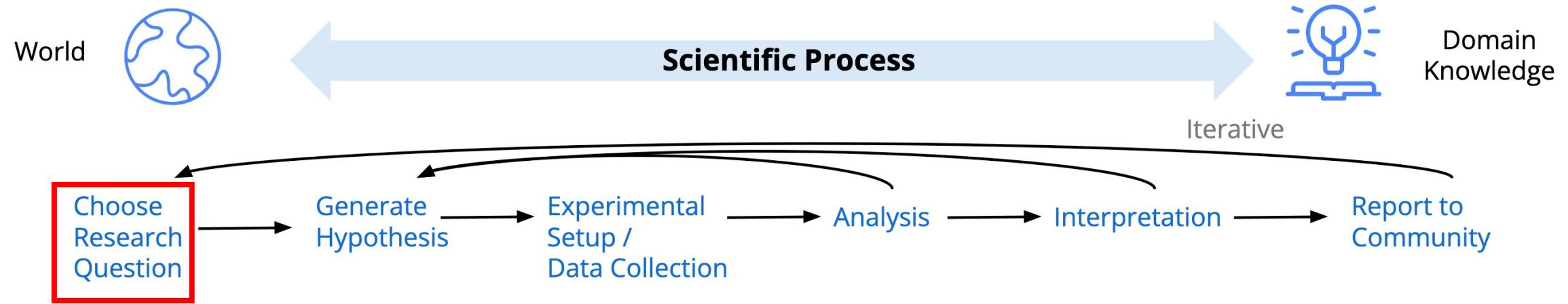
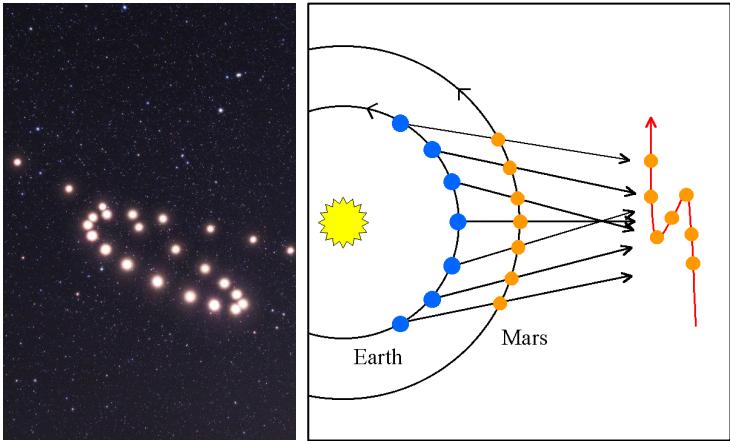
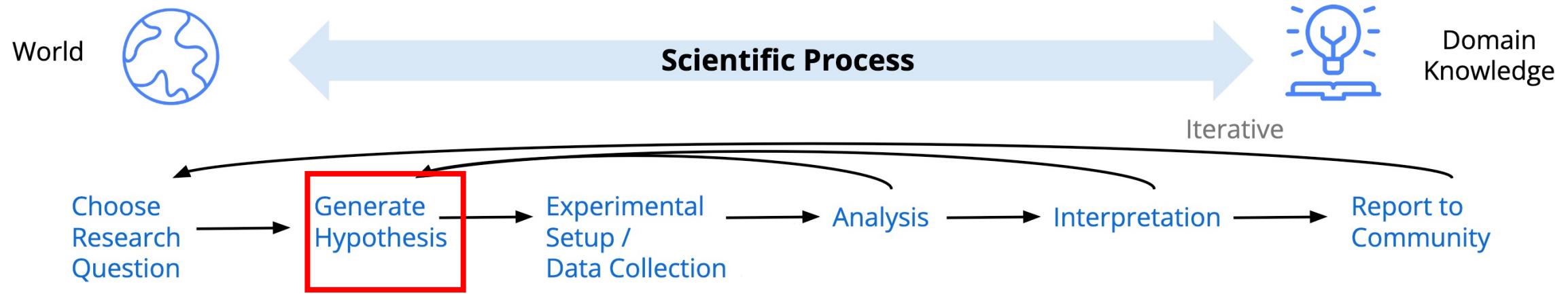


Goal: We want AI to achieve human level performance at *research in the natural sciences*.



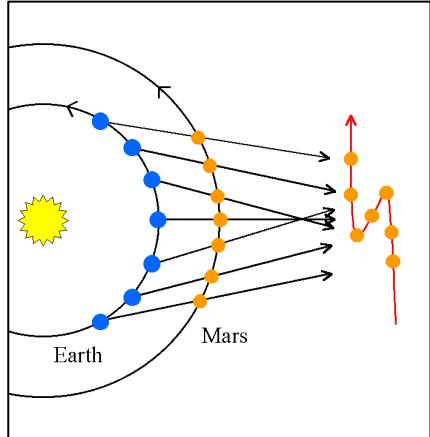
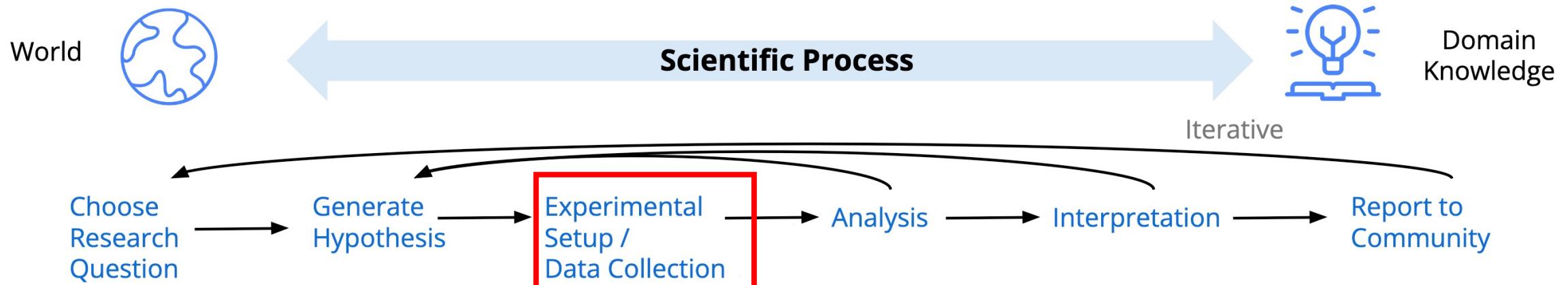
Observation: Apparent Retrograde Planetary Motion

c. The Astronomical Revolution: Copernicus- Kepler-Borelli

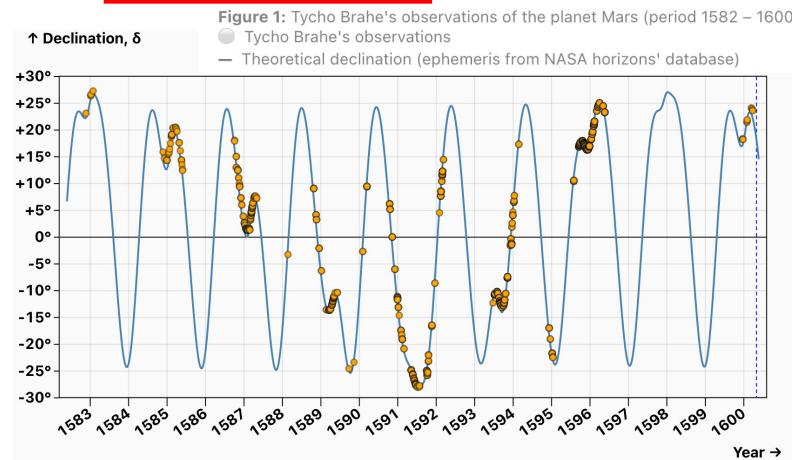


Observation: Apparent Retrograde Planetary Motion
Theory: Heliocentric Model

c. The Astronomical Revolution: Copernicus- Kepler-Borelli

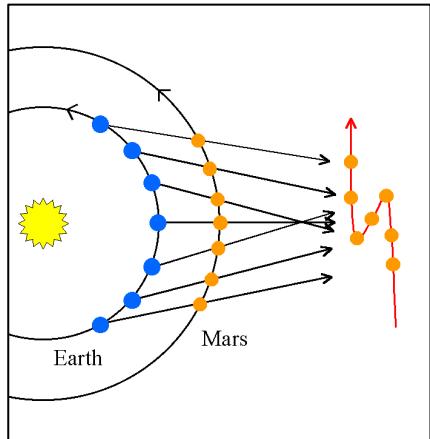
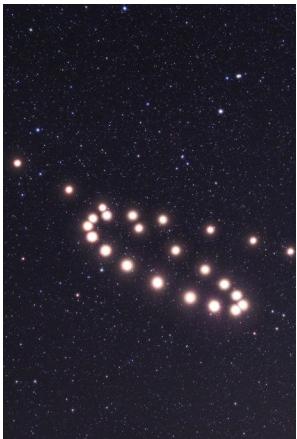
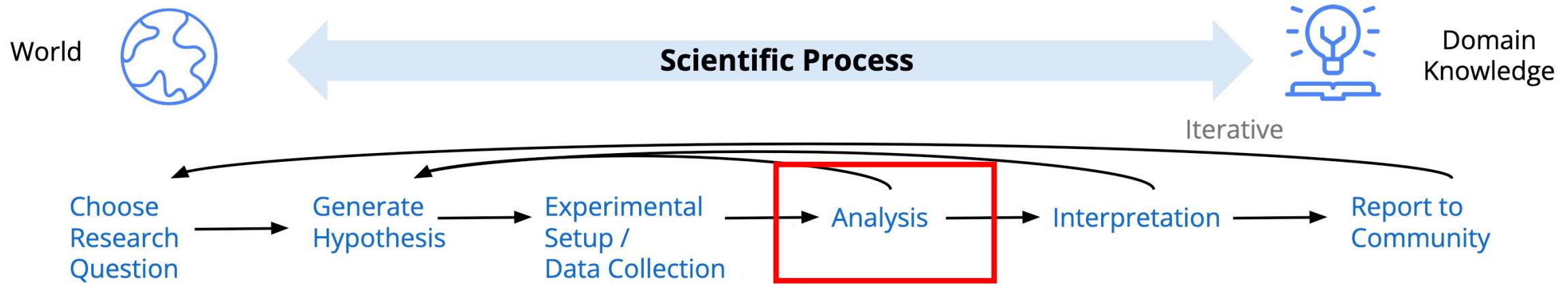


