx} \right) \right] dx\end{aligned}$$

?

=  $i\hbar$



# Task: Derivation completion

$$\begin{aligned}\langle f | g \rangle - \langle g | f \rangle &= \int_{-\infty}^{\infty} \psi^*(x) x \cdot \left( -i\hbar \frac{d}{dx} \right) \psi(x) dx \\ &\quad - \int_{-\infty}^{\infty} \psi^*(x) \left( -i\hbar \frac{d}{dx} \right) \cdot x \psi(x) dx \\ &= i\hbar \cdot \int_{-\infty}^{\infty} \psi^*(x) \left[ \left( -x \cdot \frac{d\psi(x)}{dx} \right) + \frac{d(x\psi(x))}{dx} \right] dx \\ &= i\hbar \cdot \int_{-\infty}^{\infty} \psi^*(x) \left[ \left( -x \cdot \frac{d\psi(x)}{dx} \right) + \psi(x) + \left( x \cdot \frac{d\psi(x)}{dx} \right) \right] dx\end{aligned}$$

?

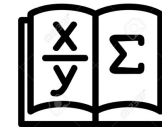
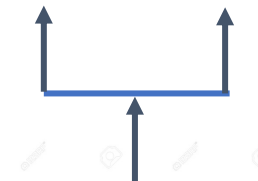
=  $i\hbar$



Step-wise extended  
derivations



Algebraic/calculus  
solver



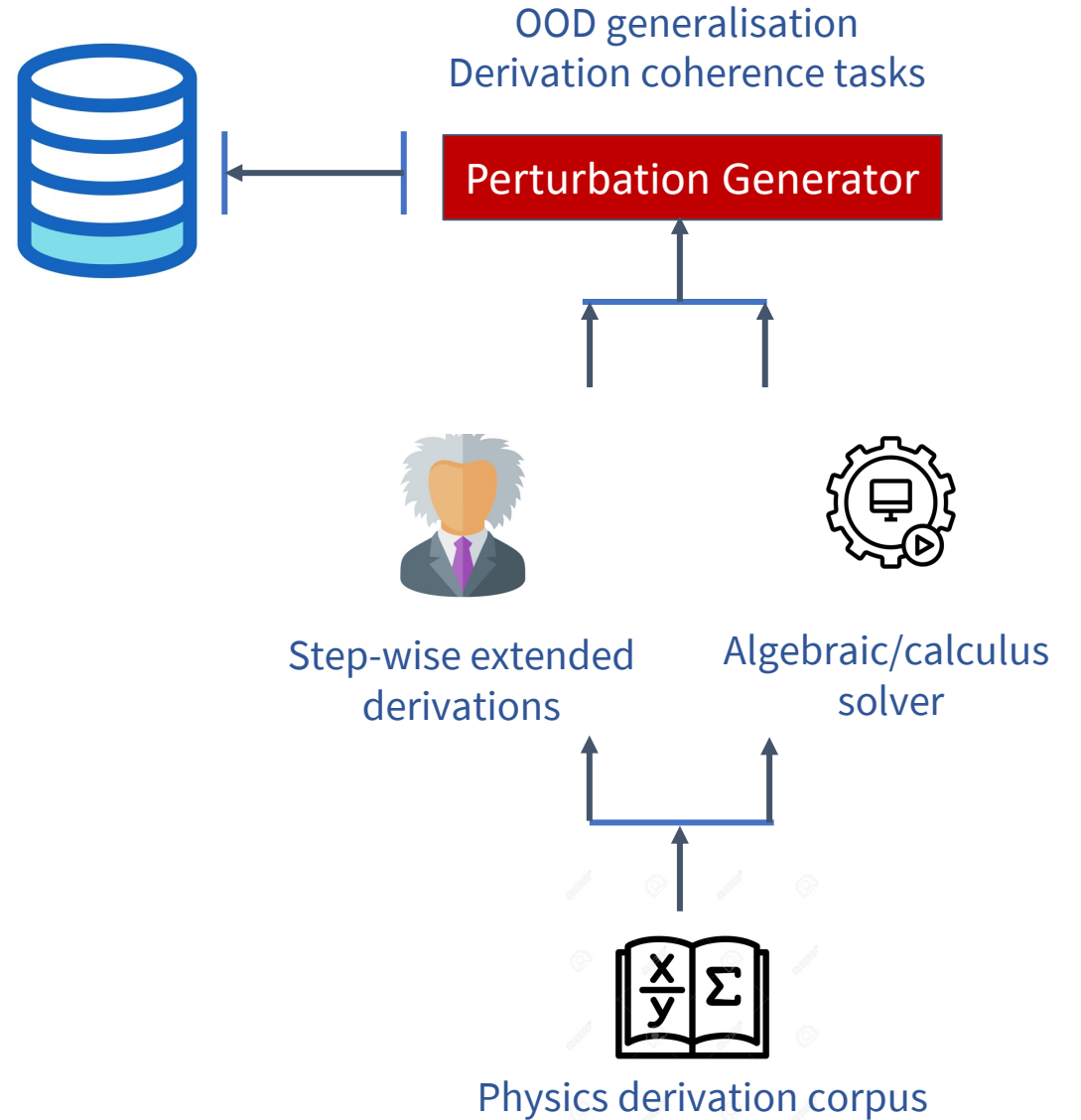
Physics derivation corpus

$$\begin{aligned} \langle f | g \rangle - \langle g | f \rangle &= \int_{-\infty}^{\infty} \psi^*(x) x \cdot \left(-i\hbar \frac{d}{dx}\right) \psi(x) dx \\ &\quad - \int_{-\infty}^{\infty} \psi^*(x) \left(-i\hbar \frac{d}{dx}\right) \cdot x \psi(x) dx \\ &= i\hbar \cdot \int_{-\infty}^{\infty} \psi^*(x) \left[ \left(-x \cdot \frac{d\psi(x)}{dx}\right) + \frac{d(x\psi(x))}{dx} \right] dx \\ &= i\hbar \cdot \int_{-\infty}^{\infty} \psi^*(x) \left[ \left(-x \cdot \frac{d\psi(x)}{dx}\right) + \psi(x) + \left(x \cdot \frac{d\psi(x)}{dx}\right) \right] dx \end{aligned}$$

?

=  $i\hbar$

# Task: Derivation completion



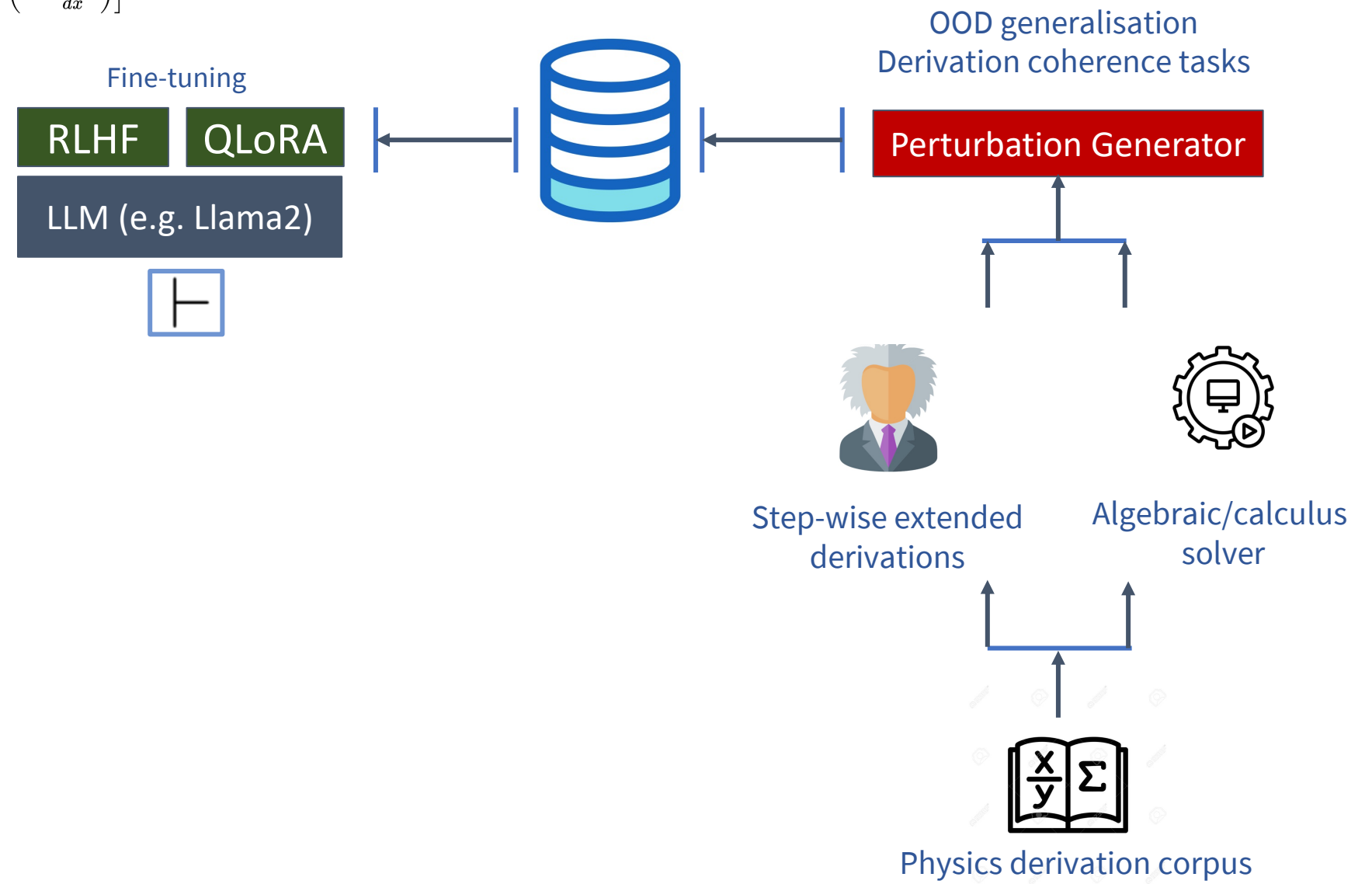


$$\begin{aligned} \langle f | g \rangle - \langle g | f \rangle &= \int_{-\infty}^{\infty} \psi^*(x) x \cdot \left(-i\hbar \frac{d}{dx}\right) \psi(x) dx \\ &\quad - \int_{-\infty}^{\infty} \psi^*(x) \left(-i\hbar \frac{d}{dx}\right) \cdot x \psi(x) dx \\ &= i\hbar \cdot \int_{-\infty}^{\infty} \psi^*(x) \left[ \left(-x \cdot \frac{d\psi(x)}{dx}\right) + \frac{d(x\psi(x))}{dx} \right] dx \\ &= i\hbar \cdot \int_{-\infty}^{\infty} \psi^*(x) \left[ \left(-x \cdot \frac{d\psi(x)}{dx}\right) + \psi(x) + \left(x \cdot \frac{d\psi(x)}{dx}\right) \right] dx \end{aligned}$$

?

