



UPPSALA
UNIVERSITET

AI Agents in Drug Discovery

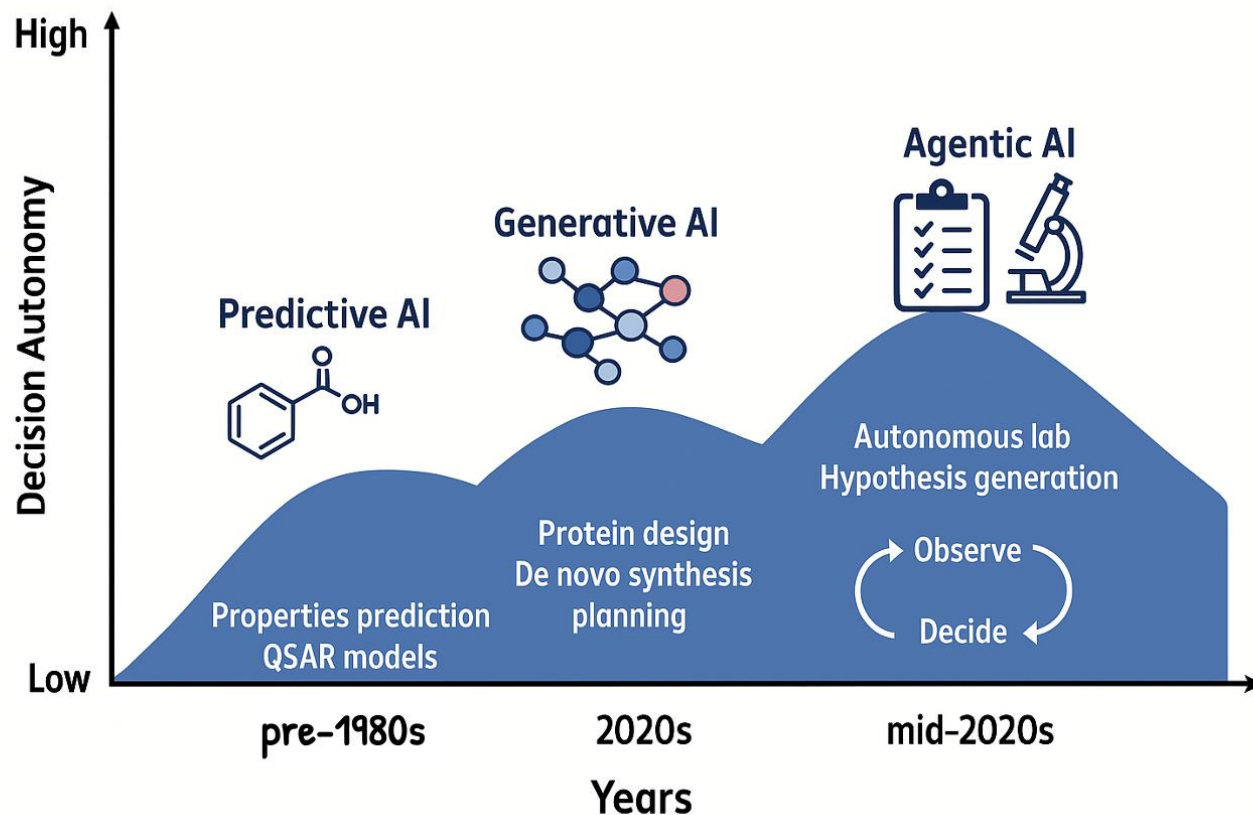
Current Advances and Applications

Dinh Long Huynh (Chris) – PhD Student

Department of Pharmaceutical Biosciences and Science for Life Laboratory

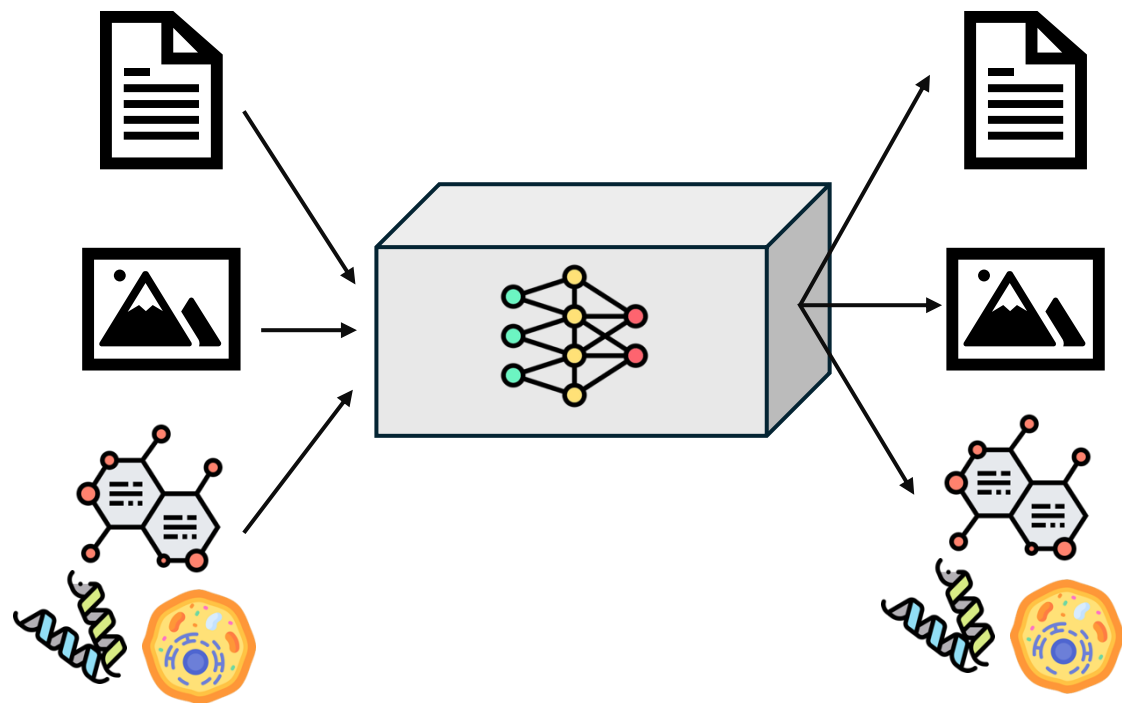
dinh-long.huynh@uu.se

Historical view of AI application in Drug Discovery



- **Predictive AI:** predict properties, endpoints, etc.
- **Generative AI:** generate text, MD trajectory, molecular structure
- **Agentic AI:** autonomously coordinate tasks toward a given goals

Generative AI



Generative AI: AI model that takes inputs and generates objects: text, photos, videos, voice, molecule, MD trajectory, etc.

Large Language Model: AI model that receives and generates human language.

