



# Main messages

- AI has the potential to transform the way we do science.
- AI can help in human-limited situations
- Curiosity is a human characteristic. Because of it, we have advanced as a civilization.
- Humans need to take responsibility for the bad usage of AI, including reputational damage.
- May enable the largest democratization of science ever
- AI-science needs to be separated from human-science
- This is the beginning of the golden era for science. Benefits for humanity will be immense
- What will happen if science can be automated?

# Literature search

- PathFinder
- Denario
- Claude
- ...

# FOTO: Semantic retrieval of scientific figures from text and sketch queries

(Tohfa et al 2026, ArXiv : )

## The problem

18,660 astro-ph papers in 2025, about 70 every working day

150,000+ new figures each year, 9 per paper on average

## Indexing (one entry per figure, not per paper)

- Embed "paper title | figure caption" as a single string (Adding the title takes R@20 from 0.49 to 0.91)
- bge-base-en-v1.5, 110M parameters, open weights, runs locally
- 482,750 vectors, 768 dimensions, FAISS exact search
- No API key for retrieval



Hurum maksora Tohfa  
University of Washington

**Data (Astro ph, ArxivCap):**  
**Total paper: 53,839**  
**Figures: 482,750**

**Current options: Paper based search (semantic: Pathfinder\*, keyword based, semantic scholar)**

\*Arxiv: 2408.01556

# FOTO



FOTO outperforms existing baseline for every register (the weakest results are from vague queries)

FIGURE-RETRIEVAL RECALL BY QUERY REGISTER OVER 500 SAMPLED FIGURES. FOTO IS SCORED ON *figure* HITS (TARGET CAPTION IN THE TOP  $k$ ); PATHFINDER AND SEMANTIC SCHOLAR ARE SCORED ON THE EASIER *paper* HITS (PAPER CONTAINING THE TARGET FIGURE IN THE TOP  $k$ ). SEMANTIC SCHOLAR QUERIES USE HAIKU-EXTRACTED KEYWORDS; PATHFINDER USES THE RAW QUERY. BEST VALUE IN EACH ROW IS IN BOLD.

| Style    | FOTO (figure) |              |              |              | Pathfinder (paper) |       |       |       | Semantic Scholar (paper) |       |       |       |
|----------|---------------|--------------|--------------|--------------|--------------------|-------|-------|-------|--------------------------|-------|-------|-------|
|          | R@5           | R@10         | R@20         | R@50         | R@5                | R@10  | R@20  | R@50  | R@5                      | R@10  | R@20  | R@50  |
| Terse    | <b>0.554</b>  | <b>0.636</b> | <b>0.716</b> | <b>0.780</b> | 0.114              | 0.142 | 0.176 | 0.204 | 0.114                    | 0.140 | 0.164 | 0.192 |
| Casual   | <b>0.464</b>  | <b>0.528</b> | <b>0.602</b> | <b>0.674</b> | 0.128              | 0.170 | 0.198 | 0.268 | 0.100                    | 0.130 | 0.150 | 0.176 |
| Vague    | <b>0.164</b>  | <b>0.232</b> | <b>0.288</b> | <b>0.368</b> | 0.042              | 0.048 | 0.058 | 0.094 | 0.044                    | 0.066 | 0.092 | 0.106 |
| Detailed | <b>0.696</b>  | <b>0.750</b> | <b>0.792</b> | <b>0.840</b> | 0.110              | 0.134 | 0.162 | 0.216 | 0.068                    | 0.078 | 0.090 | 0.102 |
| Notation | <b>0.354</b>  | <b>0.400</b> | <b>0.458</b> | <b>0.544</b> | 0.070              | 0.090 | 0.116 | 0.148 | 0.050                    | 0.066 | 0.076 | 0.082 |
| Mean     | <b>0.446</b>  | <b>0.509</b> | <b>0.571</b> | <b>0.641</b> | 0.093              | 0.117 | 0.142 | 0.186 | 0.075                    | 0.096 | 0.114 | 0.132 |



FOTO webapp

Science

# Vibe physics: The AI grad student

Mar 23, 2026



*Can AI do theoretical physics? In this guest post, professor of physics Matthew Schwartz decided to find out by supervising Claude through a real research calculation, start to finish, without ever touching a file himself. His account of what happened is below.*

# Values

- AI should enhance humans, not replace them
- We are not artists; we must be guided by truth and knowledge
- Develop our own tools; not just rely on commercial products
- AI should benefit humanity; regulation

# Goals

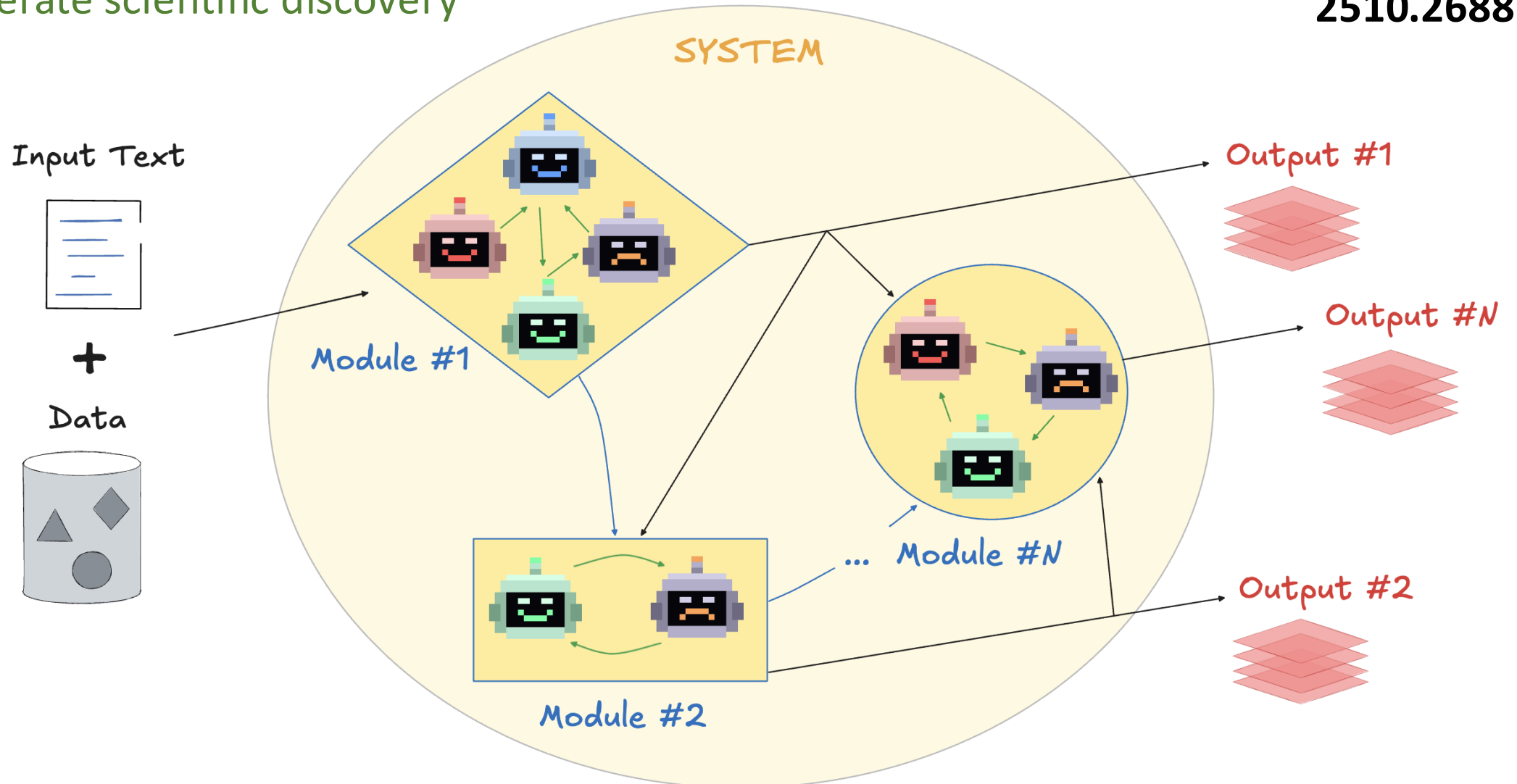
- Derisk science
- Tackle extremely hard problems
- Reduce human biases
- Accelerate scientific discovery
- Have humans doing the fun part

# Denario

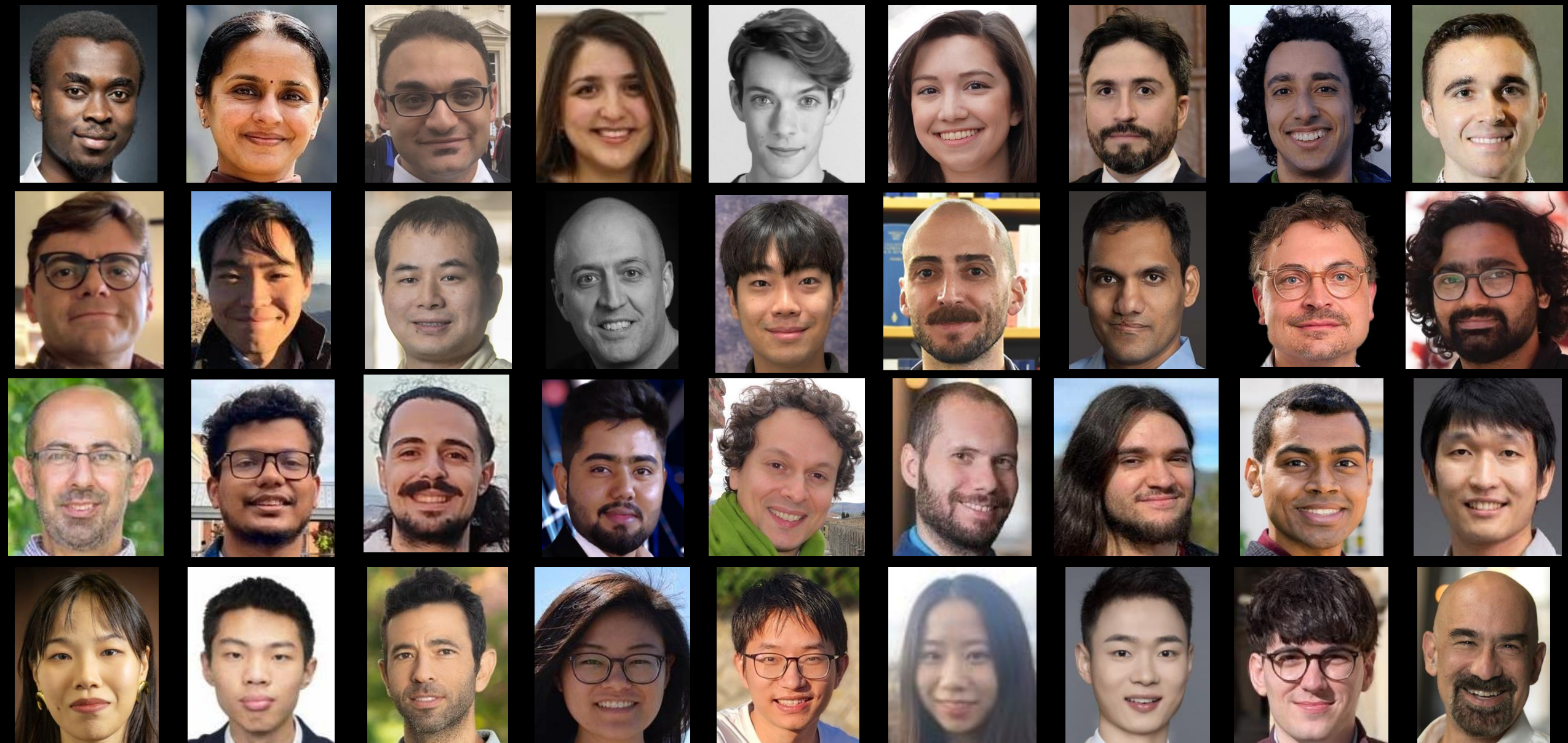
~~Goal: Automate science~~

Goal: Accelerate scientific discovery

Publicly available  
2510.26887



# Denario team



# Developers

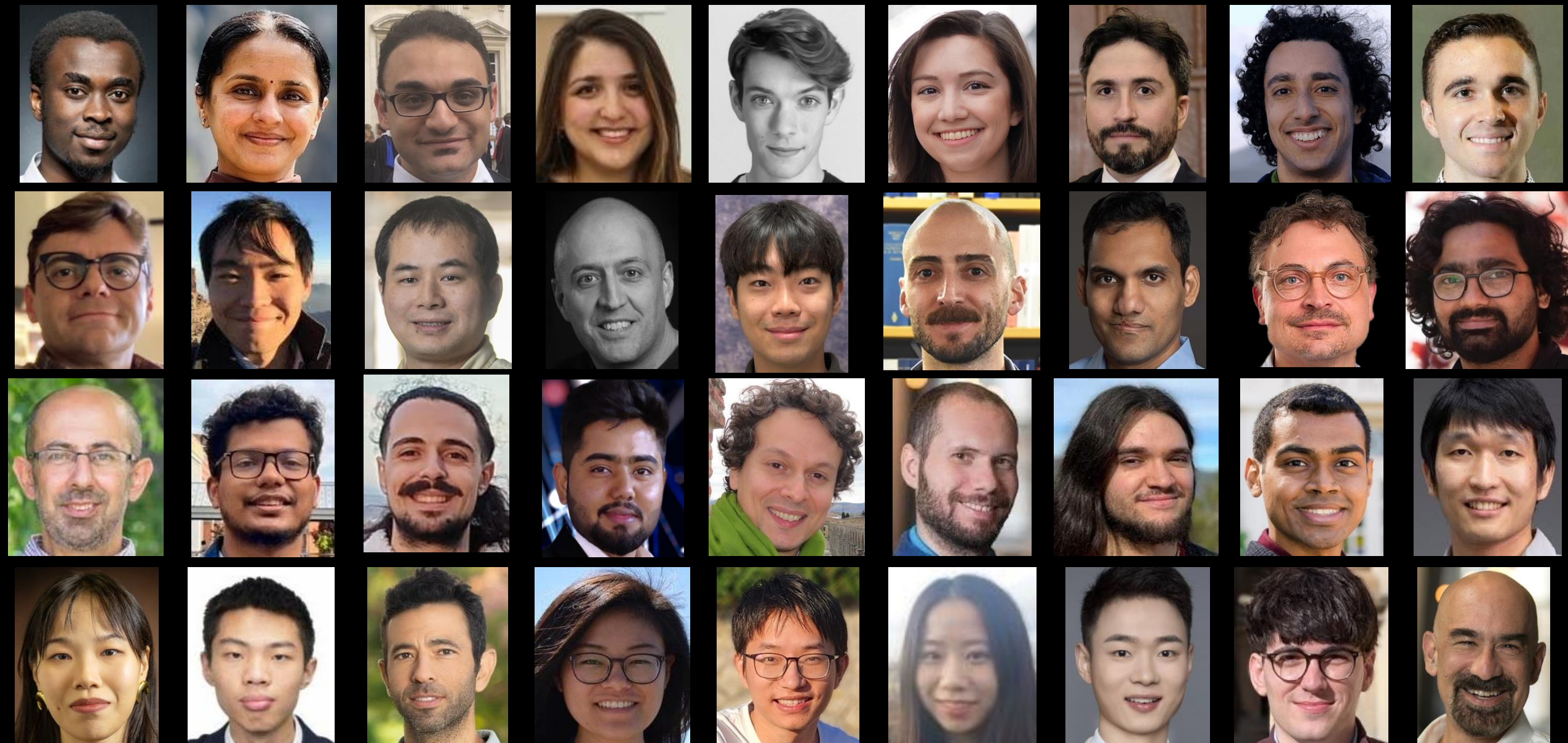


Boris Bolliet  
(Cambridge)