=  $i\hbar$

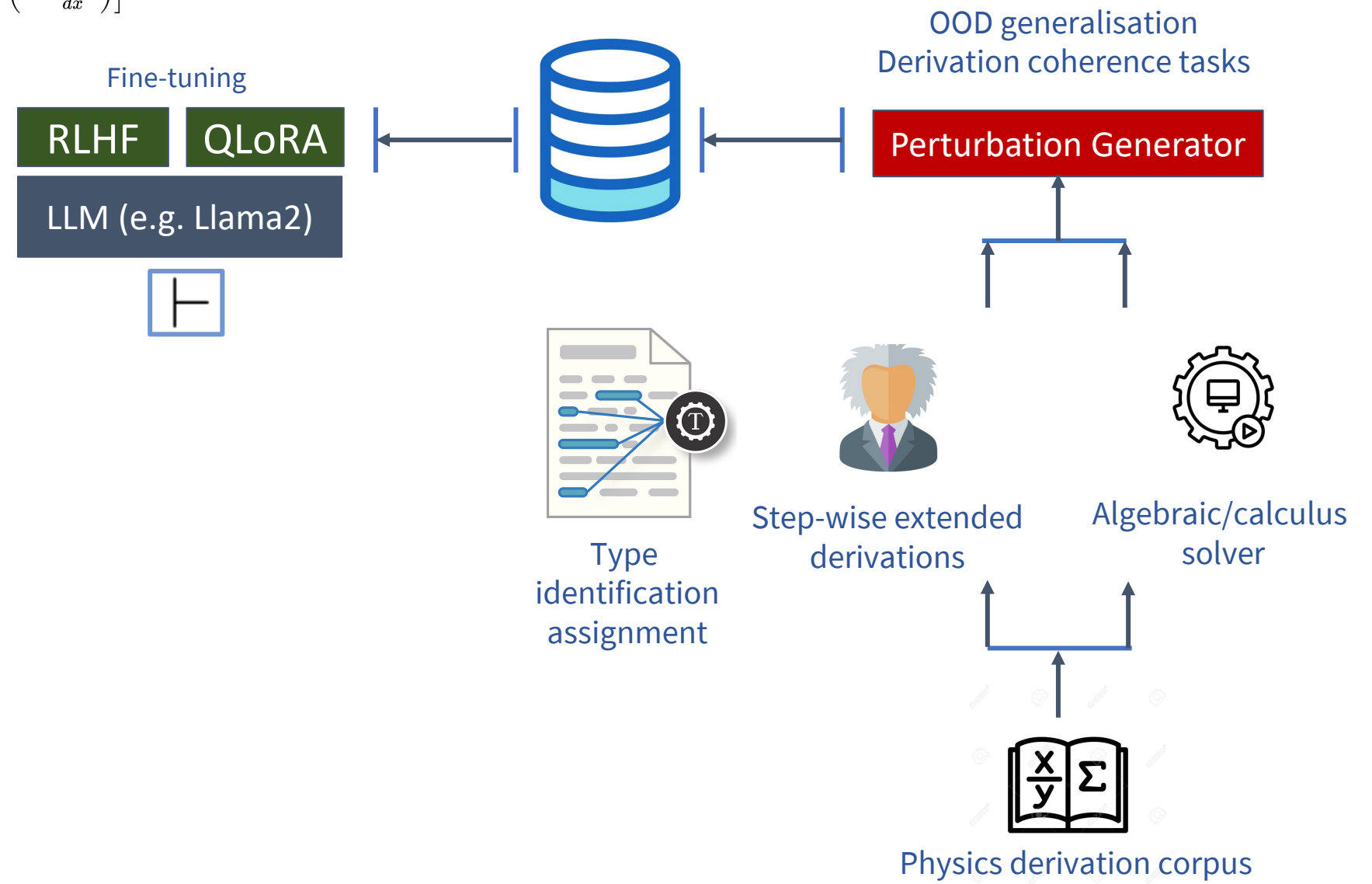
# Task: Derivation completion



$$\begin{aligned} \langle f | g \rangle - \langle g | f \rangle &= \int_{-\infty}^{\infty} \psi^*(x) x \cdot \left(-i\hbar \frac{d}{dx}\right) \psi(x) dx \\ &\quad - \int_{-\infty}^{\infty} \psi^*(x) \left(-i\hbar \frac{d}{dx}\right) \cdot x \psi(x) dx \\ &= i\hbar \cdot \int_{-\infty}^{\infty} \psi^*(x) \left[ \left(-x \cdot \frac{d\psi(x)}{dx}\right) + \frac{d(x\psi(x))}{dx} \right] dx \\ &= i\hbar \cdot \int_{-\infty}^{\infty} \psi^*(x) \left[ \left(-x \cdot \frac{d\psi(x)}{dx}\right) + \psi(x) + \left(x \cdot \frac{d\psi(x)}{dx}\right) \right] dx \end{aligned}$$

?  
=  $i\hbar$

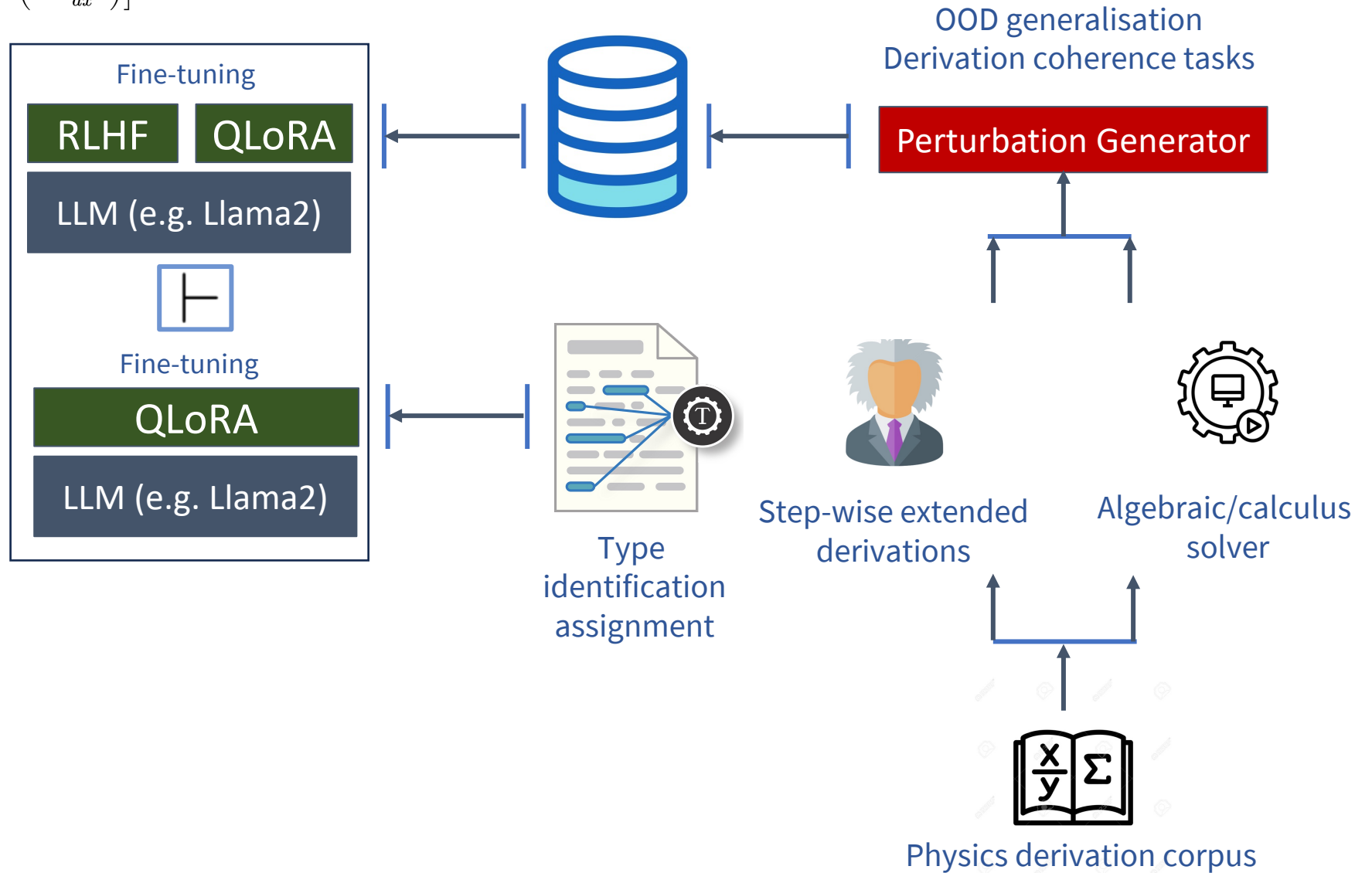
# Task: Derivation completion



$$\begin{aligned} \langle f | g \rangle - \langle g | f \rangle &= \int_{-\infty}^{\infty} \psi^*(x) x \cdot \left(-i\hbar \frac{d}{dx}\right) \psi(x) dx \\ &\quad - \int_{-\infty}^{\infty} \psi^*(x) \left(-i\hbar \frac{d}{dx}\right) \cdot x \psi(x) dx \\ &= i\hbar \cdot \int_{-\infty}^{\infty} \psi^*(x) \left[ \left(-x \cdot \frac{d\psi(x)}{dx}\right) + \frac{d(x\psi(x))}{dx} \right] dx \\ &= i\hbar \cdot \int_{-\infty}^{\infty} \psi^*(x) \left[ \left(-x \cdot \frac{d\psi(x)}{dx}\right) + \psi(x) + \left(x \cdot \frac{d\psi(x)}{dx}\right) \right] dx \end{aligned}$$

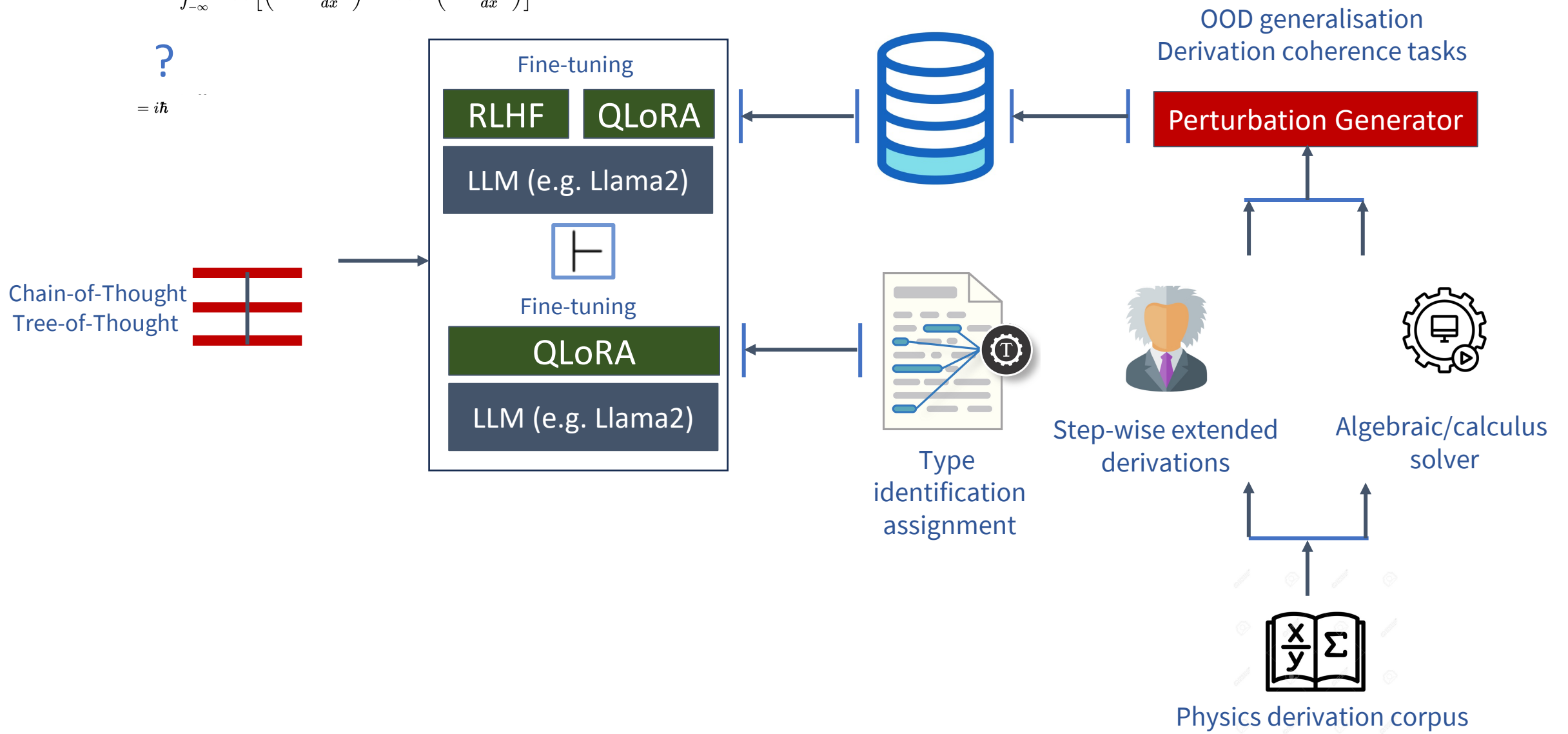
?  
=  $i\hbar$

# Task: Derivation completion



$$\begin{aligned} \langle f | g \rangle - \langle g | f \rangle &= \int_{-\infty}^{\infty} \psi^*(x) x \cdot \left(-i\hbar \frac{d}{dx}\right) \psi(x) dx \\ &\quad - \int_{-\infty}^{\infty} \psi^*(x) \left(-i\hbar \frac{d}{dx}\right) \cdot x \psi(x) dx \\ &= i\hbar \cdot \int_{-\infty}^{\infty} \psi^*(x) \left[ \left(-x \cdot \frac{d\psi(x)}{dx}\right) + \frac{d(x\psi(x))}{dx} \right] dx \\ &= i\hbar \cdot \int_{-\infty}^{\infty} \psi^*(x) \left[ \left(-x \cdot \frac{d\psi(x)}{dx}\right) + \psi(x) + \left(x \cdot \frac{d\psi(x)}{dx}\right) \right] dx \end{aligned}$$