Observation: Apparent Retrograde Planetary Motion
Theory: Heliocentric Model

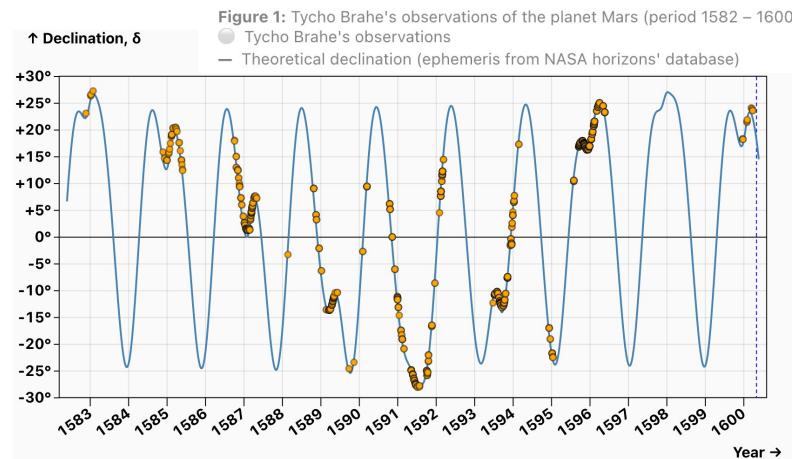


Data Collection: Sample data in regime of interest.

c. The Astronomical Revolution: Copernicus- Kepler-Borelli



Observation: Apparent Retrograde Planetary Motion
Theory: Heliocentric Model

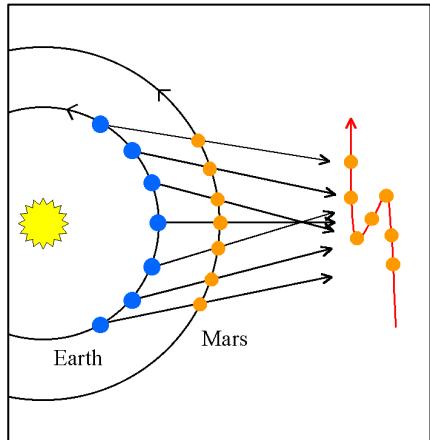
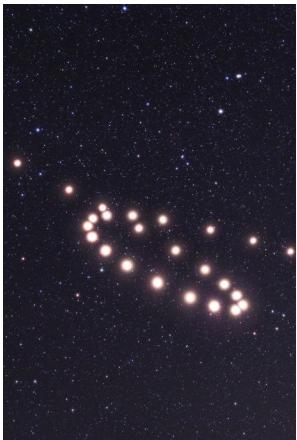
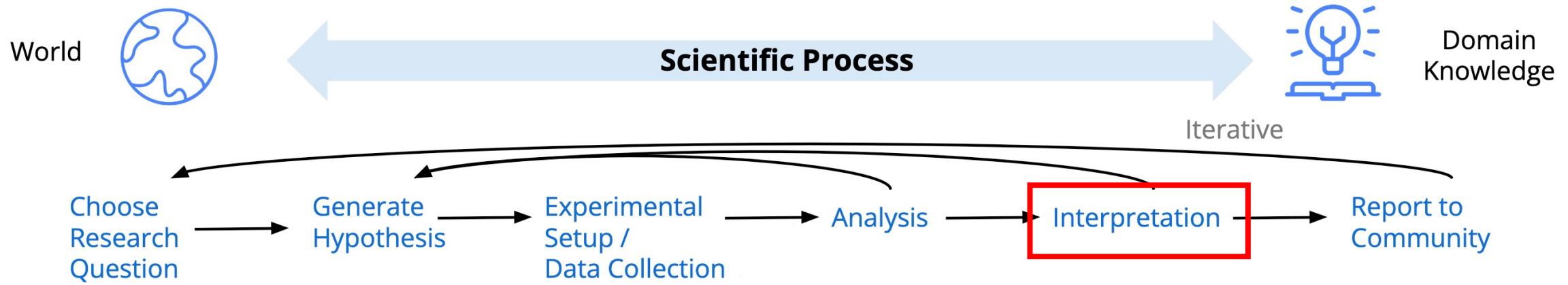


Data Collection: Sample data in regime of interest.

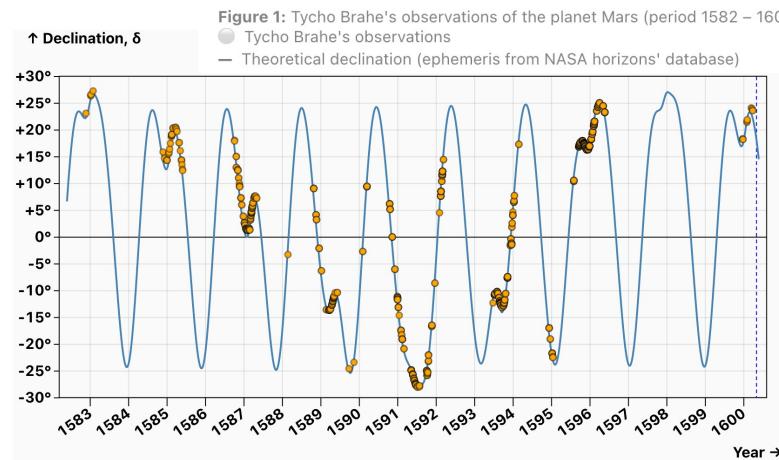
$$T^2 \propto r^3$$

Analysis: Kepler's Third Law

c. The Astronomical Revolution: Copernicus- Kepler-Borelli



Observation: Apparent Retrograde Planetary Motion
Theory: Heliocentric Model



Data Collection: Sample data in regime of interest.

$$T^2 \propto r^3$$

Analysis: Kepler's Third Law

$$mr \left(\frac{2\pi}{T} \right)^2 = G \frac{mM}{r^2}$$

Interpretation: Newton's Law of Gravitation

c. The Astronomical Revolution: Copernicus- Kepler-Borelli

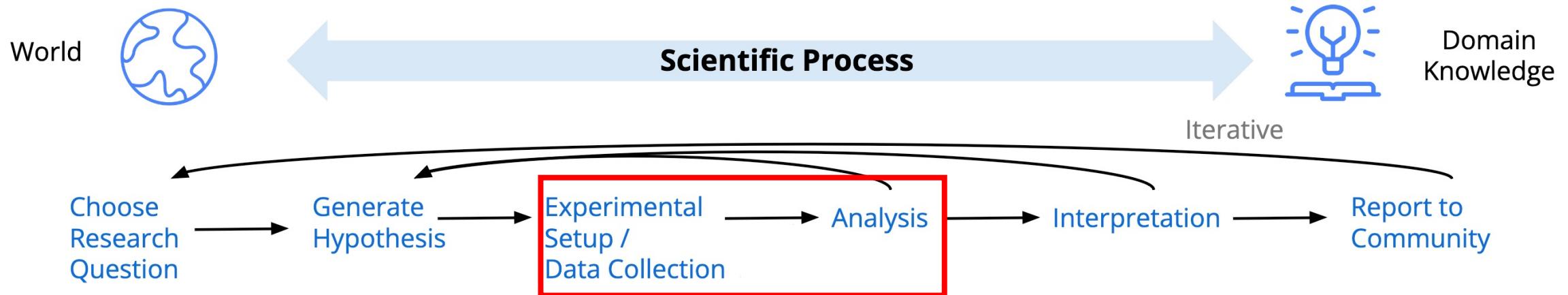
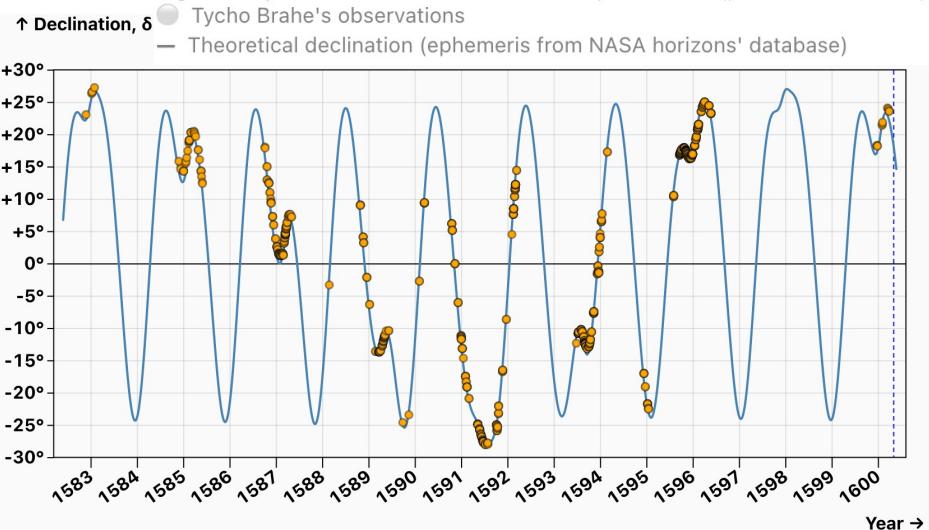


Figure 1: Tycho Brahe's observations of the planet Mars (period 1582 – 1600).



Kepler's Third Law

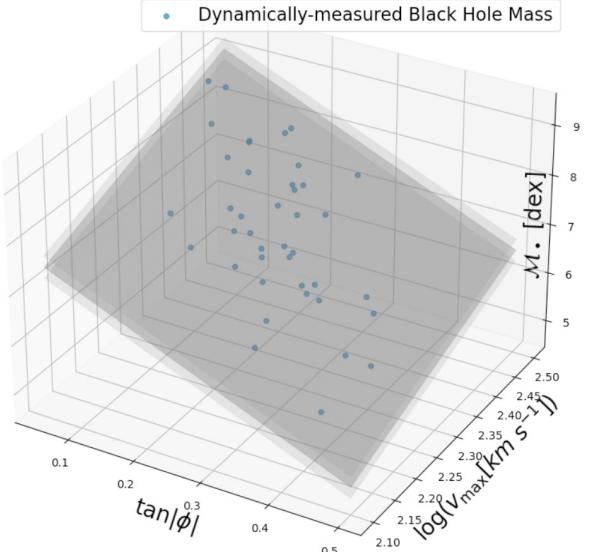
$$T^2 \propto r^3$$

c. The Astronomical Revolution: Copernicus- Kepler-Borelli

Symbolic Regression Algorithms

Symbolic Regression Algorithms

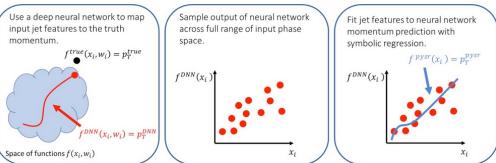
PySR's impact



Discovery of a Planar Black Hole Mass Scaling Relation for Spiral Galaxies

Benjamin L. Davis ¹, Zehao Jin ¹

¹Center for Astrophysics and Space Science, New York University Abu Dhabi



Interpretable machine learning methods applied to jet background subtraction in heavy-ion collisions

Tanner Mengel ¹, Patrick Steffanic ¹, Charles Hughes ^{1,2}, Antonio Carlos Oliveira da Silva ^{1,2}, Christine Natrass ¹

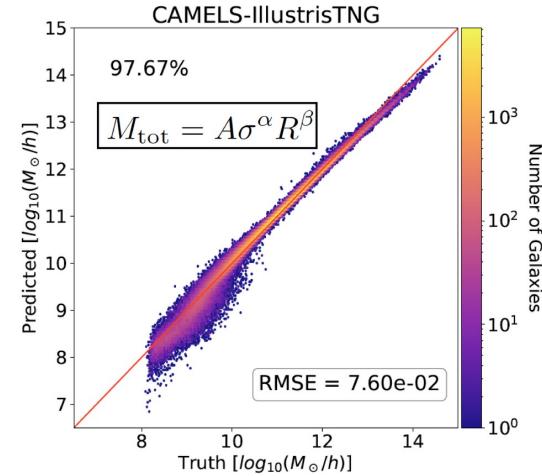
¹University of Tennessee, Knoxville, ²Iowa State University of Science and Technology