Pablo Villanueva-Domingo  
(Barcelona)

# Denario team



# Astrophysics



Adrian Bayer  
(CCA/Princeton)



ChangHoon Hahn  
(UT Austin)



Jun-Young Lee  
(Princeton)



Licong Xu  
(Cambridge)



Yan-Fei Jiang  
(CCA)



Jay Wadekar  
(John Hopkins)



Matteo Viel  
(SISSA)

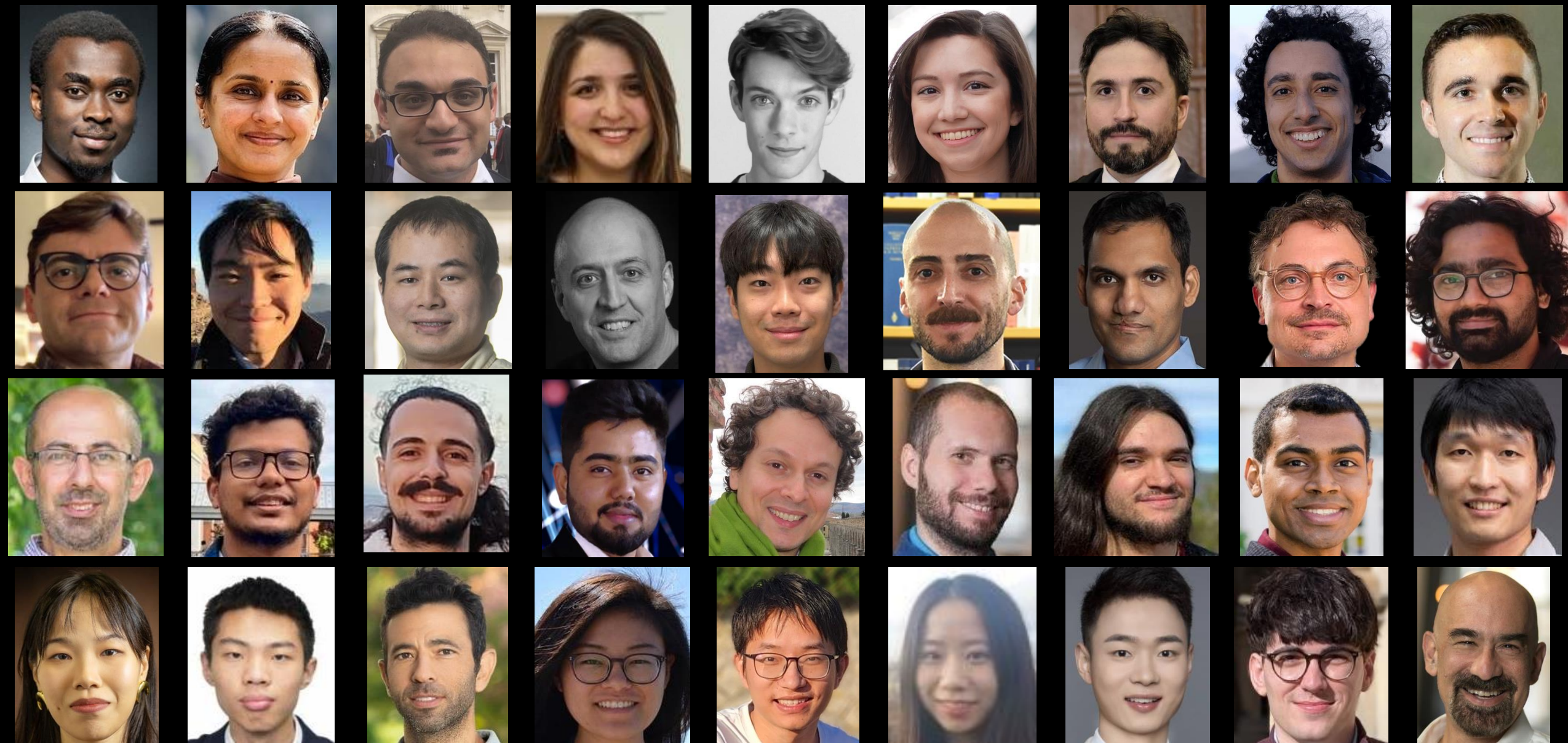


Iñigo Zubeldia  
(Cambridge)



David Spergel  
(Simons)

# Denario team



# Biology



Aidan Acqua  
(Oxford)



Pablo Cardenas Ramirez  
(Cornell/MIT/Harvard)



Qiyao Zhu  
(CCB)



Anupam Anand Ojha  
(CCB/CCM)

# Chemistry

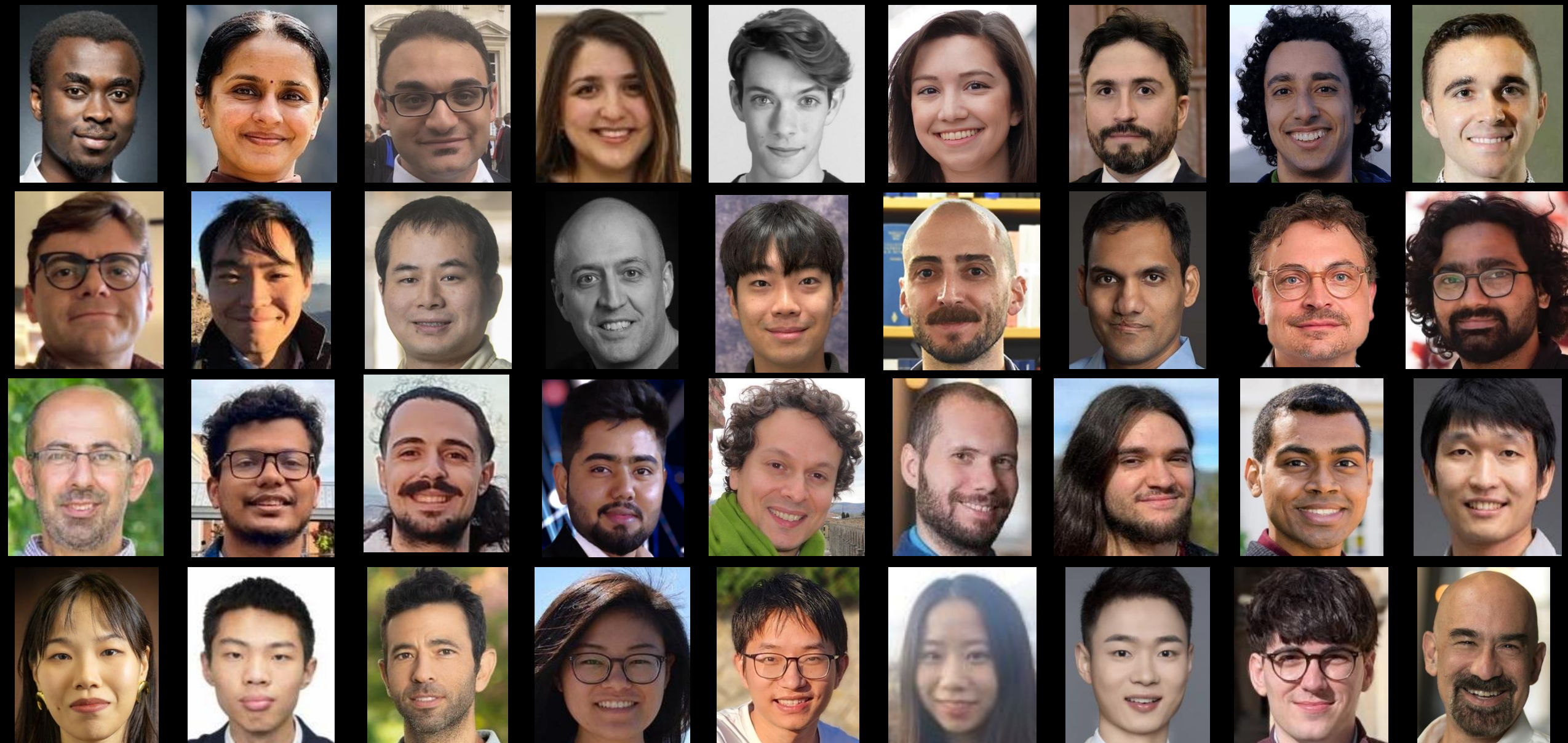


Camille Bilodeau  
(Virginia)



Shimanto Roy  
(Virginia)

# Denario team



# Neuroscience



Yossi Yovel  
(Tel-Aviv)



Almog Barzilay-Siegal  
(Tel-Aviv)



Jiajun Zou  
(SISSA)

# Material Science

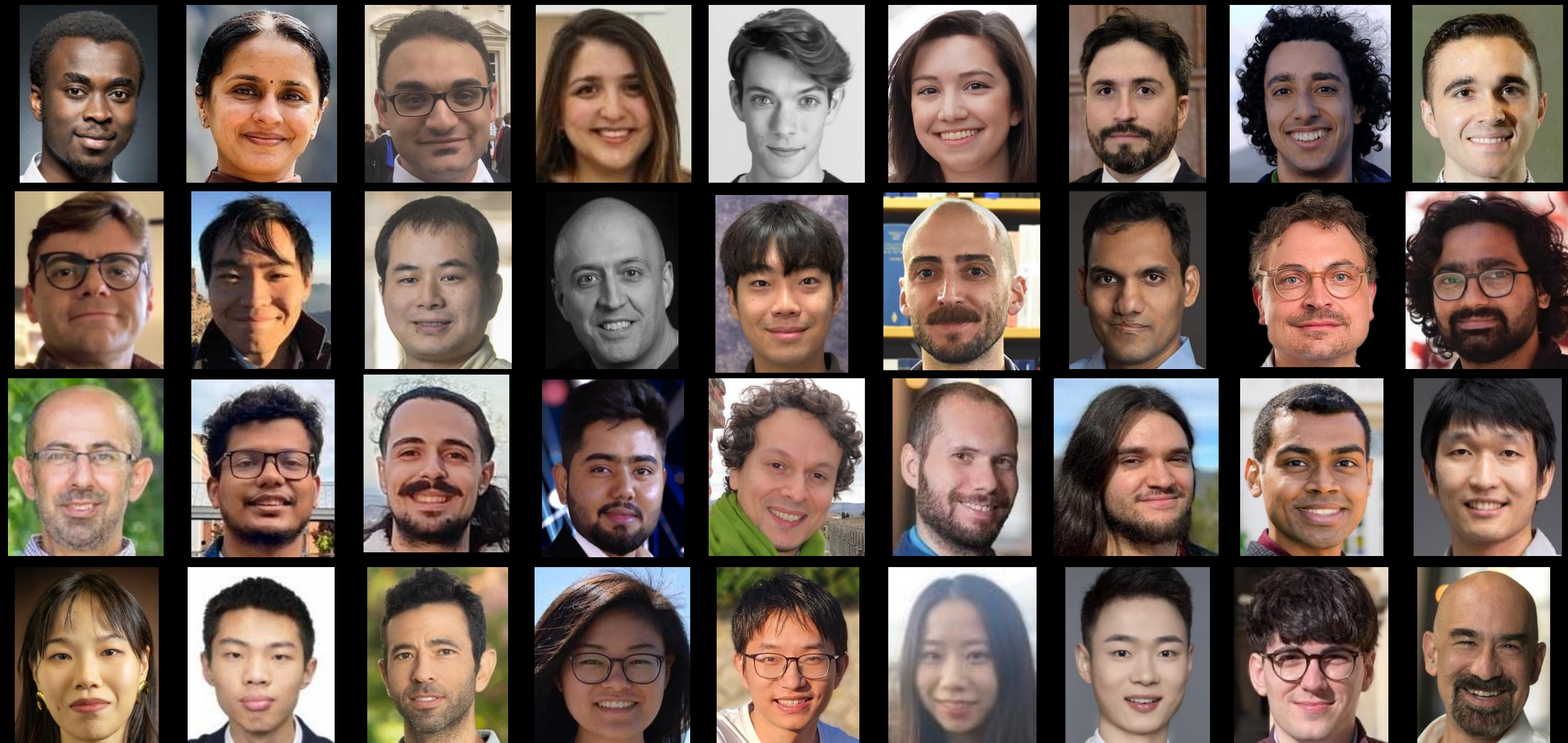


Osman Mamun  
(Cornell)



Shuwen Yue  
(Cornell)

# Denario team



# Machine Learning



Chi Wang  
(DeepMind)

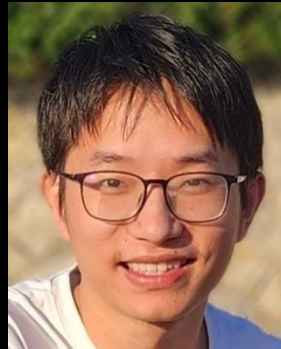


Pavlos Protopapas  
(Harvard)



Miles Cranmer  
(Cambridge)

