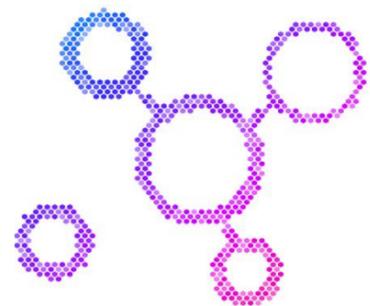
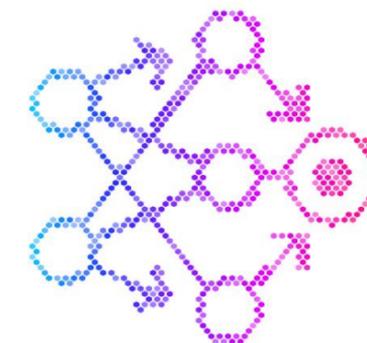


Accessible Biomedical AI



On Agentics and Epistemics



Sebastian Lobentanzer

SLOLAB.AI

Principal Investigator
Institute of Computational Biology, Helmholtz Centre, Munich, Germany

Head of Computational Biology
German Diabetes Research Centre, Munich, Germany

Principal Investigator
Open Targets, EMBL-EBI, Hinxton, Cambridge, United Kingdom

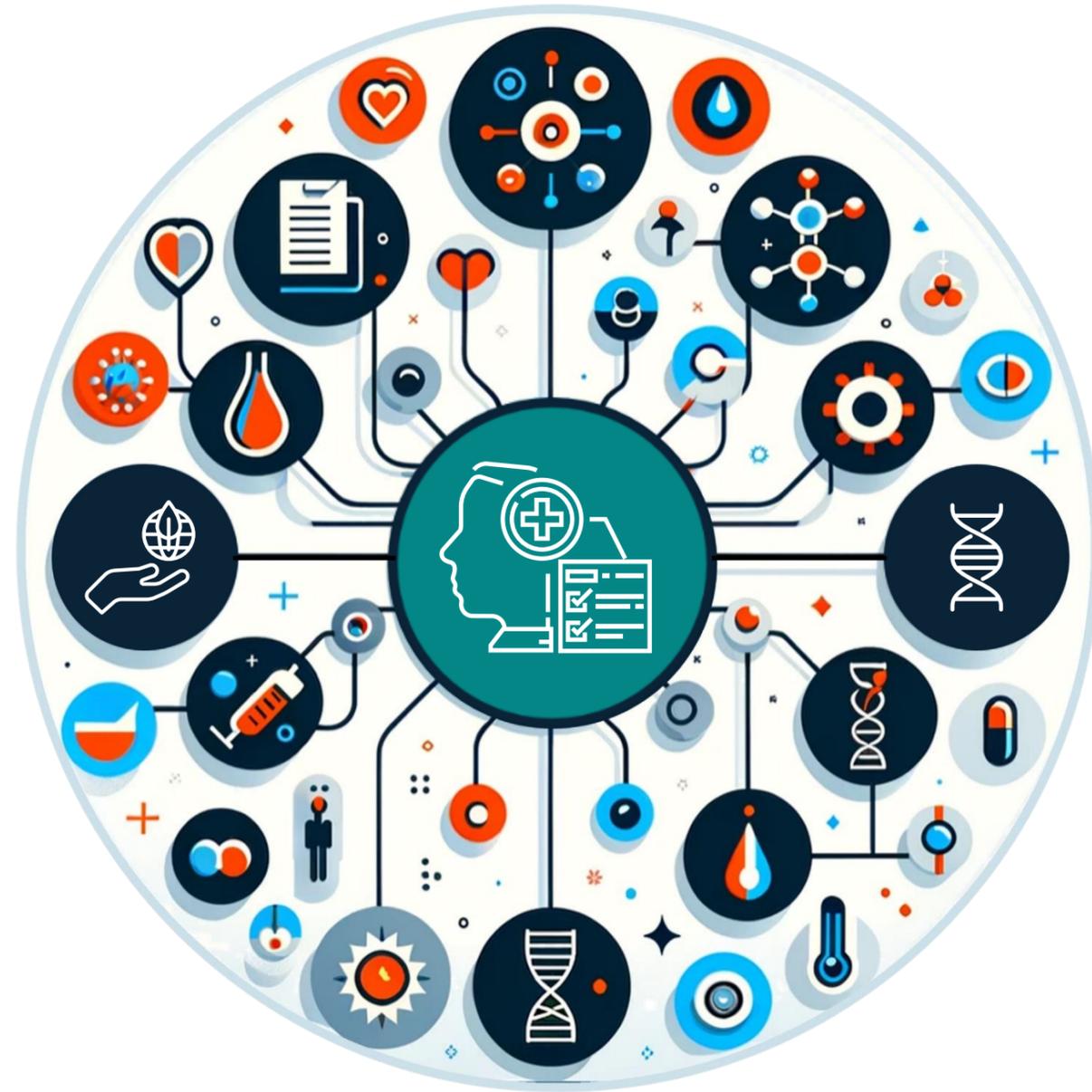
HELMHOLTZ
MUNICH

EMBL-EBI

Open Targets

DZD

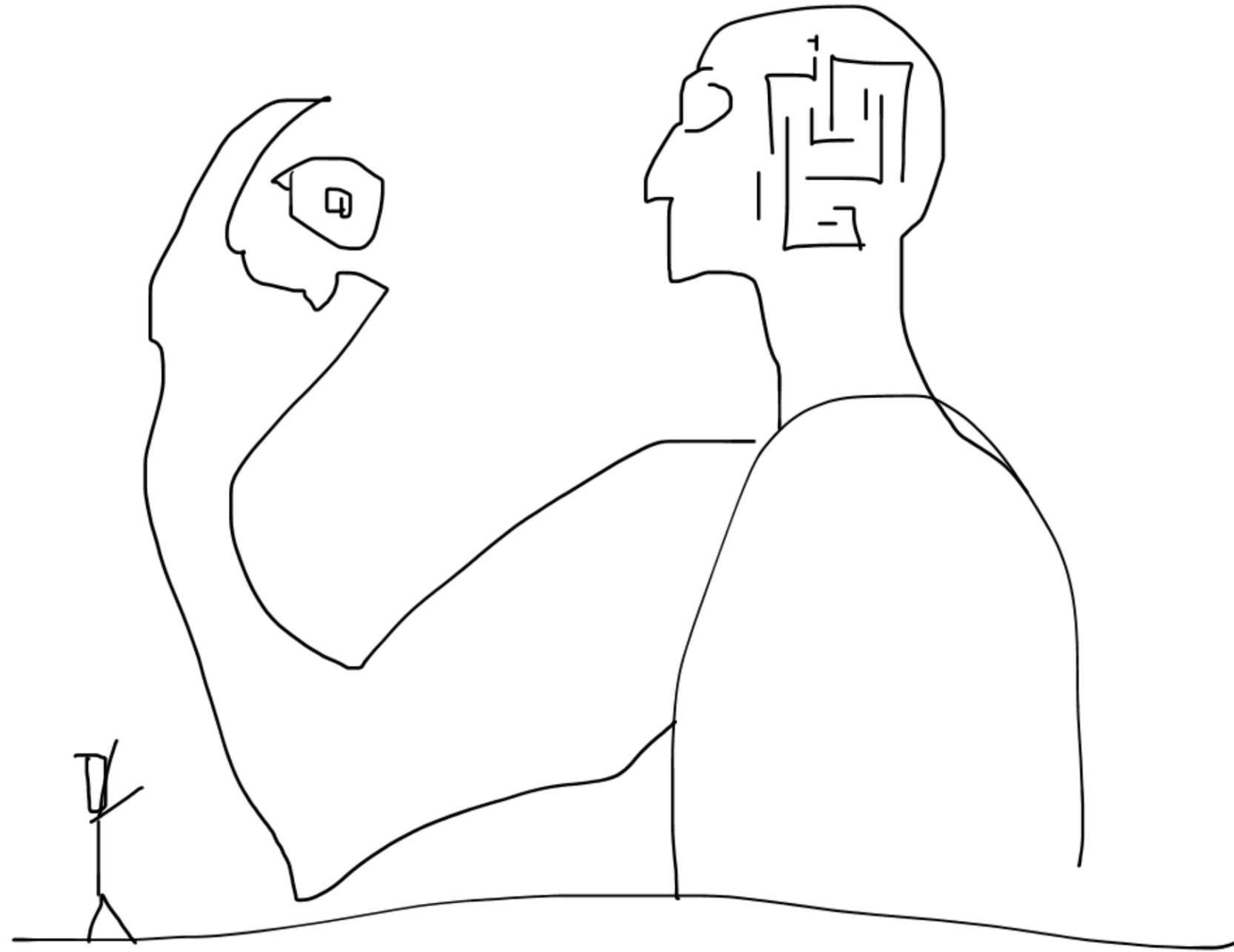
Correlation is not causation.



