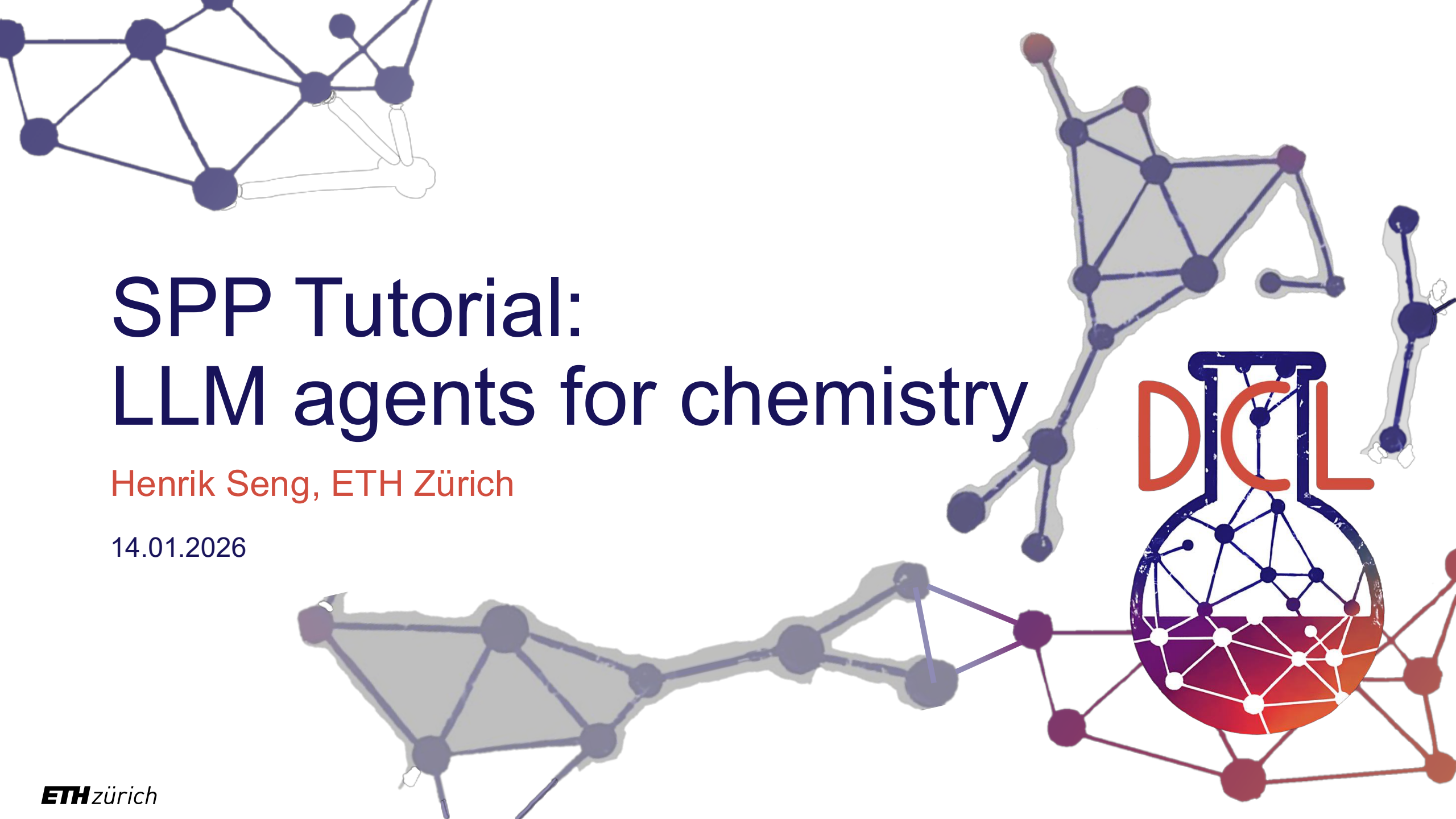
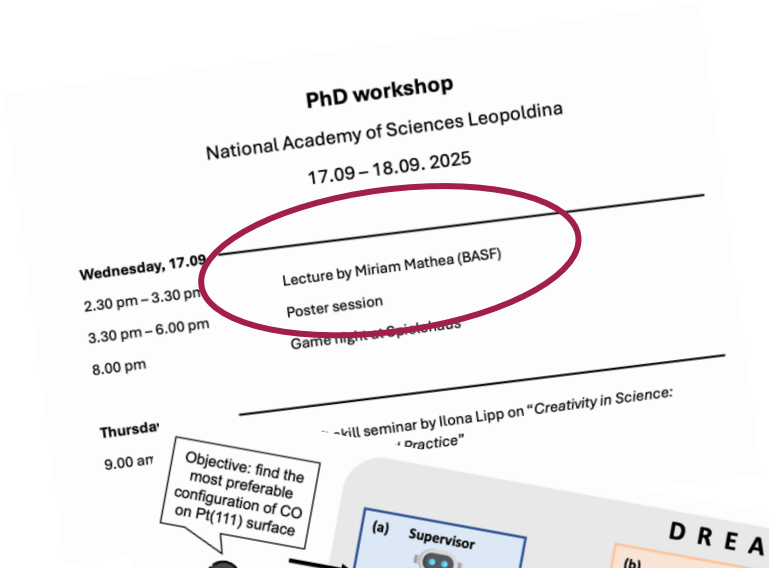
SPP Tutorial:
LLM agents for chemistry