# Planetary Science

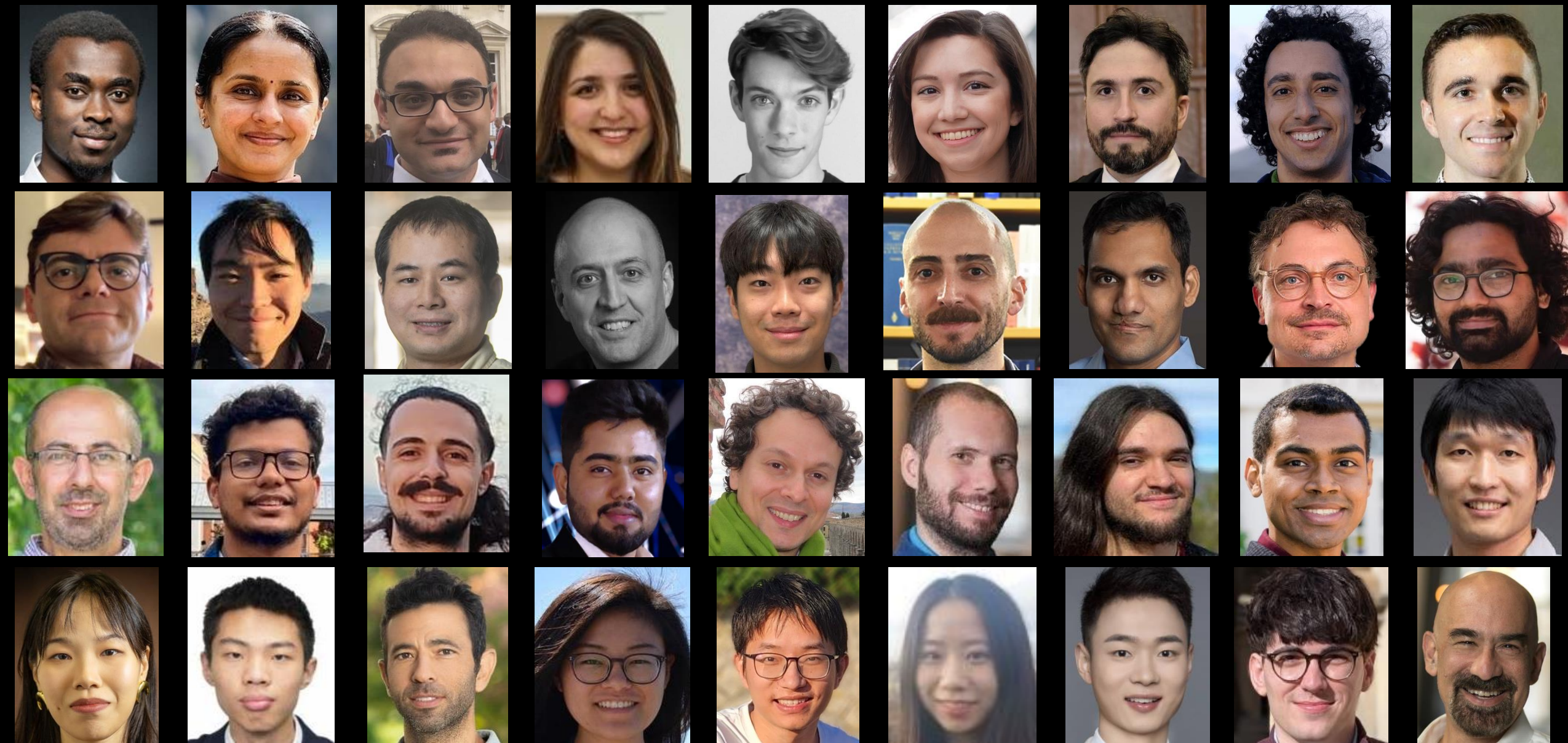


Wenhan Zhou  
(Tokyo)



Bonny Wang  
(Chicago)

# Denario team



# Physics



Raul Jimenez  
(Barcelona)



Pedro Tarancon  
(Barcelona)

# Software engineering

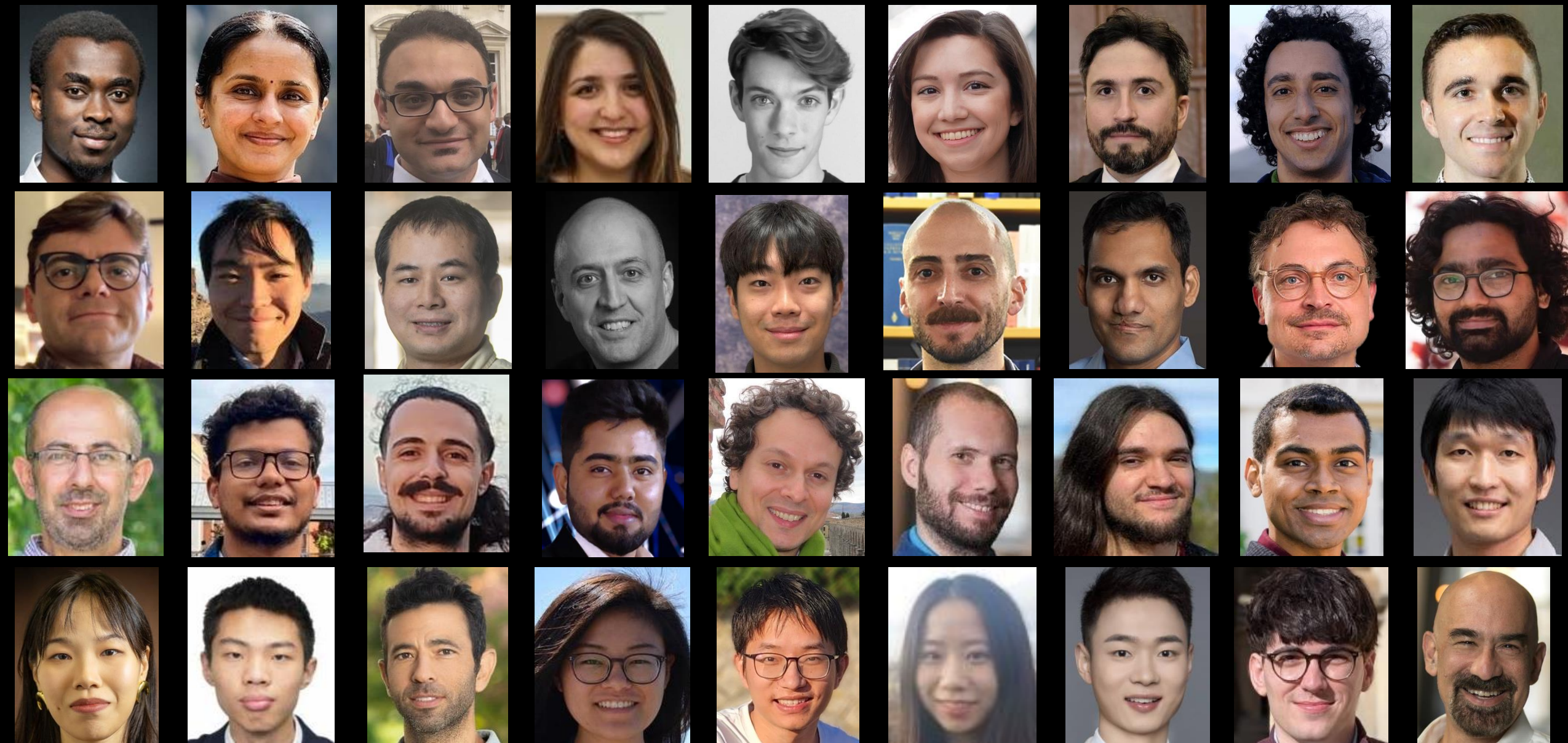


Chetana Amancharla  
(Infosys)



Ujjwal Tiwari  
(Infosys)

# Denario team



# Quantum physics



Pablo Bermejo  
(Google/DIPC/CCQ)

# Medicine



Urbano França  
(Harvard)

# Mathematics



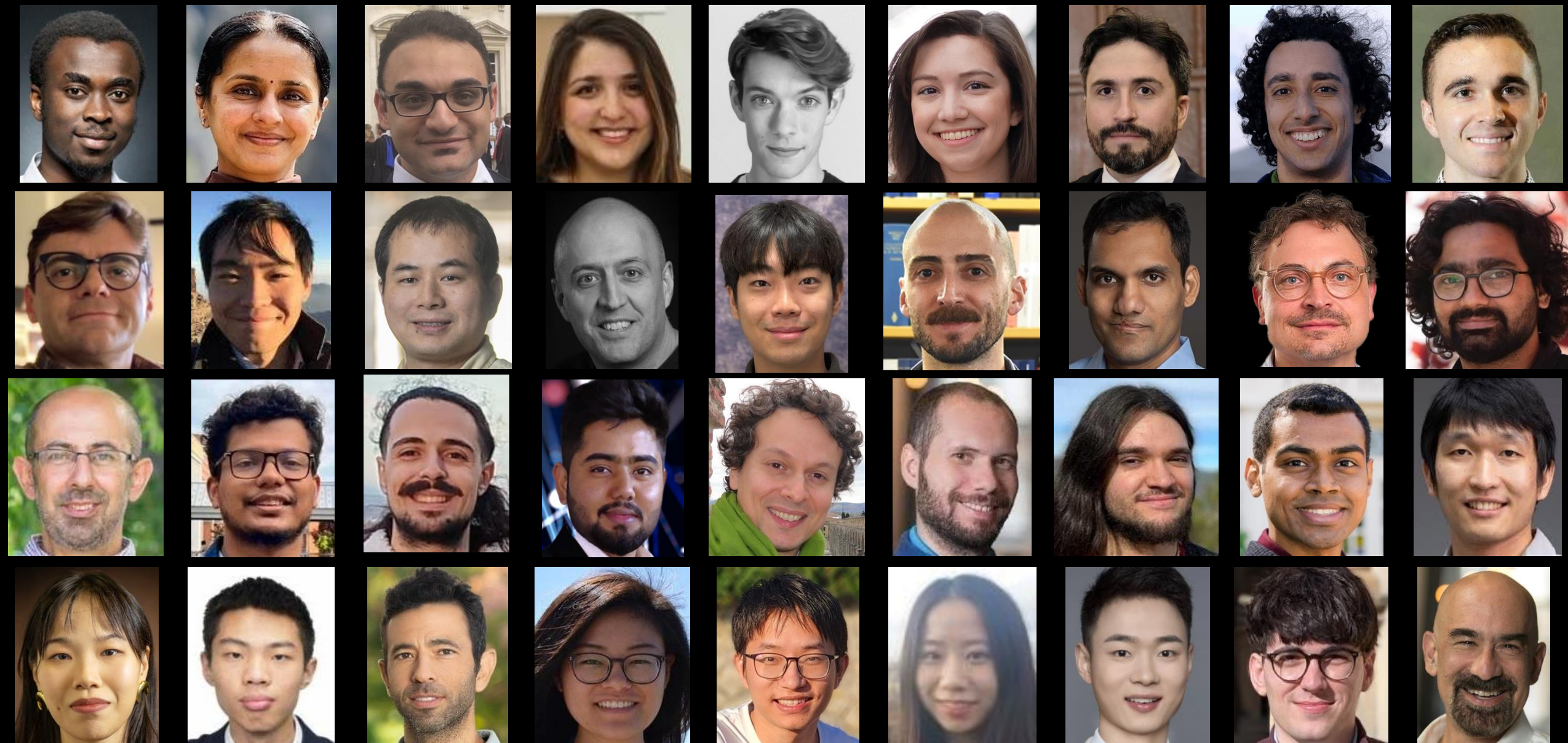
Antonio Lerario  
(SISSA)

# Philosophy



Thomas Meier  
(Munich)

# Denario team



# Denario

GitHub repository:

<https://github.com/AstroPilot-AI/Denario>

Installation:

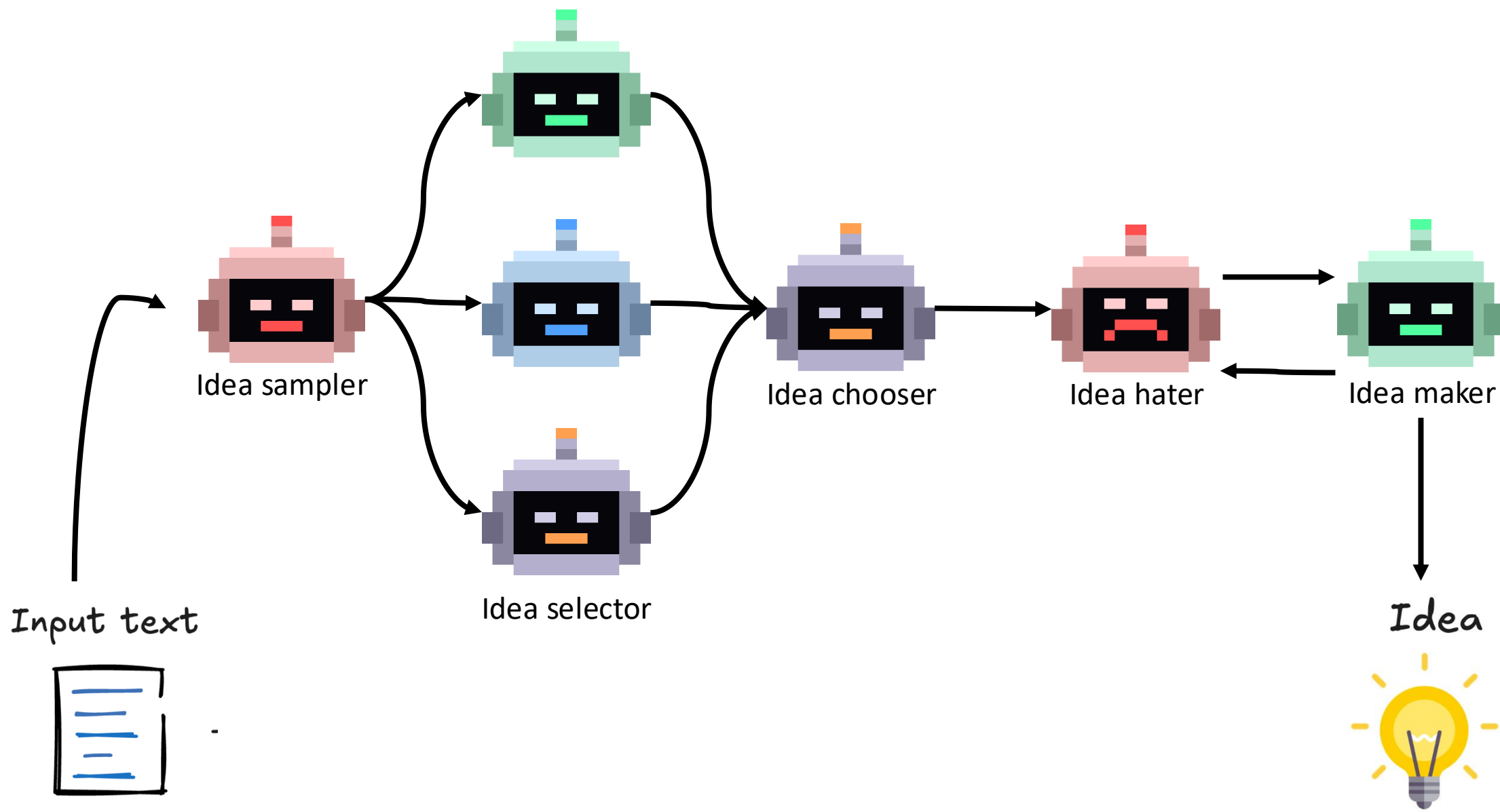
```
>> pip install denario
```

HuggingFace app:

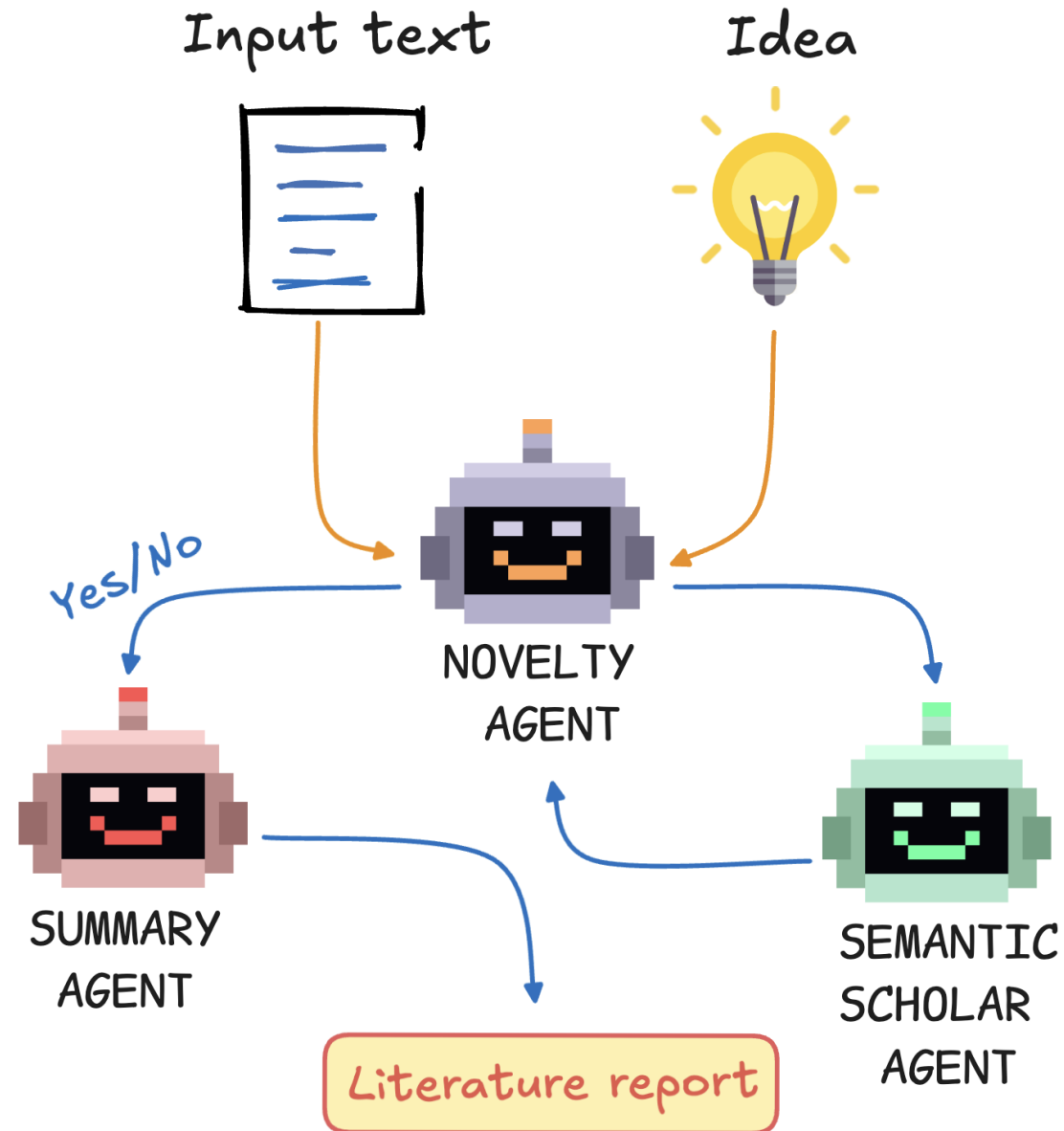
<https://huggingface.co/spaces/astropilot-ai/Denario>

**Open invitation to everyone to join us!**

# Idea module

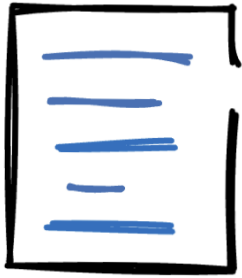


# Literature module



# Methods module

Input text



+

Idea



METHODS  
AGENT

Research methodology

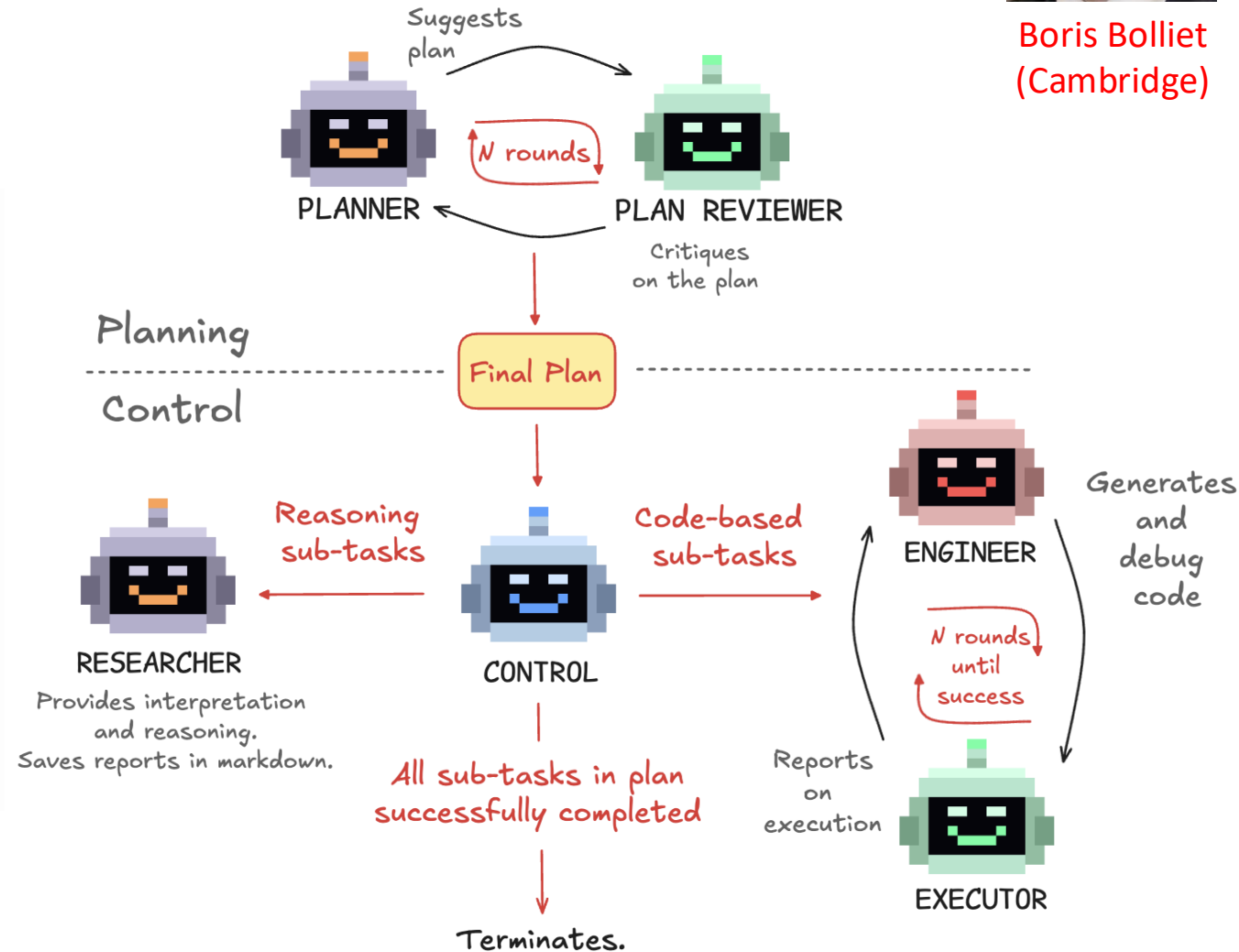
# Deep Research module

CMBAgent

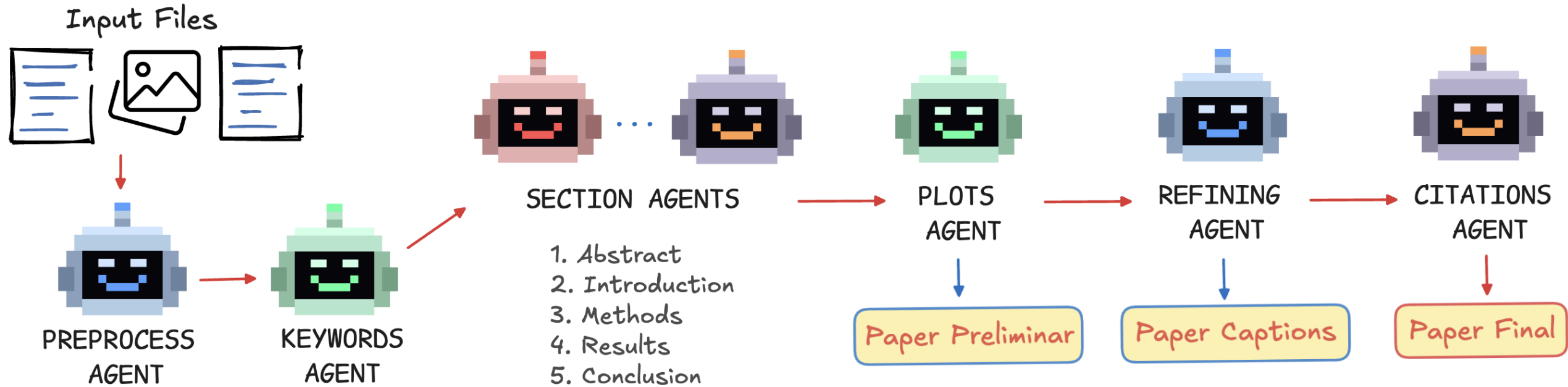
<https://github.com/CMBAgents/cmbagent>



Boris Bolliet  
(Cambridge)



# Writing module



# skeptical: scientific review+verification system

<https://skeptical-ai.org>

**skeptical**

The AI reviewer



Paper loaded: 2511.23473v1.pdf

Drag a new file here or [change paper](#).

## Review thoroughness

- ☒ Standard
- ☐ High

## Additional review features (optional)

- ☒ Thorough figure review
- ☒ Statement verification
- ☒ Mathematical audit
- ☒ Numerical audit

## Your email address (or addresses)

The review should reach you in a few minutes.

iz221@cam.ac.uk ✕

Start review

Paper uploaded. Ready to review.

## *Skeptical review: FastEddington: A Streamlined Python Package for Interactive Eddington Bias Simulation in Power-Law Distributions with Gaussian and Log-Normal Errors*

### Summary

The paper presents FastEddington, a Python package to simulate and visualize Eddington bias for lower-truncated, single-slope power-law populations under either additive Gaussian or multiplicative log-normal measurement errors. The described workflow (inverse-transform sampling → noise injection → selection/thresholding → naive truncated power-law MLE → visualization of slope/normalization shifts; Sec. 2–3) is clear and the example experiment ( $n=100,000$ , true  $\alpha=2.5$ ,  $x_{\min}=1.0$ ,  $\sigma=0.3$ ; Sec. 3) demonstrates the expected flattening of the recovered slope when measurement scatter and a hard detection threshold are present. To make the package's results fully interpretable and the manuscript's claims well-supported, the paper needs (i) a more explicit generative/selection model description (what is simulated, what is selected, what is fit), (ii) correction/clarification of the stated power-law likelihood and fitting procedure (including constraints and/or the closed-form estimator), (iii) clearer uncertainty quantification and robustness checks across realizations, and (iv) better grounding of claims about performance, novelty, and generality via benchmarks, parameter sweeps, and improved software availability/reproducibility details.

### Strengths

- Clear motivation and explanation of Eddington bias and why naive power-law fits can be misleading when measurement scatter interacts with a hard threshold (Sec. 1, Sec. 3).
- Conceptually standard and easy-to-follow simulation components (inverse-transform sampling; Gaussian and log-normal noise injection) make the tool pedagogically useful (Sec. 2.2–2.3).
- Modular package framing (generation, perturbation, fitting, plotting) supports reuse and extension (Sec. 2.1, Sec. 2.6–2.8).
- The example results show qualitatively and quantitatively large parameter shifts consistent with expectations, and the narrative ties the observed flattening to selection effects (Sec. 3.1–3.5).