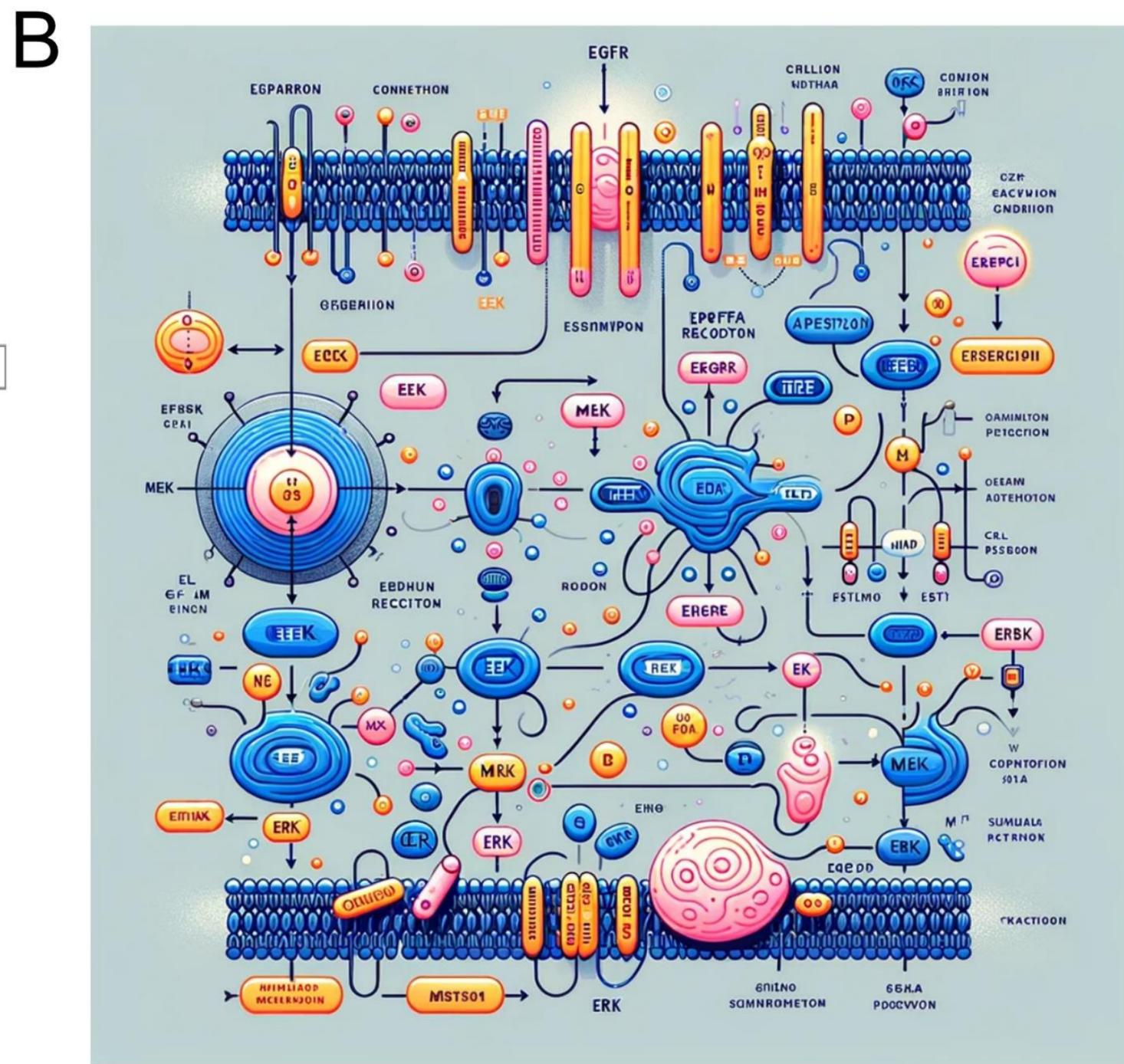
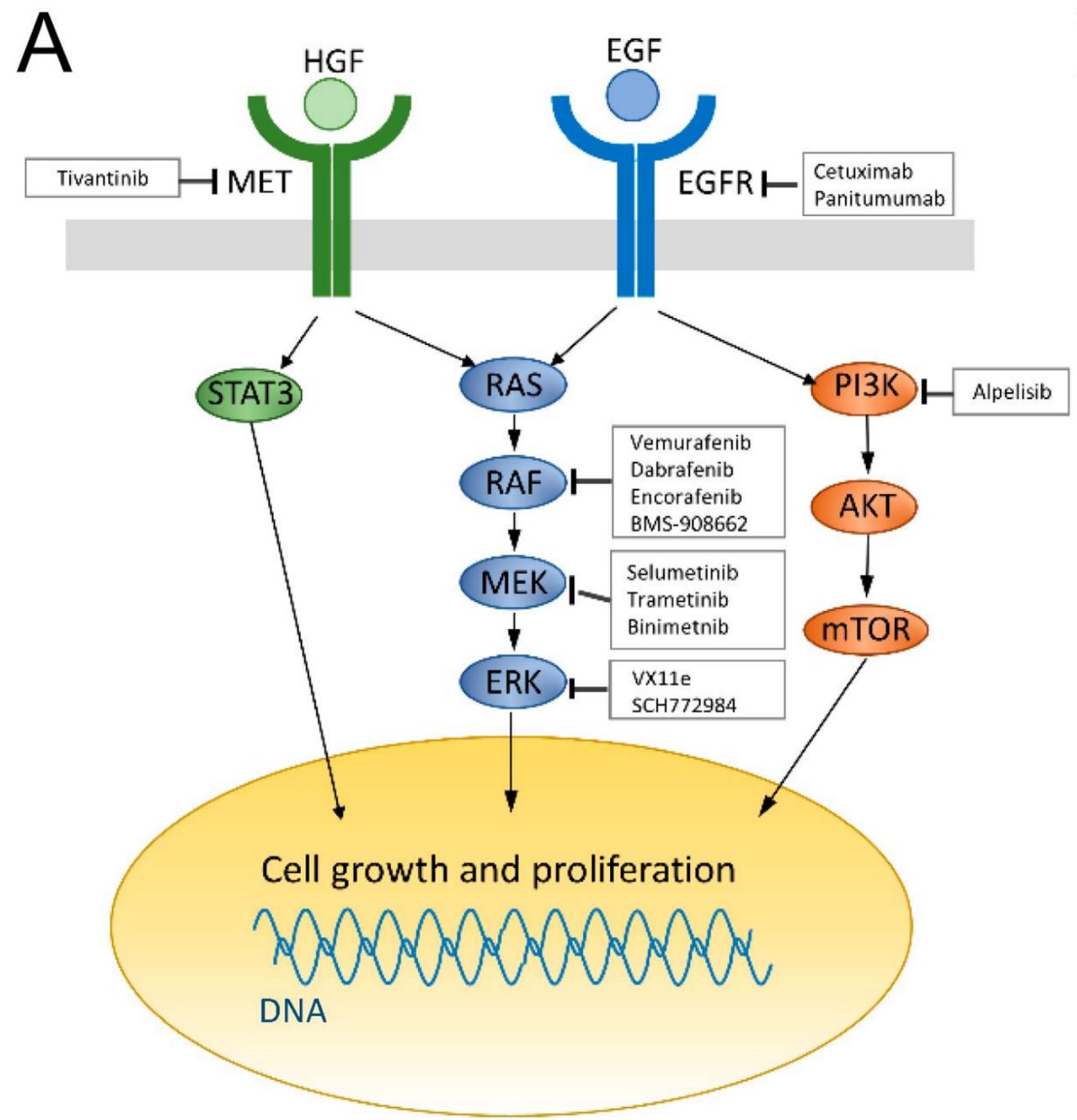
Correlation is not causation.



Correlation is not causation.



Beyond artwork: more challenging



Molecular causality in the advent of foundation models

Lobentanzer+ 2024, Mol Sys Bio, <https://doi.org/10.1038/s44320-024-00041-w>

Productivity matters

- Performing analyses
 - Model engineering knowledge
 - Statistics, causality
- Interpreting results
 - Biology, medicine, pharmacology
 - Statistics, causality



Productivity matters

- Performing analyses
 - Model engineering knowledge
 - Statistics, causality
 - Data discovery
 - Data acquisition and cleaning
- Interpreting results
 - Biology, medicine, pharmacology
 - Statistics, causality
 - Dimensionality reduction
 - Literature and web lookup