Modeling the galaxy-halo connection with machine learning

Ana Maria Delgado ¹, Digvijay Wadekar ^{2,3}, Boryana Hadzhiyska ¹, Sownak Bose ^{1,7}, Lars Hernquist ¹, Shirley Ho ^{2,4,5,6}

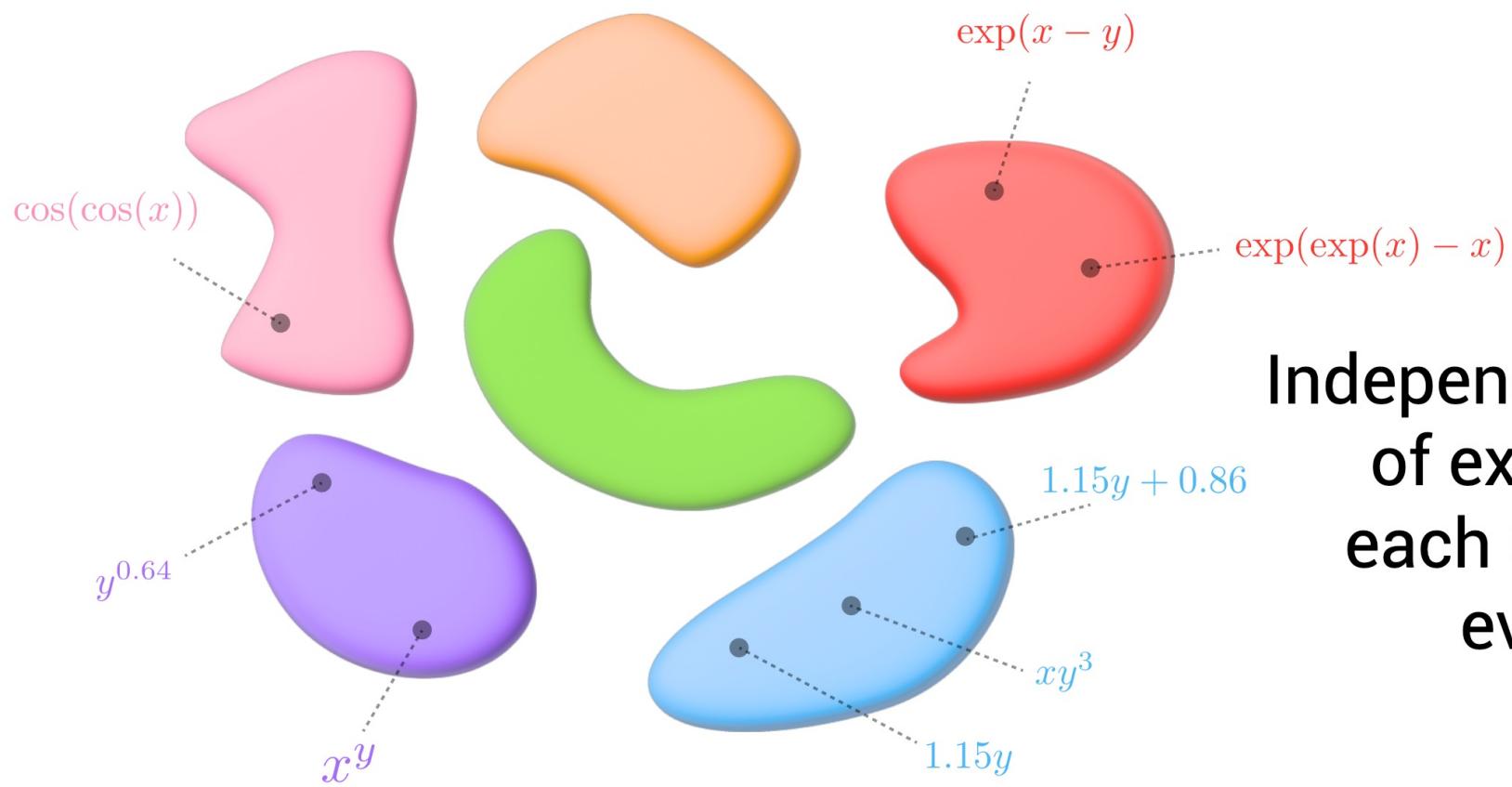
¹Center for Astrophysics | Harvard & Smithsonian, ²New York University, ³Institute for Advanced Study, ⁴Flatiron Institute, ⁵Center for Astrophysics | Harvard & Smithsonian, ⁶Carnegie Mellon University, ⁷Durham University



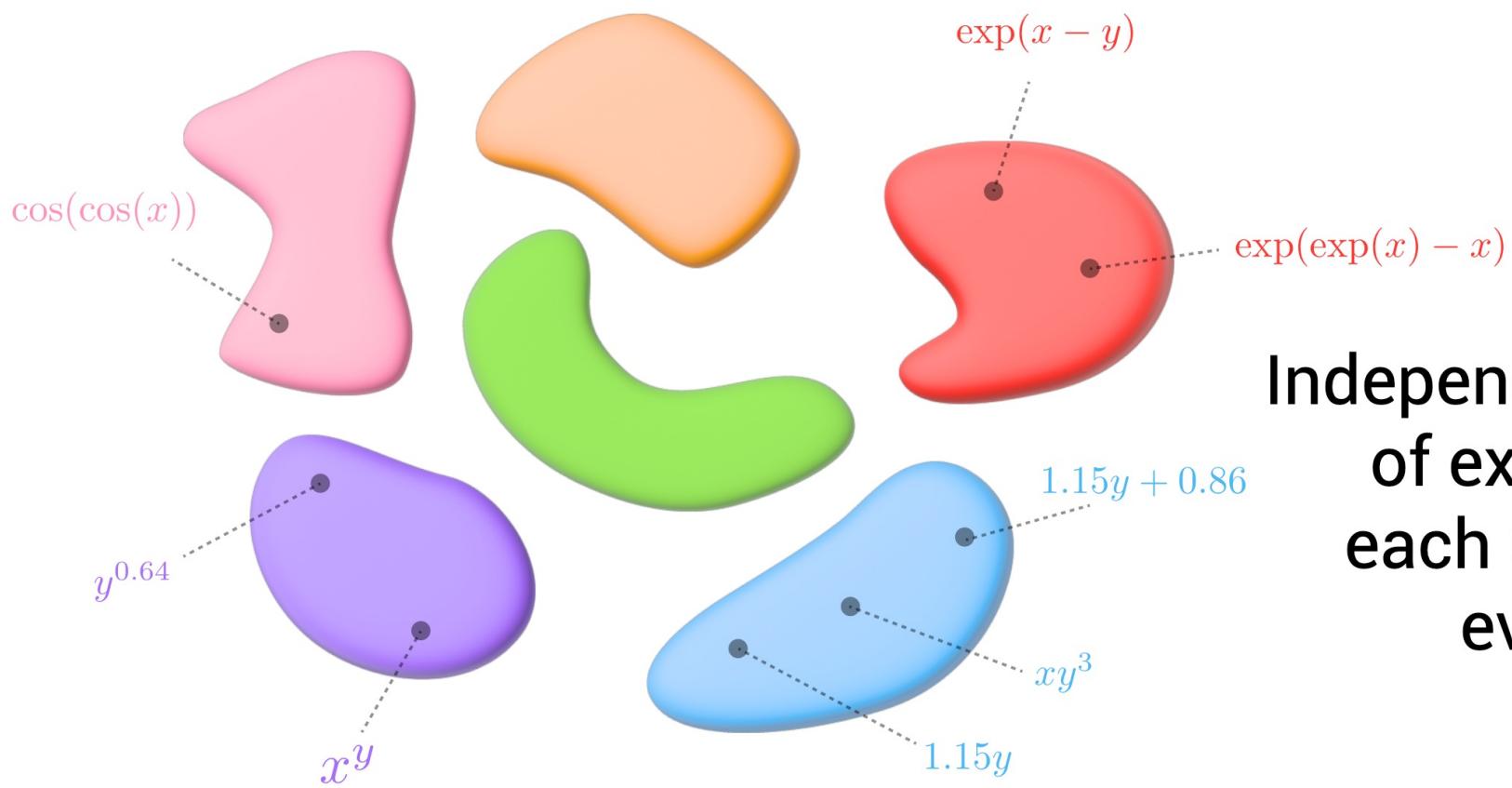
Finding universal relations in subhalo properties with artificial intelligence

Helen Shao ¹, Francisco Villaescusa-Navarro ^{1,2}, Shy Genel ^{2,3}, David N. Spergel ^{2,1}, Daniel Angles-Alcazar ^{4,2}, Lars Hernquist ⁵, Romeel Dave ^{6,7,8}, Desika Narayanan ^{9,10}, Gabriella Contardo ², Mark Vogelsberger ¹¹

¹Princeton University, ²Flatiron Institute, ³Columbia University, ⁴University of Connecticut, ⁵Center for Astrophysics | Harvard & Smithsonian, ⁶University of Edinburgh, ⁷University of the Western Cape, ⁸South African Astronomical Observatories, ⁹University of Florida, ¹⁰University of Florida Informatics Institute, ¹¹MIT



Independent “islands”
of expressions,
each undergoing
evolution



Independent “islands”
of expressions,
each undergoing
evolution

Goal: How can we increase exploration in
relevant parts of the search space?