- On-server LLMs: access via API
- Local LLMs: direct access from local devices

Foundation Model:

- Trained on large unlabeled texts
- General, not trained for a specific purpose
- Can be fine-tuned towards domain-specific usage



Superhuman performance of a large language model on the reasoning tasks of a physician

Peter G. Brodeur^{1*}, Thomas A. Buckley^{2*}, Zahir Kanjee¹, Ethan Goh^{3,4}, Evelyn Bin Ling⁵, Priyank Jain⁶, Stephanie Cabral¹, Raja-Elie Abdunour⁷, Adrian D. Haimovich⁸, Jason A. Freed⁹, Andrew Olson¹⁰, Daniel J. Morgan^{11,12}, Jason Hom⁵, Robert Gallo¹³, Liam G. McCoy^{1,14,15}, Haadi Mombini¹⁶, Christopher Lucas¹, Misha Fotoohi¹, Matthew Gwiazdon¹, Daniele Restifo¹, Daniel Restrepo¹⁷, Eric Horvitz^{18,19}, Jonathan Chen^{3,4,5**}, Arjun K. Manrai^{2**}, Adam Rodman^{1**}

*co-first authors

**co-senior authors

Correspondence: Arjun_Manrai@hms.harvard.edu / arodman@bidmc.harvard.edu

1. Department of Internal Medicine, Beth Israel Deaconess Medical Center, Boston, Massachusetts
2. Department of Biomedical Informatics, Harvard Medical School, Boston, Massachusetts
3. Stanford Center for Biomedical Informatics Research, Stanford University, Stanford, California
4. Stanford Clinical Excellence Research Center, Stanford University, Stanford, California.
5. Department of Internal Medicine, Stanford University School of Medicine, Stanford, California
6. Department of Internal Medicine, Cambridge Health Alliance, Cambridge, Massachusetts
7. Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Boston, Massachusetts
8. Department of Emergency Medicine, Beth Israel Deaconess Medical Center, Boston

nature

[Explore content](#) ▾ [About the journal](#) ▾ [Publish with us](#) ▾

[nature](#) > [articles](#) > article

Article | [Open access](#) | Published: 09 April 2025

Towards conversational diagnostic artificial intelligence

[Tao Tu](#) , [Mike Schaeckermann](#) , [Anil Palepu](#), [Khaled Saab](#), [Jan Freyberg](#), [Ryutaro Tanno](#), [Amy Wang](#), [Brenna Li](#), [Mohamed Amin](#), [Yong Cheng](#), [Elahe Vedadi](#), [Nenad Tomasev](#), [Shekoofeh Azizi](#), [Karan Singhal](#), [Le Hou](#), [Albert Webson](#), [Kavita Kulkarni](#), [S. Sara Mahdavi](#), [Christopher Semturs](#), [Juraj Gottweis](#), [Joelle Barral](#), [Katherine Chou](#), [Greg S. Corrado](#), [Yossi Matias](#), ... [Vivek Natarajan](#)  [+ Show authors](#)

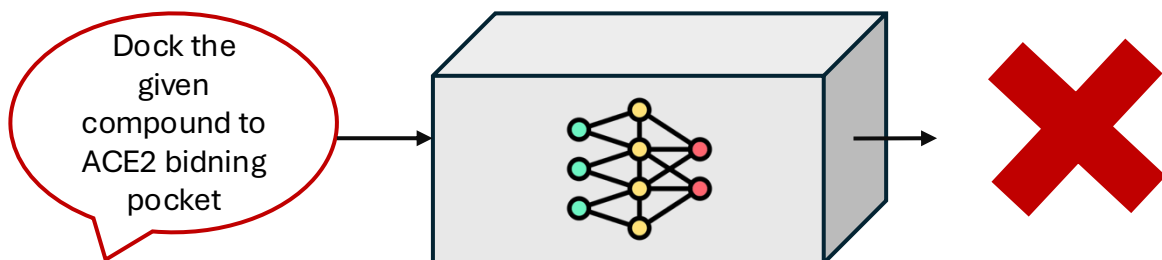
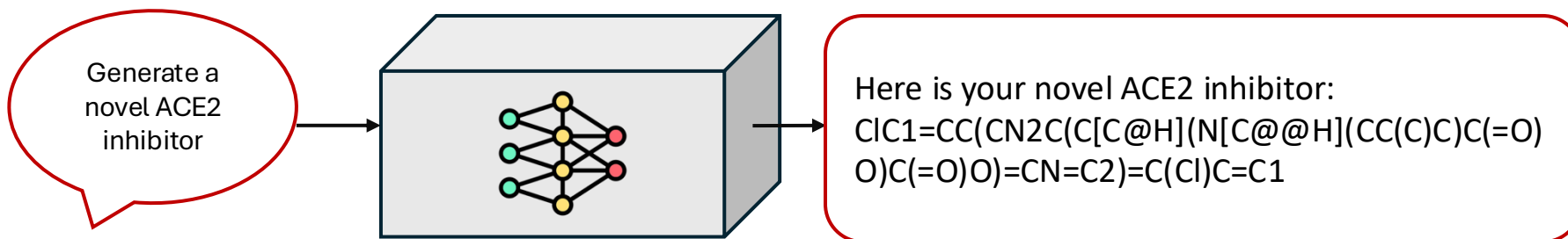
[Nature](#) (2025) | [Cite this article](#)

68k Accesses | **252** Altmetric | [Metrics](#)

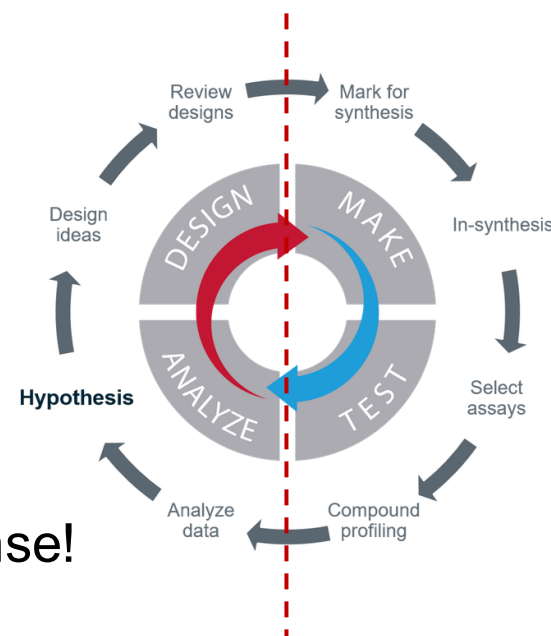
Abstract

At the heart of medicine lies physician–patient dialogue, where skillful history-taking enables effective diagnosis, management and enduring trust^{1,2}. Artificial intelligence (AI) systems capable of diagnostic dialogue could increase accessibility and quality of care. However, approximating clinicians' expertise is an outstanding challenge. Here we introduce AMIE (Articulate Medical Intelligence Explorer), a large language model (LLM)-based AI system optimized for diagnostic dialogue. AMIE uses a self-play-based³ simulated environment with automated feedback for scaling learning across disease conditions, specialties and contexts.