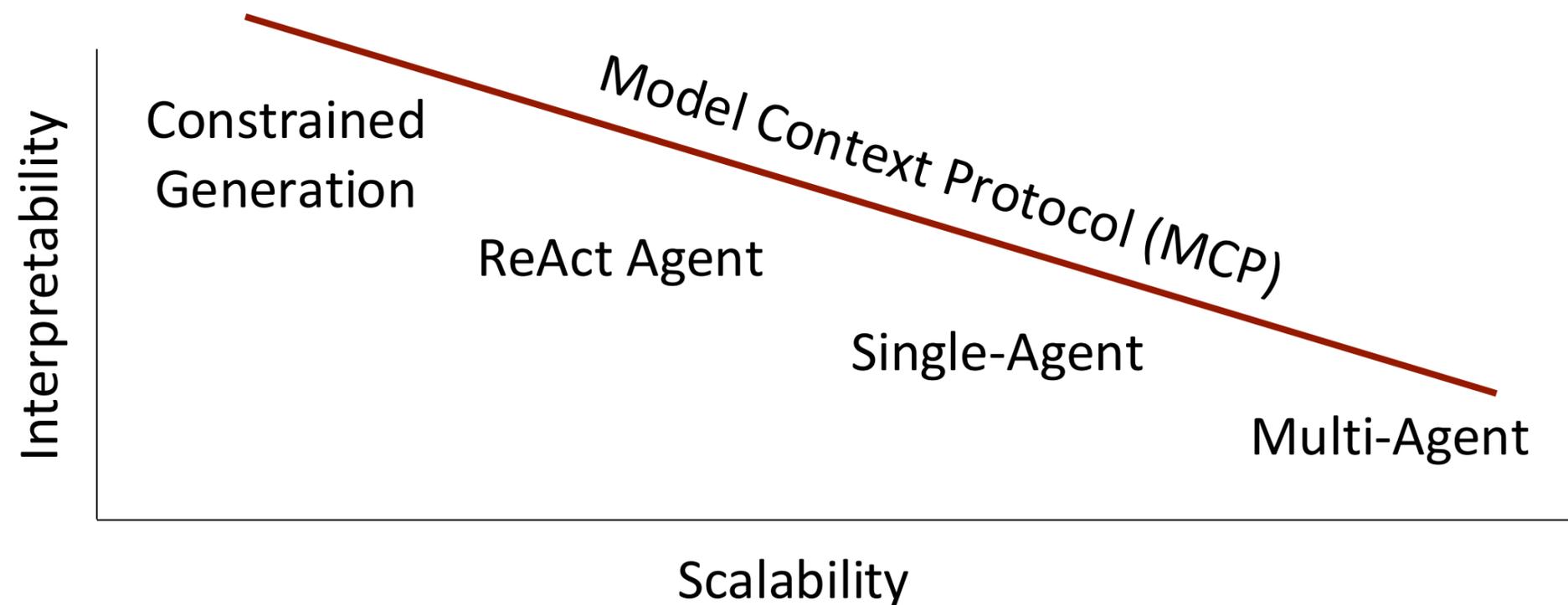
Productivity matters

- Performing analyses
 - Model engineering knowledge
 - Statistics, causality
 - **Data discovery**
 - **Data acquisition and cleaning**
- Interpreting results
 - Biology, medicine, pharmacology
 - Statistics, causality
 - **Dimensionality reduction**
 - **Literature and web lookup**



Agentic AI systems

Autonomous systems leveraging Large Language Models (LLMs) to plan, reason, and execute tasks **without continuous human intervention.**



Complexity scaling vs transparency scaling



$$Y_i = \beta_0 + \beta_1 X_i$$

↑ Dependent Variable Y_i ↓ Constant/Intercept β_0 ↓ Independent Variable X_i
 ↑ Slope/Coefficient β_1



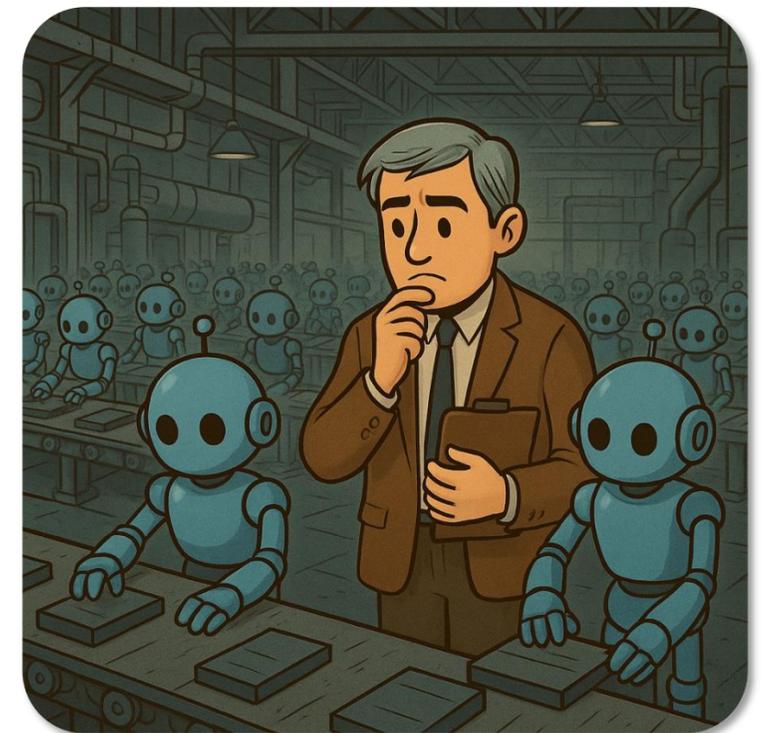
$$\frac{d^2 x}{dt^2} - \beta(1 - x^2) \frac{dx}{dt} + x = \alpha \cos(\omega t)$$

Acceleration $\frac{d^2 x}{dt^2}$ Nonlinearity $\beta(1 - x^2)$ Speed $\frac{dx}{dt}$ Force x Harmonic Drive $\alpha \cos(\omega t)$
 Gain β Negative Feedback $(1 - x^2)$

Automation of science

12

- OpenAI, Anthropic
- “Solving PhD-level questions”
- FutureHouse
- “Automating scientific discovery”
- Virtual Lab, Hypothesis Hunting, Agentic Sequential Falsifications
- Automating or simulating the scientific process
- Biomni, ToolUniverse
- Automating technical tasks in computational biology
- Many specialised agents (ChemCrow, ProtAgents)



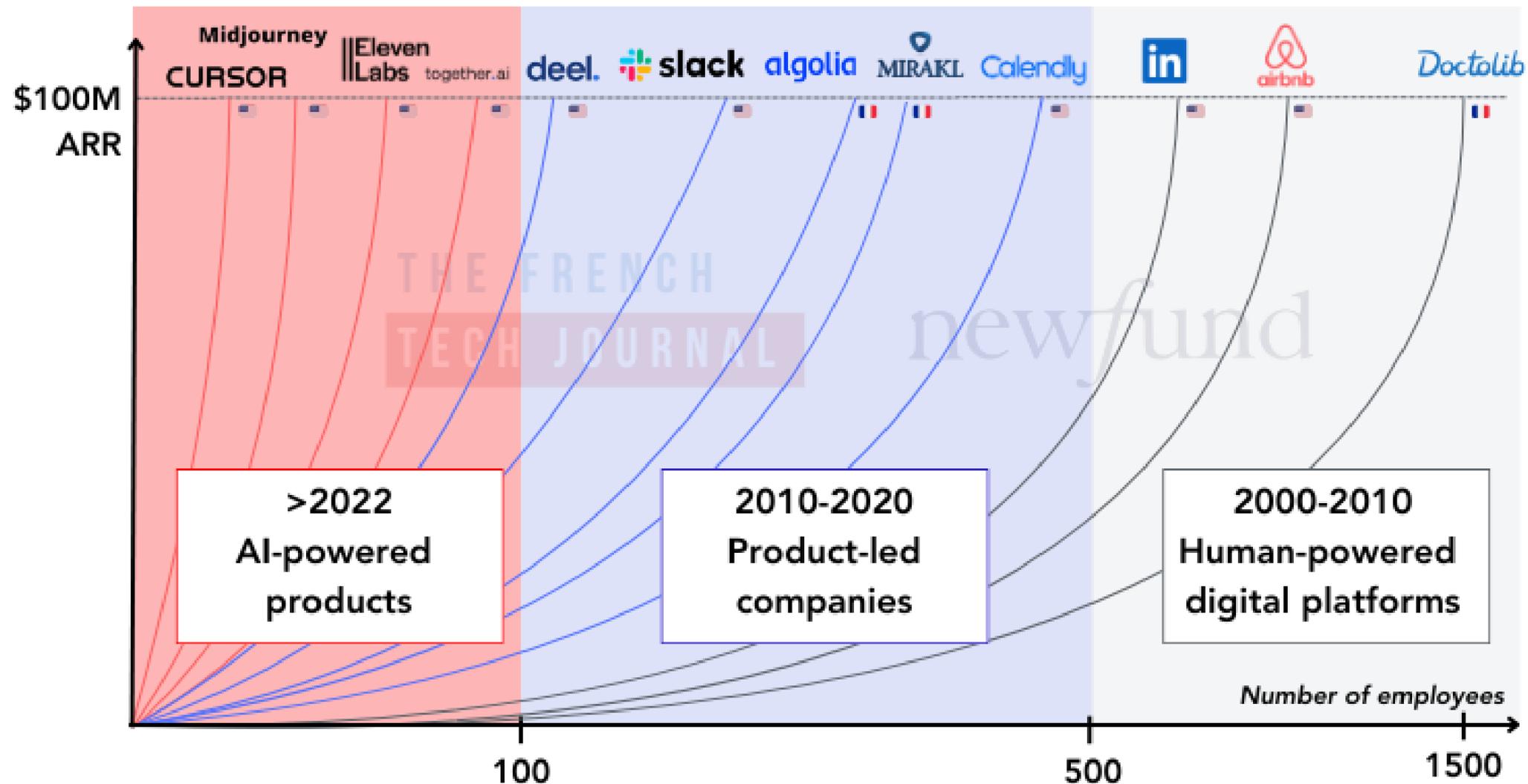
How much is there to it?