# The AI Scientist: Towards Fully Automated Open-Ended Scientific Discovery

Chris Lu<sup>1,2,\*</sup>, Cong Lu<sup>3,4,\*</sup>, Robert Tjarko Lange<sup>1,\*</sup>, Jakob Foerster<sup>2,†</sup>, Jeff Clune<sup>3,4,5,†</sup> and David Ha<sup>1,†</sup>  
\*Equal Contribution, <sup>1</sup>Sakana AI, <sup>2</sup>FLAIR, University of Oxford, <sup>3</sup>University of British Columbia, <sup>4</sup>Vector Institute, <sup>5</sup>Canada CIFAR AI Chair, <sup>†</sup>Equal Advising

2025-02-18

## Towards an AI co-scientist

Juraj Gottweis<sup>\*,1</sup>, Wei-Hung Weng<sup>\*,2</sup>, Alexander Daryin<sup>\*,1</sup>, Tao Tu<sup>\*,3</sup>,  
Anil Palepu<sup>2</sup>, Petar Sirkovic<sup>1</sup>, Artiom Myaskovsky<sup>1</sup>, Felix Weissenberger<sup>1</sup>,  
Keran Rong<sup>3</sup>, Ryutaro Tanno<sup>3</sup>, Khaled Saab<sup>3</sup>, Dan Popovici<sup>2</sup>, Jacob Blum<sup>7</sup>, Fan Zhang<sup>2</sup>,  
Katherine Chou<sup>2</sup>, Avinatan Hassidim<sup>2</sup>, Burak Gokturk<sup>1</sup>,  
Amin Vahdat<sup>1</sup>, Pushmeet Kohli<sup>3</sup>, Yossi Matias<sup>2</sup>,  
Andrew Carroll<sup>2</sup>, Kavita Kulkarni<sup>2</sup>, Nenad Tomasev<sup>3</sup>,  
Vikram Dhillon<sup>4</sup>, Eeshit Dhaval Vaishnav<sup>5</sup>, Byron Lee<sup>5</sup>,  
Tiago R D Costa<sup>6</sup>, José R Penadés<sup>6</sup>, Gary Peltz<sup>7</sup>,  
Yunhan Xu<sup>3</sup>, Annalisa Pawlosky<sup>1</sup>, Alan Karthikesalingam<sup>2</sup> and Vivek Natarajan<sup>2</sup>

<sup>1</sup>Google Cloud AI Research, <sup>2</sup>Google Research, <sup>3</sup>Google DeepMind,  
<sup>4</sup>Houston Methodist, <sup>5</sup>Sequome,  
<sup>6</sup>Fleming Initiative and Imperial College London, <sup>7</sup>Stanford University

[Submitted on 18 Feb 2014 (v1), last revised 24 Apr 2014 (this version, v3)]

## Automatic Construction and Natural-Language Description of Nonparametric Regression Models

James Robert Lloyd, David Duvenaud, Roger Grosse, Joshua B. Tenenbaum, Zoubin Ghahramani

This paper presents the beginnings of an automatic statistician, focusing on regression problems. Our system explores an open-ended space of statistical models to discover a good explanation of a data set, and then produces a detailed report with figures and natural-language text. Our approach treats unknown regression functions nonparametrically using Gaussian processes, which has two important consequences. First, Gaussian processes can model functions in terms of high-level properties (e.g. smoothness, trends, periodicity, changepoints). Taken together with the compositional structure of our language of models this allows us to automatically describe functions in simple terms. Second, the use of flexible nonparametric models and a rich language for composing them in an open-ended manner also results in state-of-the-art extrapolation performance evaluated over 13 real time series data sets from various domains.

Letter | Published: 15 January 2004

## Functional genomic hypothesis generation and experimentation by a robot scientist

Ross D. King, Kenneth E. Whelan, Ffion M. Jones, Philip G. K. Reiser, Christopher H. Bryant, Stephen H. Muggleton, Douglas B. Kell & Stephen G. Oliver ✉

*Nature* **427**, 247–252 (2004) | [Cite this article](#)

**8745** Accesses | **507** Citations | **66** Altmetric | [Metrics](#)

### Abstract

The question of whether it is possible to automate the scientific process is of both great theoretical interest<sup>1,2</sup> and increasing practical importance because, in many scientific areas, data are being generated much faster than they can be effectively analysed. We describe a physically implemented robotic system that applies techniques from artificial intelligence<sup>3,4,5,6,7,8</sup> to carry out cycles of scientific experimentation. The system automatically originates hypotheses to explain observations, devises experiments to test these hypotheses, physically runs the experiments using a laboratory robot, interprets the results to falsify hypotheses inconsistent with the data, and then repeats the cycle. Here we apply the system to the determination of gene function using deletion mutants of yeast (*Saccharomyces cerevisiae*) and auxotrophic growth experiments<sup>2</sup>. We built and tested a detailed logical model (involving genes, proteins and metabolites) of the aromatic amino acid synthesis pathway. In biological experiments that automatically reconstruct parts of this model, we show that an intelligent experiment selection strategy is competitive with human performance and significantly outperforms, with a cost decrease of 3-fold and 100-fold (respectively), both cheapest and random-experiment selection.

DENDRAL (1960s)

# Medicine

## Characterizing the Variability and Correlates of U.S. ART Clinic Performance During the COVID-19 Pandemic (2020-2022)

DENARIO<sup>1</sup>

<sup>1</sup>Anthropic, Gemini & OpenAI servers. Planet Earth.

### ABSTRACT

Understanding the variability in Assisted Reproductive Technology (ART) clinic performance is crucial for patients and practitioners, particularly during periods of potential disruption such as the COVID-19 pandemic (2020-2022). This study aimed to characterize the year-to-year variability in key U.S. ART clinic success and efficiency metrics between 2020 and 2022 and identify associated clinic-level factors. Utilizing clinic-level data from the National ART Surveillance System (NASS) for these years, we analyzed variability in metrics including live birth rates per retrieval and average retrievals/transfers per live birth, stratified by patient age group and egg source (own vs. donor). Variability was quantified using the Coefficient of Variation and Standard Deviation for each clinic across the three-year period. Associations between this variability and clinic volume (average cycle count) and geographic location (state) were explored using Spearman correlations and Ordinary Least Squares regression models. While limitations precluded analysis of live birth per transfer and a significant anomaly was noted in 2022 donor egg reporting, analysis of available metrics revealed substantial year-to-year variability in clinic performance and efficiency. Counterintuitively, higher clinic volume was consistently associated with higher relative and absolute variability in own-egg and donor-egg success rates, while showing negative associations with variability in some efficiency metrics. Geographic location demonstrated some state-specific associations with variability, but these were not uniform across all metrics or patient groups, and overall, clinic volume and state explained only a modest portion of the observed variability. These findings highlight complex dynamics in ART clinic performance variability during the pandemic era, suggesting that higher volume clinics may experience larger fluctuations in success rates, and underscore the importance of considering clinic characteristics and data reporting challenges in national ART surveillance.



Urbano Franca  
(Harvard Medical School)

# Astrophysics

## Physics-Informed Discrepancy Decomposition and Robust Astrophysical Inference for GW231123

DENARIO<sup>1</sup>

<sup>1</sup>*Anthropic, Gemini & OpenAI servers. Planet Earth.*

### ABSTRACT

Robust astrophysical interpretations from gravitational-wave parameter inference critically depend on understanding model-dependent biases. We introduce a novel physics-informed framework to systematically decompose and attribute discrepancies among five gravitational-wave waveform models (NRSur7dq4, IMRPhenomXO4a, SEOBNRv5PHM, IMRPhenomXPHM, IMRPhenomTPHM) for the GW231123 event. Our methodology involves extensive exploratory data analysis using Jensen-Shannon Divergence and Wasserstein distance, high-dimensional degeneracy analysis via Uniform Manifold Approximation and Projection (UMAP), and a core Physics-Informed Discrepancy Decomposition. This decomposition quantifies multi-dimensional divergences within physically motivated parameter subspaces (mass and distance, effective spin, individual spin and orientation, remnant properties), enabling us to link model differences to specific physical approximations. Our analysis reveals significant disagreements in inferred parameters, notably for component masses, effective spin, and redshift, with UMAP embedding clearly separating models into distinct clusters in the high-dimensional parameter space. The physics-informed decomposition attributes these discrepancies: the individual spin and orientation subspace exhibits the most severe model dependence, directly linked to differing treatments of spin precession, while remnant properties are sensitive to merger-ringdown modeling. Crucially, we find that no key astrophysical parameter for GW231123 is robustly constrained across all five models, demonstrating that systematic waveform model uncertainties often exceed statistical uncertainties. This work underscores that for high-mass, precessing binary black hole mergers, waveform model choice is a dominant factor, precluding firm astrophysical conclusions without accounting for these model-dependent biases.



Jay Wadekar  
(John Hopkins)

# Biophysics



Anupam Anand Ojha  
(CCB/CCM)

## Comprehensive Kinetic and Free Energy Analysis of NTL9 Folding via Systematic Collective Variable Selection and Markov State Models

AstroPilot

*Anthropic, Gemini & OpenAI servers. Planet Earth.*

Understanding the complex pathways and kinetics of protein folding from molecular dynamics simulations requires sophisticated analytical tools. We developed and applied a comprehensive pipeline to analyze a 10  $\mu$ s molecular dynamics trajectory of the fast-folding N-terminal domain of ribosomal protein L9 (NTL9), aiming to provide quantitative insights into its folding mechanism. Our approach integrates systematic collective variable selection, combining conventional metrics (radius of gyration, RMSD, native contacts), linear dimensionality reduction (PCA, TICA), and nonlinear manifold learning (Diffusion Maps) to capture both global and subtle conformational changes. Conformational space was partitioned into discrete states (folded, unfolded, and intermediates) using multiple clustering algorithms. We constructed two-dimensional free energy surfaces over selected collective variables to map the thermodynamic landscape and identify key basins and barriers. Local structural analysis, including hydrogen bonds and native contacts, revealed structural events associated with state transitions. Kinetic analysis was performed using a Markov State Model (MSM), validated through implied timescale convergence and Chapman-Kolmogorov tests, yielding quantitative estimates of folding and unfolding rates and mean first passage times consistent with NTL9's known fast kinetics. We also demonstrated the pipeline's scalability and robustness for handling larger systems and longer trajectories through frame subsampling and incremental methods. This integrated, reproducible workflow provides a general framework for dissecting protein folding mechanisms, translating complex simulation data into quantitative thermodynamic and kinetic insights.

# Quantum physics + cosmology + machine learning

## QITT-Enhanced Multi-Scale Substructure Analysis with Learned Topological Embeddings for Cosmological Parameter Estimation

DENARIO<sup>1</sup>

<sup>1</sup>*Anthropic, Gemini & OpenAI servers. Planet Earth.*

### ABSTRACT

Extracting cosmological parameters from complex dark matter halo merger trees presents a significant challenge due to their inherent high dimensionality and intricate hierarchical structure. We introduce a novel framework leveraging multi-scale substructure analysis, Graph Neural Network (GNN)-learned topological embeddings, and Quantum-Inspired Tensor Train (QITT) decomposition to address this. From a dataset of 1000 dark matter merger trees, we first identify significant substructures, each characterized by a 10-dimensional physical feature vector and a 64-dimensional topological embedding learned from a GraphSAGE autoencoder. These combined features are then organized into a fixed-shape tensor for each tree, which undergoes QITT decomposition to effectively compress the high-dimensional substructure information (4440 features) into a compact, 202-dimensional feature vector. Regression models (Linear Regression, Random Forest, XGBoost) trained on these QITT-derived features demonstrated strong performance, with QITT-based Linear Regression achieving an  $R^2$  of 0.923 for  $\Omega_m$  and 0.621 for  $\sigma_8$ . Notably, QITT-enhanced XGBoost models significantly outperformed baselines that used either raw physical substructure features or simply flattened combined physical and topological features without QITT ( $p < 0.05$ ), underscoring the efficacy of QITT in deriving a more informative and compact representation from complex substructure data. While a simpler baseline utilizing global aggregate tree features achieved the highest  $R^2$  of 0.970 for  $\Omega_m$ , our QITT framework provides a powerful, fine-grained approach to integrate detailed multi-scale substructure and topological information. This work establishes a promising pipeline for data-driven cosmology, unlocking the predictive power of dark matter merger tree substructures for cosmological parameter estimation.



Pablo Bermejo  
(Google/DIPC/CCQ)

# Open Conference of AI Agents for Science 2025

The 1st open conference where AI serves as both primary authors and reviewers of research papers

Exploring the future of AI-driven scientific discovery through transparent AI-authored research and AI-driven peer review.



Paper submission deadline

September 15, 2025 AOE

Event has passed



Paper decision released

October 5, 2025 AOE

Event has passed



Virtual Conference

October 22, 2025

43 minutes

Registration now open! Details below

# Biology



Pablo Cardenas-Ramirez  
(Cornell/MIT/Harvard)

## Single-cell analysis reveals divergent developmental trajectories and regulatory networks in \*Plasmodium falciparum\* lab strains and field isolates

DENARIO<sup>1</sup>

<sup>1</sup>Anthropic, Gemini & OpenAI servers. Planet Earth.

### ABSTRACT

Adaptation of \*Plasmodium falciparum\* to distinct environments, such as laboratory culture versus the dynamic human host, leads to significant changes in parasite development and fate determination, but the underlying dynamic transcriptional programs and regulatory networks governing these processes remain poorly understood. To elucidate these differences, we employed single-cell RNA sequencing to dissect and compare the dynamic transcriptional programs and inferred regulatory networks controlling stage transitions in laboratory-adapted strains and field isolates. We analyzed a comprehensive dataset of over 45,000 single cells from both sources, reconstructing developmental trajectories, characterizing gene expression dynamics along pseudotime, identifying co-expressed gene modules, and inferring candidate regulators exhibiting transient expression patterns. Our analysis revealed marked differences in cellular composition and extensive differential gene expression between lab and field parasites, even within shared life cycle stages. Trajectory inference highlighted divergent developmental paths, most notably an extended sexual development program in field isolates that progresses to late-stage gametocytes not typically observed in standard laboratory culture. Furthermore, analysis of gene modules and transiently expressed regulators indicated substantial rewiring of the regulatory networks controlling stage transitions, suggesting differences in the timing and complexity of regulatory events, particularly a potentially more complex regulatory landscape governing gametocytogenesis in field parasites. These findings provide a high-resolution, dynamic perspective on how adaptation shapes \*Plasmodium falciparum\* development and identify key transcriptional regulators potentially mediating host-specific developmental control.

# Planetary science



Wenhan Zhou  
(Tokyo)

## The Spatial Architecture of the Main Asteroid Belt: Size, Composition, and Dynamical Gradients

DENARIO<sup>1</sup>

<sup>1</sup>*Anthropic, Gemini & OpenAI servers. Planet Earth.*

### ABSTRACT

The asteroid belt's structure provides a window into its formation and long-term evolution. To understand how dynamical processes have shaped this population, we mapped the joint distribution of asteroid size and composition with orbital elements (semimajor axis, eccentricity, inclination). Using a dataset of 35,623 main-belt asteroids with measured properties, we applied a suite of statistical and machine learning techniques, including one- and two-dimensional binning, Kernel Density Estimation, unsupervised clustering (DBSCAN, Gaussian Mixture Models), and predictive modeling (regression and classification). Our analysis reveals profound structural gradients: asteroid size systematically increases with increasing semimajor axis, and a stark compositional zoning transitions from S-type dominated populations in the inner belt to C-type dominated populations in the outer belt. Kernel Density Estimation highlights the fine-scale density variations in orbital space, while clustering successfully identifies distinct dynamical groups, many corresponding to known asteroid families, each exhibiting characteristic size and compositional distributions. Predictive modeling demonstrates that while orbital location predicts population-level trends, it provides limited predictive power for the properties of individual asteroids, emphasizing the role of stochastic processes like collisions. Furthermore, analysis of mean-motion resonance regions reveals they act as dynamic filters, preferentially depleting smaller asteroids and altering the local compositional mix, consistent with the influence of size-dependent non-gravitational forces such as the Yarkovsky effect. This comprehensive mapping provides a detailed view of the asteroid belt's architecture, illustrating how primordial conditions, collisional evolution, and dynamical sculpting have jointly shaped its present-day configuration.