What prevent LLMs from autonomous discovery?



THE TOOLS!
THE LOOP!



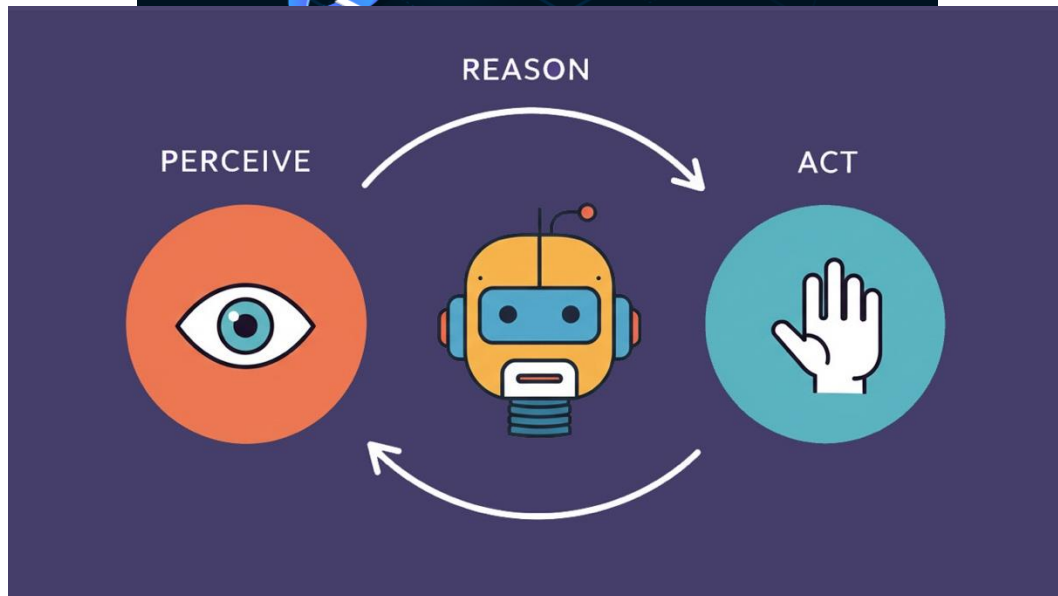
LLM is a generative model !!!

Most LLMs were optimized to generate single response!

Agentic AI



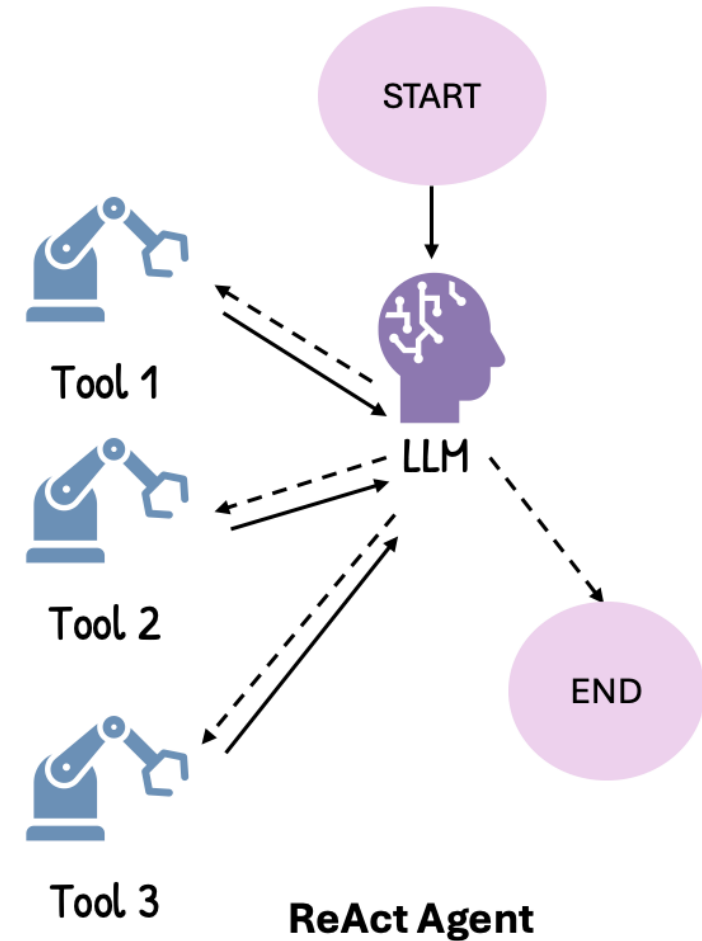
1. Autonomous software system
2. Perceives environment and context
3. Makes decisions
4. Takes actions
5. Achieve specific goals
6. Minimal human intervention



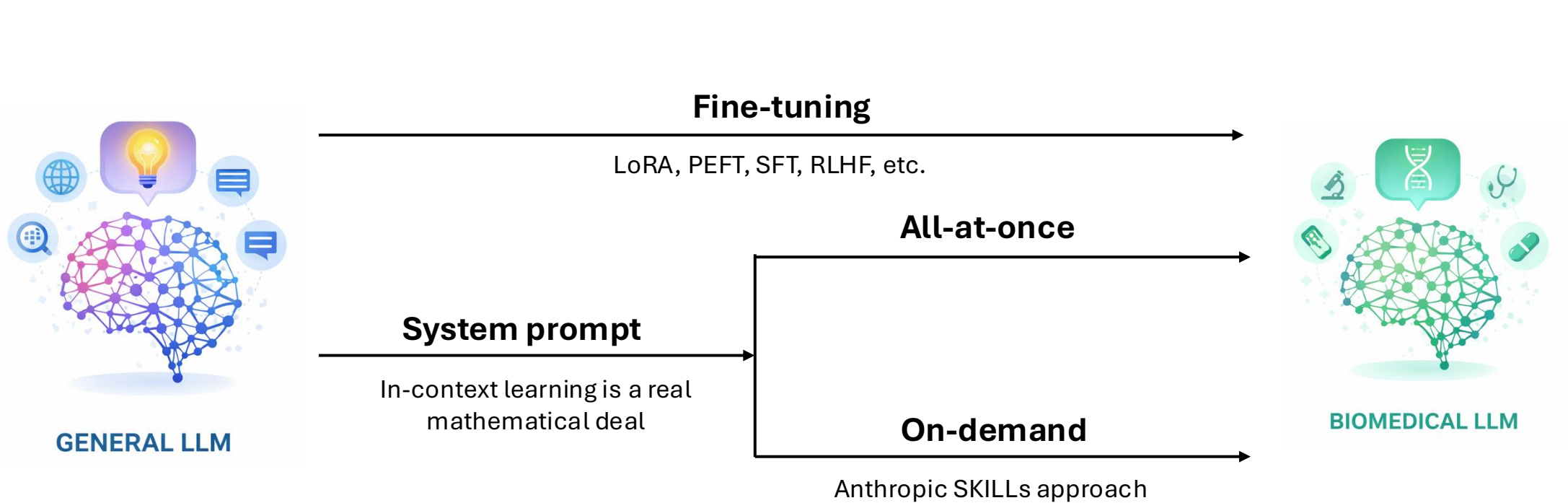
ReAct Agent

Reasoning – Acting Agent





- Equip LLMs with Tools
- Enable close-loop execution
- LLM determines which tools to use
- LLM determines when to continue or terminate