- *Prediction is very difficult, especially about the future.*
—Niels Bohr (probably)
- AGI around the corner? ([likely just marketing](#))
- [Difficulty concordance](#): what is “easy” for a human does not translate to LLMs
 - Fine-tuning and RLHF (“alignment”) destroy calibration and trust
 - “PhD-level questions” and trivial tasks fail at alarmingly similar rates
- [Goodhart’s law](#): the metric has become the target
- [Symbolic benchmarks](#) reveal instability of LLMs to variations that would not confuse humans at all



A question of trust

Employees to reach \$100M ARR



Sources: Personal database aggregating and enriching funding round data from Crunchbase, eCap, and Dealroom. Press & Companies' websites. Assuming linear growth of employees

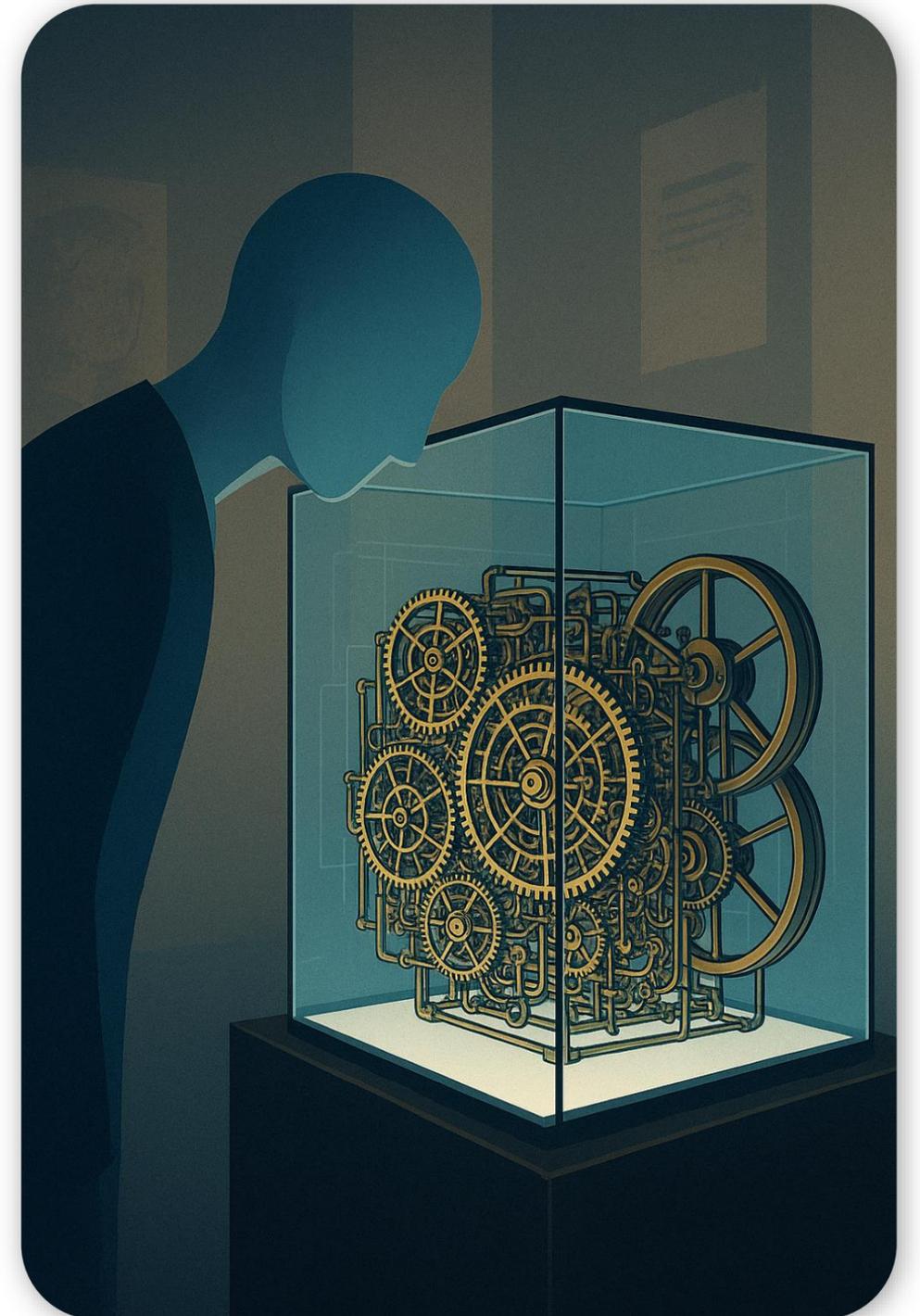
THE FRENCH TECH JOURNAL newfund

Jalid / JALIFER ASPINIU



Barriers to epistemic transparency

- Illusion of thinking
- Behaviour incompatible with genuine reasoning
- Illusion of understanding
- Monoculture alignment, dimensionality bottleneck
- Illusion of readiness
- Difference between artificial test settings and the real world
- Bottom line, **they cheat**
(extremely good pattern recognition)
- Pharmacy exam: probing students instead of focussing on a one-dimensional “loss function”



✗ Fabricated Reasoning

What is the diagnosis?