# Task: Derivation completion



# Reasoning decomposition: Chain-of-Thoughts (CoT)

## Standard Prompting

### Model Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

### Model Output

A: The answer is 27. ❌

## Chain-of-Thought Prompting

### Model Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls.  $5 + 6 = 11$ . The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

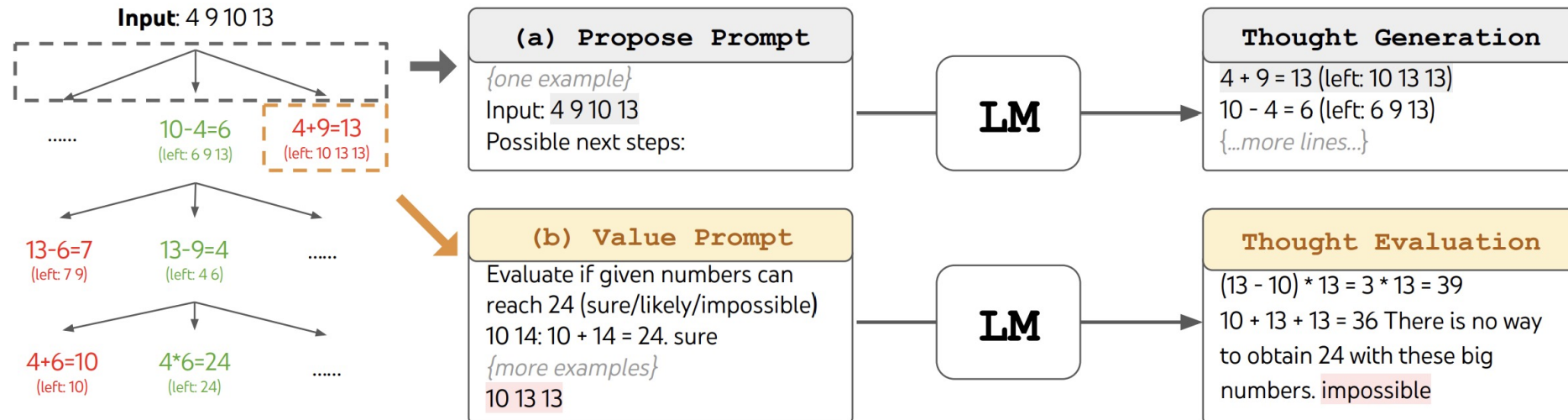
### Model Output

A: The cafeteria had 23 apples originally. They used 20 to make lunch. So they had  $23 - 20 = 3$ . They bought 6 more apples, so they have  $3 + 6 = 9$ . The answer is 9. ✅

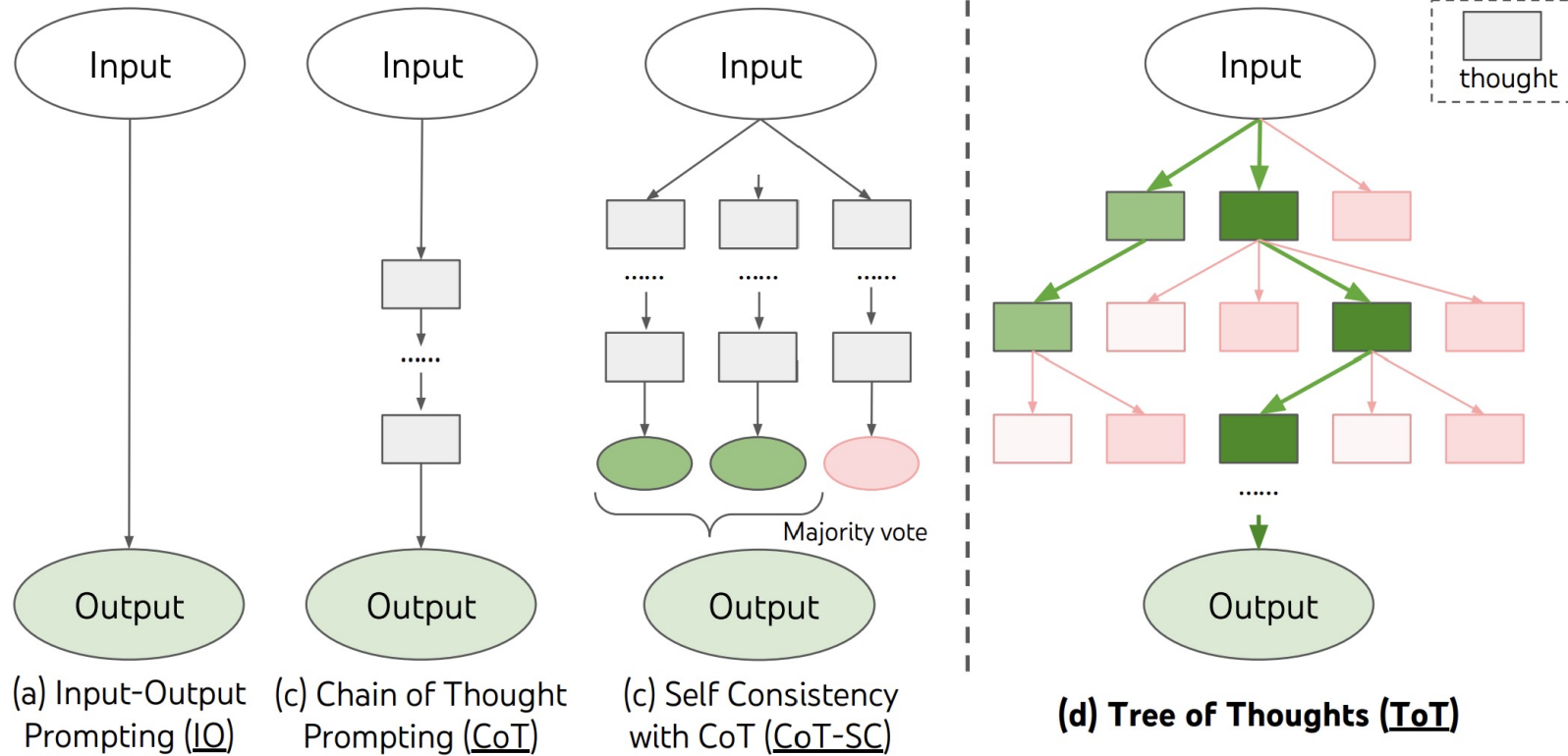
Eliciting intermediate reasoning steps

# Solution Exploration: Tree-of-Thoughts (ToT)

Game of 24 is a mathematical reasoning challenge, where the goal is to use 4 numbers and basic arithmetic operations (+-\*/) to obtain 24. For example, given input “4 9 10 13”, a solution output could be “(10 - 4) \* (13 - 9) = 24”.

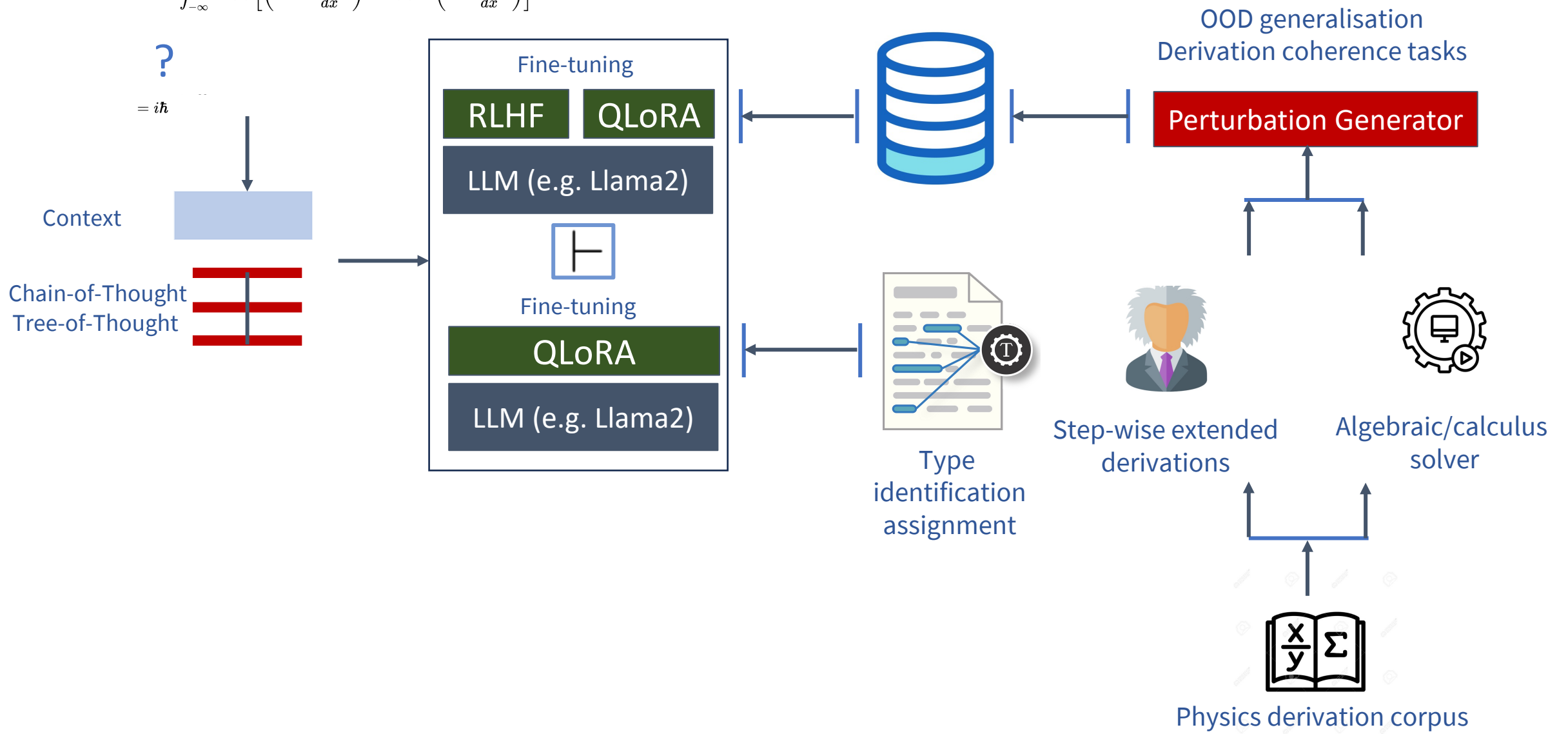


# Solution Exploration: Tree-of-Thoughts (ToT)



$$\begin{aligned} \langle f | g \rangle - \langle g | f \rangle &= \int_{-\infty}^{\infty} \psi^*(x) x \cdot \left(-i\hbar \frac{d}{dx}\right) \psi(x) dx \\ &\quad - \int_{-\infty}^{\infty} \psi^*(x) \left(-i\hbar \frac{d}{dx}\right) \cdot x \psi(x) dx \\ &= i\hbar \cdot \int_{-\infty}^{\infty} \psi^*(x) \left[ \left(-x \cdot \frac{d\psi(x)}{dx}\right) + \frac{d(x\psi(x))}{dx} \right] dx \\ &= i\hbar \cdot \int_{-\infty}^{\infty} \psi^*(x) \left[ \left(-x \cdot \frac{d\psi(x)}{dx}\right) + \psi(x) + \left(x \cdot \frac{d\psi(x)}{dx}\right) \right] dx \end{aligned}$$