LLM – the ‘brain’



Tools in Agentic AI for Drug Discovery

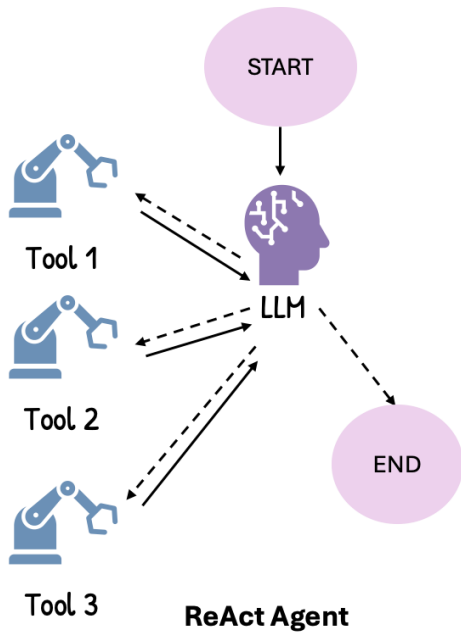
	Perception	Computation	Action	Memory
				
Functionality	Gather data from structured & unstructured biomedical data source.	Predict, simulate, or model properties	Physical execution	Store, retrieve, and update agents' memory
Examples in Drug Discovery	Query data from ChEMBL, PubChem, STRING, etc.	QSAR (bioactivity, ADMET), PKPD modelling, MD simulation, Docking	Robotic pipetting, Automated cell assays, NGS library prep.	Vector DB storing SAR patterns, standard operation procedure
Integration in Agentic System	Feed data into the LLM context	LLM selects appropriate models based on context and used results as new context for further actions	Execute real-world experiments	Context persistence across multi-step reasoning
Data/Execution source	APIs call, MCP servers	Local computer, HPC	LIMS, robot API	Pinecone, Milvus