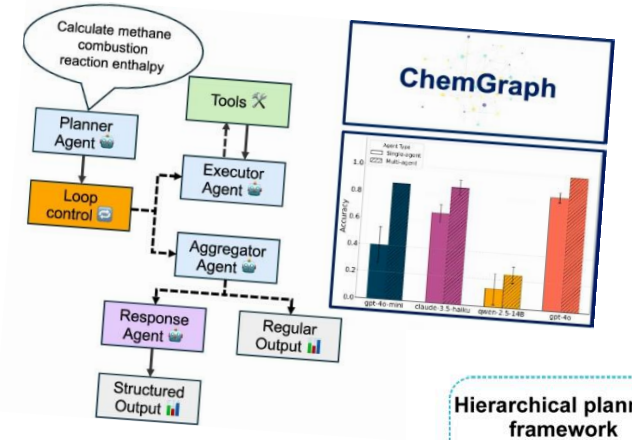
Henrik Seng, ETH Zürich

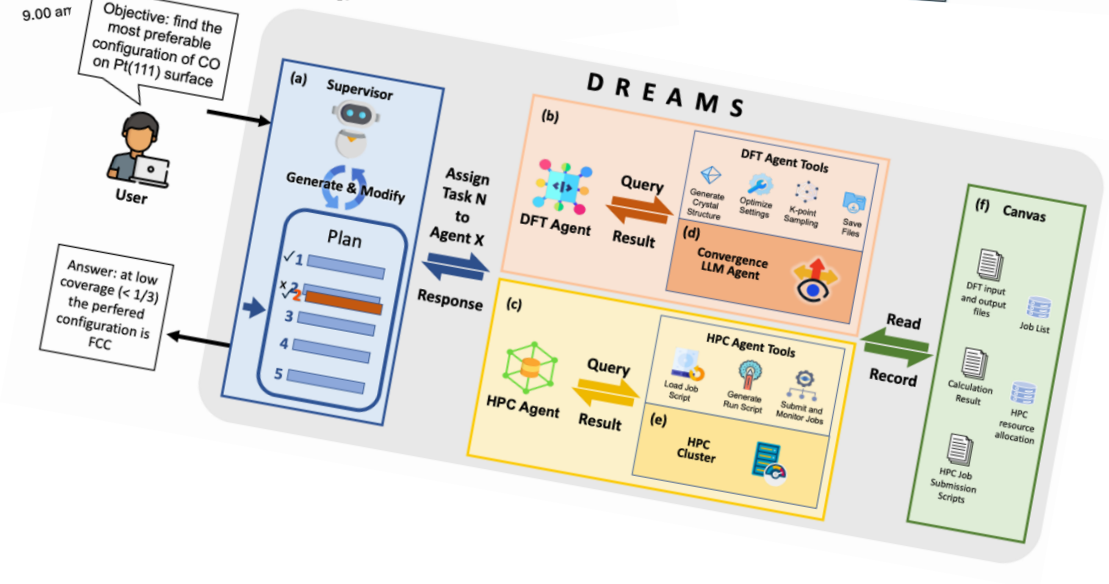
14.01.2026

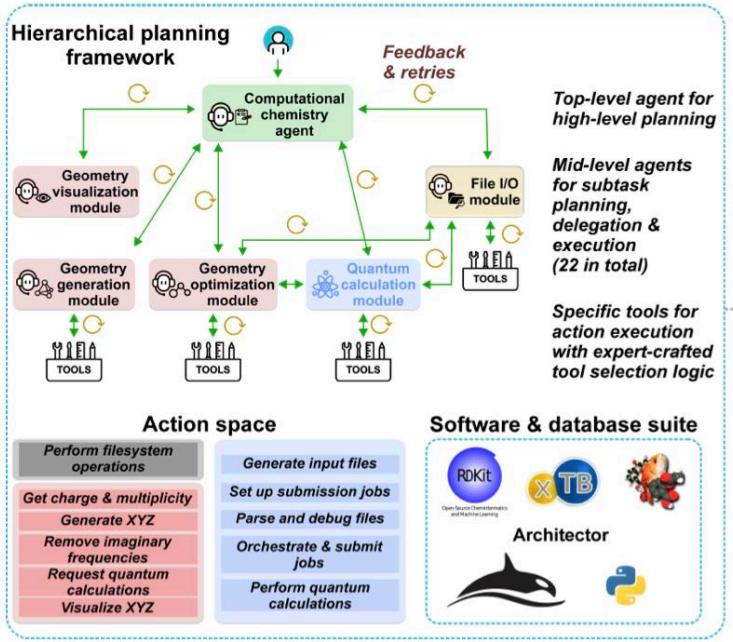


We hear a lot about LLM agents in chemistry

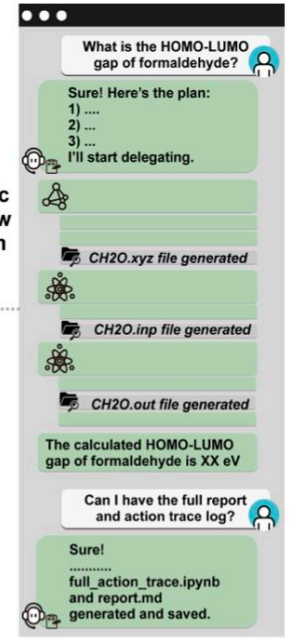