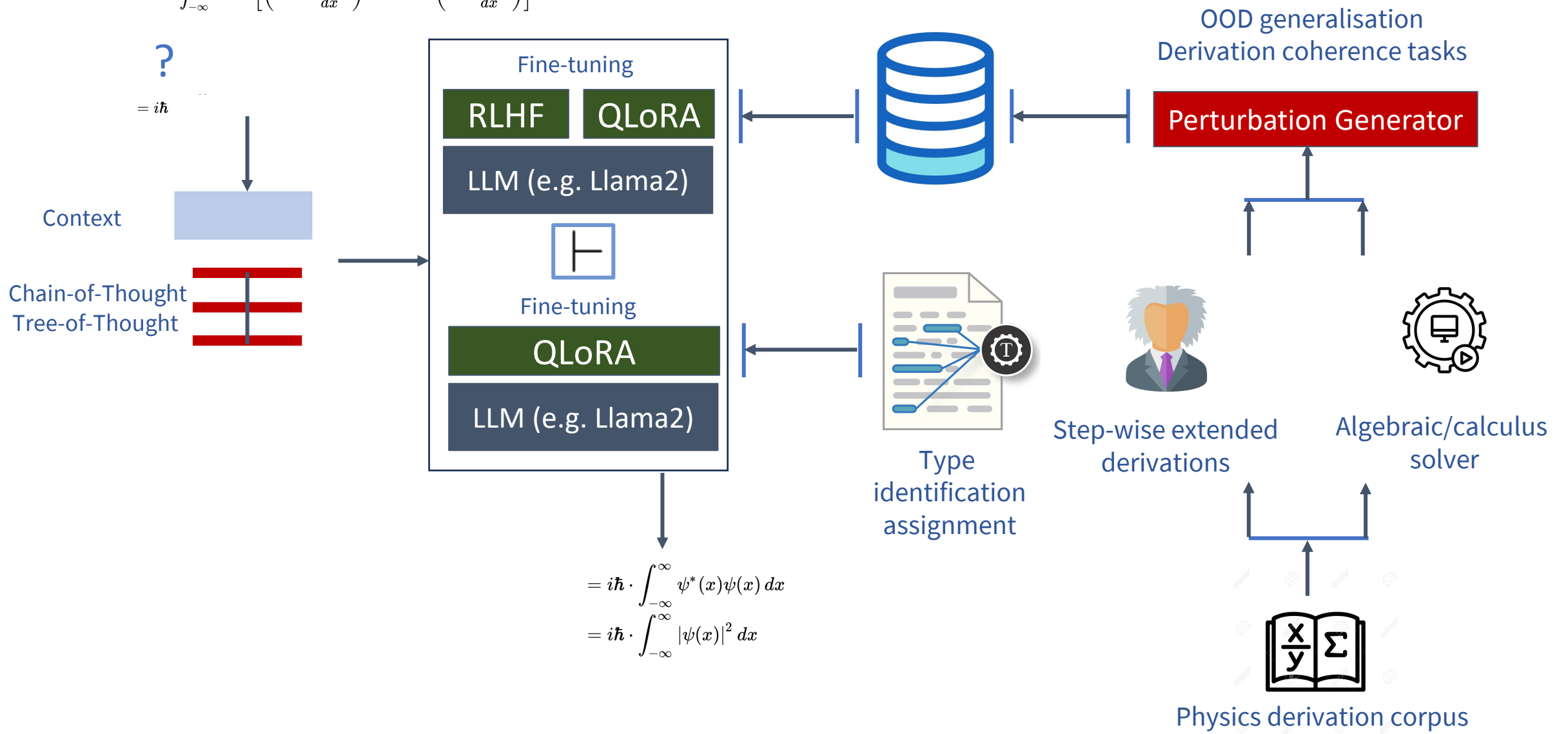
# Task: Derivation completion





$$\begin{aligned}
\langle f | g \rangle - \langle g | f \rangle &= \int_{-\infty}^{\infty} \psi^*(x) x \cdot \left(-i\hbar \frac{d}{dx}\right) \psi(x) dx \\
&\quad - \int_{-\infty}^{\infty} \psi^*(x) \left(-i\hbar \frac{d}{dx}\right) \cdot x \psi(x) dx \\
&= i\hbar \cdot \int_{-\infty}^{\infty} \psi^*(x) \left[ \left(-x \cdot \frac{d\psi(x)}{dx}\right) + \frac{d(x\psi(x))}{dx} \right] dx \\
&= i\hbar \cdot \int_{-\infty}^{\infty} \psi^*(x) \left[ \left(-x \cdot \frac{d\psi(x)}{dx}\right) + \psi(x) + \left(x \cdot \frac{d\psi(x)}{dx}\right) \right] dx
\end{aligned}$$

# Task: Derivation completion



$$\begin{aligned}
&= i\hbar \cdot \int_{-\infty}^{\infty} \psi^*(x)\psi(x) dx \\
&= i\hbar \cdot \int_{-\infty}^{\infty} |\psi(x)|^2 dx
\end{aligned}$$

OOD generalisation  
Derivation coherence tasks

**Perturbation Generator**

Type  
identification  
assignment

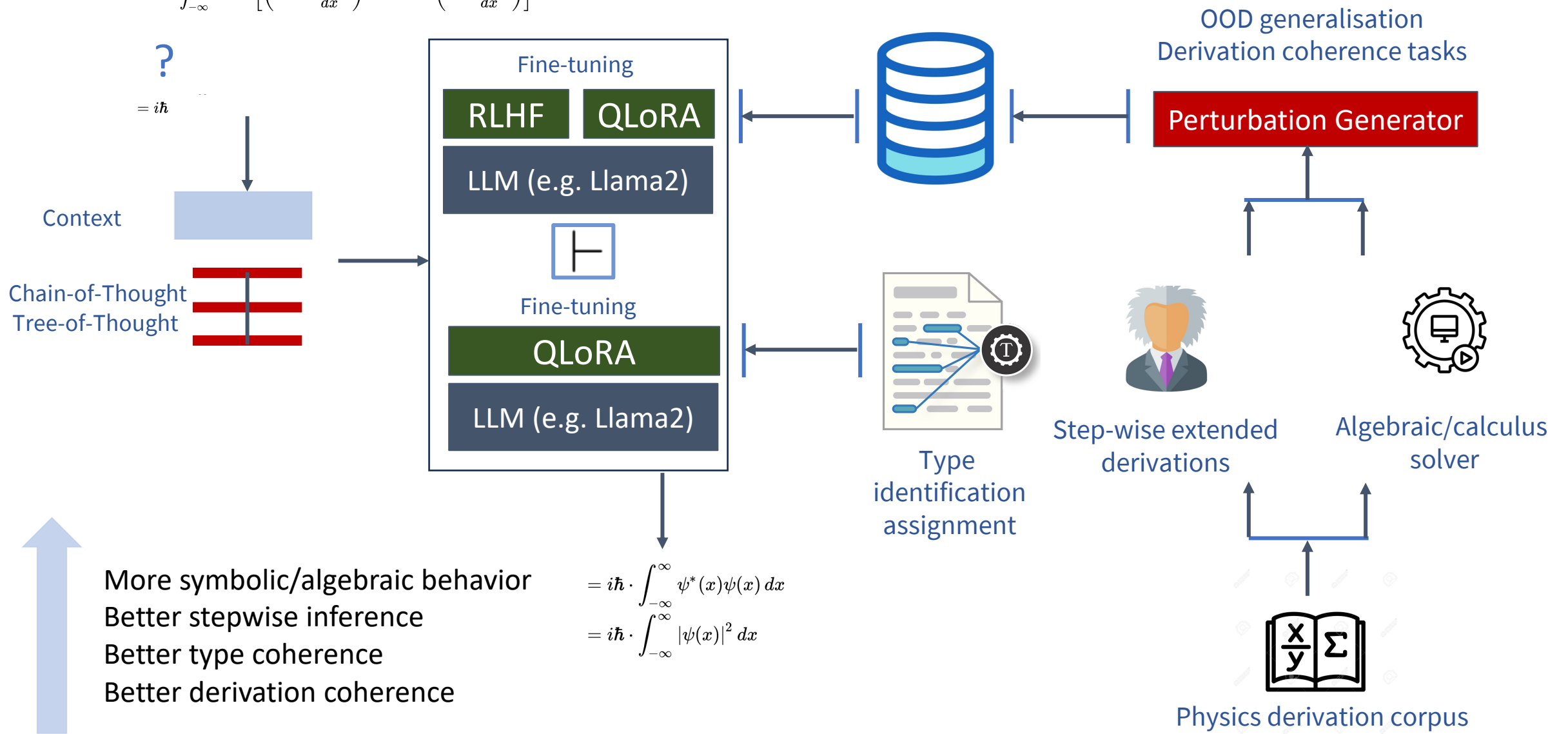
Step-wise extended  
derivations

Algebraic/calculus  
solver

Physics derivation corpus

$$\begin{aligned} \langle f | g \rangle - \langle g | f \rangle &= \int_{-\infty}^{\infty} \psi^*(x) x \cdot \left(-i\hbar \frac{d}{dx}\right) \psi(x) dx \\ &\quad - \int_{-\infty}^{\infty} \psi^*(x) \left(-i\hbar \frac{d}{dx}\right) \cdot x \psi(x) dx \\ &= i\hbar \cdot \int_{-\infty}^{\infty} \psi^*(x) \left[ \left(-x \cdot \frac{d\psi(x)}{dx}\right) + \frac{d(x\psi(x))}{dx} \right] dx \\ &= i\hbar \cdot \int_{-\infty}^{\infty} \psi^*(x) \left[ \left(-x \cdot \frac{d\psi(x)}{dx}\right) + \psi(x) + \left(x \cdot \frac{d\psi(x)}{dx}\right) \right] dx \end{aligned}$$

# Task: Derivation completion



A close-up, profile view of a man wearing a brown tweed hat and a matching coat. He is smoking a dark wooden pipe with a curved stem. The background is a textured, greyish-brown wall. The text "Abductive Reasoning" is overlaid in white, sans-serif font across the center of the image.

# Abductive Reasoning

# Abductive Reasoning

- First introduced by Peirce (1903).
- Inference to the best explanation.
- “Abduction is the mechanism via which we generate hypotheses about what we observe.”
- Dialogues closely with assumed background knowledge.

Veen, Creative leaps in theory: **the might of abduction** (2021)

What Sherlock does well.

# Abductive Natural Language Inference (ANLI)

Inference to the best explanation  
(facts, evidence)

**Claim:** Specialized cells protect the human body from disease-causing microbes by producing chemicals that destroy the microbes.

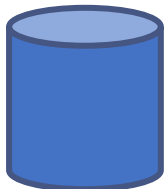
**True | False**

Why? (Explanation)

Multi-hop  
Multi-premise

Specialized cells are a source of chemicals that destroy disease-causing microbes.

disease-causing microbes have a negative impact on the body.



~10.000 facts

# Abductive Natural Language Inference (ANLI)

Inference to the best explanation  
(facts, evidence)

**Claim:** Specialized cells protect the human body from disease-causing microbes by producing chemicals that destroy the microbes.

**True | False**

Why? (Explanation)

Multi-hop  
Multi-premise

Specialized cells are a source of chemicals that destroy disease-causing microbes.

disease-causing microbes have a negative impact on the body.