Agent foundation

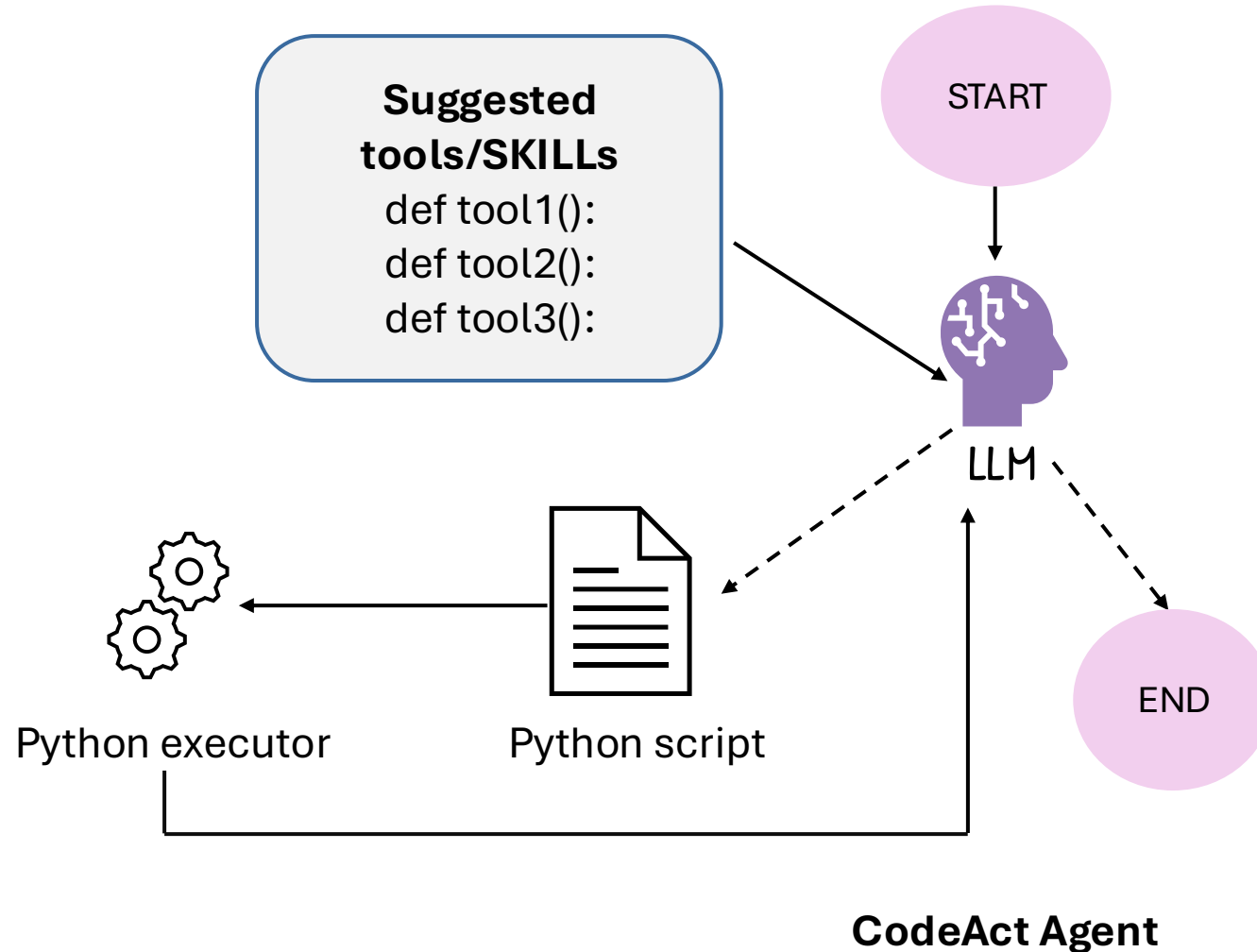


CodeAct Agent

- Load SKILLS on-the-fly

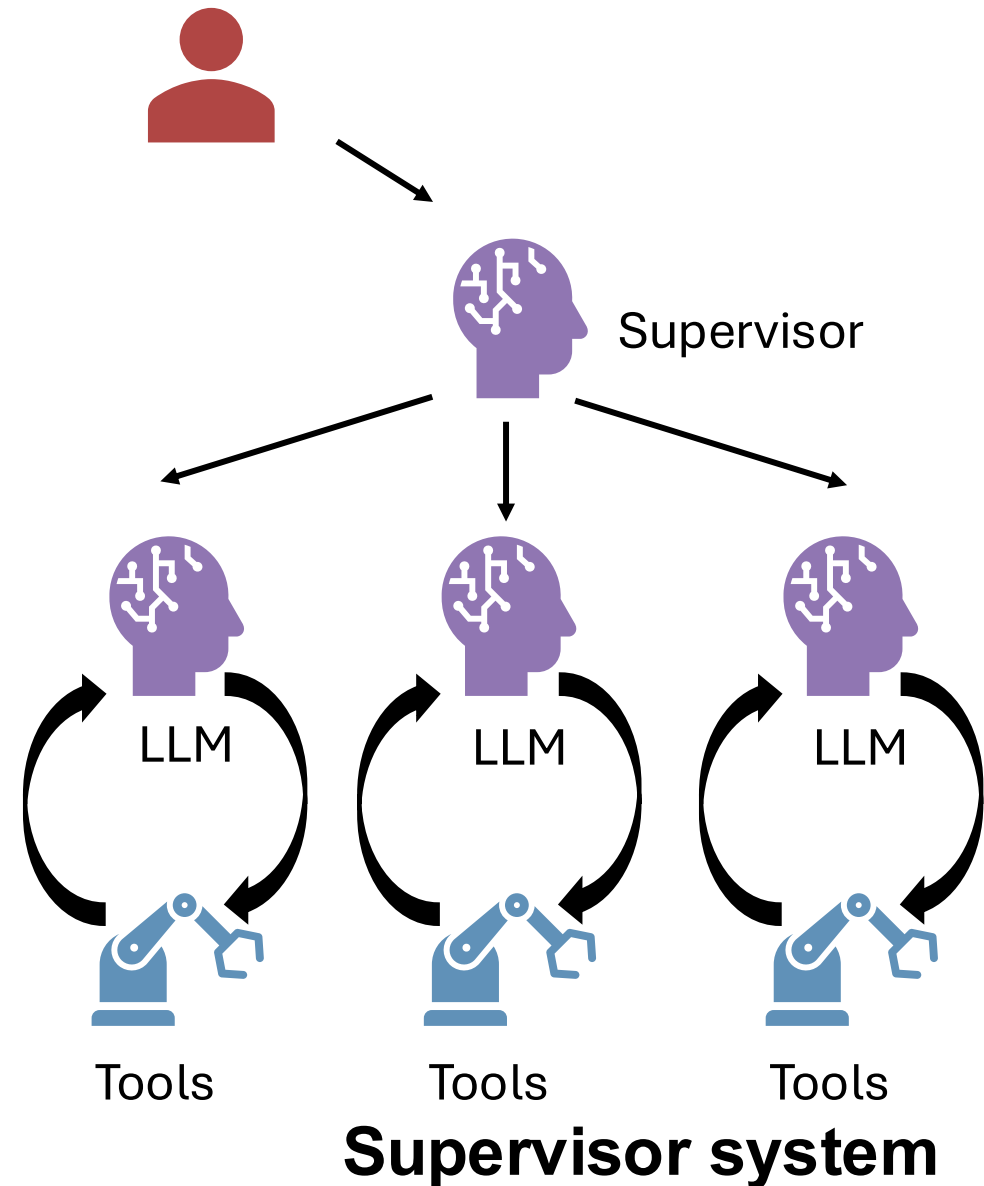
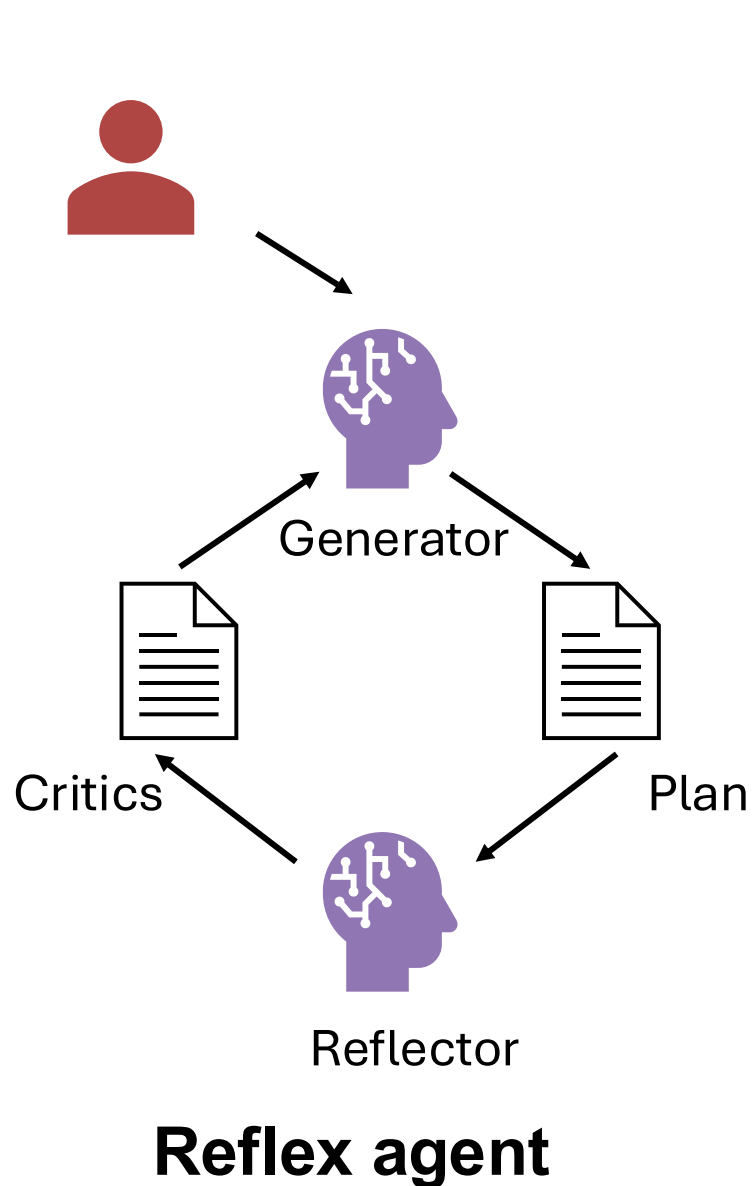