TEXAS
The University of Texas at Austin



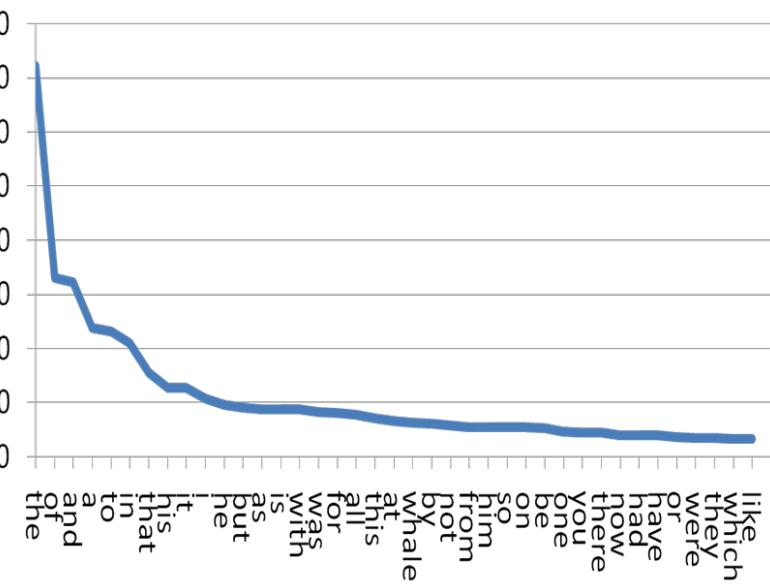
UNIVERSITY OF
CAMBRIDGE



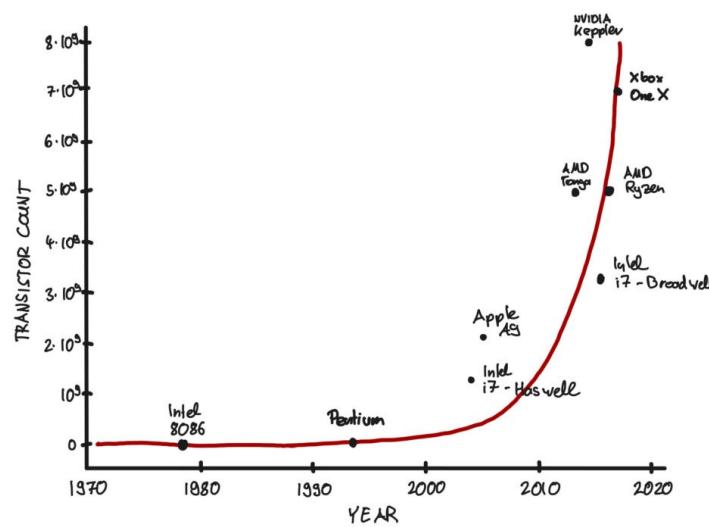
LaSR: Symbolic Regression with a Learned Concept Library

Arya Grayeli*, Atharva Sehgal*, Omar Costilla-Reyes, Miles Cranmer, Swarat Chaudhuri

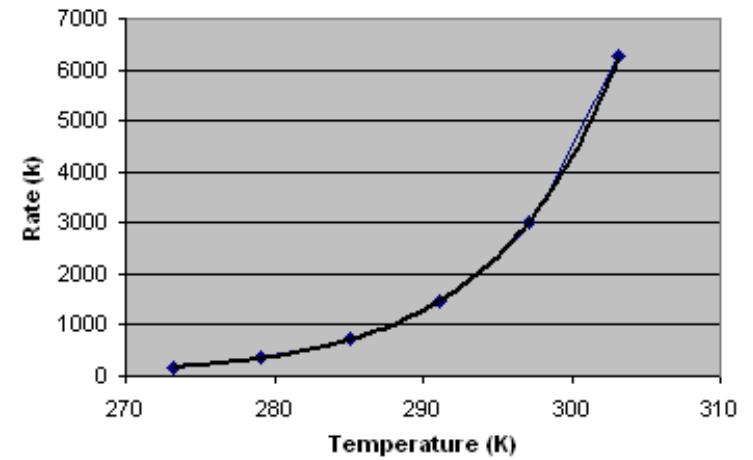
Zipf's Law



Moore's Law



Arrhenius' Equation



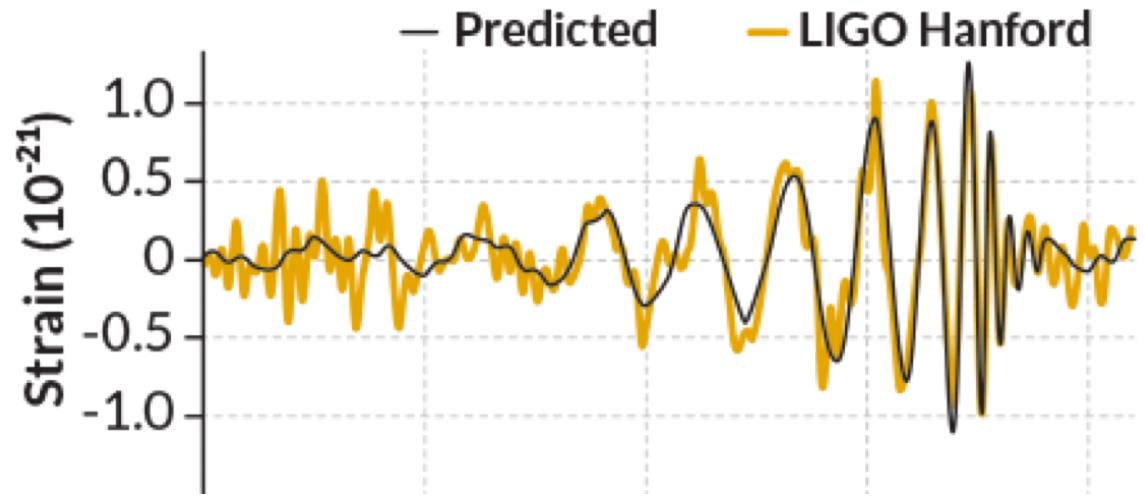
$$y = ax^k + \epsilon \Leftrightarrow \text{"Power Law Trend"}$$



Concepts (by Physicist or LLM)

“Wave strain diminishes as distance increases”

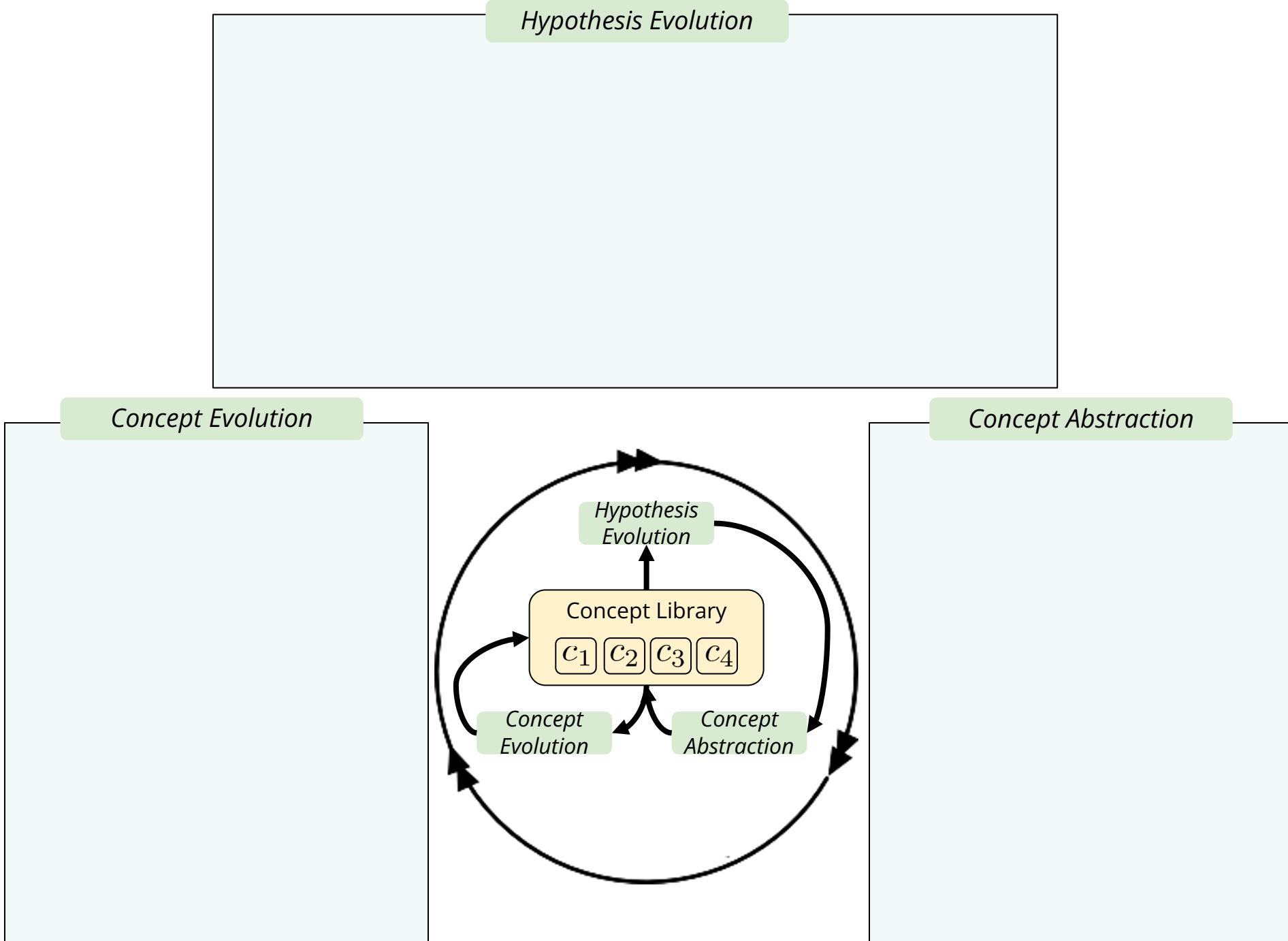
“Wave strain has extraordinarily small magnitude”