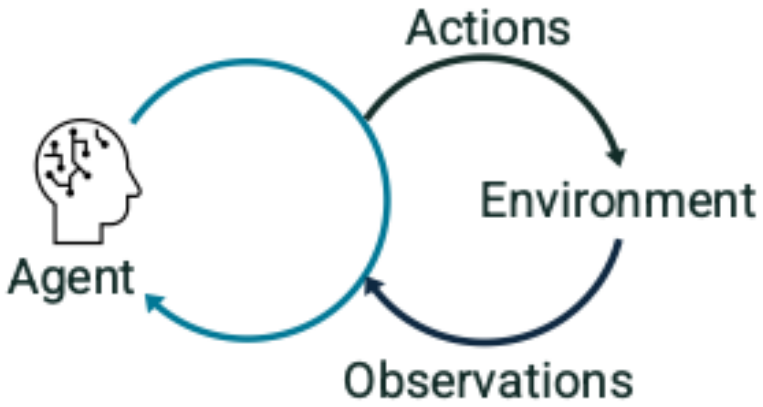
Dynamic workflow creation



Yunheng et al *Cell* 8, 7, 102263 **2025**
Pham et al *Commun. Chem.* **2026**
Wang et al *arxiv:2507.14267* **2025**

What can AI agents do?

Everything ?!