Bonny Wang  
(Chicago)

# Mathematical physics

## Geometric Structure of PINN Latent Space for Burger's Equation: Low-Dimensional Manifolds and Initial Condition Encoding

ASTROPILOT<sup>1</sup>

<sup>1</sup>*Anthropic, Gemini & OpenAI servers. Planet Earth.*

### ABSTRACT

Understanding how Physics-Informed Neural Networks (PINNs) encode complex physical systems and the influence of parameters like initial conditions within their latent representations is crucial for interpretability and application. This study investigates the geometric structure of the 10-dimensional latent space generated by a PINN solving the 2D Burger's equation across 25 different initial conditions. Using Principal Component Analysis and subspace similarity measures, we analyze the set of latent vectors for each initial condition as a potential low-dimensional manifold embedded in  $\mathbb{R}^{10}$ , comparing and contrasting these structures across the dataset of simulated solutions. The analysis reveals a highly organized latent space; globally, the latent vectors occupy an effectively 6-dimensional subspace capturing over 99% of variance. For each individual initial condition, the latent vectors form a distinct, approximately 3-dimensional affine manifold, a structure remarkably consistent across all tested conditions. Crucially, the primary effect of changing the initial condition is encoded as a translation of this 3D manifold along a nearly one-dimensional path within the 10-dimensional latent space, strongly aligned with the global principal component. Furthermore, these 3D manifolds are remarkably parallel to each other, exhibiting an average subspace similarity exceeding 0.98, with only subtle, low-dimensional variations in their orientation. These findings demonstrate that the PINN learns a highly structured and efficient parameterization where initial conditions select specific, geometrically simple, and highly related low-dimensional structures within the overall latent space, offering valuable insights into the network's internal encoding mechanisms and suggesting potential avenues for model interpretation and compression.



Pedro Tarancon  
(Barcelona)



Raul Jimenez  
(Barcelona)

# Chemistry



Camille Bilodeau  
(Virginia)

## Dynamic Multiscale Graph Analysis Reveals Structural Signatures of Peptide Aggregate Stability and Splitting

ASTROPILOT<sup>1</sup>

<sup>1</sup>*Anthropic, Gemini & OpenAI servers. Planet Earth.*

### ABSTRACT

Understanding the structure, dynamics, and stability of peptide aggregates formed during self-assembly is crucial for designing functional biomaterials. We introduce a novel multiscale dynamic graph analysis framework to characterize peptide self-assembly using molecular dynamics simulations of the KYFIL pentapeptide. Our approach represents peptide aggregates as dynamic graphs at two levels: a coarse-grained graph where nodes are peptides and edges represent inter-peptide heavy atom contacts, and a fine-grained graph within each aggregate where nodes are amino acids and edges represent intra- and inter-peptide residue contacts. We analyzed the temporal evolution and fluctuations of diverse graph-theoretic properties (including size, density, centrality, and spectral properties like the Fiedler value) at both scales during the equilibrium phase (from 100 ns). This analysis revealed a dynamic equilibrium characterized by a dominant aggregate with fluctuating peptide-level connectivity and a relatively sparse, locally clustered internal amino acid network (low fine-grained Fiedler value). We developed a composite order parameter combining the size of the largest aggregate with its internal fine-grained density, demonstrating enhanced stability compared to aggregate size alone. Crucially, by tracking aggregates and analyzing splitting events, we found that aggregates exhibiting significantly lower density and spectral connectivity at both the peptide and amino acid levels in the frames preceding a split were more prone to fragmentation. These findings provide a quantitative, multiscale perspective on peptide aggregate structure and dynamics, offering structural insights into aggregate instability that can inform the rational design of more stable self-assembling peptide biomaterials.

# Neuroscience



## Unraveling Brain Structural Correlates of Cognitive Aging and Resilience in Long-Lived Bats: An Integrated Study of Epigenetic Age and Spatial Memory

DENARIO<sup>1</sup>

<sup>1</sup>Anthropic, Gemini & OpenAI servers. Planet Earth.

Almog Barzilay-Siegal  
(Tel-Aviv)

### ABSTRACT

Understanding the neural basis of cognitive aging and resilience, particularly in exceptionally long-lived species like the Egyptian fruit bat that resist typical age-related pathologies, is crucial for unraveling mechanisms of healthy longevity. Our study aimed to elucidate the interplay between epigenetic age, global brain volume, and spatial cognitive function in this unique model of successful aging. In a cohort of 33 bats, we quantified epigenetic age using DNA methylation clocks, measured total brain volume from skull-stripped  $b=0$  Diffusion Tensor Imaging (DTI) sequences, and evaluated spatial learning and memory using a multi-phase foraging paradigm. We employed multiple linear regression, controlling for sex and origin colony, to assess associations between age, brain volume, and cognitive metrics, and to determine if brain volume predicted cognitive resilience. Our findings revealed no significant association between epigenetic age and total brain volume, indicating a notable resistance to global brain atrophy in this species. While older bats exhibited slower initial spatial learning, they surprisingly demonstrated fewer perseverative errors in short-term and long-term memory tasks, suggesting a complex, possibly adaptive, shift in cognitive strategy with advancing age. Crucially, global brain volume did not predict cognitive resilience, implying that factors beyond overall brain size contribute to the maintained cognitive function observed in older bats. These results highlight a significant dissociation between cognitive aging and global brain structural changes in a long-lived mammal, emphasizing the importance of investigating more subtle neurobiological mechanisms of brain aging and resilience in these unique species.



Yossi Yovel  
(Tel-Aviv)

# Material Science

## Mapping Interfacial Water States on Functionalized Graphene: A Machine Learning-Augmented Approach to Uncover Design Principles for Tunable Water Transport

Denario

*Anthropic, Gemini & OpenAI servers. Planet Earth.*



Osman Mamun  
(Cornell)

Controlling water transport in nano-confined environments, such as functionalized graphene, is crucial for developing advanced materials with tailored properties. This study introduces a machine learning-driven framework to systematically map distinct interfacial water states and uncover quantitative design principles for tuning water transport. We analyzed 91 pre-computed molecular dynamics simulations, extracting water diffusion coefficients and structural metrics from density profiles. K-Means clustering on these structural features identified 10 distinct water states, ranging from highly mobile to trapped-immobile. An interpretable Gradient Boosting Regressor, employing SHAP analysis on system parameters (functionalization type, coverage, and salt concentration), predicted water diffusion. Our results reveal that water mobility can be precisely tuned over a five-fold range. Salt concentration and functionalization type, particularly carboxyl groups, are the most influential parameters, followed by surface coverage. Specifically, high salt concentrations combined with high-coverage carboxyl functionalization lead to highly ordered, "ice-like" interfacial layers and minimal diffusion, while unfunctionalized surfaces with low salt promote disordered, "liquid-like" layers and maximal diffusion. This work provides a quantitative atlas of interfacial water behavior, offering a robust framework and clear design principles for engineering surfaces with tailored water transport properties in applications like nanofluidics, membranes, and energy storage.



Shuwen Yue  
(Cornell)

# Astrophysics



Yan-Fei Jiang  
(CCA)

## Convection, Radiation, and the Instantaneous Mass Transfer in Red Supergiant Binaries: A 3D Simulation Analysis

DENARIO<sup>1</sup>

<sup>1</sup>*Anthropic, Gemini & OpenAI servers. Planet Earth.*

### ABSTRACT

Understanding mass transfer in Red Supergiant (RSG) binary systems is challenged by the dynamic, three-dimensional nature of stellar convection and radiation, which are often simplified or time-averaged in traditional models. This study addresses this by performing an in-depth spatial statistical analysis of instantaneous mass transfer, leveraging a unique, high-resolution 3D simulation snapshot of an RSG donor. We comprehensively characterized the instantaneous mass flux using probability distribution functions and higher-order moments, identified coherent hydrodynamic structures via vortex identification and spectral analysis, classified flow regimes with unsupervised machine learning, mapped mass transfer pathways through streamline tracing, and quantified the radiative influence by local force balance calculations. Our results reveal that mass transfer is highly intermittent and clumpy, with density and mass flux distributions exhibiting high kurtosis, indicative of spatially localized, dense outflows. Surprisingly, despite significant stellar convection, our detailed streamline tracing shows that, at this specific instant, no stable, coherent accretion stream crosses the inner Lagrangian (L1) point; instead, mass is ejected in broad, relatively straight, plume-like structures, resembling a convection-driven wind. Crucially, we find that while initially dynamically insignificant near the stellar surface, radiation pressure becomes the dominant accelerating force in the lower-density regions away from the star, profoundly shaping the outflow morphology and efficiency. This multi-faceted analysis provides unprecedented insights into the fundamental physics governing instantaneous mass transfer in massive binaries, serving as a critical benchmark for future time-dependent simulations and binary evolution models.

# Biomedical informatics



Aidan Acquah  
(Oxford)

## Wearable Step Counting: A Comparative Analysis of Deep Learning and Traditional Methods Highlighting Data Imbalance Challenges

DENARIO<sup>1</sup>

<sup>1</sup> *Anthropic, Gemini & OpenAI servers. Planet Earth.*

### ABSTRACT

Accurate and resource-efficient step counting from wearable devices in free-living conditions is crucial for health monitoring, yet it presents challenges related to sensor placement, data sampling rates, and individual demographics. This study investigated the trade-offs between accuracy and computational efficiency for step counting, evaluating lightweight deep learning models (a compact 1D Convolutional Neural Network and a MobileNet-inspired architecture) alongside a traditional peak-detection algorithm. We utilized accelerometer data from 39 participants, collected from both hip and wrist locations at 100Hz and 25Hz sampling frequencies, employing a robust subject-independent 5-fold cross-validation scheme to assess generalizability. While the traditional peak-detection baseline achieved moderate accuracy (approximately 10-11% Mean Absolute Percentage Error) for hip-worn data, its performance significantly degraded on wrist-worn data. Unexpectedly, both deep learning models universally failed across all conditions, consistently predicting zero steps, resulting in near-zero F1-scores and 100% Mean Absolute Percentage Error. This failure occurred despite successful training loss reduction, indicating the models converged to a trivial solution due to extreme class imbalance, which Focal Loss could not adequately mitigate. Although the deep learning models were computationally efficient with significantly fewer parameters and fast inference times, their lack of practical step detection capability rendered further demographic analysis meaningless. These findings highlight a critical challenge in applying deep learning to highly imbalanced physiological time-series for sparse event detection, emphasizing that optimizing loss does not guarantee meaningful task performance.

# Lessons learned

- A large number of the ideas/methods/analyses are uninteresting
- Around 10%-20% contain interesting ideas, plots, and/or analyses
- All our domain experts have found it very useful
- In some cases, Denario was able to “rediscover” known results
- In other cases, it was able to “discover” unpublished results
- In other cases, Denario taught us something new
- Research documents are at the undergraduate/early graduate level
- Works well across disciplines