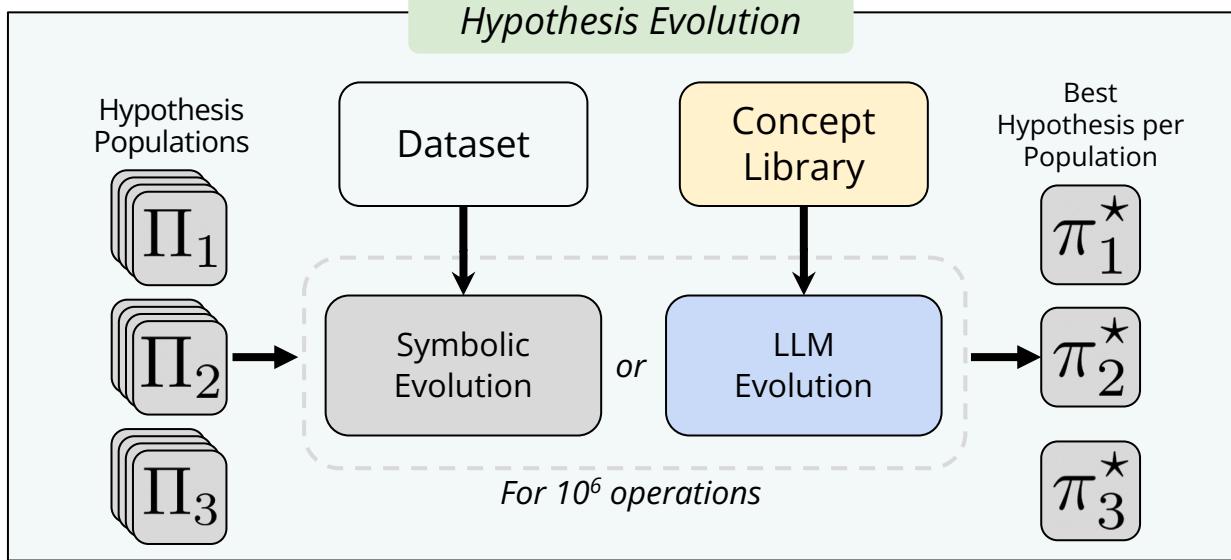
Guide the
search for

$$h = \frac{2G}{c^4} \frac{1}{r} \frac{\partial^2 Q}{\partial t^2}$$

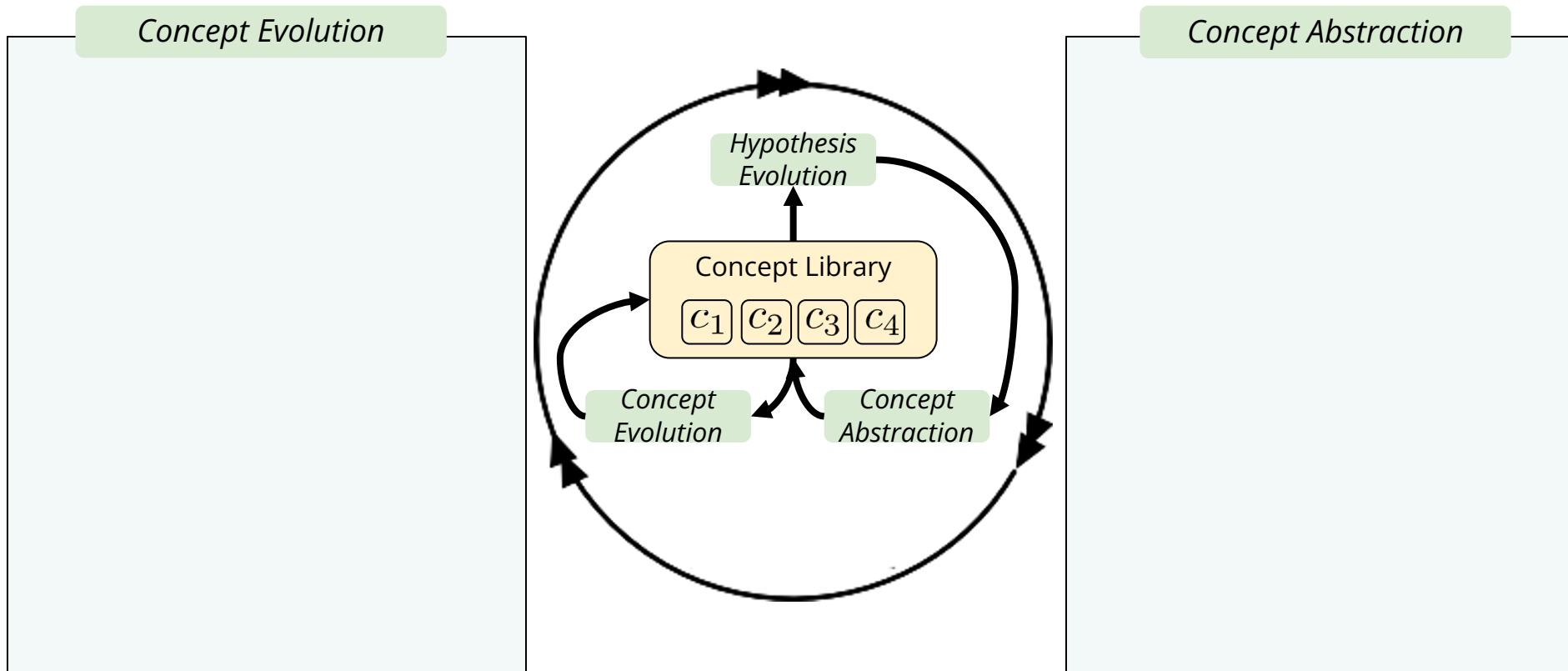
LaSR

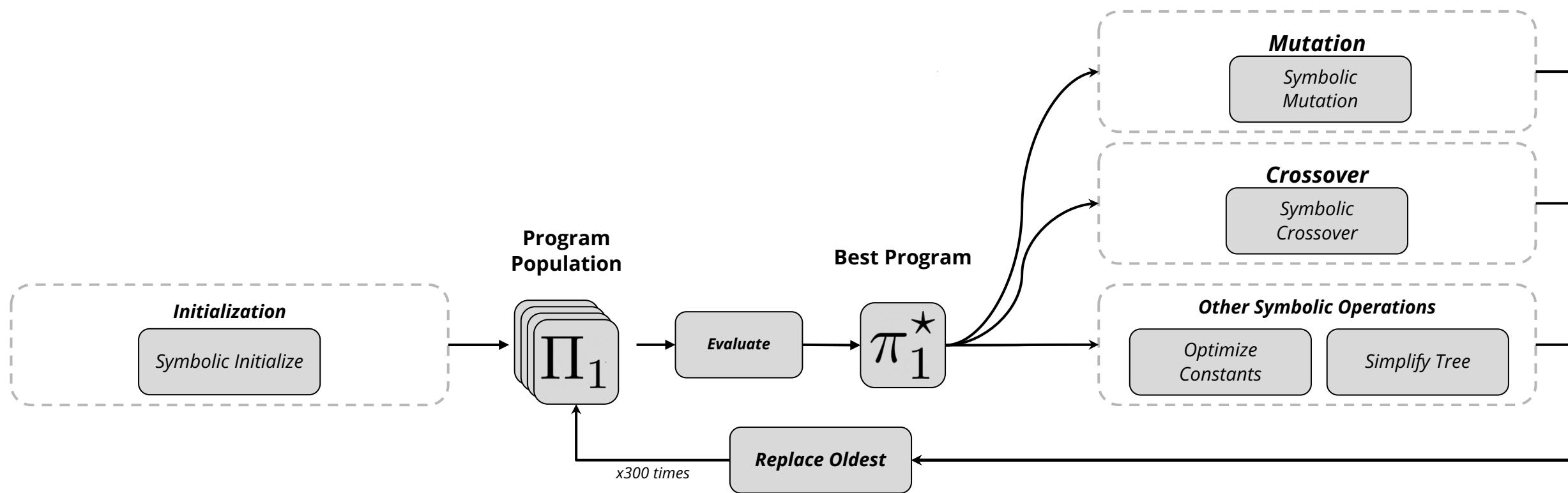


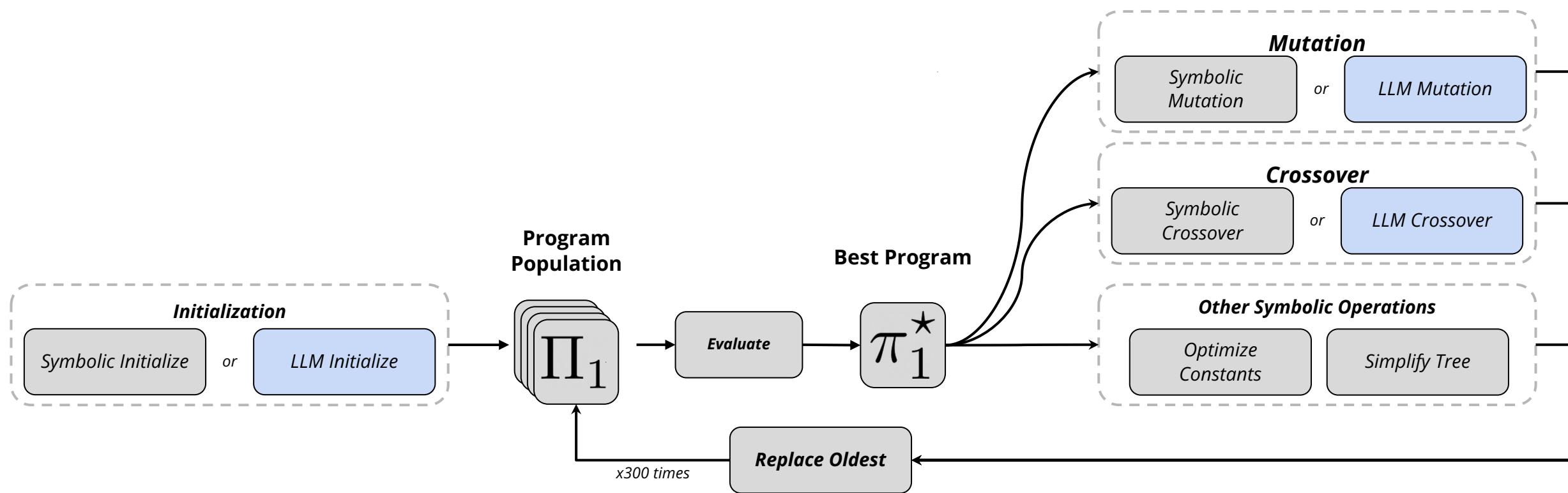
LaSR

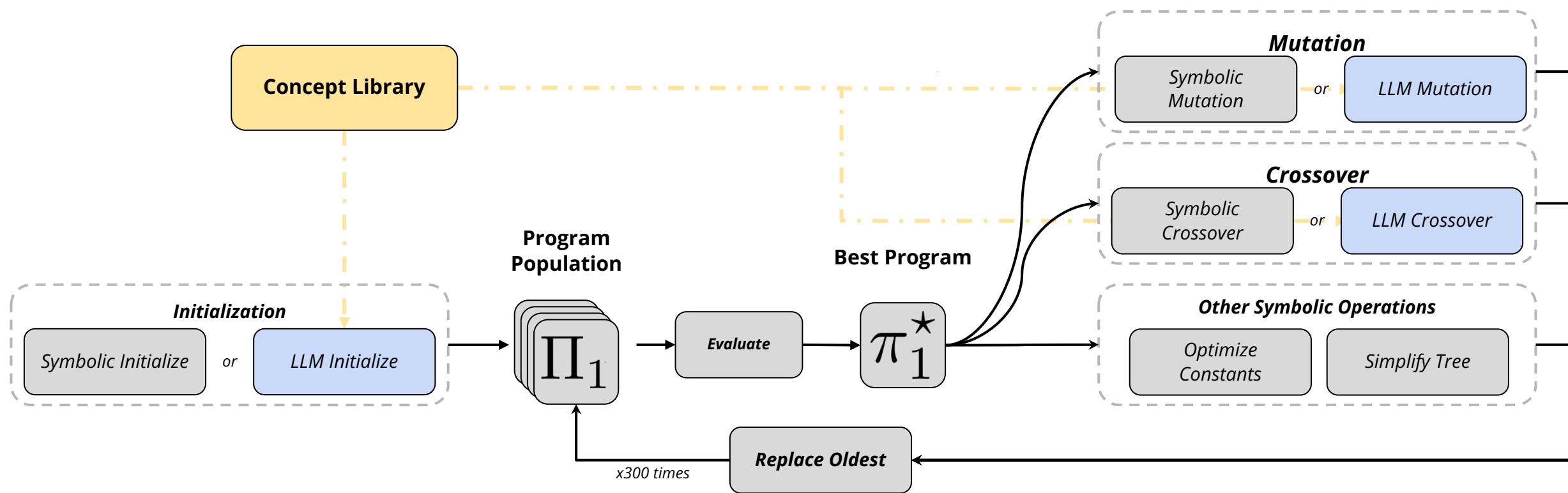


LLM Evolution provides neural guidance (over a language prior)