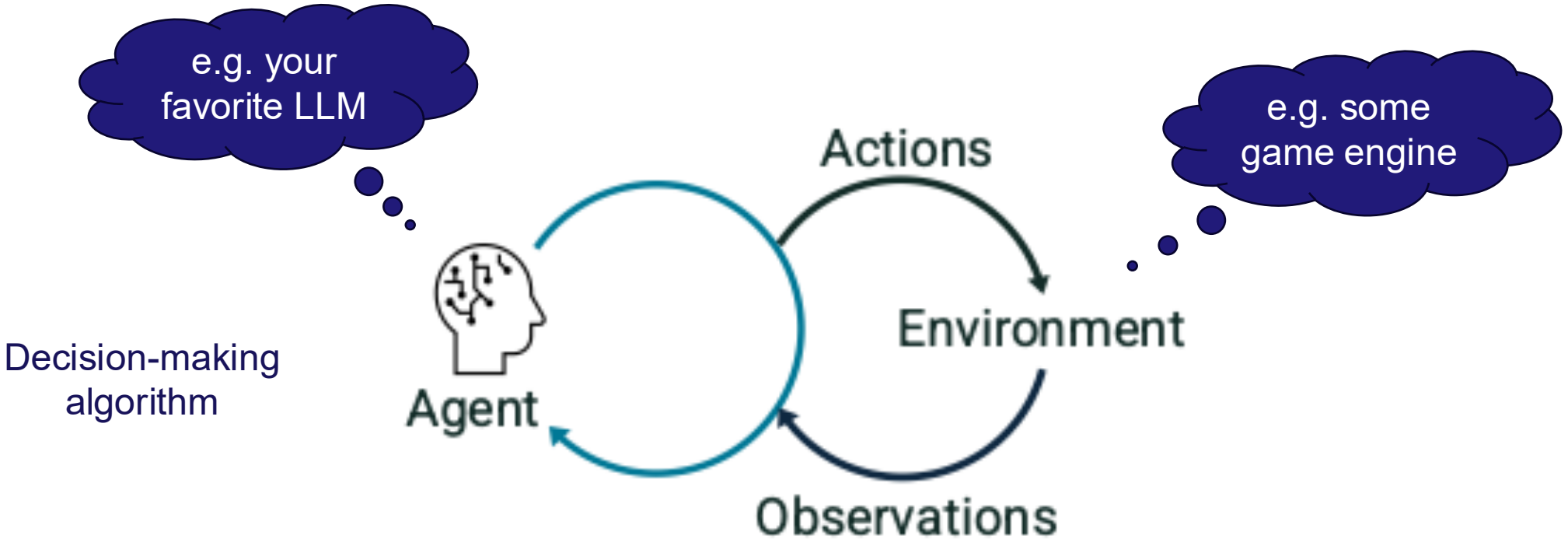


= Sequential decision process

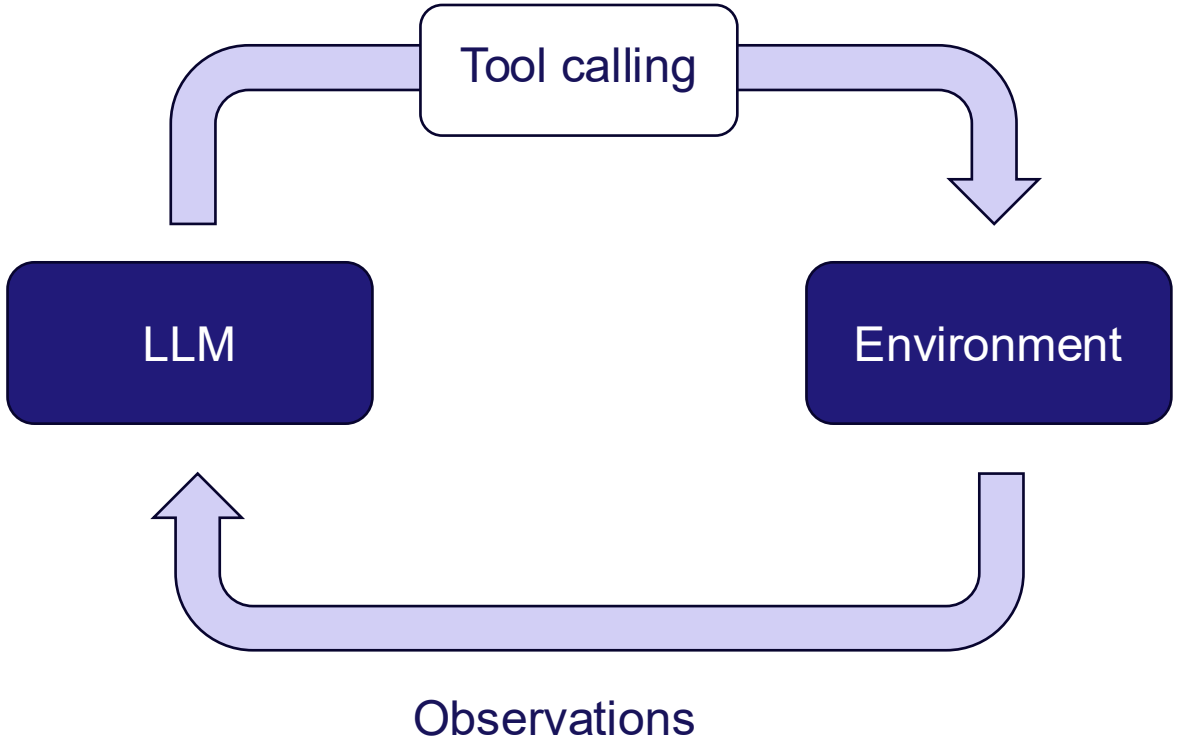
https://geekflare.com/blog/what-is-ai-agent/