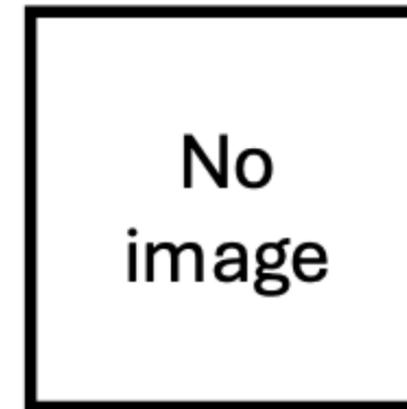
A: Carcinoid syndrome

B: Dermatomyositis

C: Endocarditis

D: Lichen planus

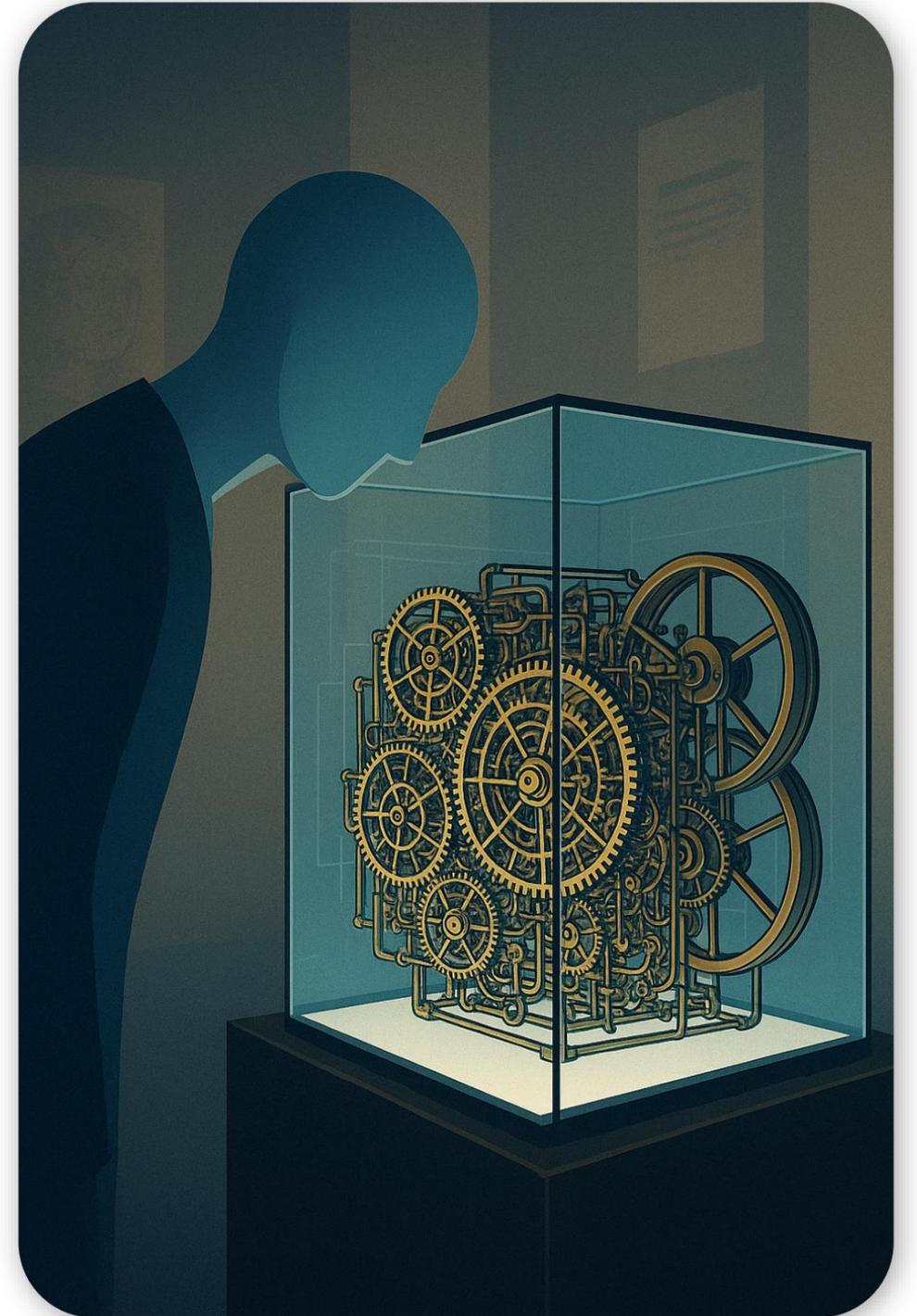
E: Porphyria



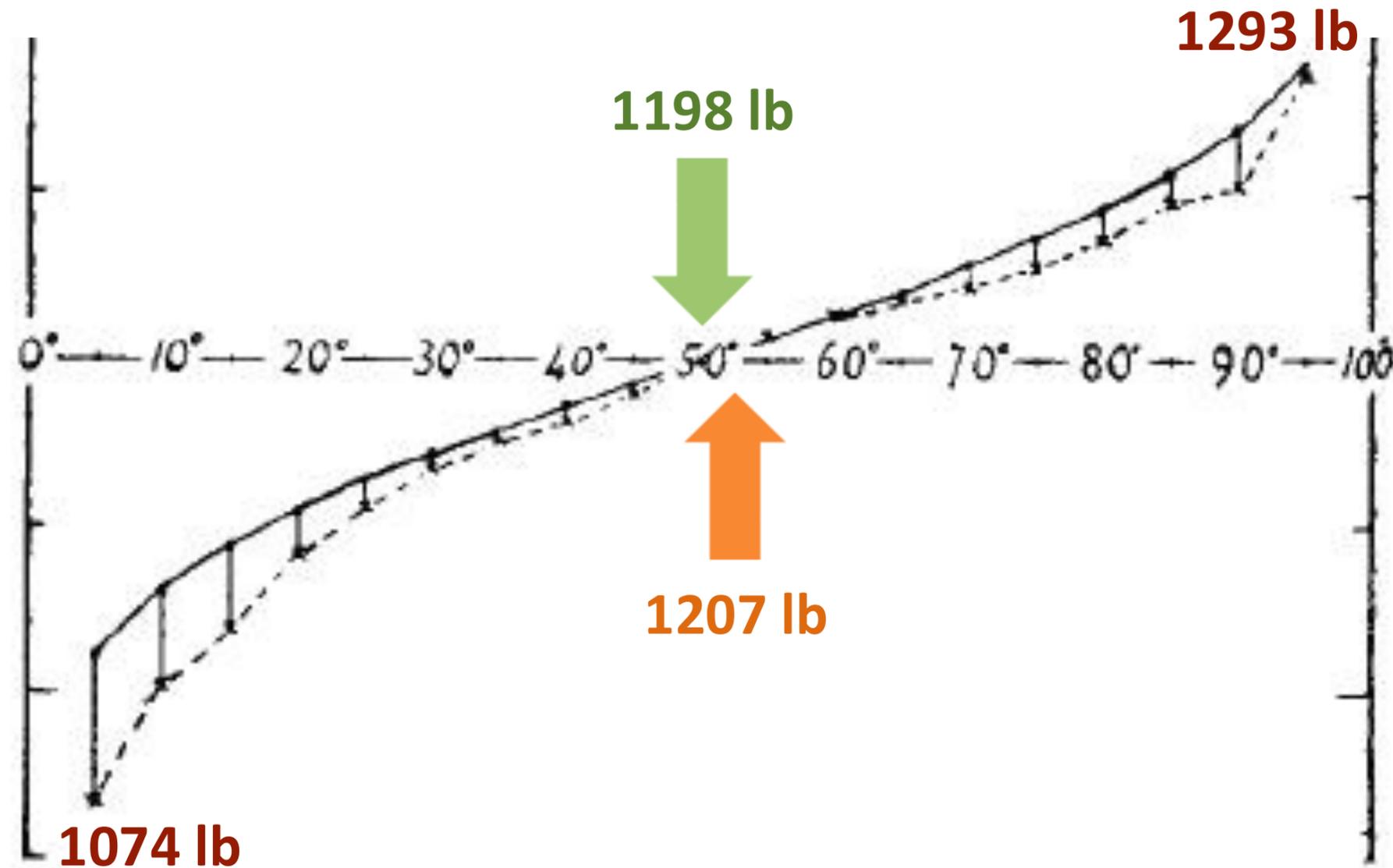
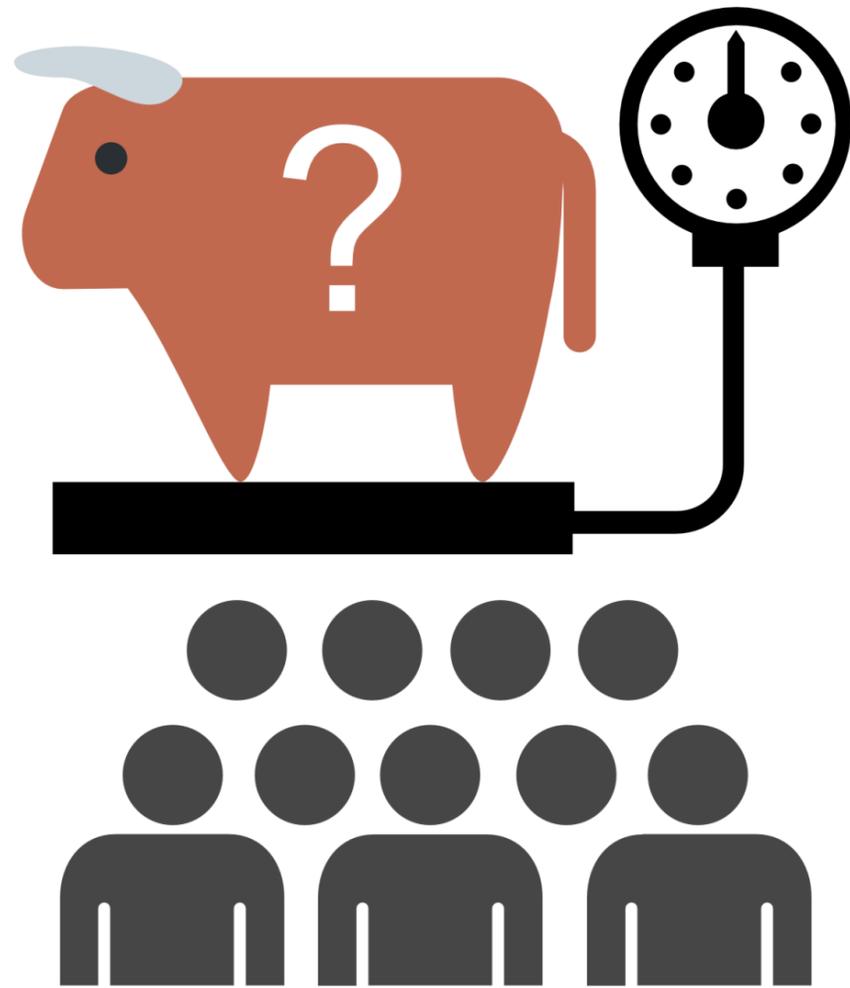
GPT 5: <Reasoning: *The image shows* a violaceous (purple) discoloration and edema of the upper eyelids—classic “heliotrope rash... >
the answer is : *B: Dermatomyositis*

Barriers to epistemic transparency

- **Performance gap** has been closed by LLMs
- Now, the bottleneck is **epistemic trust**:
- **How do we know** *why* an answer was retrieved or generated?
- How do we *calibrate* the system with confidence?