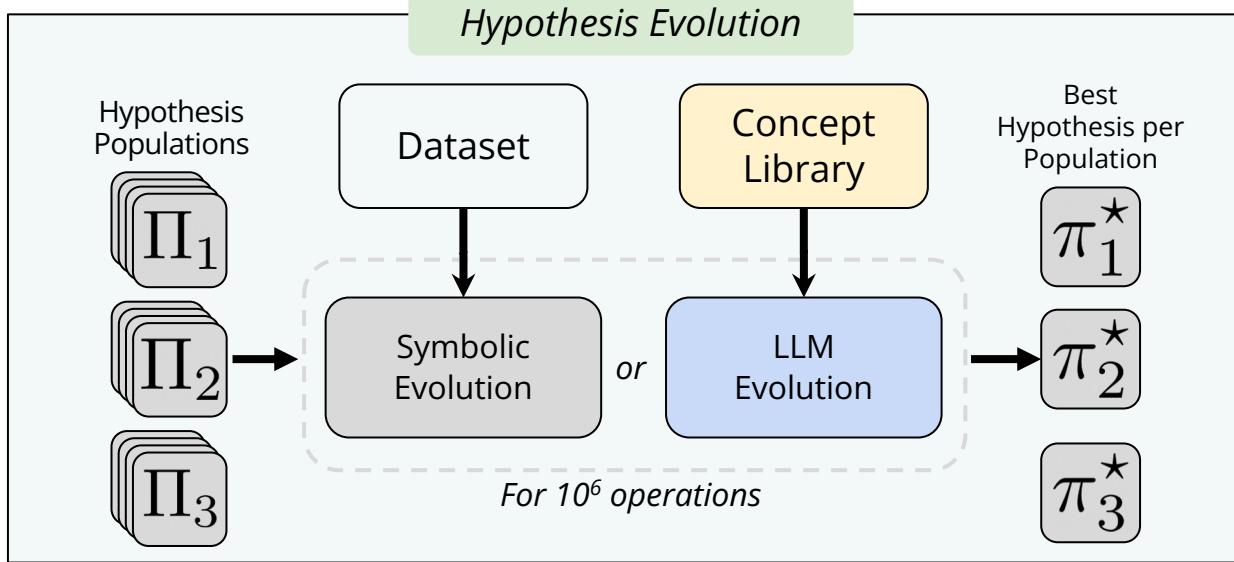




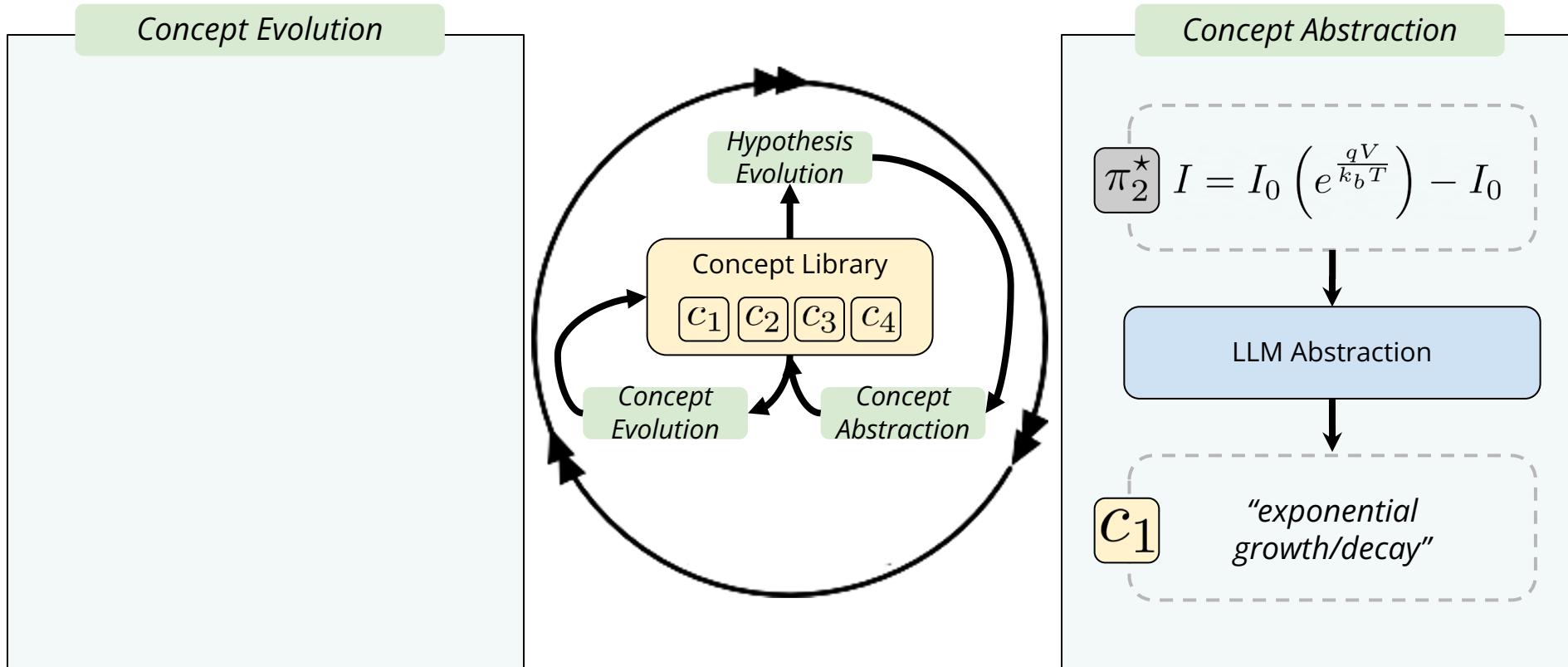




LaSR

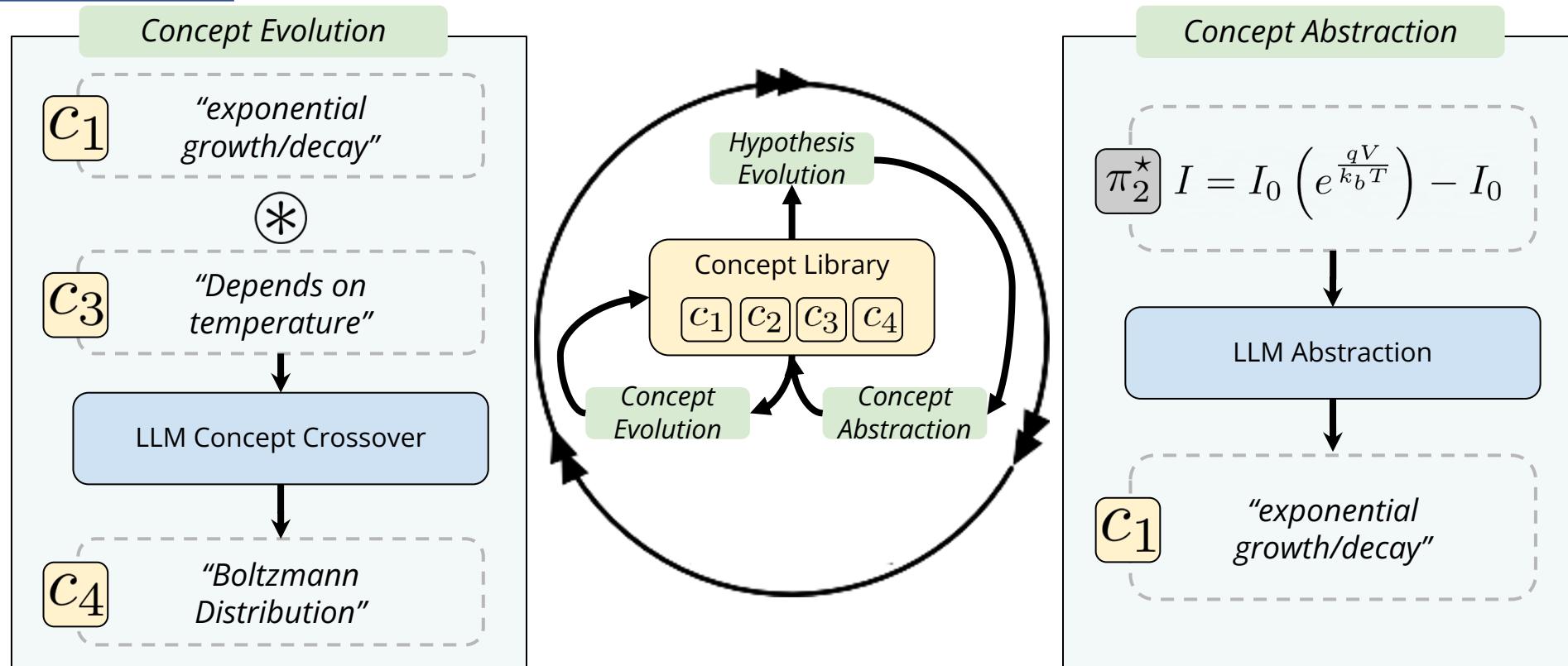
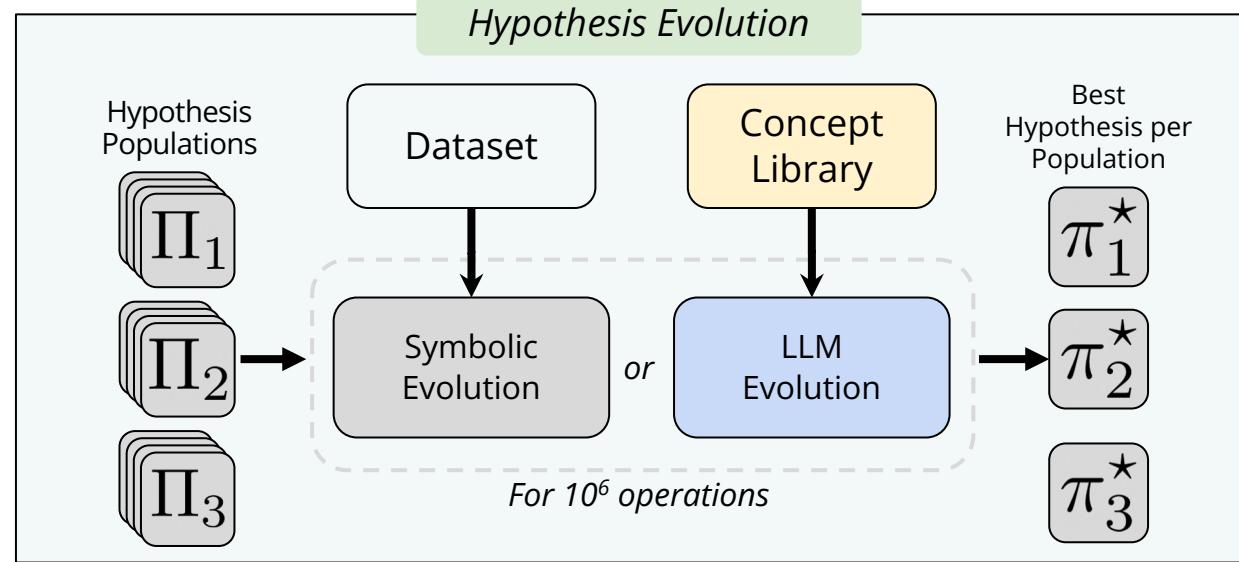