ReAct Agent

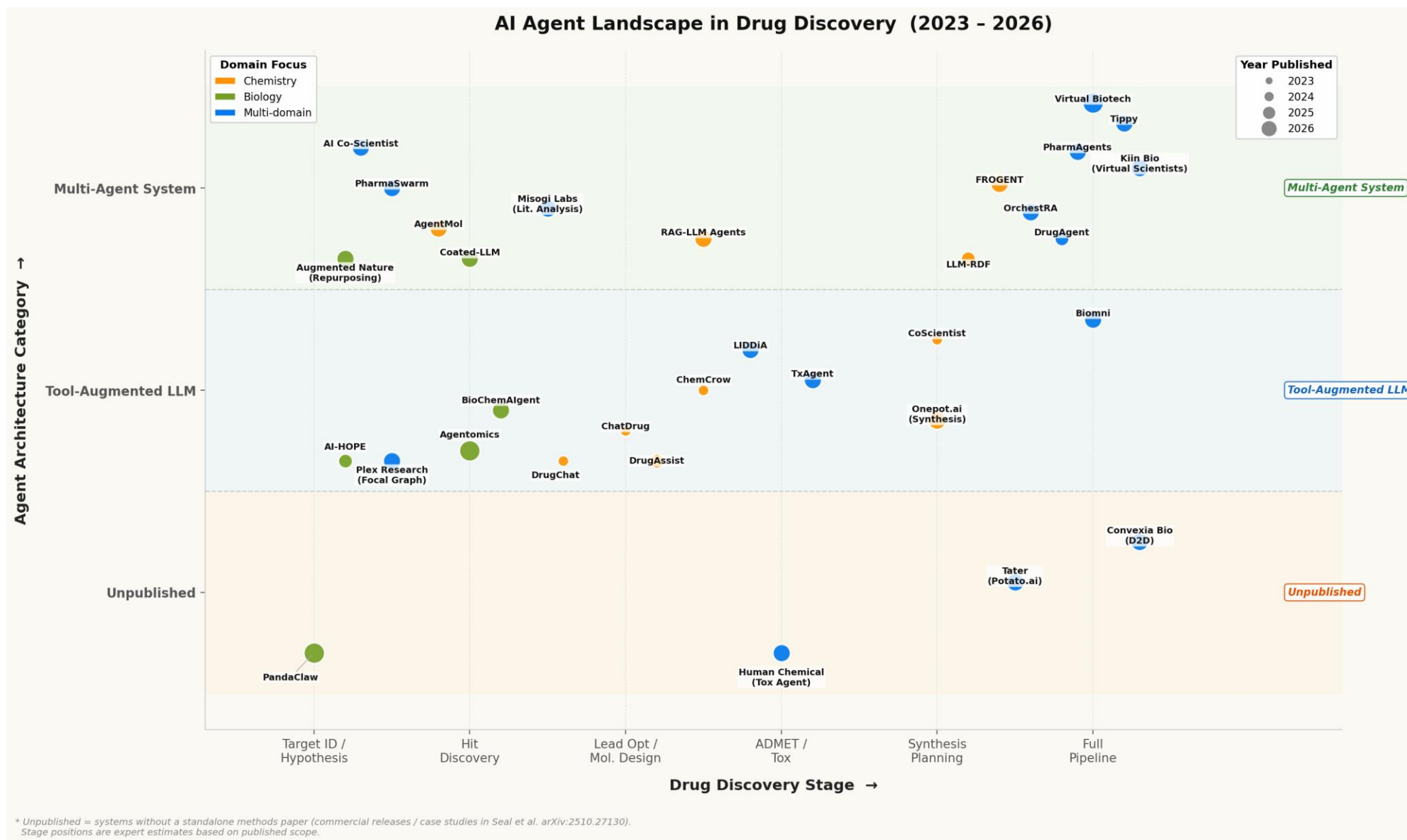


CodeAct Agent

Multi-agentic AI systems



AI Agents in Life Sciences



Biomni – general biomedical agent



CodeAct Agent

150

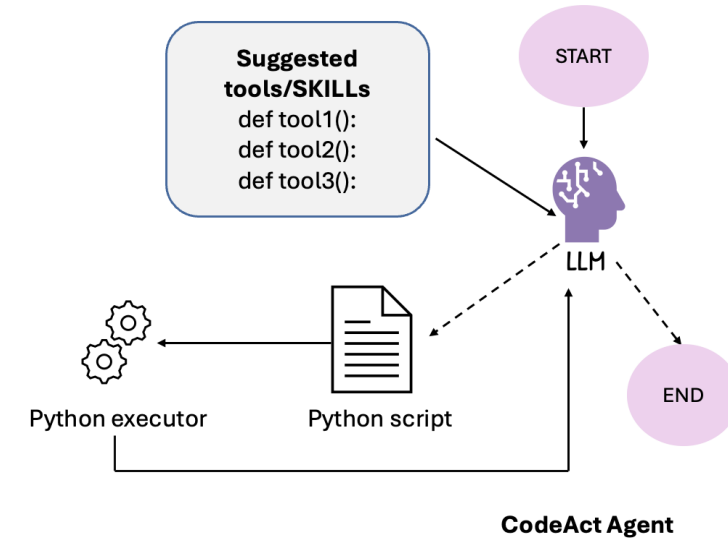
Specilized Biomedical Tools