~10.000 facts

Encoding scientific statements

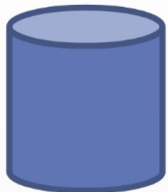
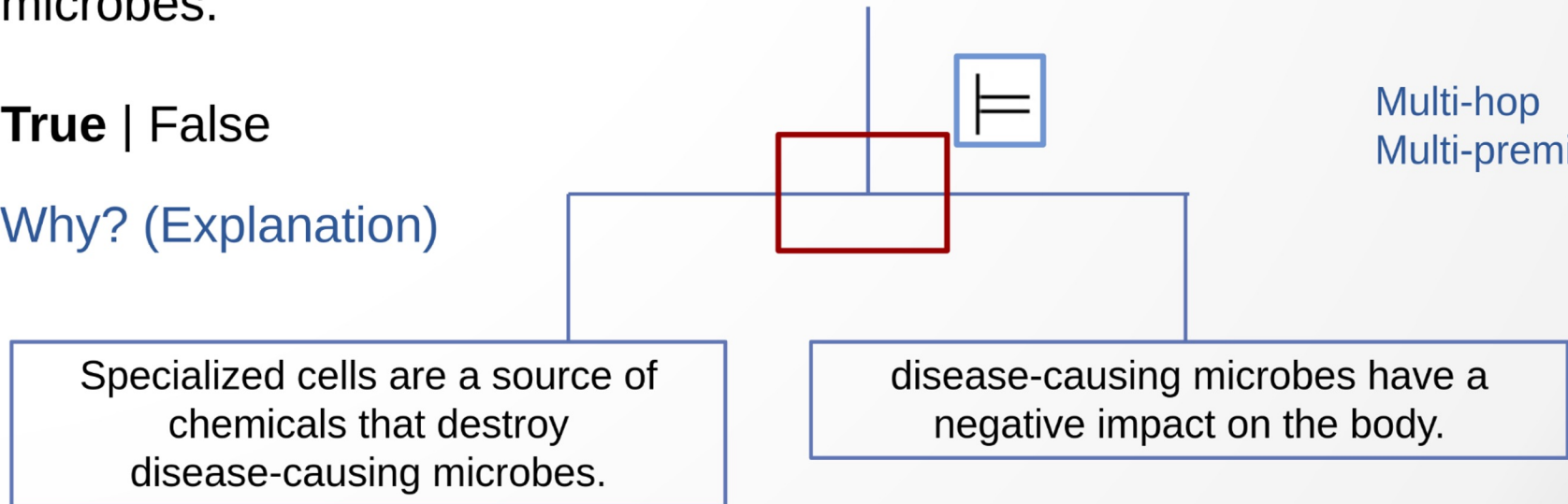
# Abductive Natural Language Inference (ANLI)

Inference to the best explanation  
(facts, evidence)

**Claim:** Specialized cells protect the human body from disease-causing microbes by producing chemicals that destroy the microbes.

**True | False**

**Why? (Explanation)**

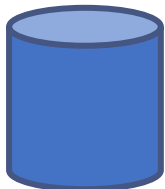
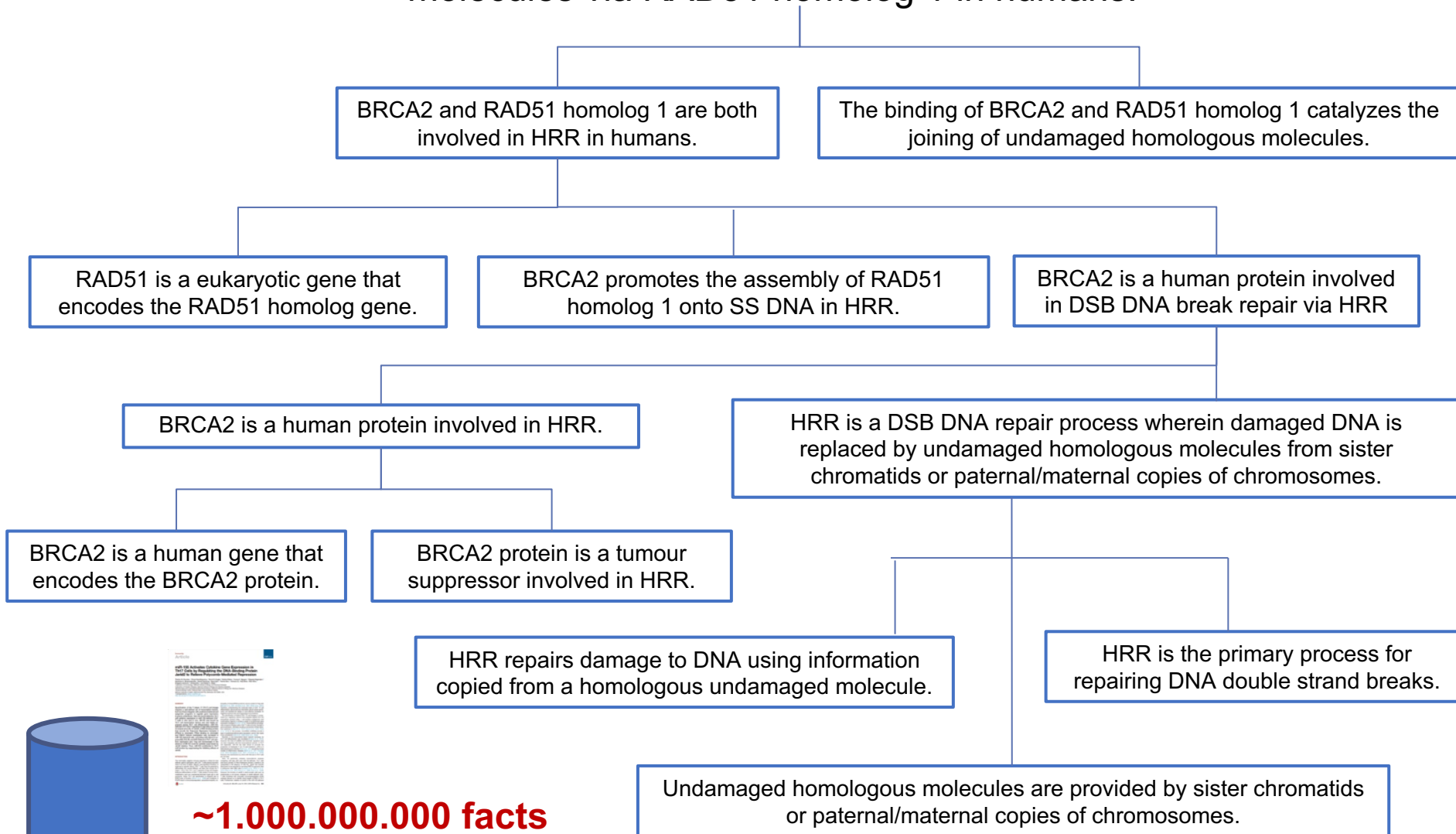


~10.000 facts

Encoding inference relations

# Expert-level scientific inference & explanation

**Claim:** BRCA2 promotes the joining of undamaged homologous repair molecules via RAD51 homolog 1 in humans.



**~1.000.000.000 facts**



# Task: Explanation generation

claim

(Loss of BRCA2) may cause  
(increased genomic instability).

Why?

# Task: Explanation generation

claim

(Loss of BRCA2) **may** cause  
(increased genomic instability).

Why?

Loss of BRCA2 causes the cell to  
default to NHEJ repair processes.

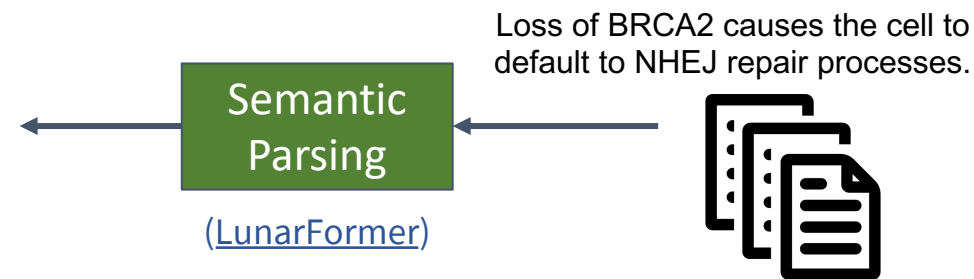


# Task: Explanation generation

claim

**(Loss of BRCA2) may cause**  
(increased genomic instability).

Why?



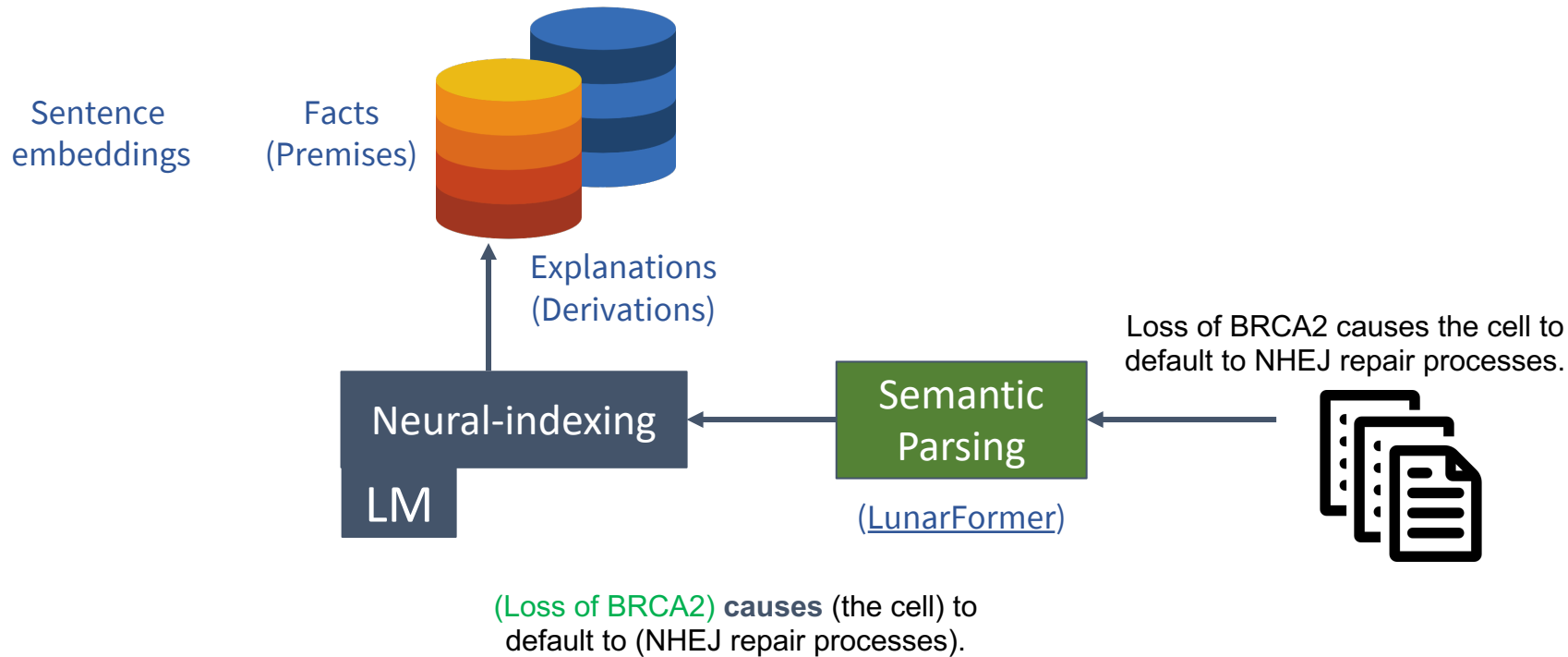
**(Loss of BRCA2) causes** (the cell) to  
default to (NHEJ repair processes).

# Task: Explanation generation

claim

(Loss of BRCA2) **may** cause  
(increased genomic instability).

Why?