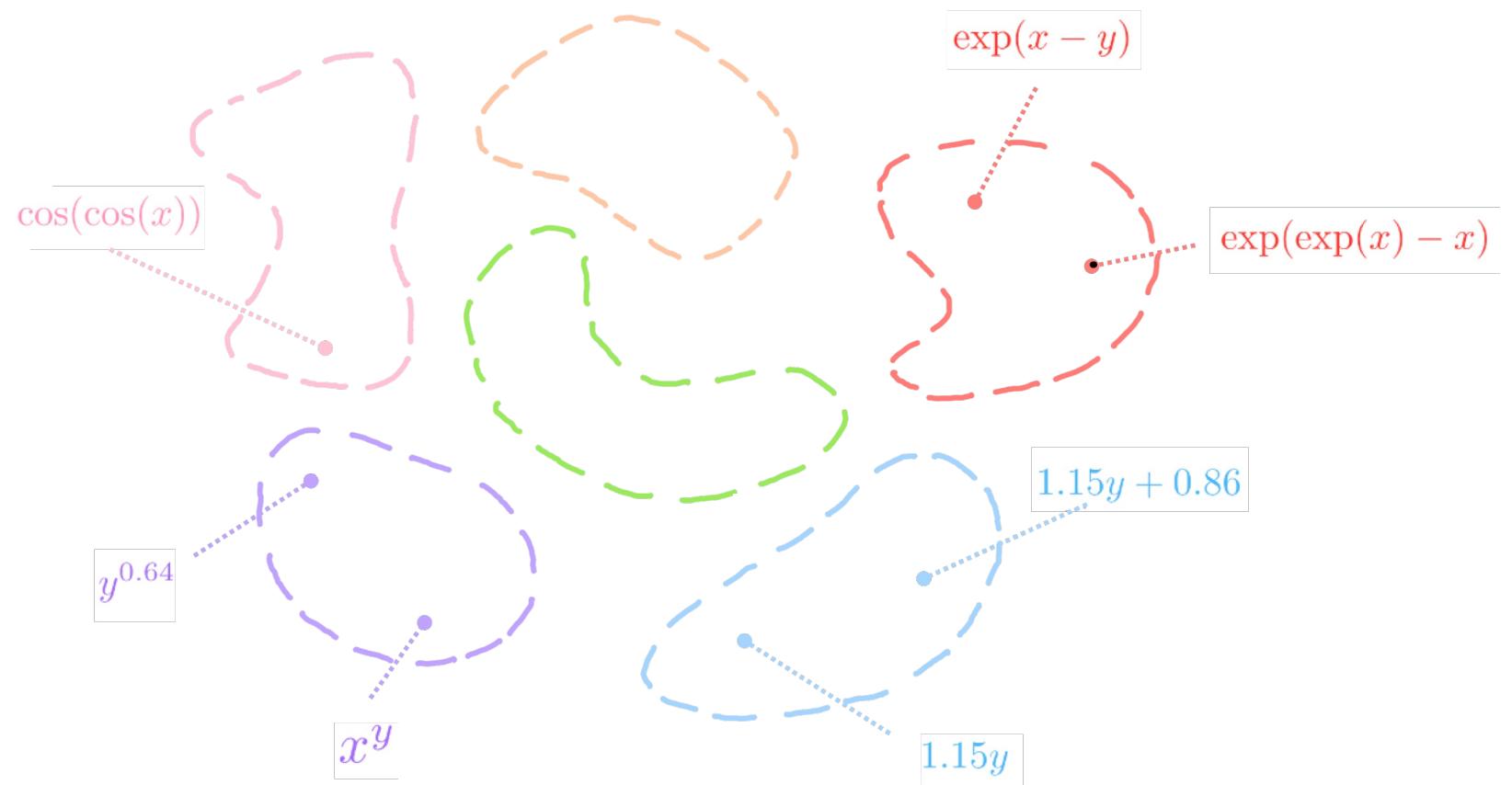
LLM Abstraction
induces *useful** abstractions.