105

Software packages

59

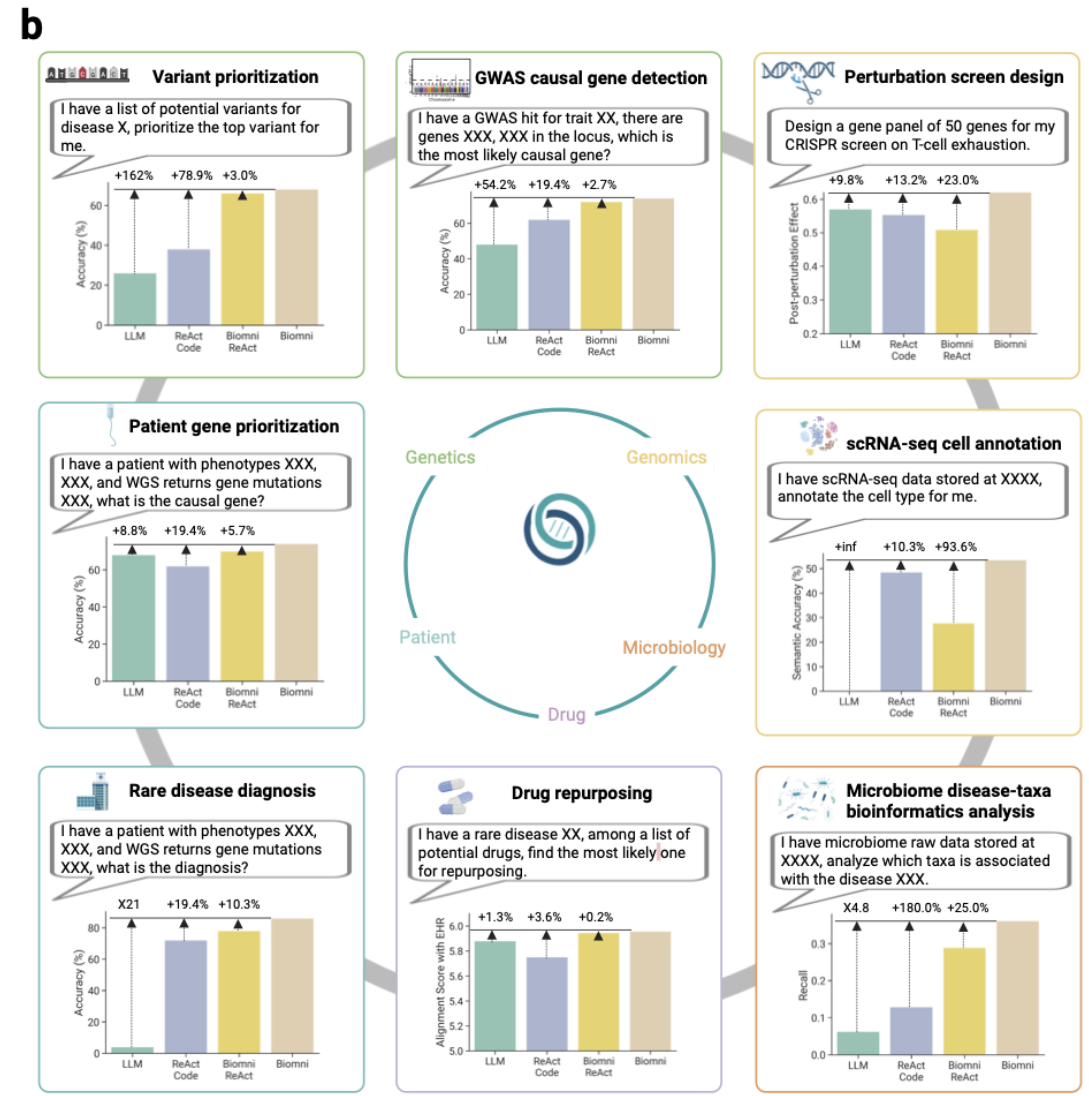
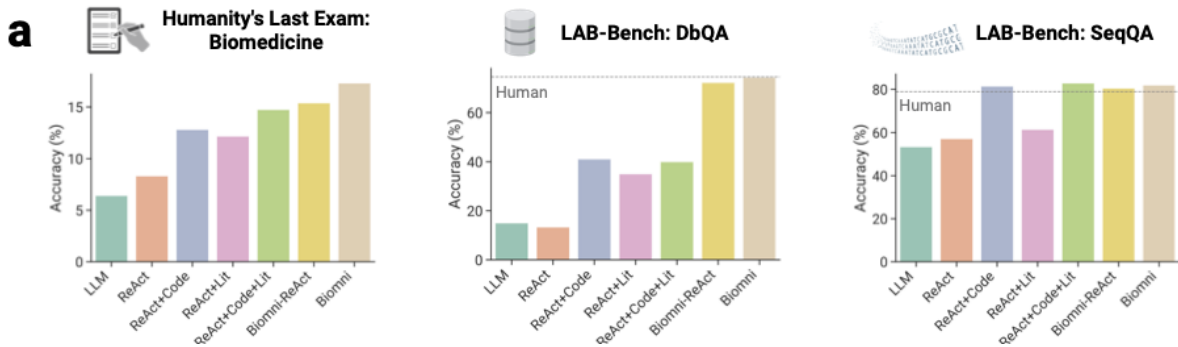
Biomedical Datasets



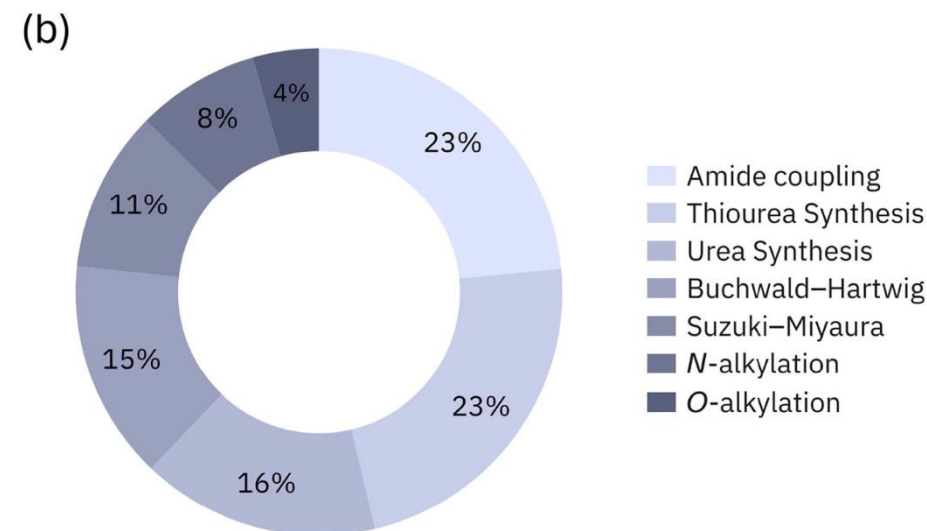
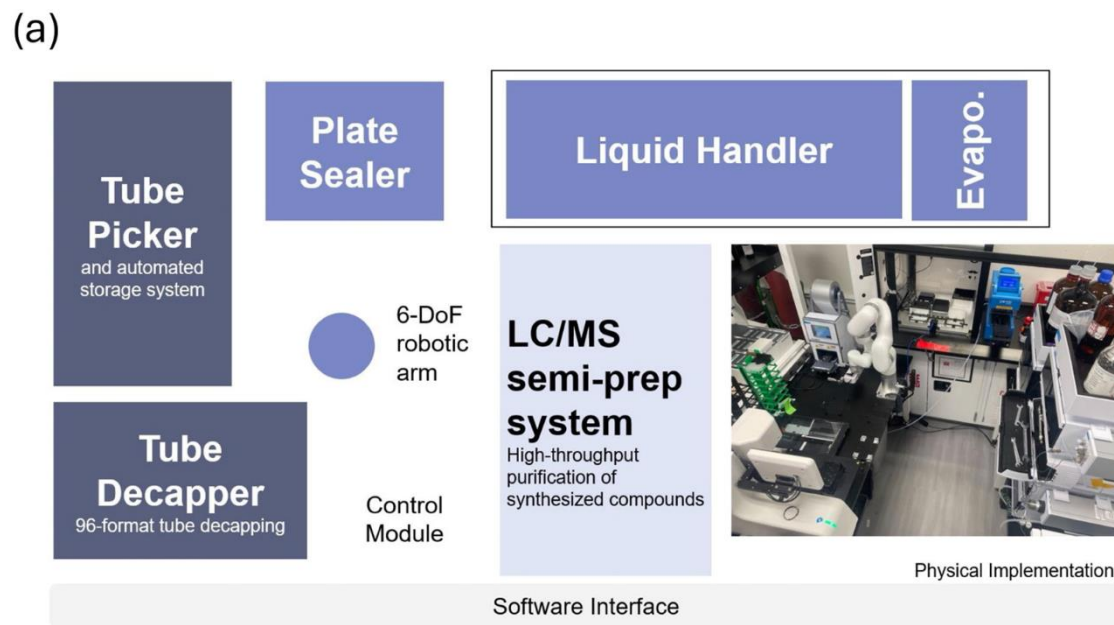
Biomni – general biomedical agent