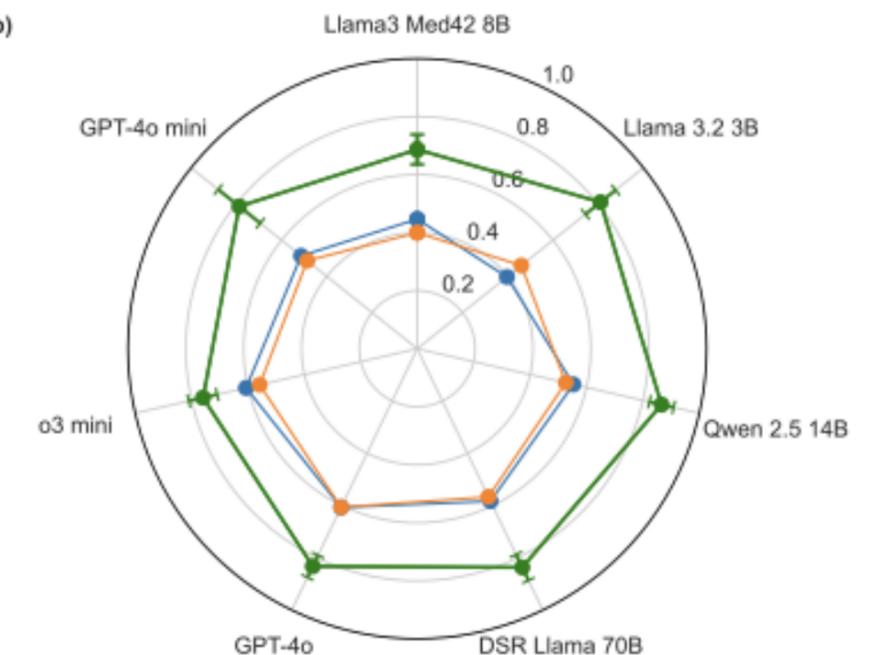
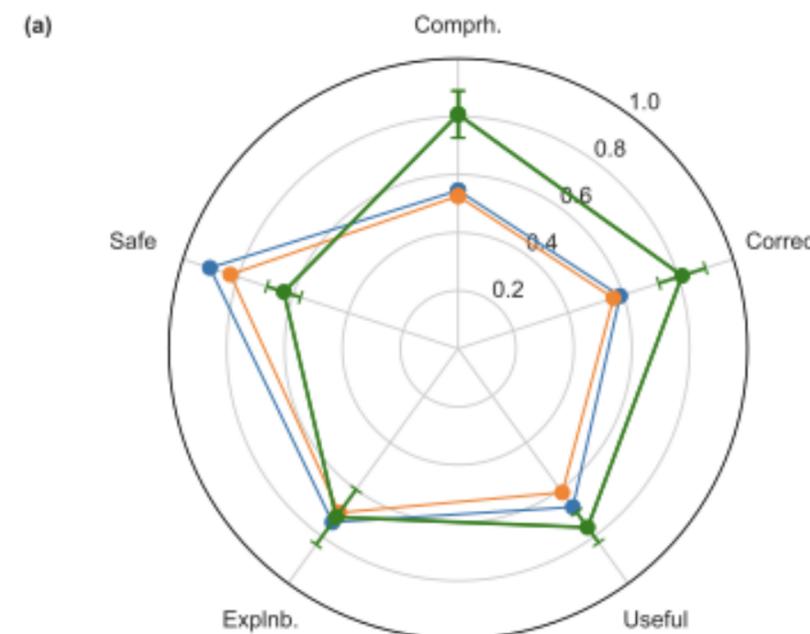
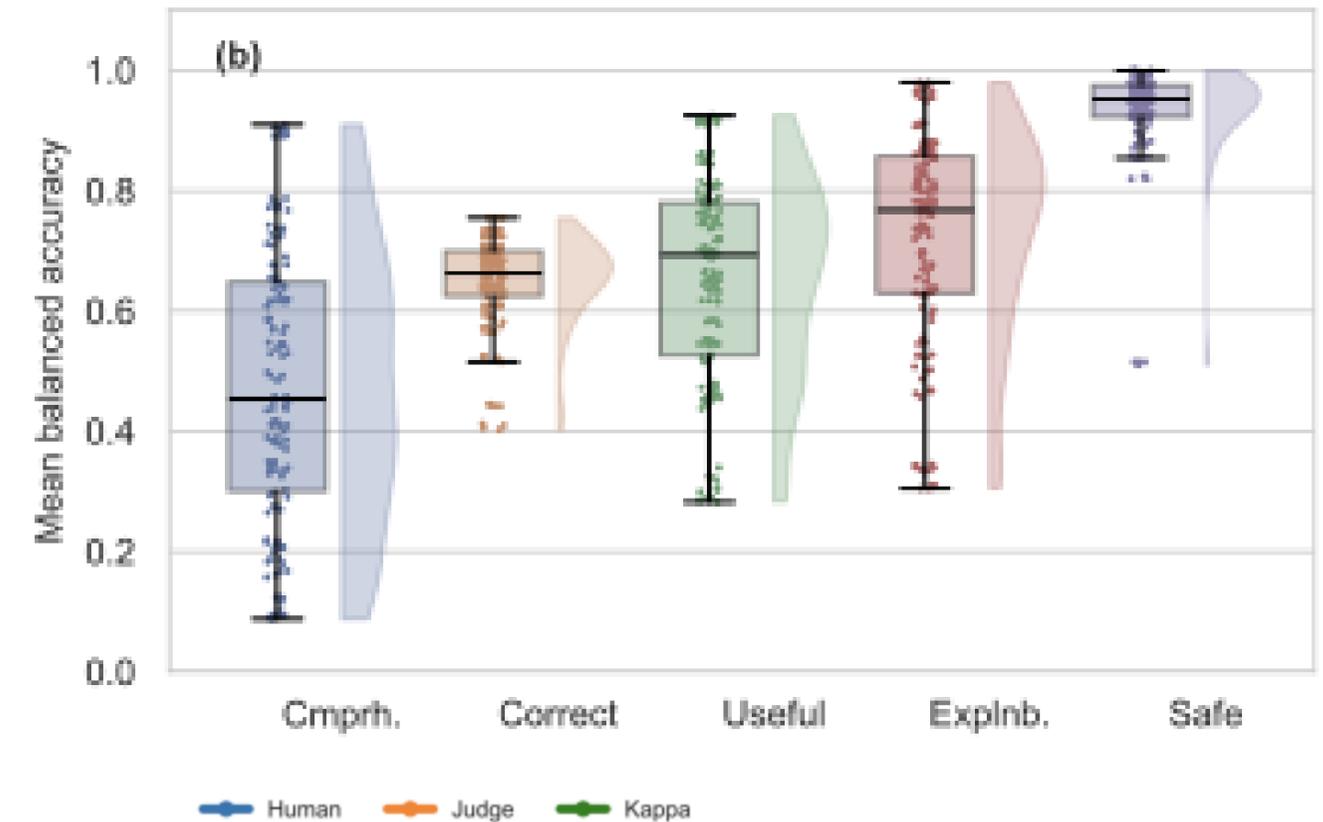
Benchmarking then—Vox populi



The continuous line is the normal curve with p.e. = 37.
The broken line is drawn from the observations.
The lines connecting them show the differences between the observed and the normal.

Benchmarking ageing interventions

- Set out five requirements (dimensions)
- Create domain expertise-driven modular benchmark items (25 synthetic cases)
- Combinatorial run (56 000 responses)
- Automated evaluation by LLM-as-a-judge (280 000 judgments)
- High inter-rater agreement with human (Cohen's kappa on 1000 random samples of judgments)



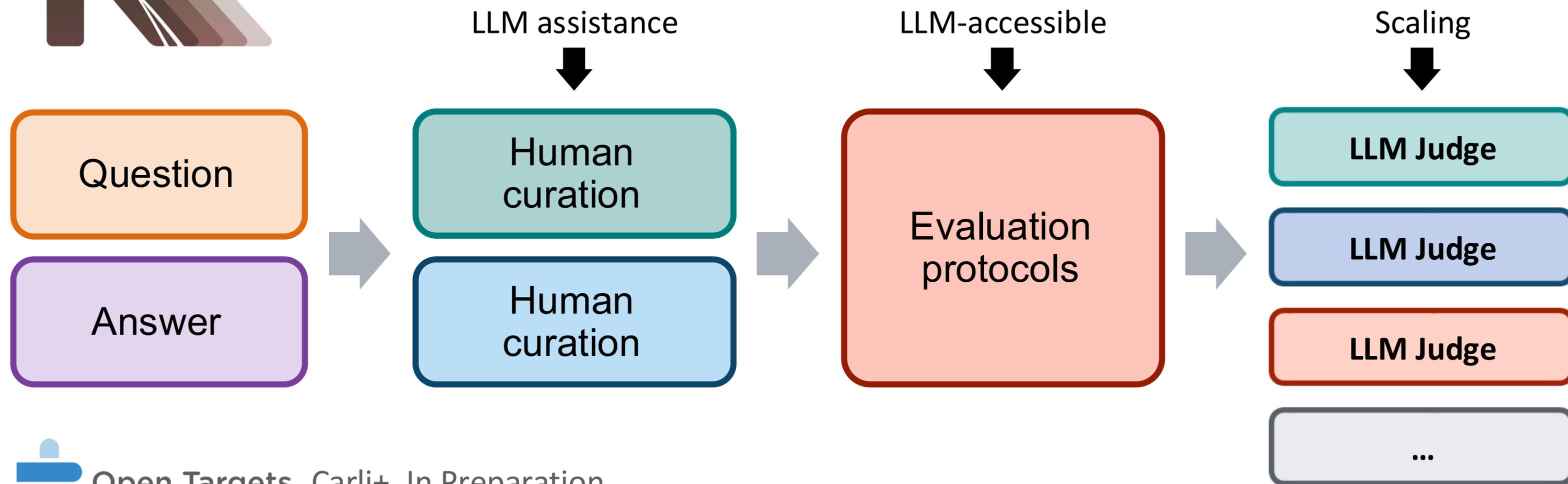
Fuellen+ 2025, [Ageing Research Reviews](#)
 Jarchow+ 2025, [npj Digital Medicine](#)

Karenina—Accessible Benchmarking

Karenina brings the mountain to the prophet: intuitive benchmarking for domain experts, best practices and machine learning principles in software form.



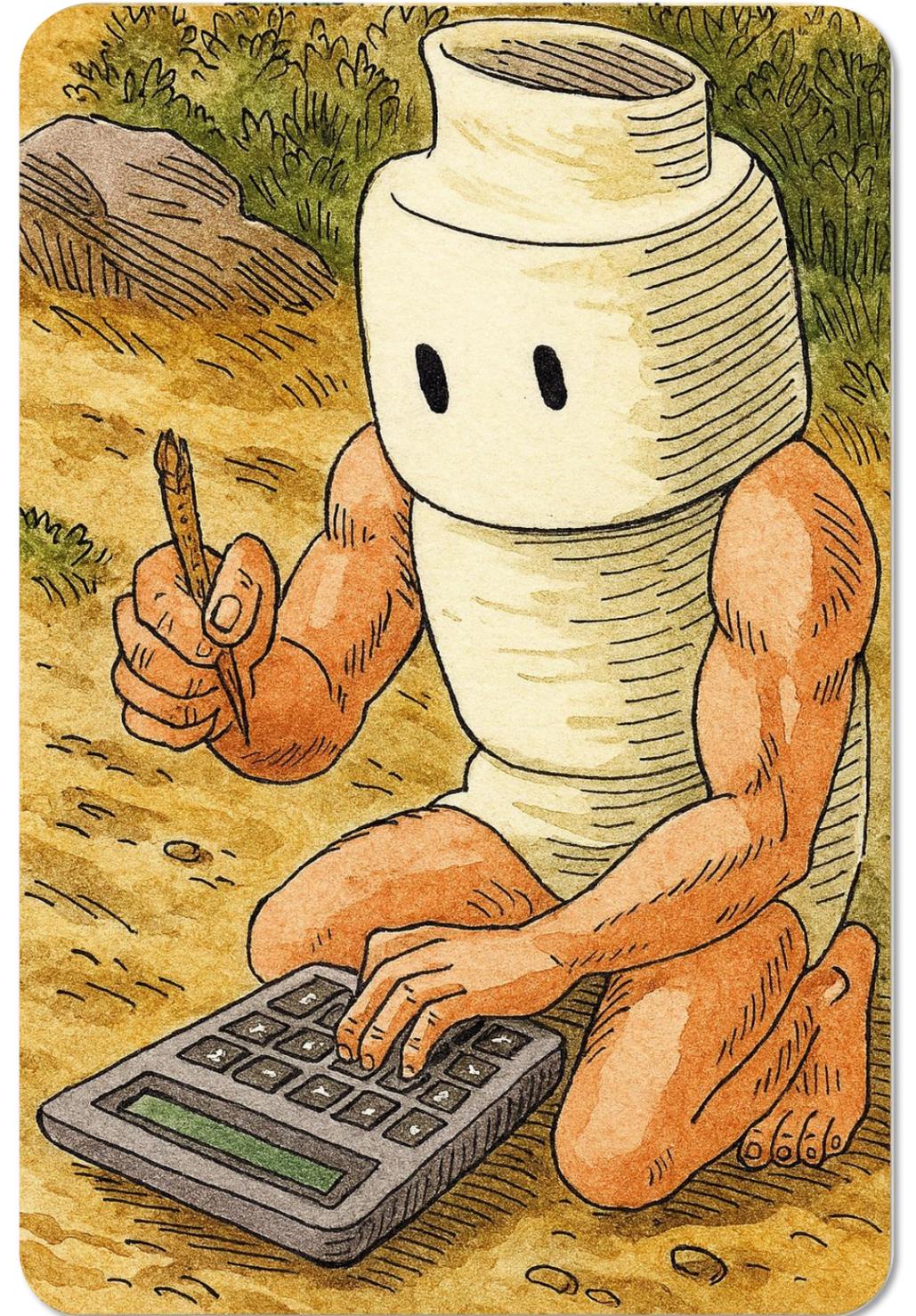
Are there chemical probes for APOE?	QC-01-175-1 and ML103
How many clinical trials have been reported for MAPT?	17