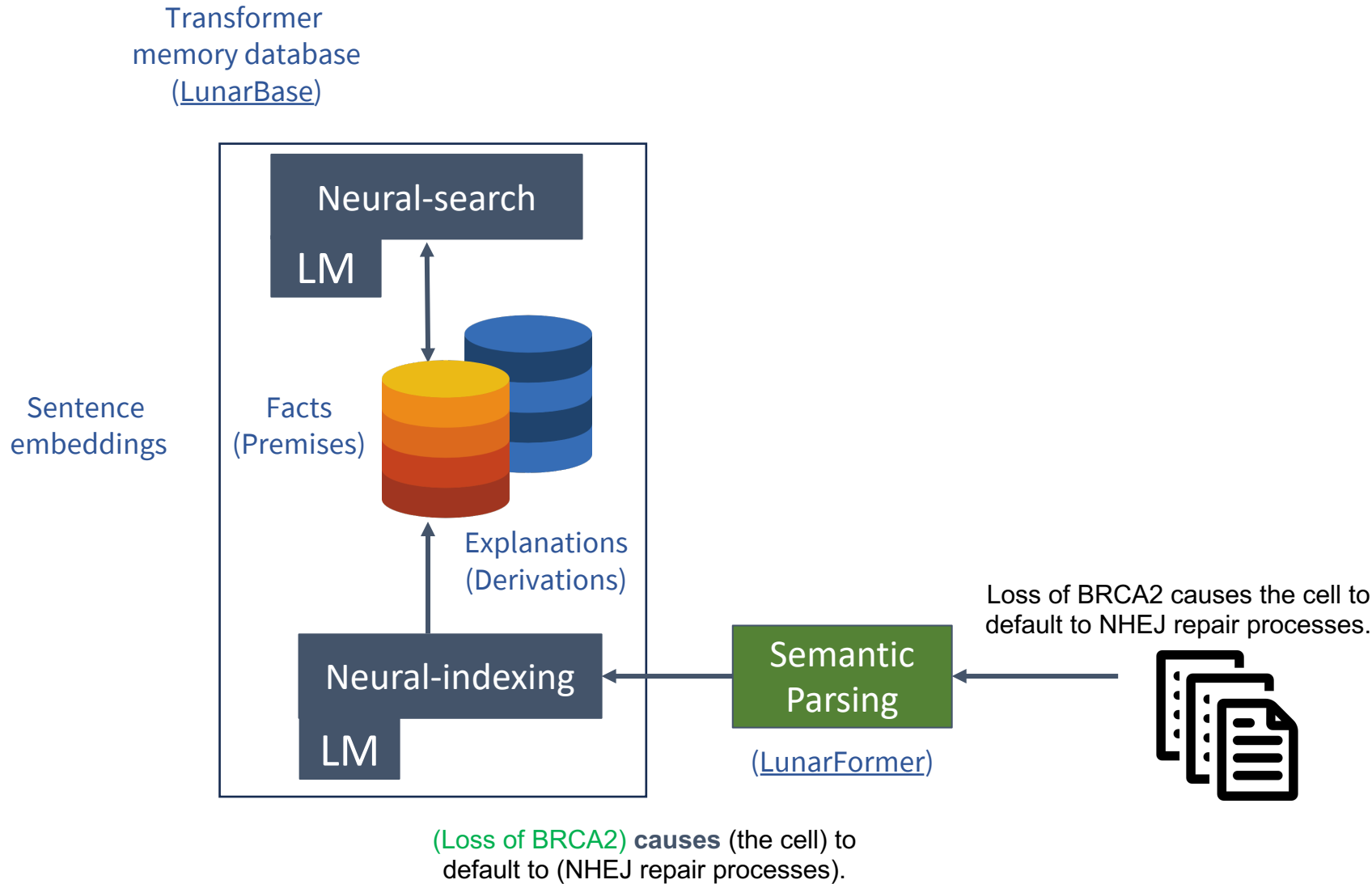


# Task: Explanation generation

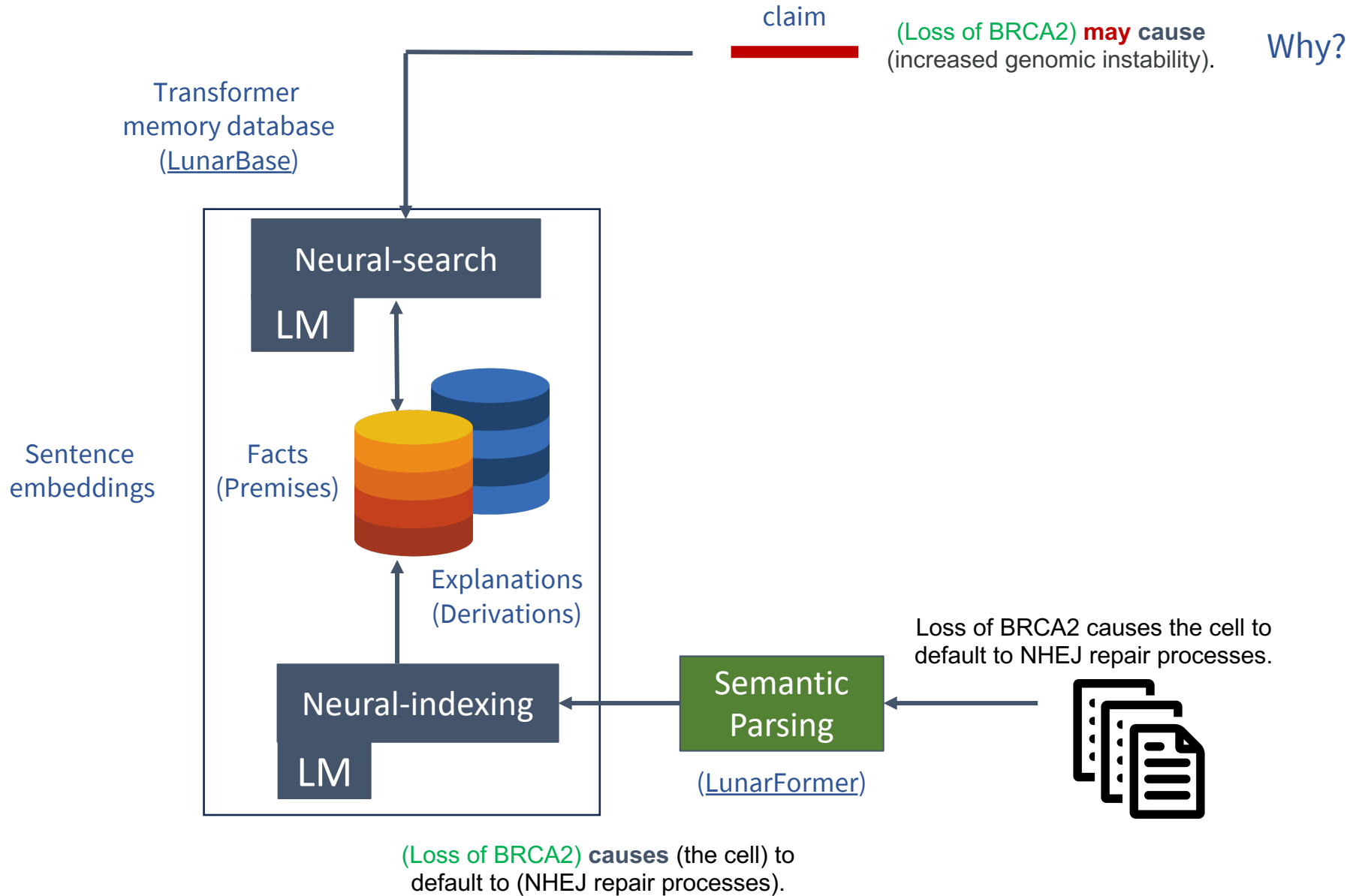
claim

(Loss of BRCA2) **may** cause  
(increased genomic instability).

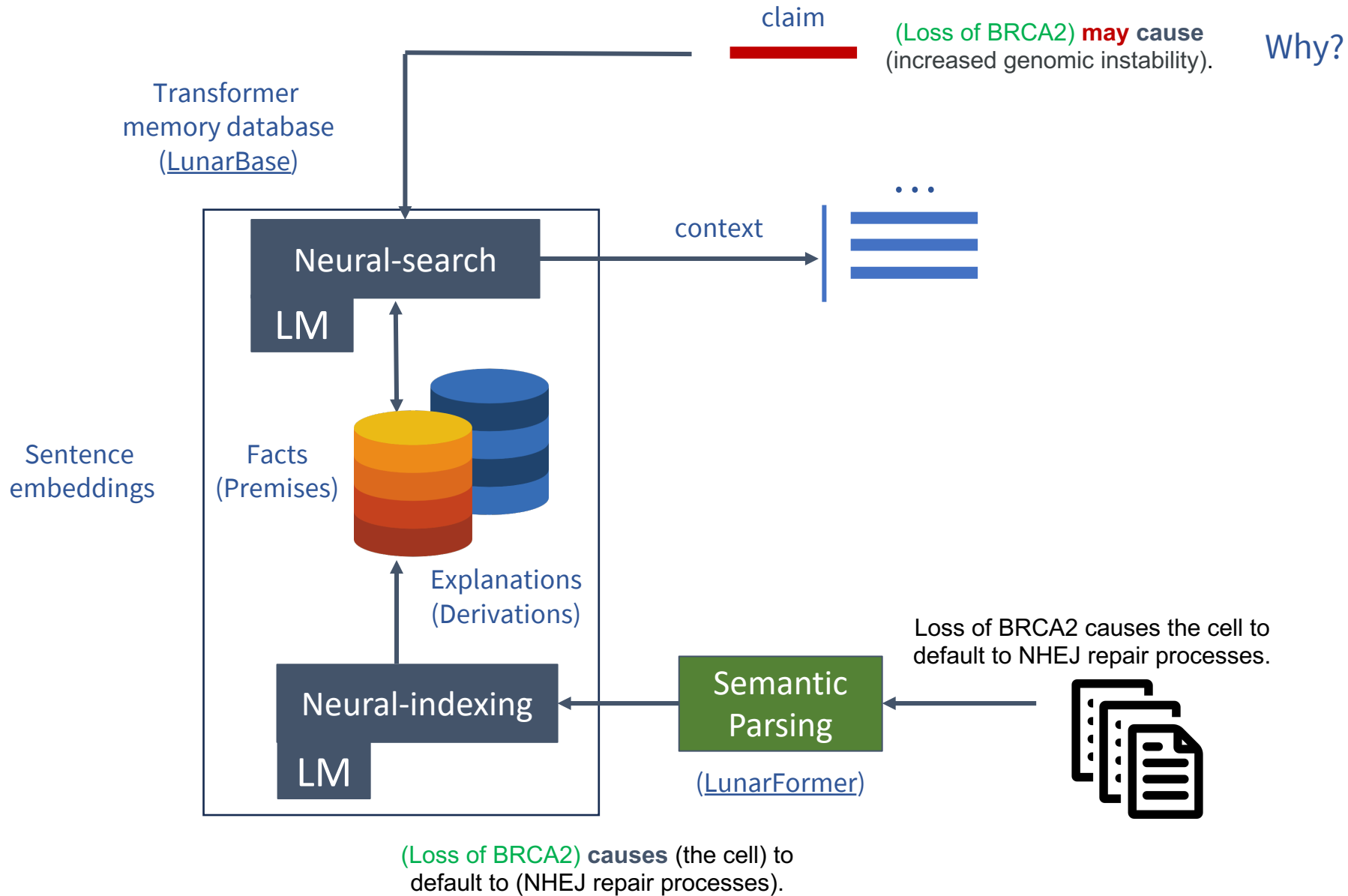
Why?