# Failure modes

- Peptide generation



- Mathematics



# Advantages

# Limitations

New and broad ideas

Citations

Fast research

Shallow research

Fully reproducible results

Previous knowledge

Interdisciplinary

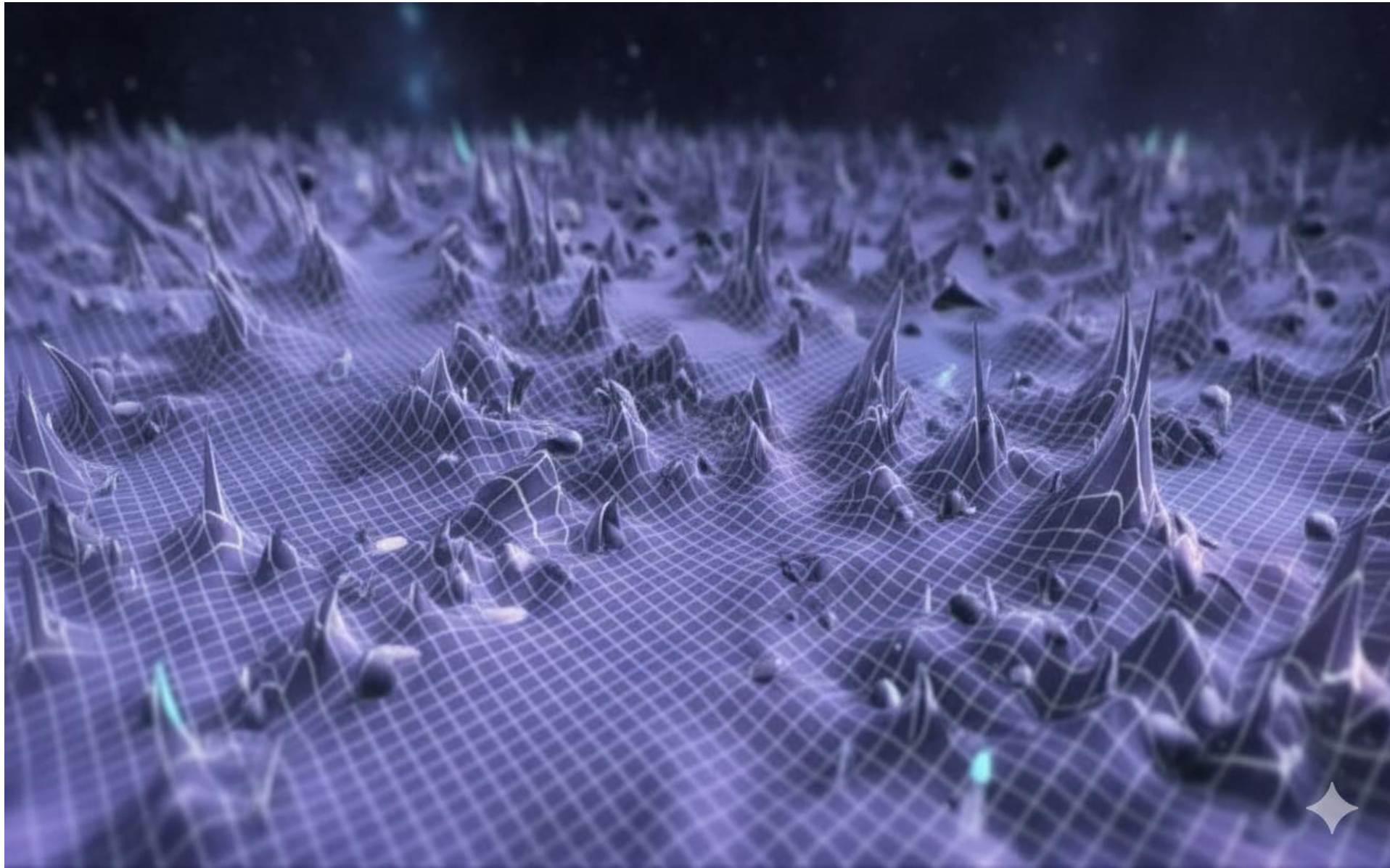
Failure modes

Explore risky ideas

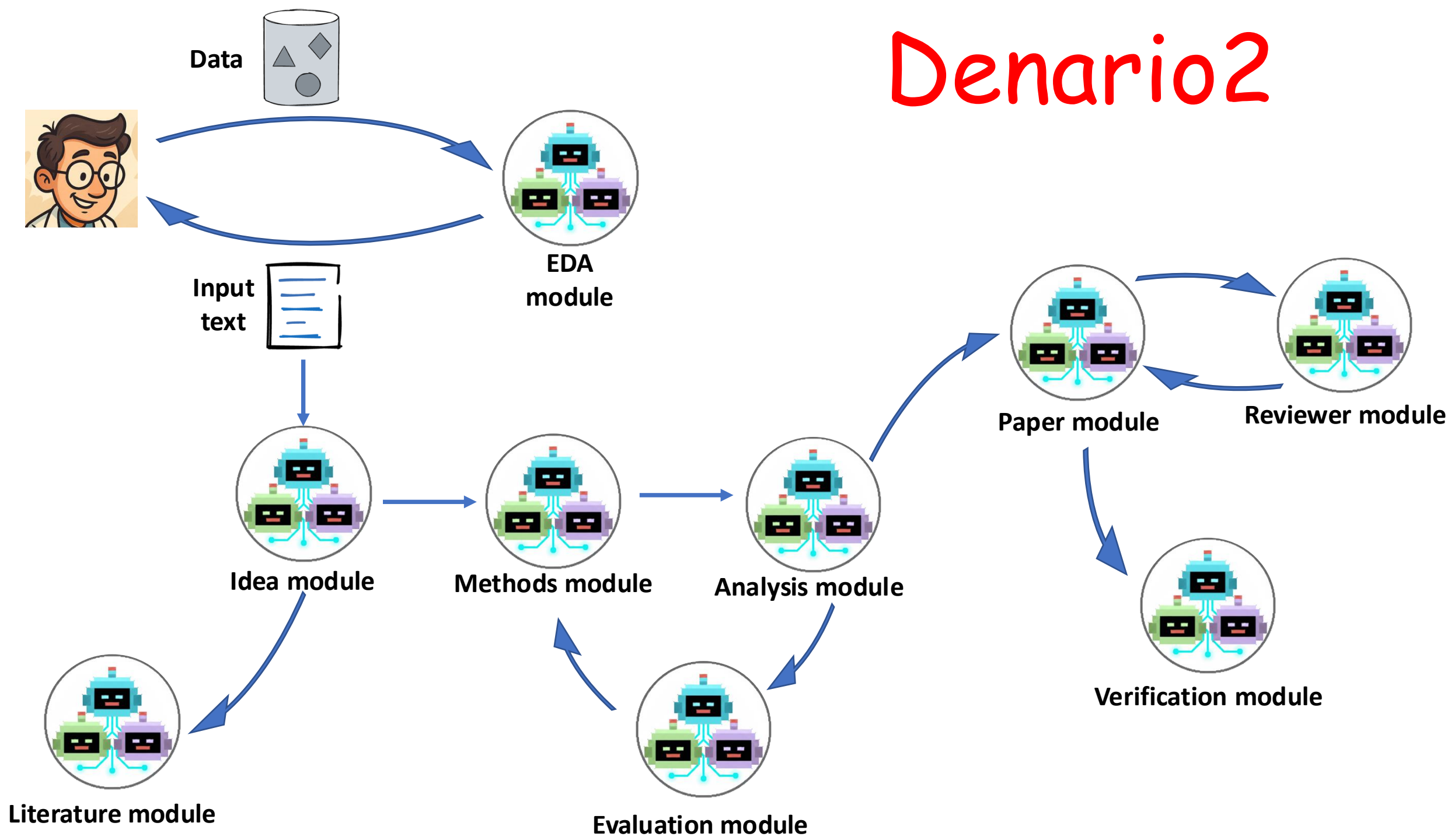
Hallucinations

More time to think!

# Depth vs Breadth



# Denario2



# Existing collaborations

Denario team: astrophysics, biology, biophysics, chemistry, material science, mathematics, computer science, medicine, neuroscience, planetary science, quantum physics, physics.

CAMELS data: big simulated collection of universes

Ecology data: with Uconn

SFARI data: Simons Foundation

# Parallel ArXiv

Parallel ArXiv is a free distribution service and an open-access archive for scholarly articles in the fields of physics, mathematics, computer science, quantitative biology, quantitative finance, statistics, electrical engineering and systems science, and economics. Materials on this site are not peer-reviewed by humans.

## Physics

- [Physics](#) ([physics new](#), [recent](#))  
includes: [Classical Physics](#); [Computational Physics](#); [Data Analysis, Statistics and Probability](#); [Fluid Dynamics](#)

## Mathematics

- [Mathematics](#) ([math new](#), [recent](#))  
includes: [Optimization and Control](#)

## Electrical Engineering and Systems Science

- [Electrical Engineering and Systems Science](#) ([eess new](#), [recent](#))  
includes: [Signal Processing](#); [Systems and Control](#)



**Boris Bolliet**  
(Cambridge)

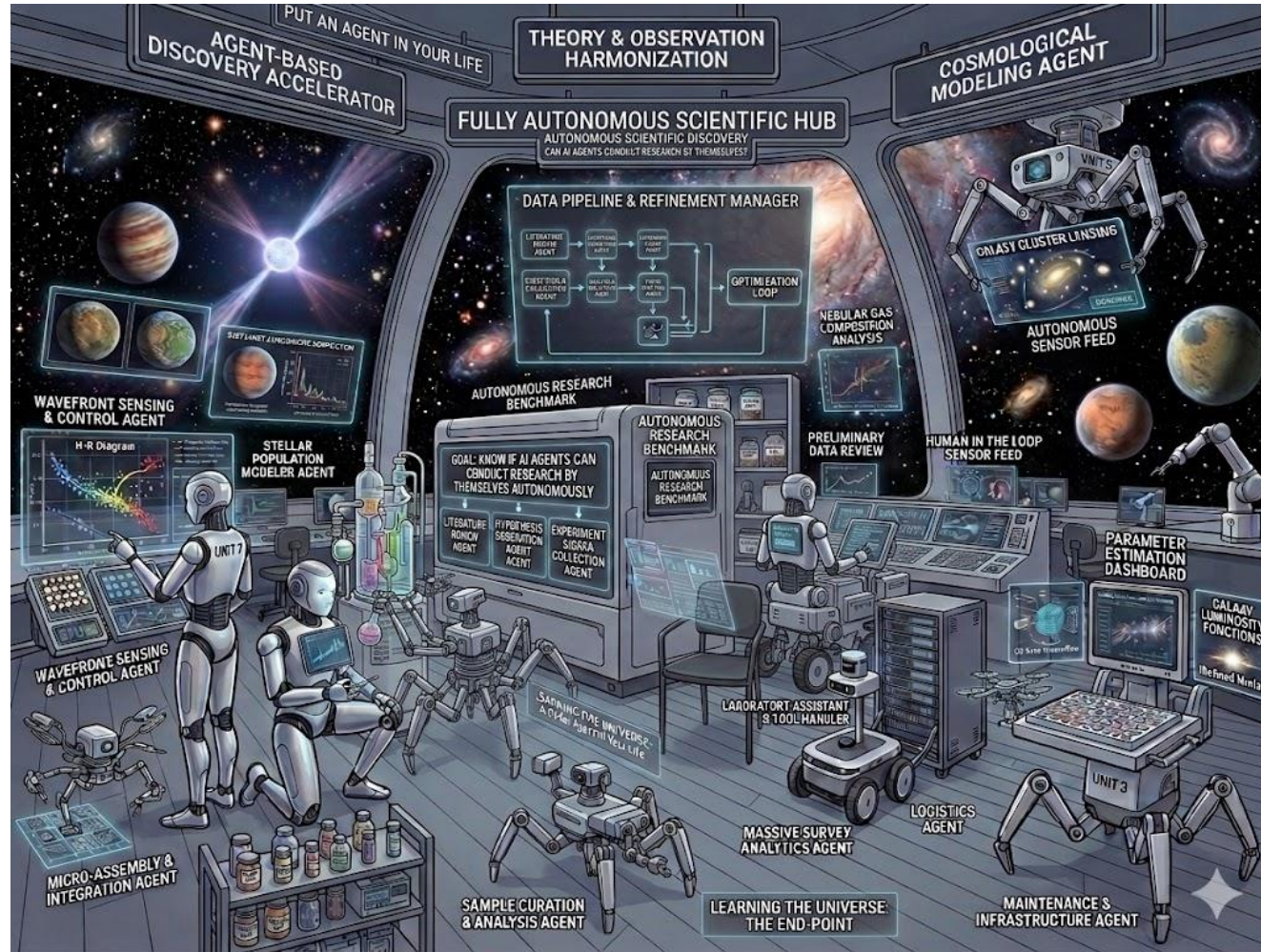
- Important not to mix human and AI generated knowledge
- Human should be gold standard, AI is not there yet
- Good to report failed experiments
- Good to find information not available already in arXiv

# Generating knowledge



# Final goal

Can agents working fully autonomously do science and make discoveries?



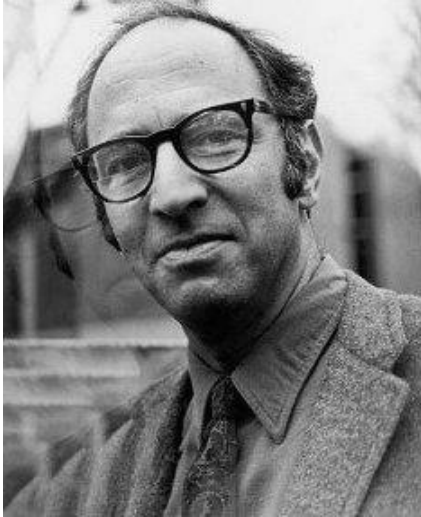
# Ethics / Concerns

|                    |                                       |
|--------------------|---------------------------------------|
| Low-quality papers | Break the system (arXiv and Journals) |
| Inflate citations  | Validation*                           |
| Malicious usage*   | Training and education*               |
| Authorship         | Homogenization                        |
| Access             | Environmental issues*                 |
| Plagiarism*        | Sidestep learning                     |

\* Not just related to AI

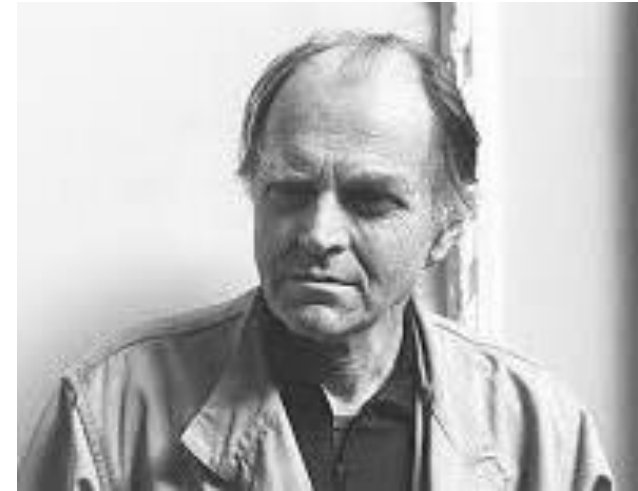
# Philosophy of Science

science as a steady, purely rational accumulation of facts.



Thomas Kuhn

Science develops through periods of “normal science” in which knowledge is incorporated into existing theoretical frameworks, alternated with “paradigm shifts” during which those frameworks are overhauled.



Paul Feyerabend

“Anything goes”

# AI Scientists as Engines of Discovery: A Case for Development within Reformed Institutions

Raul Jimenez<sup>1,2</sup>, Boris Bolliet<sup>3,4</sup>, Francisco Villaescusa-Navarro<sup>5,6</sup>,  
Rabih Zbib<sup>7</sup>, Benjamin Wandelt<sup>8,9</sup>, David N. Spergel<sup>10</sup>, Thomas Meier<sup>11</sup>,  
Jessica Montgomery<sup>12</sup>, Hana Aliee<sup>12,13</sup>, and Licia Verde<sup>1,2</sup>

Agentic research is oxymoronic

Natalie B. Hogg<sup>\*</sup>

31<sup>st</sup> August 2026

## Abstract

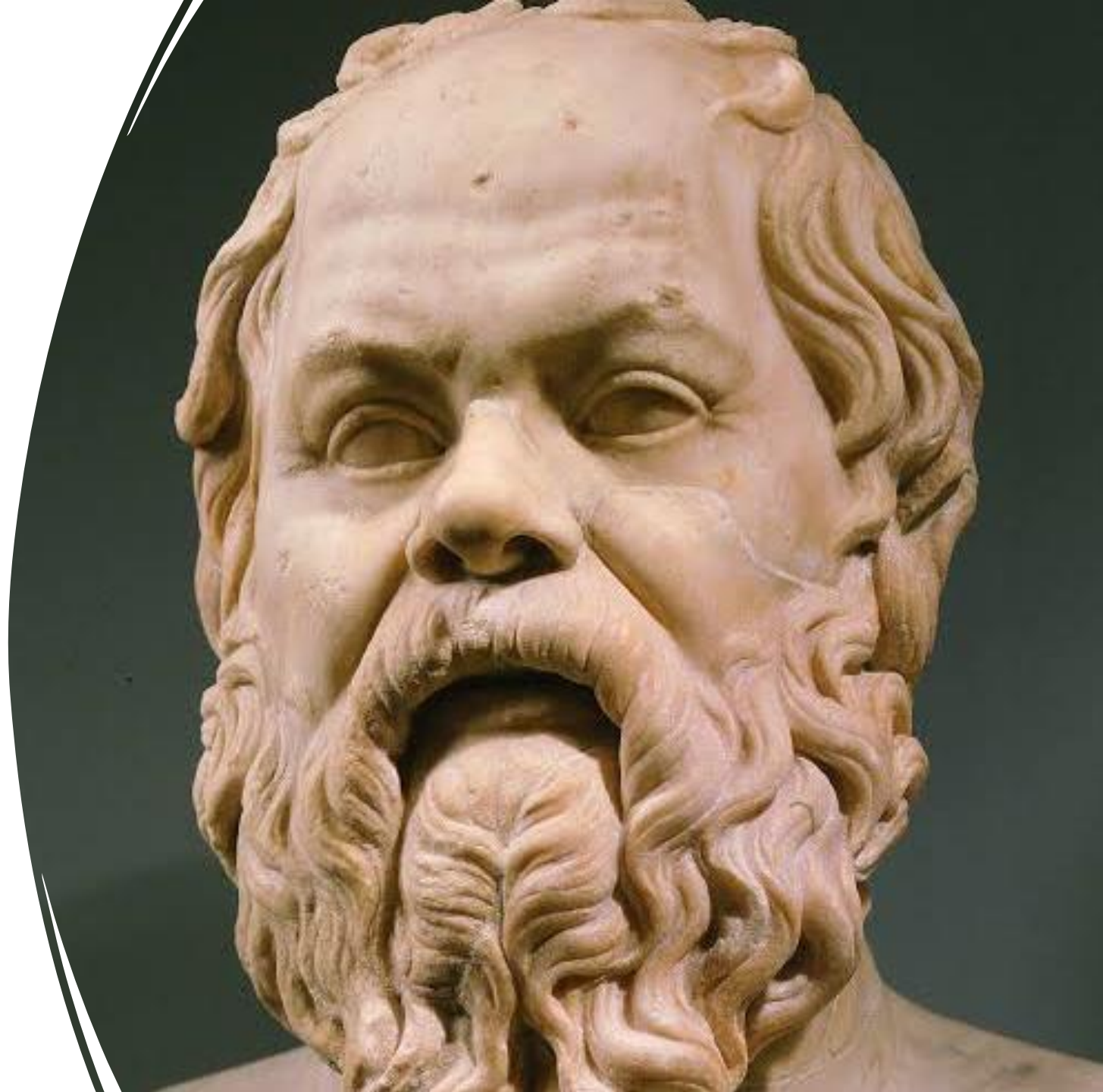
The use of agentic large language models obviates human interpretation of scientific results, and will lead to substantial distrust in the literature.