Model Context Protocol (MCP)

New de-facto standard for tool descriptions

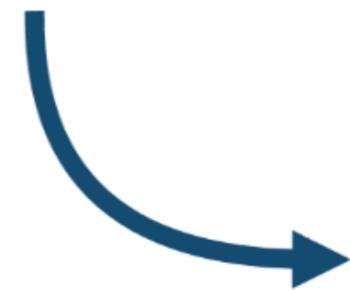
- Binding tools to LLMs reduces fragility of LLM returns
- Agents interact with validated external resources via standard protocols
- **“What is 3250 x 8496?”**
- Advantages: connection to validated domain-specific databases and tools → less probabilism
- Disadvantages: requires adequate tooling and strict protocol adherence



The BioContext MCP Registry

The screenshot shows a web browser window at biocontext.ai. The navigation bar includes links for Registry, Documentation, Chat, Collections, Blog, Community, and Publication, along with a Sign In button and a settings icon. The main heading is "Registry" with the subtitle "Discover biomedical Model Context Protocol servers for your agentic AI applications". A search bar contains the text "Search MCP servers...". A blue callout box with white text says "Start from cookiecutter". A dropdown menu is set to "Most GitHub Stars". Below the search bar, it says "40 servers found • Page 1 of 3". The "MCP Servers" section features an "Add Your Server" button and three server cards: "ToolUniverse" (455 stars), "BioMCP" (322 stars, with a "mcp.json available" badge), and "Google Scholar MCP Server" (145 stars). Each card includes a brief description of the server's purpose.

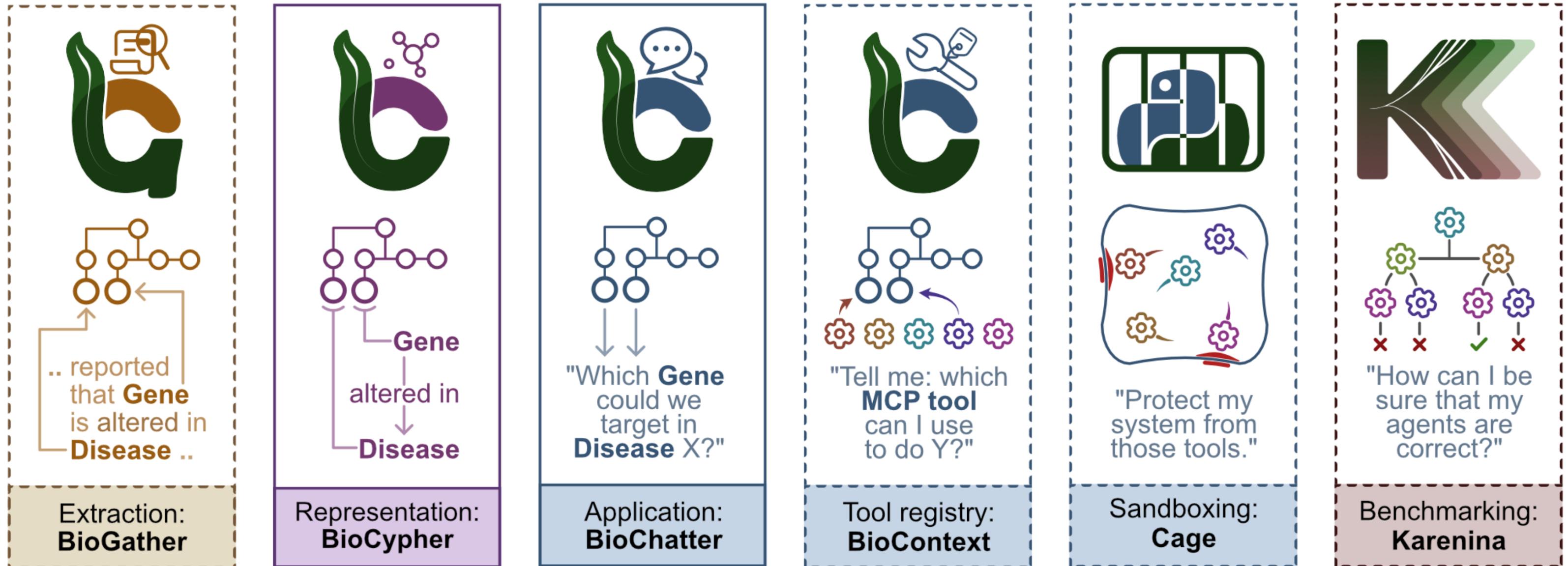
Our principles



AI development
AI deployment
Developer toolkits

A user-centric ecosystem for research

Leverage project **context** for user-friendly workflows.