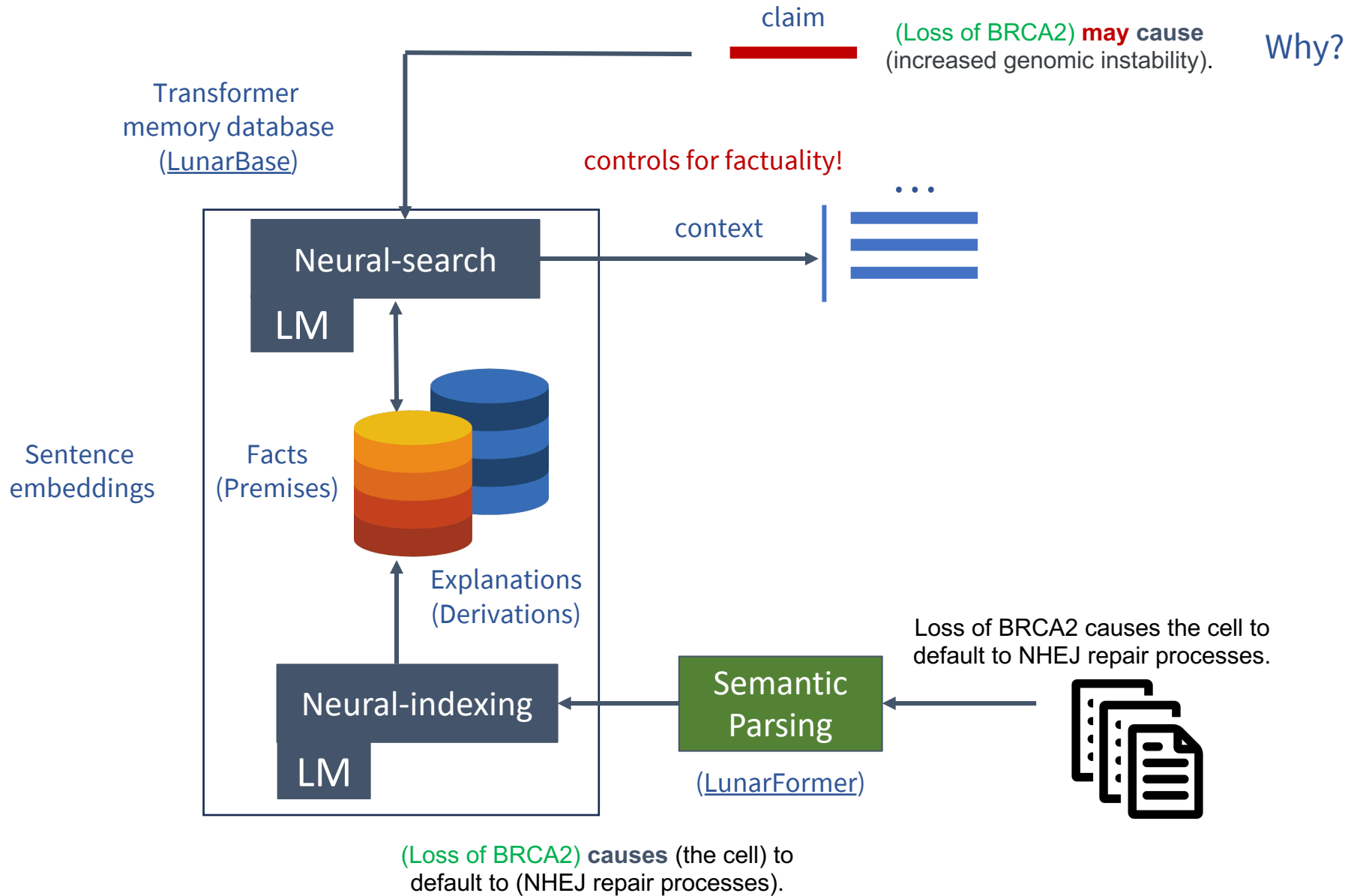
# Task: Explanation generation



# Task: Explanation generation

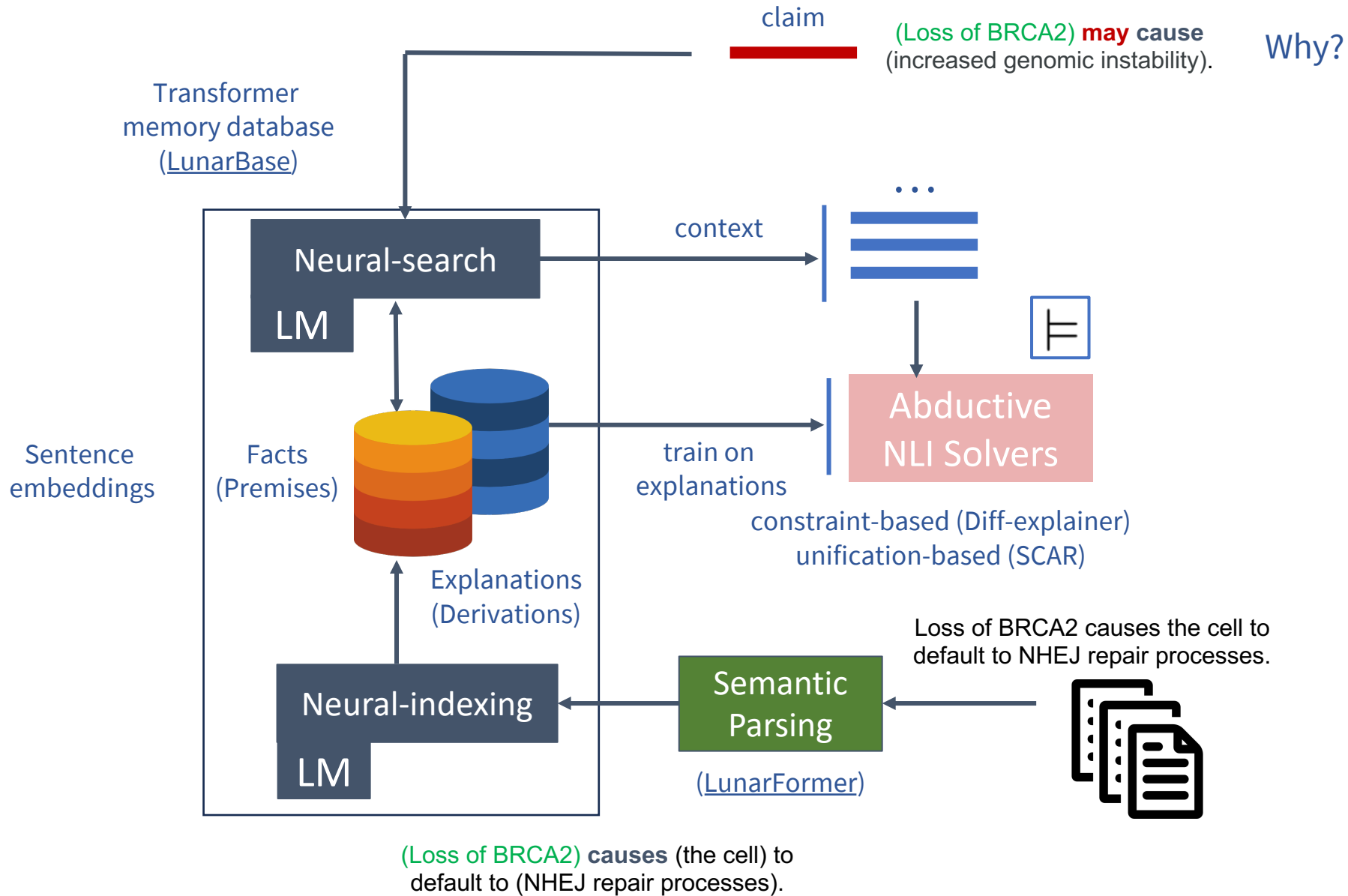


# Task: Explanation generation

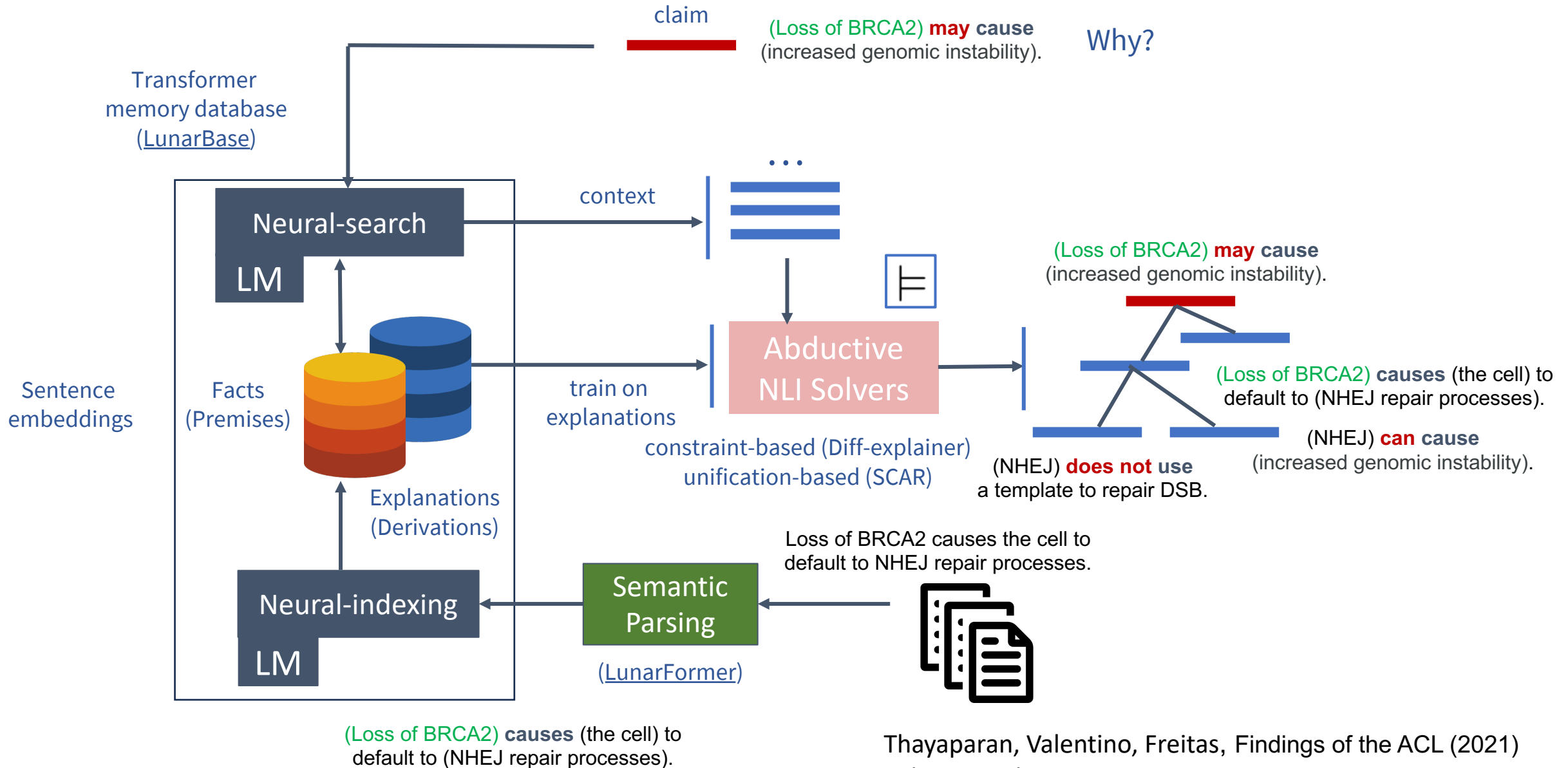




# Task: Explanation generation



# Task: Explanation generation



# Automating & scaling-up evidence collection

**Cytokine release syndrome (CRS):** Significant adverse event of T cell-engaging therapies.

**Need:** Predictive models for CRS

**Problem:** Lack of patient-level datasets.

Can one explore relevant evidence in the literature?

19hs

38hs



~ 460 papers

17 highly aligned papers

Parameter extraction

Reference	Year	Source	Study	Trial Phase	Patients evaluated, N*	Cancer type(s)	CAR antigen target	Co-stim domain	CRS grading scale
1 Jacobson et al. [37]	2022	Lancet Oncol	Clinical Trial: ZUMA-5	II	148	r/r indolent NHL	CD19	ND	Lee et al. 2014 criteria
2 Hong et al. [38]	2021	Bone Marrow Transplant	Chinese Clinical Trial	R**	41	ALL, MM, NHL (COVID-19 - not included)	ND	ND	CTCAE v 5.0
3 Yan et al. [39]	2021	Front. Immunol.	Clinical Trial		142	r/r ALL, Lymphoma, MM	CD19, CD19+BCMA, CD19+CD20	ND	Lee et al. 2014 criteria
4 Topp et al. [40]	2021	BJHaem	Clinical Trial: ZUMA-4		41	DLBCL, PMBCL, TFL, HGBCCL	CD19	ND	Modified criteria of Lee and colleagues
5 Shah et al. [41]	2021	Lancet	Clinical Trial: ZUMA-3	II	55	r/r B-ALL	CD19	ND	Lee et al. 2014 criteria
6 Liu et al. [12]	2021	Translational Oncology	Chinese Clinical Trial	I	9	DLBCL, TFL, FL	CD19	ND	Lee et al. 2014 criteria
7 Sang et al. [13]	2020	Cancer Med.	Clinical Trial	II	21	r/r DLBCL	CD19, CD20	4-1BB/CD3	Lee et al. 2014 criteria
8 Yan et al. [42]	2019	Lancet Haematol	Clinical Trial	II	21	MM	CD19, BCMA	4-BB, 4-1BB	Modified criteria of Lee and colleagues and NCI CTCAE v4.03
9 Zhao et al. [43]	2018	J Hematol Oncol	Clinical Trial	I	57	MM	BCMA	4-1BB	Modified criteria of Lee and colleagues and NCI CTCAE v4.03
10 Neelapu et al. [44]	2017	N Eng J Med	Clinical Trial: ZUMA-1	II	111	DLBCL, PMBCL, TLF	CD19	CD28	Lee et al. 2014 criteria‡
11 Hay et al. [24]	2017	Blood	Clinical Trial	I/II	133	r/r B-ALL, CLL, NHL	CD19	4-1BB	Lee et al. 2014 criteria
12 Turtle et al. [45]	2017	J Clin Oncol	Clinical Trial	I/II	24	CLL	CD19	ND	Modified criteria of Lee and colleagues and NCI CTCAE v4.03
13 Hu et al. [15]	2017	Clin Cancer Res	Chinese Clinical Trial		15	r/r ALL	CD19	4-1BB/CD3	Modified criteria of Lee and colleagues and NCI CTCAE v4.03
14 Teachey et al. [18]	2016	Cancer Discov	Clinical Trial		51	r/r ALL	CD19	4-1BB	Custom CRS grading scale (Modified criteria of Lee and Davila)
15 Porter et al. [16]	2015	Sci Transl Med	Pilot Clinical Trial		14	r/r CLL	CD19	4-1BB/CD3	Penn Grading System
16 Davila et al. [5]	2014	Sci Transl Med	Clinical Trial	I	16	r/r B-ALL	CD19	ND	Davila et al. criteria