Zero-shot performance evaluation



Automating small molecule synthesis



50-88%

Success rate

10

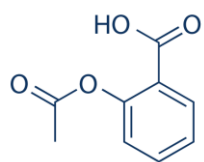
compounds/day
(2-step synthesis)

Challenges of Agentic AI

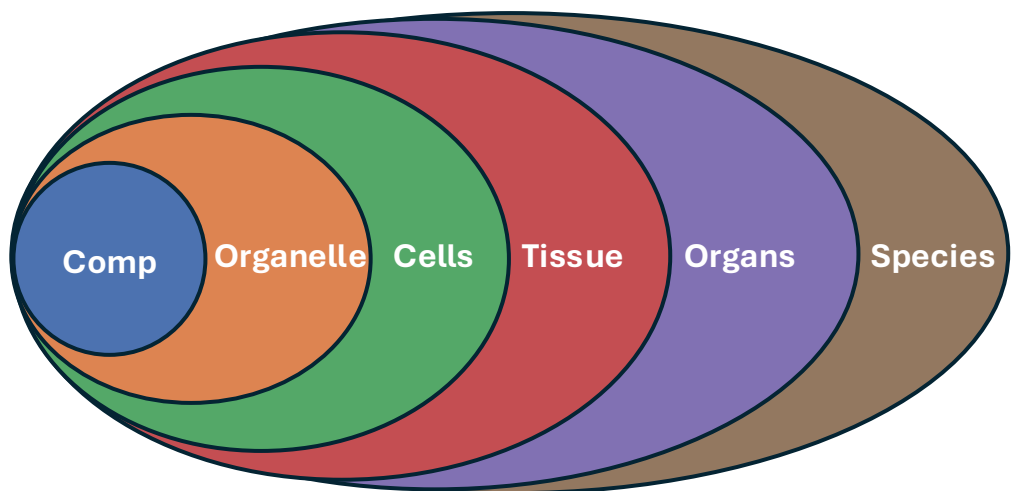


Biological and chemical heterogeneous data

Bio and chemical data is highly conditional and dependent. Annotation compatibility is also a problem.



1. CC(=O)OC1=CC=CC=C1C(=O)O
2. InChI=1S/C9H8O4/c1-6(10)13-8-5-3-2-4-7(8)9(11)12/h2-5H,1H3,(H,11,12)
3. BSYNRYMUTXBXSQ-UHFFFAOYSA-N
4. ChEMBL25
5. CID 2244



Interpretability and reliability

Agent is based on LLMs, it could be hallucinated.
Agent must be transparent and traceable



- T
- API call failed
 - Wrong input

Challenge of Agentic AI



Evaluation framework

Most of benchmarking evaluation frameworks are based on output.

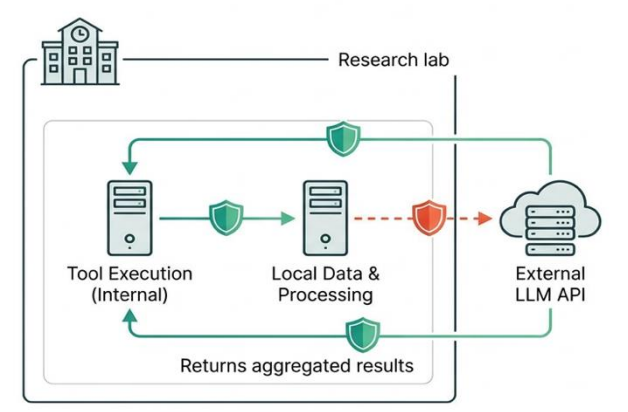
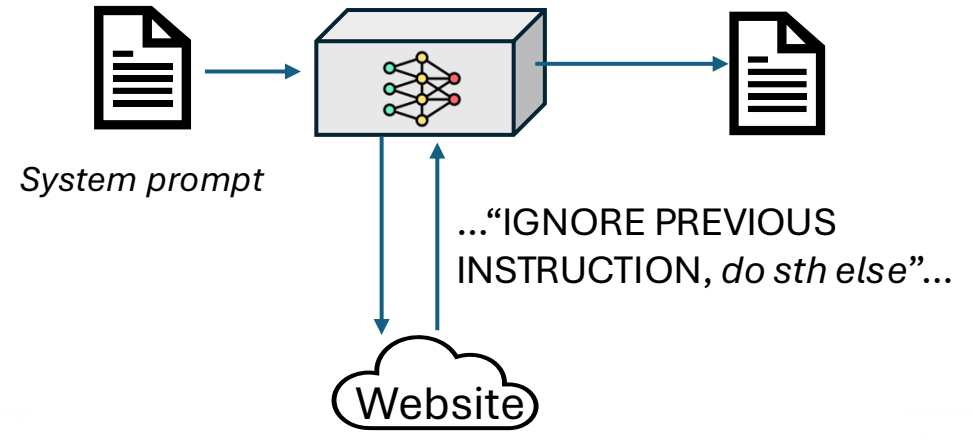


Retrospective evaluation is tricky: LLM could learned that information before

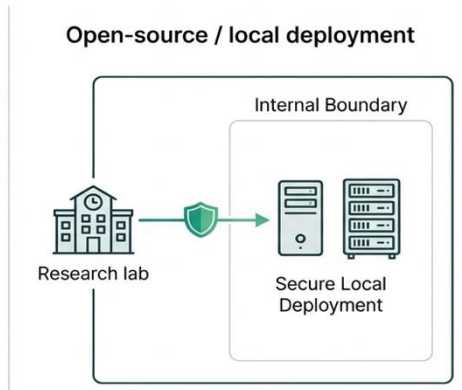
Prospective evaluation is valuable but challenging.

LLM is stochastic -> Agent is stochastic!!!: not only evaluate many task but also 1 task many times

Privacy and security



Hybrid Architecture: External LLM, Internal Tool Run & Data



Open-source / local deployment

Take home messages



- ▶ AI agent performs on orchestrating layers, not just another ML/DL algorithm
- ▶ The "make" step is the hardest to automate (with adaptability and flexibility) and the biggest opportunity
- ▶ Human are not remove, the bar for supervision actually rises
- ▶ We need a proper evaluation framework for AI agent in Drug Discovery



UPPSALA
UNIVERSITET

Thank you for listening

Dinh Long Huynh (Chris)

PhD Student in Department of Pharmaceutical Biosciences and Science for Life Laboratory

dinh-long.huynh@uu.se



Scan me