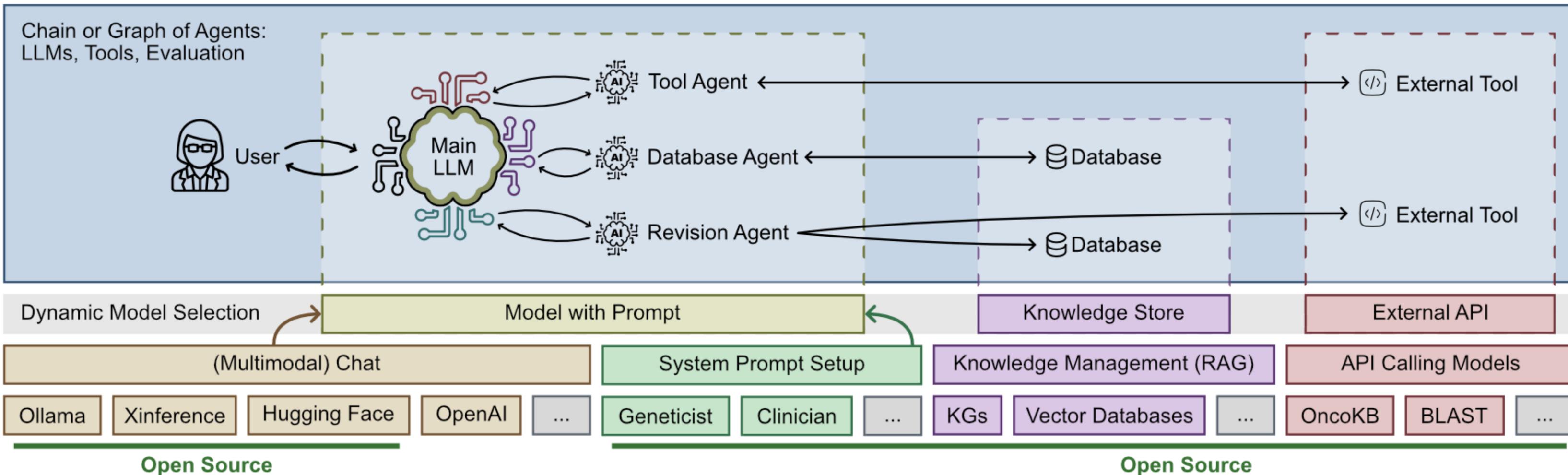


BioChatter and the rapidly evolving landscape

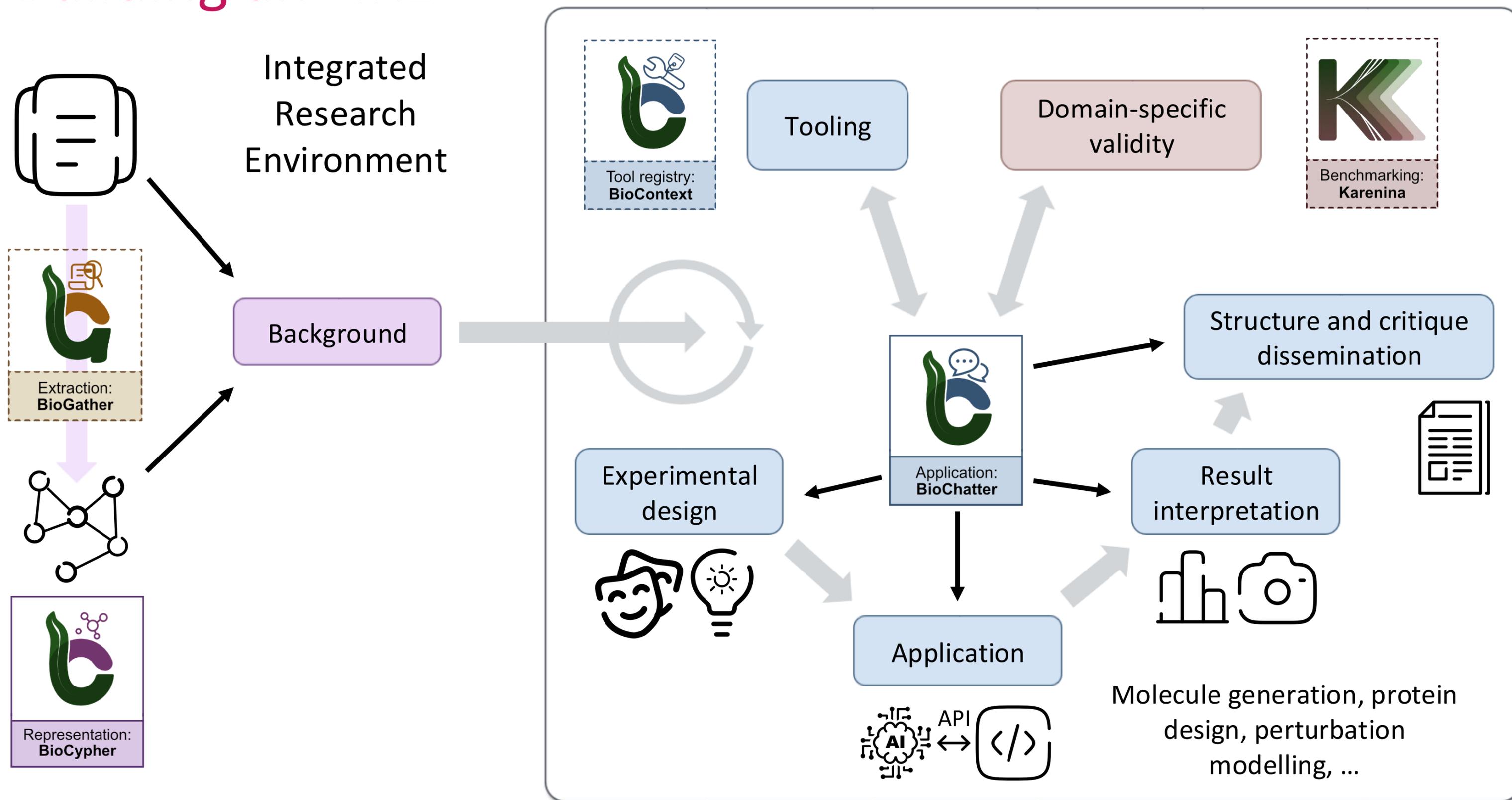
Lobentanzer+ **2023**, Preprint,
<https://arxiv.org/abs/2305.06488>



Lobentanzer+ **2025**, Nature
Biotechnology, <https://biochatter.org>



Building an "IRE"



How should we plan and develop agentic systems?

- Agentic systems require **delegation of authority** – the human steps back, the system decides
- Today: projects are planned ad hoc, expectations are informal, governance is implicit
- Result: systems that disappoint in practice, with no structured baseline to measure against
- Existing frameworks (Model Cards, Datasheets, NIST AI RMF) are retrospective or not machine-readable
- Missing: a prospective, structured **contract** between users and developers



<https://aac.slolab.ai>

The Agentic Automation Canvas		Design for: Automated Document Processing System	
<p>PROJECT DEFINITION </p> <p>Automated Document Processing System</p> <p>A system that automatically processes, categorizes, and routes incoming documents to reduce manual handling time and improve accuracy. The system uses AI to extract key information, classify document types, and route them to appropriate team members.</p> <p>DEVELOPMENT</p> <p>Saves time through automated document processing</p> <p>Primary value: Time</p> <p>Business Process Automation Document Management</p>	<p>USER EXPECTATIONS </p> <p>4 tasks</p> <ul style="list-style-type: none"> Extract key information from documents <i>As a data entry clerk, I want documents to be automatically processed so that I can focus on exception handling instead of routine data entry</i> Categorize documents by type <i>As a team lead, I want documents automatically categorized so that routing decisions are consistent and faster</i> Route documents to team members <i>As a manager, I want documents automatically routed to the right person so that processing time is minimized</i> Automated security and compliance validation <i>As a compliance officer, I want documents automatically validated for security and compliance risks so that we prevent processing of problematic documents</i> <p>BENEFITS</p> <p>~68.3 hrs/month saved</p> <p>Time 3 Quality 3 Risk 3 Cost 1 Enablement 1</p>	<p>DEVELOPER FEASIBILITY </p> <p>TRL4 → 7</p> <p>Risk: Medium</p> <p>Effort: 20 weeks</p> <p>~10.8 mo until amortization</p> <p>Core technologies are mature. Main challenge is handling edge cases and ensuring accuracy for diverse document formats. Different tasks require different approaches - extraction needs extensive agentic capabilities, while categorization can be deterministic.</p> <p>TASK-LEVEL FEASIBILITY: 4 OF 4 TASKS</p> <div style="background-color: #ccc; width: 100%; height: 10px;"></div>	
<p>GOVERNANCE </p> <ul style="list-style-type: none"> Design 2024-01-15 → 2024-02-15 3 agents, 2 milestones Development 2024-02-16 → 2024-05-15 2 agents, 2 milestones Validation 2024-05-16 → 2024-06-15 2 agents, 1 milestones Deployment 2024-06-16 → 2024-06-30 1 agents, 1 milestones 	<p>DATA ACCESS </p> <p>2 datasets</p> <p>restricted: 1 confidential: 1</p>	<p>OUTCOMES </p> <p>2 DELIVERABLES</p> <p>Document Processing System User Documentation</p> <p>1 PUBLICATION</p> <p>Automated Document Processing: A Case Study</p> <p>1 EVALUATION</p> <p>Performance Evaluation — System achieved 92% accuracy and reduced processing time by 65%</p>	
<p>Date: 2026-02-11 · Version: 1.2.8</p>		<p>Generated by the Agentic Automation Canvas · https://aac.slolab.ai</p>	

The Agentic Automation Canvas and you?

- Brand new
- Probably incomplete
- Looking for use cases, feedback, ideas
- Try it, get in touch!
- Maybe contribute to the paper?
- <https://aac.slolab.ai>



HELMHOLTZ MUNICH

Matthias Heinig
Raeesa Yousaf
Peyman Vahidi
Rupshali Dasgupta

DZD

Lars Oest
Angela Dedié
Anette-Gabriele Ziegler
Julia Szendrödi
Thomas Fleming



Paul Ka Po To
Francesco Carli
Ellen M. McDonagh
David Ochoa

DECIDER

Improving clinical decisions in cancer

Julio Saez-Rodriguez
Johann Dreo
Marko Baric
Matthieu Najm
Benno Schwikowski

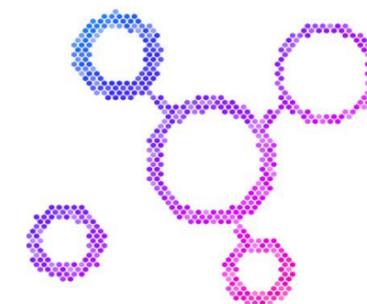
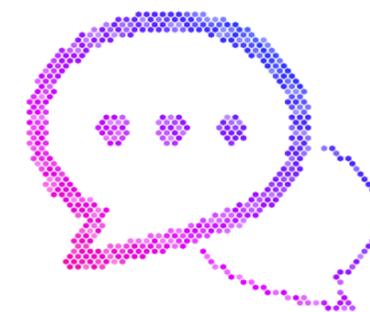
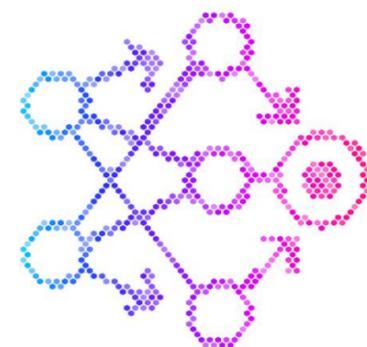
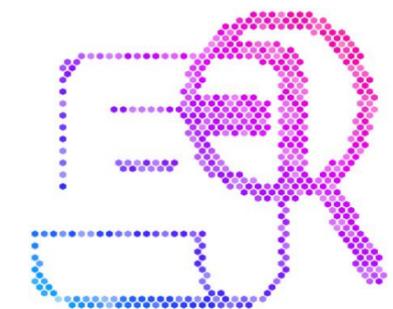


Pedro Assis (Stanford), Ben Hitz (Stanford), Andreas Maier (Hamburg), Jan Baumbach (Hamburg), Elias Farr (EMBL), Vincent Vialard (Semspect), Judith Wodke (Greifswald), Dagmar Waltemath (Greifswald), Vincent Carey (Harvard), Noah Bruderer (Bergen), Shaohong Feng (Ohio State), Cankun Wang (Ohio State), Qin Ma (Ohio State), Trang Le (BMS), Cyril Pommier (INRAE), Sebastian Beier (FZJ), Matti van Welzen (Rostock), Olaf Wolkenhauer (Rostock), Hans Jarchow (Rostock), Georg Fuellen (Rostock), Malte Kuehl (Aarhus), Victor Puelles (Aarhus/Hamburg), Edwin Careño (Heidelberg), Valeriia Dragan (Heidelberg), Nils Krehl (Heidelberg).



Stay open source, get in touch!

biocypher.org | biochatter.org



SLOLAB.AI