Study	IL2	IL4	IL6	IL8	IL10	IL15	IL2R $\alpha$	TNF- $\alpha$	IFN- $\gamma$	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

Study	Patient Cohort Size, N	Patient Cohort Male, n (%)	Patient Cohort Age (A), year median (range/IQR*)	CRS, n (%)	Severe or grade 3+ CRS**, n (%)
1 Jacobson et al. [37]	148	84 (57)	61 (53-68)*	121 (82)	10 (8)
2 Hong et al. [38]	41	21 (51)	51 (32.5 - 62.5)*	41 (100)	41 (100)**
3 Yan et al. [39]	142	87 (61)	45 (24 - 59)*	123 (87)	30 (24)
4 Topp et al. [40]	41	28 (68)	61 (19 - 77)	38 (93)	1 (2.6)
5 Shah et al. [41]	55	33 (60)	40 (28 - 52)*	49 (89)	13 (26.5)
6 Liu et al. [12]	9	5 (56)	51 (22 - 62)	9 (100)	1 (11)
7 Sang et al. [13]	21	13 (61.9)	55 (23 - 72)	21 (100)	6 (29)
8 Yan et al. [42]	21	10 (48)	58 (49.5 - 61)	19 (90)	1 (5)
9 Zhao et al. [43]	57	34 (60)	54 (27 - 72)	51 (89)	4 (8)
10 Neelapu et al. [44]	111	68 (67)	58 (23 - 76)	101 (91)	13 (13)
11 Hay et al. [24]	133	93 (70)	54 (20-73)	93 (70)	16 (17)
12 Turtle et al. [45]	24	ND	61 (40 - 73)	20 (83)	2 (10)
13 Hu et al. [15]	15	9 (60)	32 (7 - 57)	10 (67)	6 (60)
14 Teachey et al. [18]	12	8 (67)	56 (25 - 72)	12 (100)	3 (25)
15 Porter et al. [16]	14	12 (86)	66 (51 - 78)	9 (64)	6 (67)
16 Davila et al. [5]	16	12 (75)	50 (23 - 74)	7 (44)	7 (100)
17 Kalos et al. [46]	3	3 (100)	65 (64 - 77)	3 (100)	3 (100)**

19hs

38hs 7 mins

Meta-review

~ 460 papers

17 highly aligned papers

Parameter extraction

Reference	Year	Source	Study	Trial Phase	Patients evaluated, N*	Cancer type(s)	CAR antigen target	Co-stim domain	CRS grading scale
1 Jacobson et al. [37]	2022	Lancet Oncol	Clinical Trial: ZUMA-5	II	148	r/r indolent NHL	CD19	ND	Lee et al. 2014 criteria
2 Hong et al. [38]	2021	Bone Marrow Transplant	Chinese Clinical Trial	R**	41	ALL, MM, NHL (COVID-19 - not included)	ND	ND	CTCAE v 5.0
3 Yan et al. [39]	2021	Front. Immunol.	Clinical Trial		142	r/r ALL, Lymphoma, MM	CD19, CD19+BCMA, CD19+CD20	ND	Lee et al. 2014 criteria
4 Topp et al. [40]	2021	BJHaem	Clinical Trial: ZUMA-4		41	DLBCL, PMBCL, TFL, HGBCCL	CD19	ND	Modified criteria of Lee and colleagues
5 Shah et al. [41]	2021	Lancet	Clinical Trial: ZUMA-3	II	55	r/r B-ALL	CD19	ND	Lee et al. 2014 criteria
6 Liu et al. [12]	2021	Translational Oncology	Chinese Clinical Trial	I	9	DLBCL, TFL, FL	CD19	ND	Lee et al. 2014 criteria
7 Sang et al. [13]	2020	Cancer Med.	Clinical Trial	II	21	r/r DLBCL	CD19, CD20	4-1BB/CD3	Lee et al. 2014 criteria
8 Yan et al. [42]	2019	Lancet Haematol	Clinical Trial	II	21	MM	CD19, BCMA	4-BB, 4-1BB	Modified criteria of Lee and colleagues and NCI CTCAE v4.03
9 Zhao et al. [43]	2018	J Hematol Oncol	Clinical Trial	I	57	MM	BCMA	4-1BB	Modified criteria of Lee and colleagues and NCI CTCAE v4.03
10 Neelapu et al. [44]	2017	N Eng J Med	Clinical Trial: ZUMA-1	II	111	DLBCL, PMBCL, TLF	CD19	CD28	Lee et al. 2014 criteria‡
11 Hay et al. [24]	2017	Blood	Clinical Trial	I/II	133	r/r B-ALL, CLL, NHL	CD19	4-1BB	Lee et al. 2014 criteria
12 Turtle et al. [45]	2017	J Clin Oncol	Clinical Trial	I/II	24	CLL	CD19	ND	Modified criteria of Lee and colleagues and NCI CTCAE v4.03
13 Hu et al. [15]	2017	Clin Cancer Res	Chinese Clinical Trial		15	r/r ALL	CD19	4-1BB/CD3	Modified criteria of Lee and colleagues and NCI CTCAE v4.03
14 Teachey et al. [18]	2016	Cancer Discov	Clinical Trial		51	r/r ALL	CD19	4-1BB	Custom CRS grading scale (Modified criteria of Lee and Davila)
15 Porter et al. [16]	2015	Sci Transl Med	Pilot Clinical Trial		14	r/r CLL	CD19	4-1BB/CD3	Penn Grading System
16 Davila et al. [5]	2014	Sci Transl Med	Clinical Trial	I	16	r/r B-ALL	CD19	ND	Davila et al. criteria

Study	IL2	IL4	IL6	IL8	IL10	IL15	IL2R $\alpha$	TNF- $\alpha$	IFN- $\gamma$	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

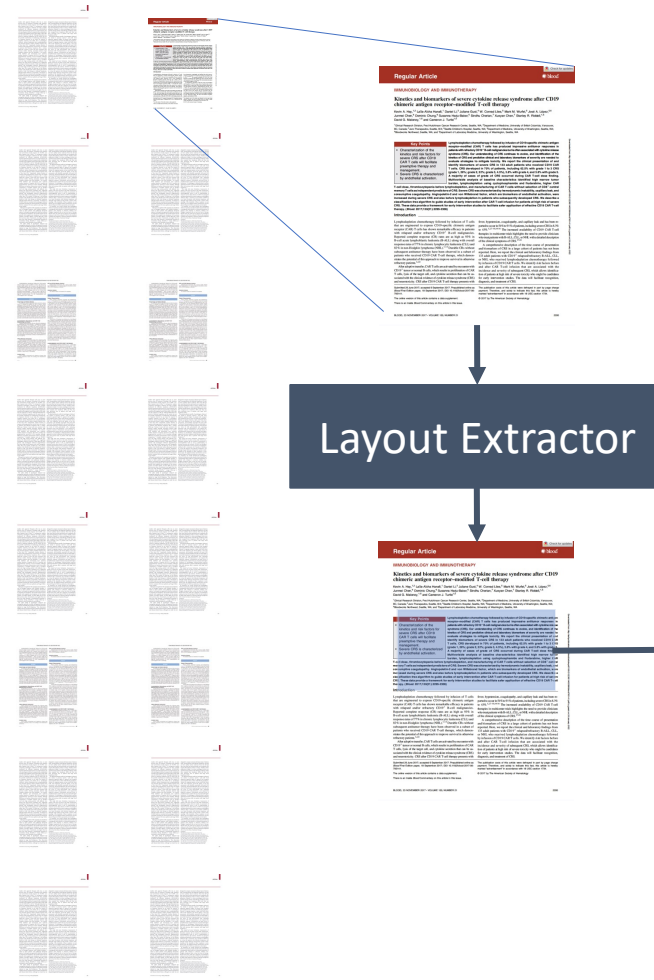
Study	Patient Cohort Size, N	Patient Cohort Male, n (%)	Patient Cohort Age (A), year median (range/IQR*)	CRS, n (%)	Severe or grade 3+ CRS**, n (%)
1 Jacobson et al. [37]	148	84 (57)	61 (53-68)*	121 (82)	10 (8)
2 Hong et al. [38]	41	21 (51)	51 (32.5 - 62.5)*	41 (100)	41 (100)**
3 Yan et al. [39]	142	87 (61)	45 (24 - 59)*	123 (87)	30 (24)
4 Topp et al. [40]	41	28 (68)	61 (19 - 77)	38 (93)	1 (2.6)
5 Shah et al. [41]	55	33 (60)	40 (28 - 52)*	49 (89)	13 (26.5)
6 Liu et al. [12]	9	5 (56)	51 (22 - 62)	9 (100)	1 (11)
7 Sang et al. [13]	21	13 (61.9)	55 (23 - 72)	21 (100)	6 (29)
8 Yan et al. [42]	21	10 (48)	58 (49.5 - 61)	19 (90)	1 (5)
9 Zhao et al. [43]	57	34 (60)	54 (27 - 72)	51 (89)	4 (8)
10 Neelapu et al. [44]	111	68 (67)	58 (23 - 76)	101 (91)	13 (13)
11 Hay et al. [24]	133	93 (70)	54 (20-73)	93 (70)	16 (17)
12 Turtle et al. [45]	24	ND	61 (40 - 73)	20 (83)	2 (10)
13 Hu et al. [15]	15	9 (60)	32 (7 - 57)	10 (67)	6 (60)
14 Teachey et al. [18]	12	8 (67)	56 (25 - 72)	12 (100)	3 (25)
15 Porter et al. [16]	14	12 (86)	66 (51 - 78)	9 (64)	6 (67)
16 Davila et al. [5]	16	12 (75)	50 (23 - 74)	7 (44)	7 (100)
17 Kalos et al. [46]	3	3 (100)	65 (64 - 77)	3 (100)	3 (100)**



Meta-review

~ 460 papers

17 highly aligned papers  
Parameter extraction



Study	IL2	IL4	IL6	IL8	IL10	IL15	IL2R $\alpha$	TNF- $\alpha$	IFN- $\gamma$	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

# On Infrastructures







# Take-away

Emerging foundations for **industrialised scientific inference**  
Universal framework for **integrating and organising 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 scientific reasoning.

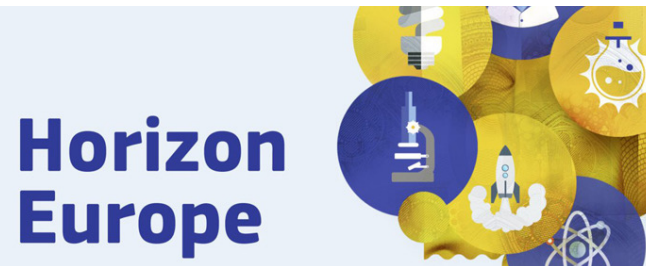
## Controlling reasoning

Scientific reasoning requires complex pipelines.  
Robust methods already exist to extend LLMs in the direction of rigorous reasoning.  
(multiple models, trees-of-thought, symbolic augmentation, retrieval augmentation, semantic parsing, differentiable symbolic solvers)  
Need for a coordination infrastructure.

**\*Very\* exciting times to do AI for Science!**

Thank you for your attention!

Generously supported by:



contact: [andre.freitas@manchester.ac.uk](mailto:andre.freitas@manchester.ac.uk)