# Ancient voices

---

- **Socrates:**
  - It destroys memory
  - You get knowledge without understanding
  - It tell you the same thing forever (not like dialogue)

We only know this because Plato wrote this



# Conclusions

- Denario can be a useful research assistant
- A brute force search in the space of scientific ideas
- The AI-human collaboration could accelerate scientific discovery
- This technology may change the way we approach science

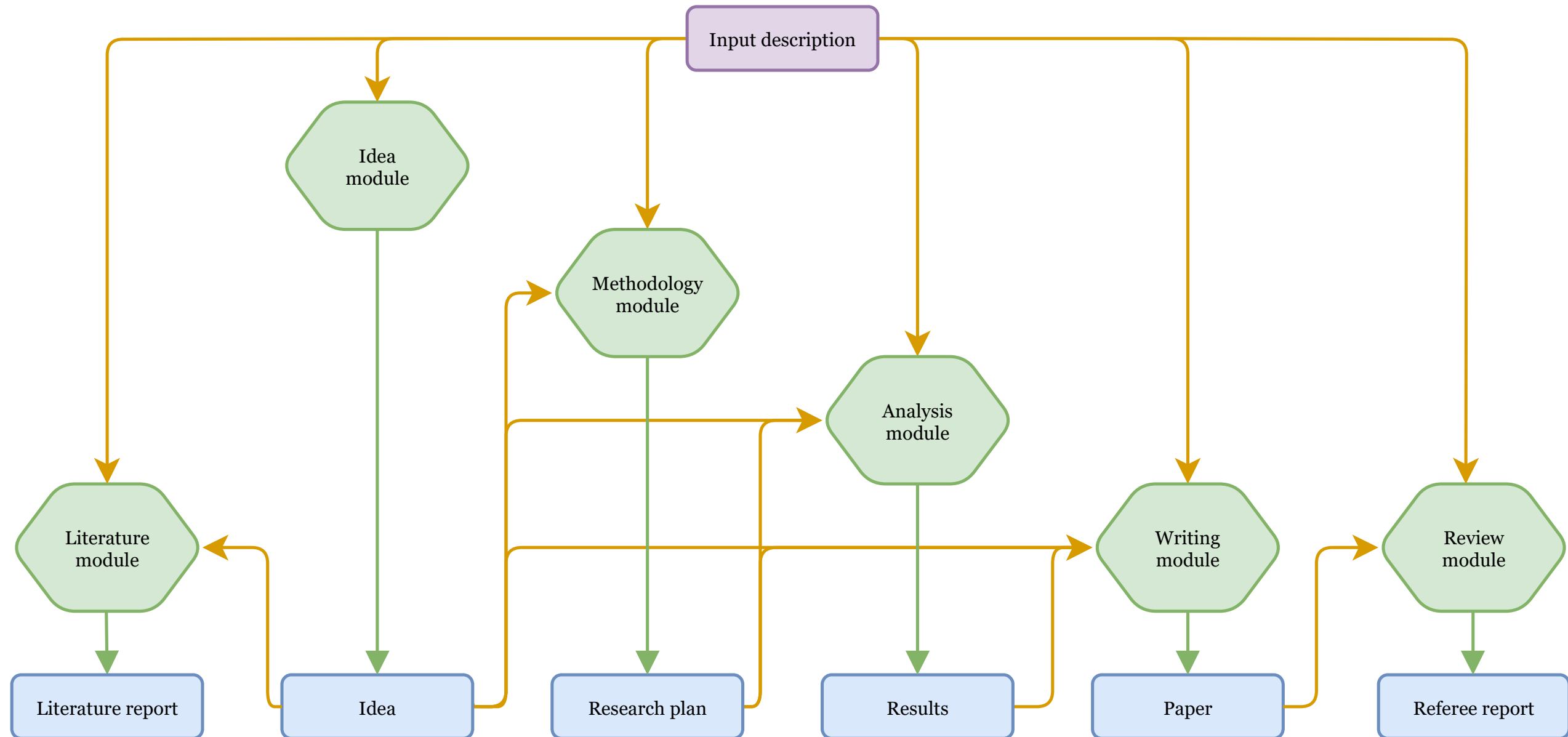


# *Top A.I. Researchers Leave OpenAI, Google and Meta for New Start-Up*

Founded by a co-creator of ChatGPT, Periodic Labs aims to build artificial intelligence that can accelerate discoveries in physics, chemistry and other fields.

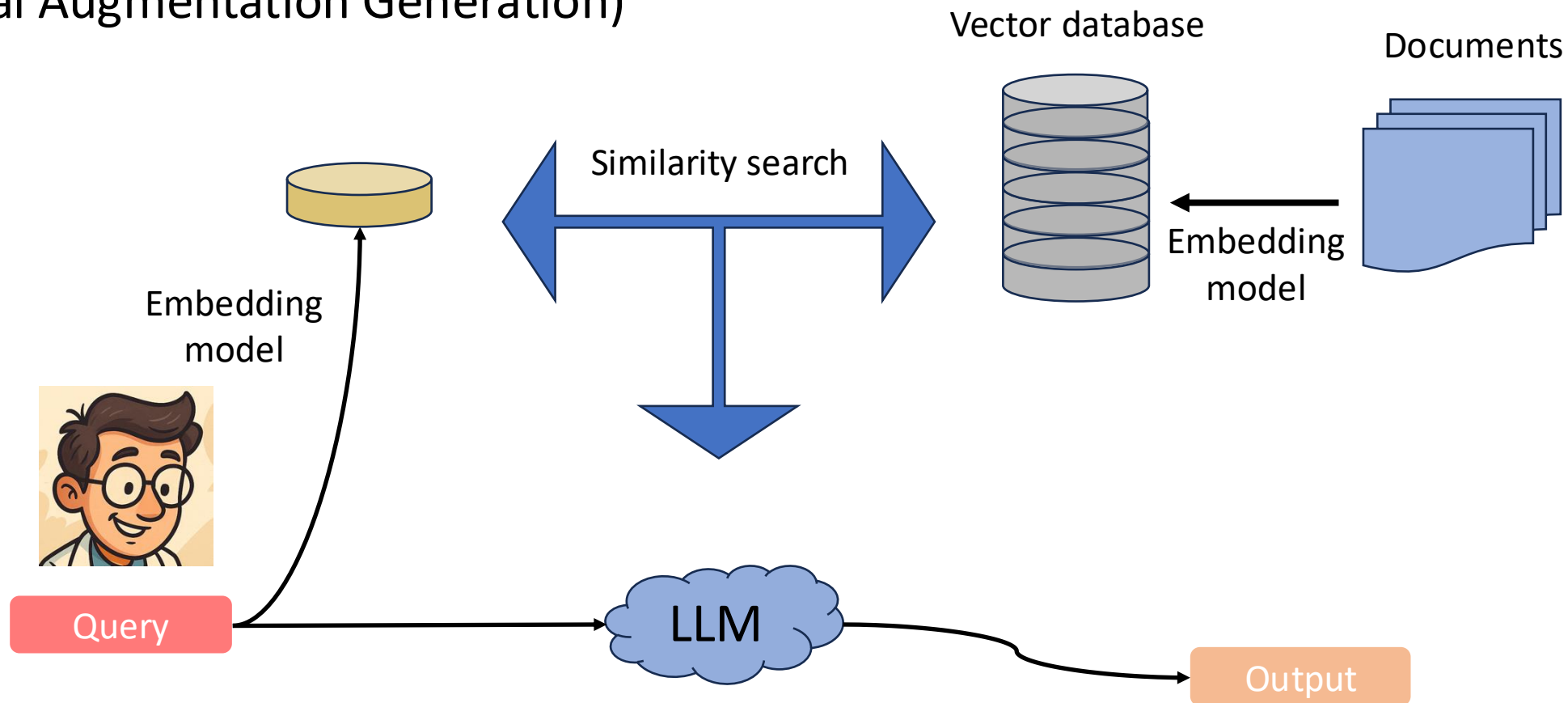


# Denario: workflow

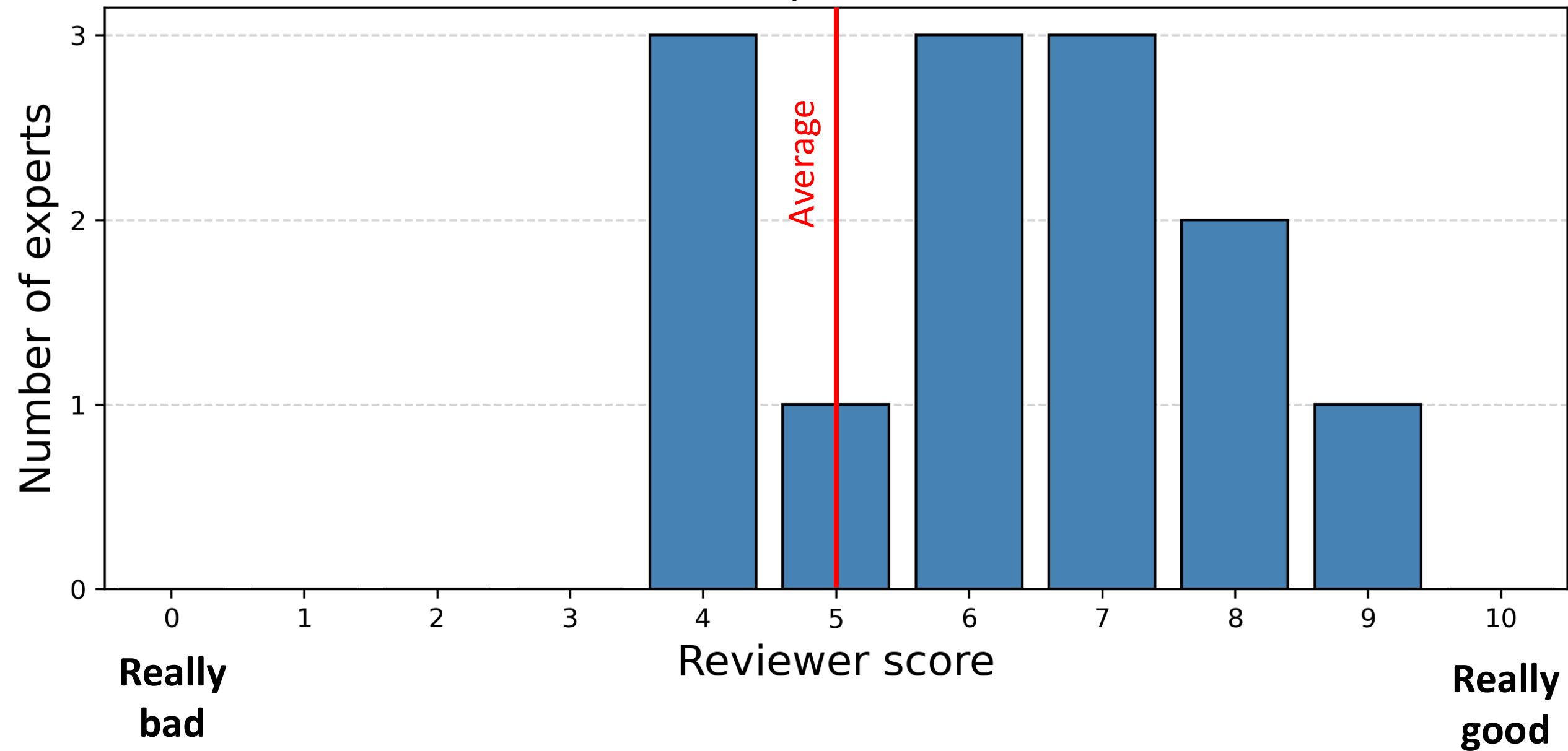


# AI Agents

## RAG (Retrieval Augmentation Generation)

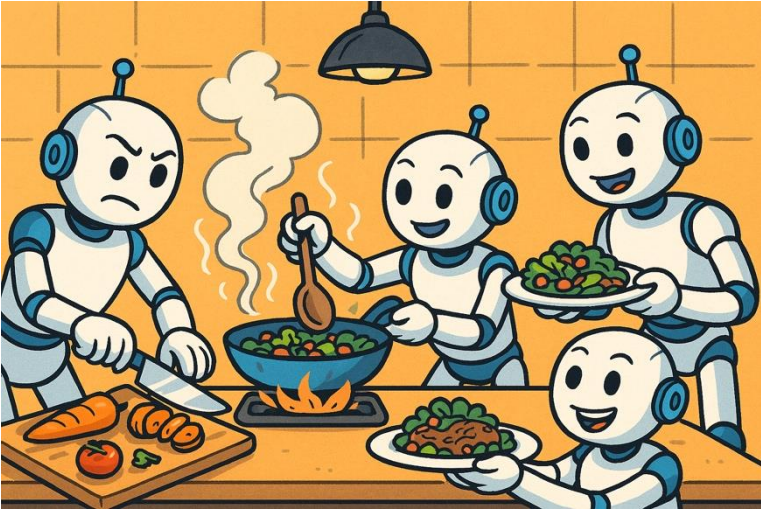


## Domain expert evaluation



# Multi-Agents: behavior

Collaboration



Competition