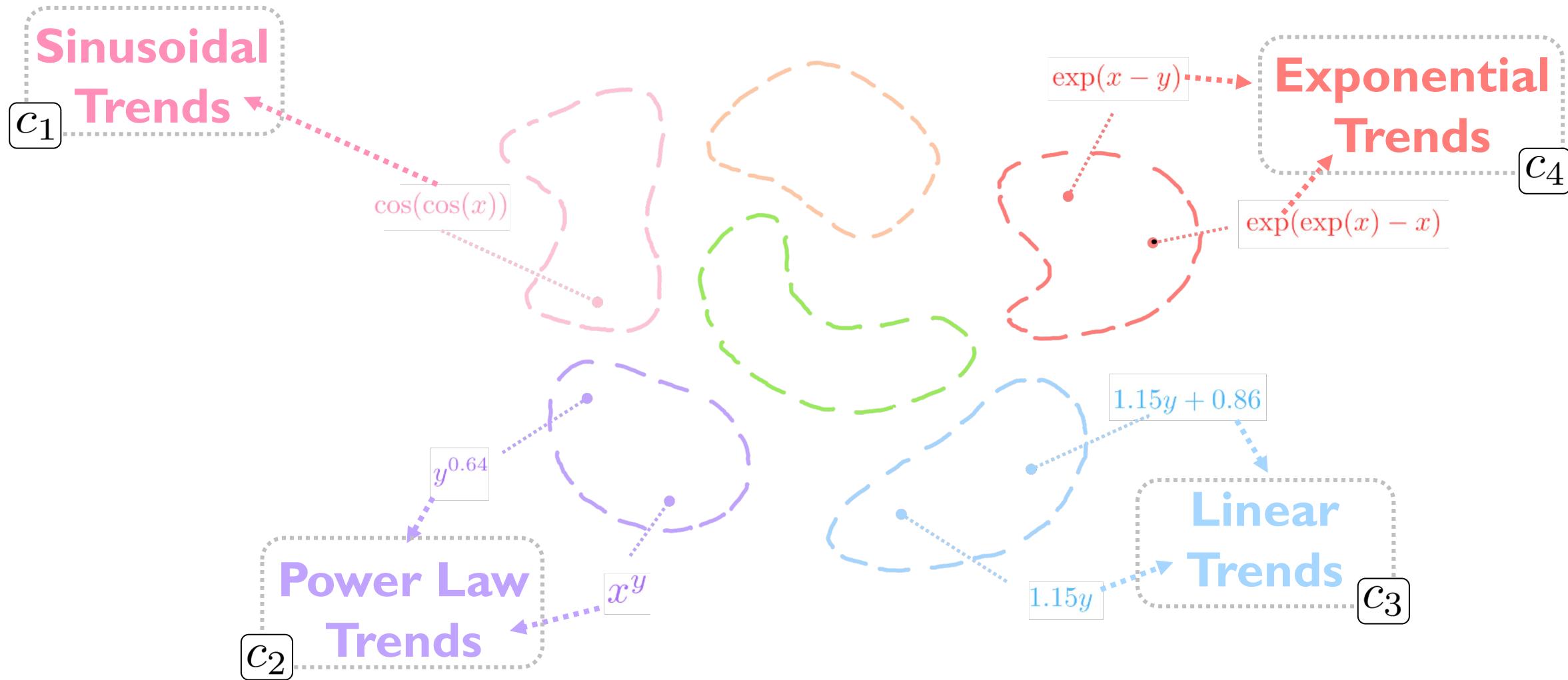


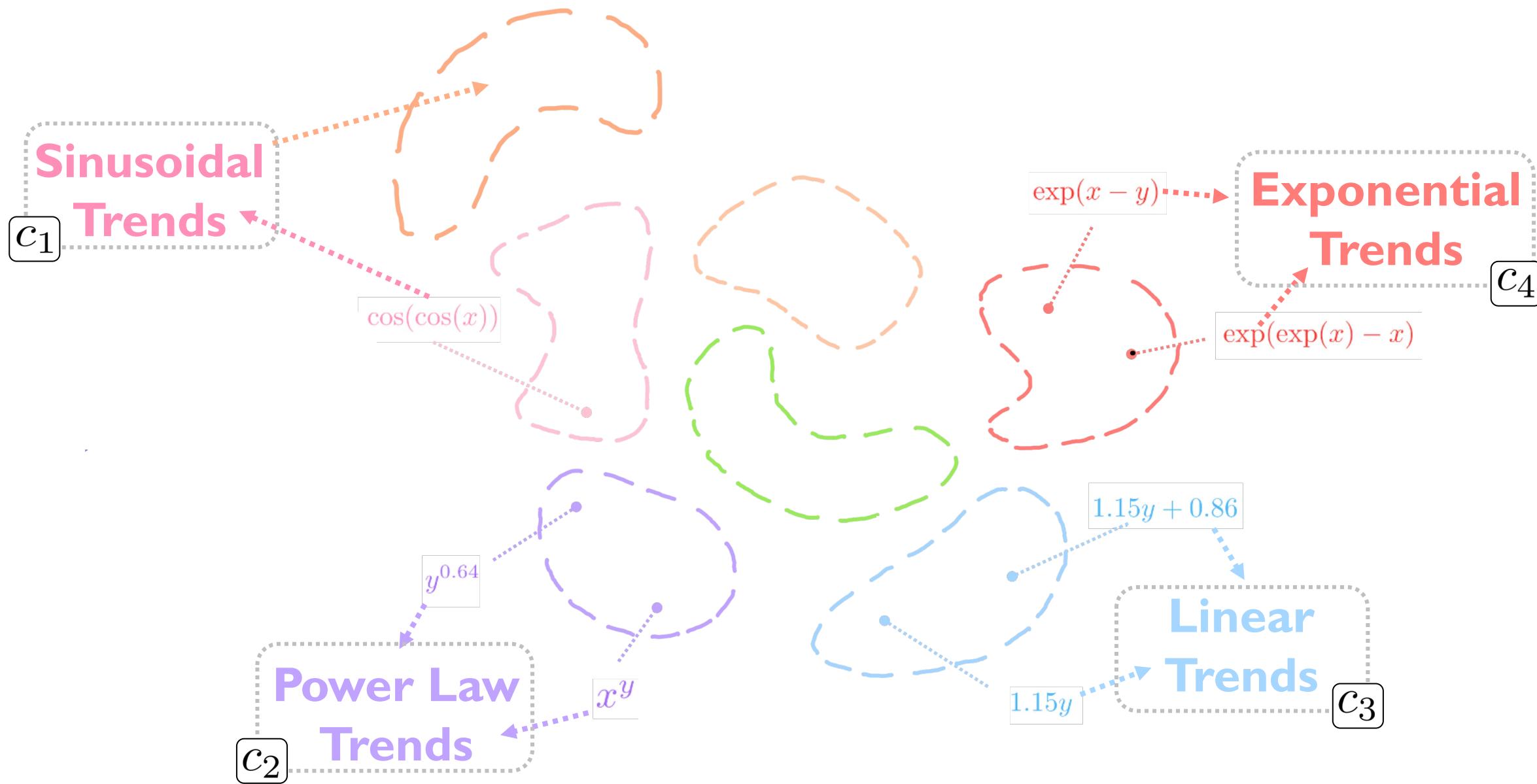
LaSR

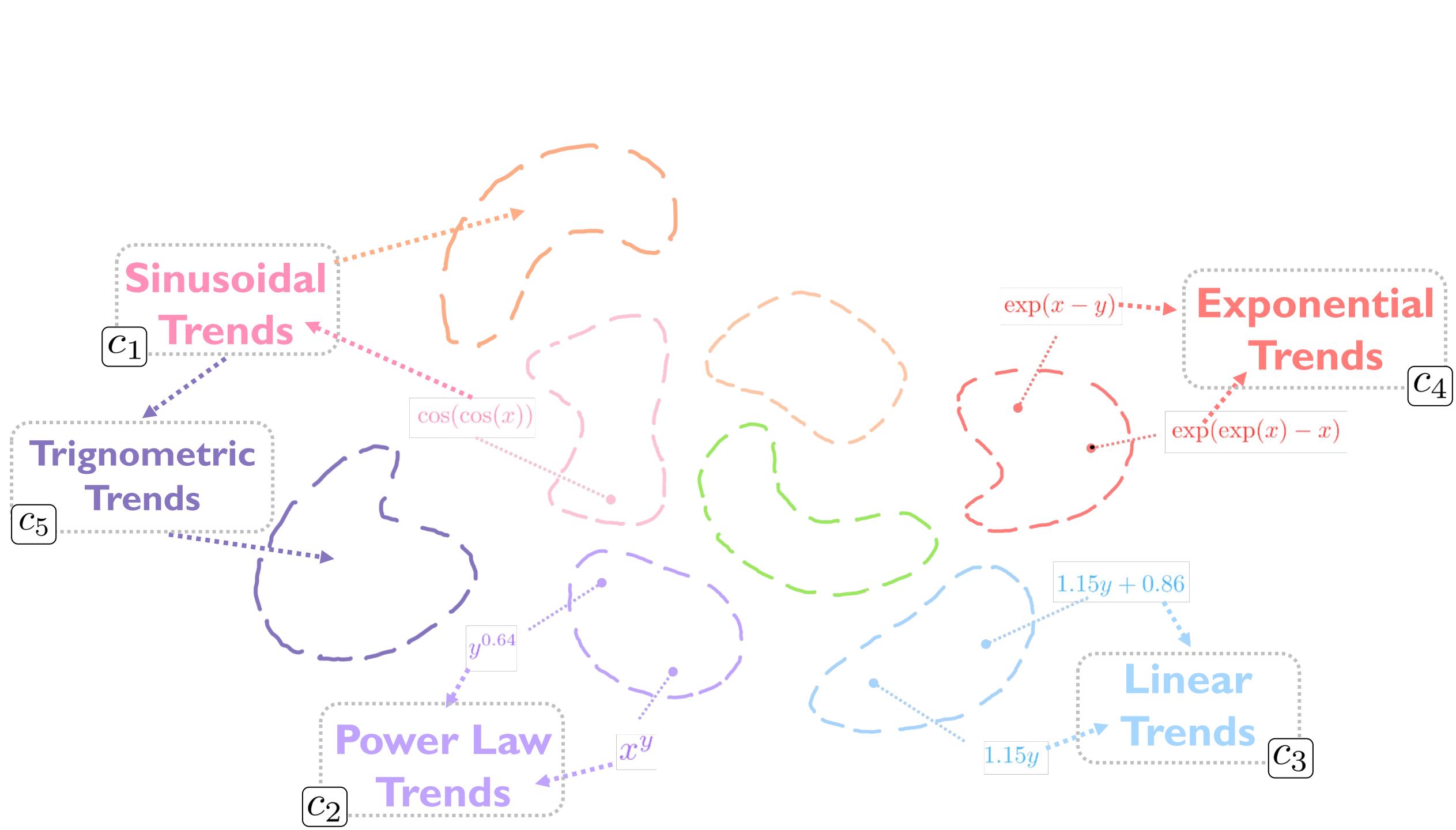
LLM Concept
Crossover evolves
all concepts.











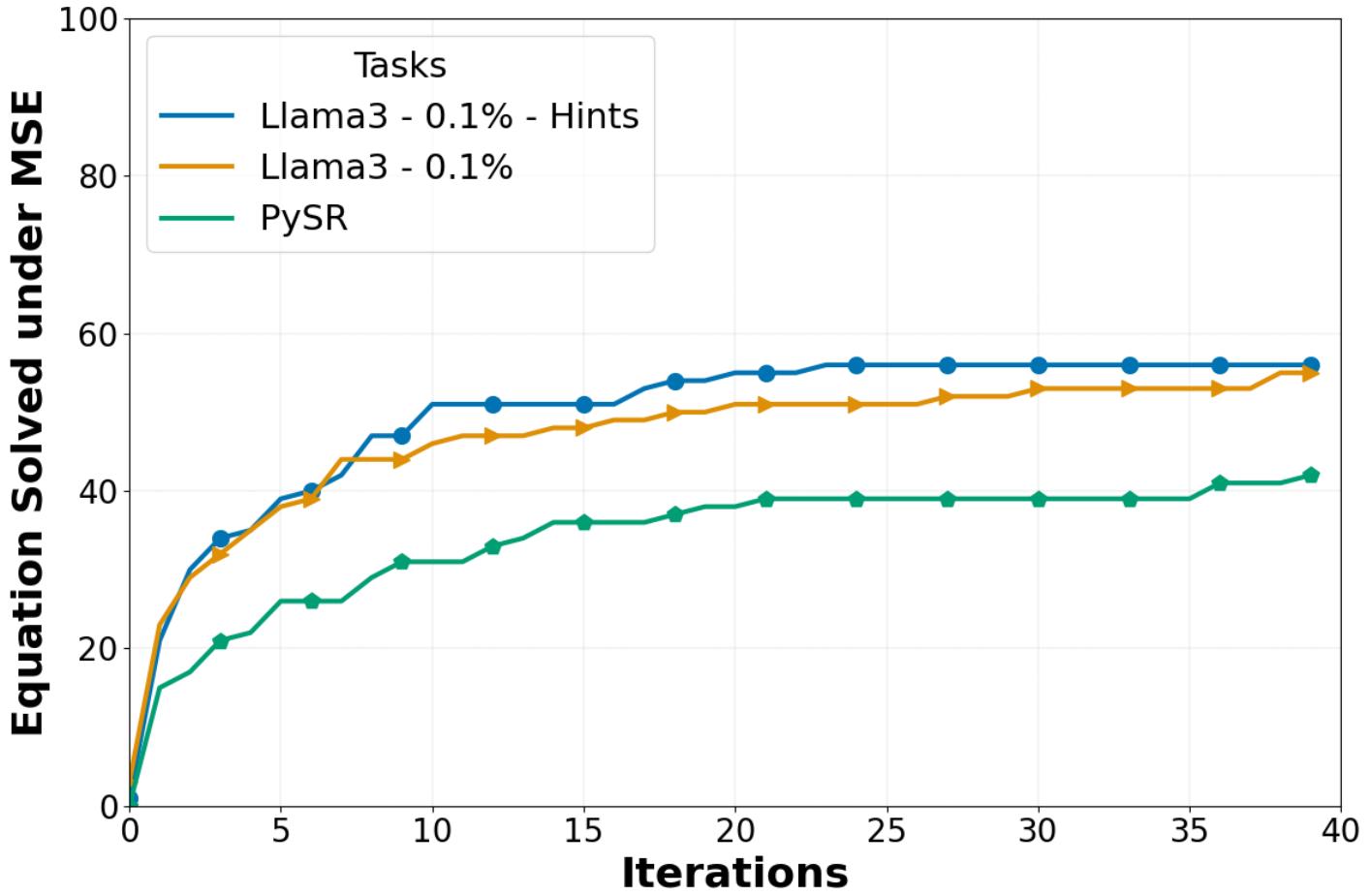
- **Concept Guidance accelerates scientific discovery.**
- **LaSR outperforms PySR even with local language models (llama-3-7b, 1%)**

GLearn	AFP	AFP-FE	DSR	uDSR	AIFeynman	PySR	LaSR
20/100	24/100	26/100	23/100	40/100	38/100	59/100	59 + 7/100

Table 1: Results on 100 Feynman equations from [41]. We report exact match solve rate for all models. LASR achieves the best exact match solve rate using the same hyperparameters as PySR [8].

Type of Solve	PySR	LASR (Llama3-8B)			LASR (GPT-3.5)
		$p = 1\%$	$p = 5\%$	$p = 10\%$	$p = 1\%$
Exact Solve	59/100	63/100	65/100	65/100	66/100
Almost Solve	7/100	6/100	9/100	12/100	13/100
Close	16/100	13/100	14/100	11/100	9/100
Not Close	18/100	18/100	12/100	13/100	13/100

Table 2: Evaluation results on Feynman dataset by cascading LASR’s LLM backbone (llama3-8b, gpt-3.5-turbo) and changing the probability of calling the model ($p = [0.01, 0.05, 0.10]$) in the order of increasing concept guidance. LASR outperforms PySR even with minimal concept guidance using an open-source LLM.



$$F = \frac{1}{4\pi\epsilon} \frac{q_1 q_2}{r^2}$$

Eq 10: Coulomb's Law

- Inverse Square Law
 - Directly proportional to charges
 - Force symmetric w.r.t charges

PySR's Solution

- Reduces to ground truth after 10 steps of simplification.
 - Unwieldy
 - Fitting more constants => more optimization errors

$$F = \frac{1}{4\pi\epsilon} \frac{q_1 q_2}{r^2}$$

Eq 10: Coulomb's Law

- Inverse Square Law
- Directly proportional to charges
- Force symmetric w.r.t charges

$$\begin{aligned}
 F &= \frac{q_1}{\left(\frac{r}{q_2}\right) \left(r + \frac{1.9181636 \times 10^{-5}}{q_2}\right) \epsilon} \cdot 0.07957782 \\
 &= \frac{q_1}{\left(\frac{r}{q_2}\right) \left(r + \frac{1.9181636 \times 10^{-5}}{q_2}\right) \epsilon} \cdot \frac{1}{4\pi} && \text{(Substitute constant)} \\
 &= \frac{q_1 q_2}{r \left(r + \frac{1.9181636 \times 10^{-5}}{q_2}\right) \epsilon} \cdot \frac{1}{4\pi} && \text{(Simplify denominator)} \\
 &\approx \frac{q_1 q_2}{r(r)\epsilon} \cdot \frac{1}{4\pi} && \text{(Negligible. } \frac{1.9181636 \times 10^{-5}}{q_2} \approx 0)
 \end{aligned}$$

LaSR's Solution

- Reduces to ground truth after 4 steps of simplification
- Smaller models synthesize simpler equations!

$$F = \frac{1}{4\pi\epsilon} \frac{q_1 q_2}{r^2}$$

Eq 10: Coulomb's Law

- Inverse Square Law
- Directly proportional to charges
- Force symmetric w.r.t charges

Iteration	Discovered Concept
2	<i>The good mathematical expressions exhibit [...] with a focus on power functions and trigonometric functions [...]</i>
6	<i>The good mathematical expressions exhibit [...] symmetry or regularity [...]</i>
24	<i>The good mathematical expressions have [...] with a specific pattern of division and multiplication</i>

LaSR's Concepts (Limitations)

- Cannot guarantee factuality or correctness.
- Good concepts depend on LLM training.
Concepts can mislead scientists.