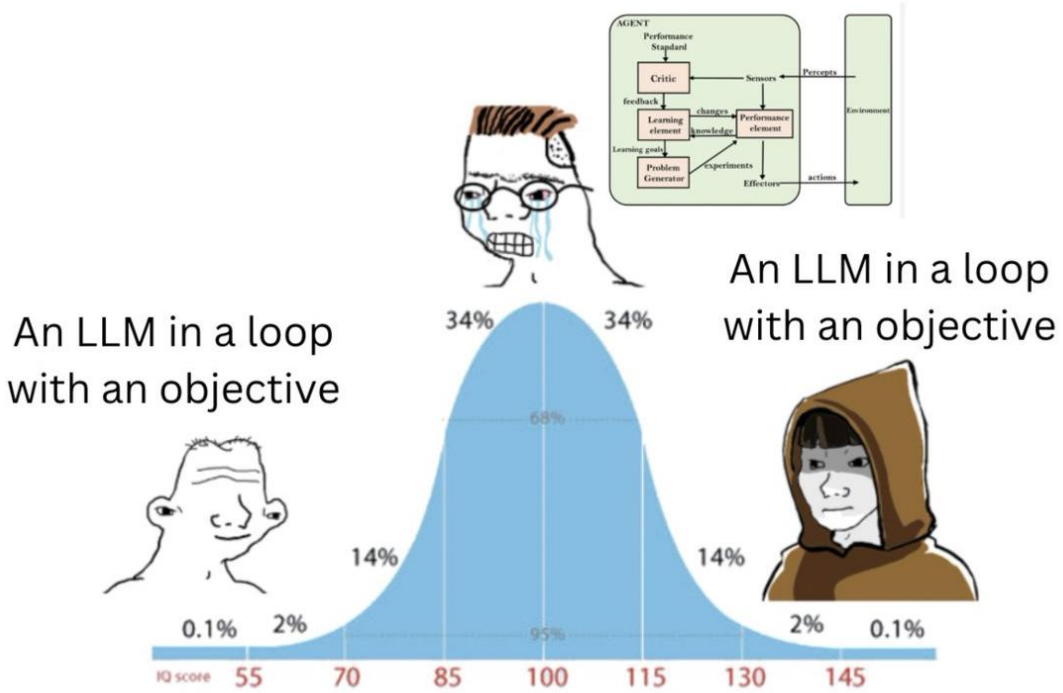
Krause et *al. Probabilistic Artificial Intellicence* **2025** arXiv:2502.05244v1

What is an AI agent?



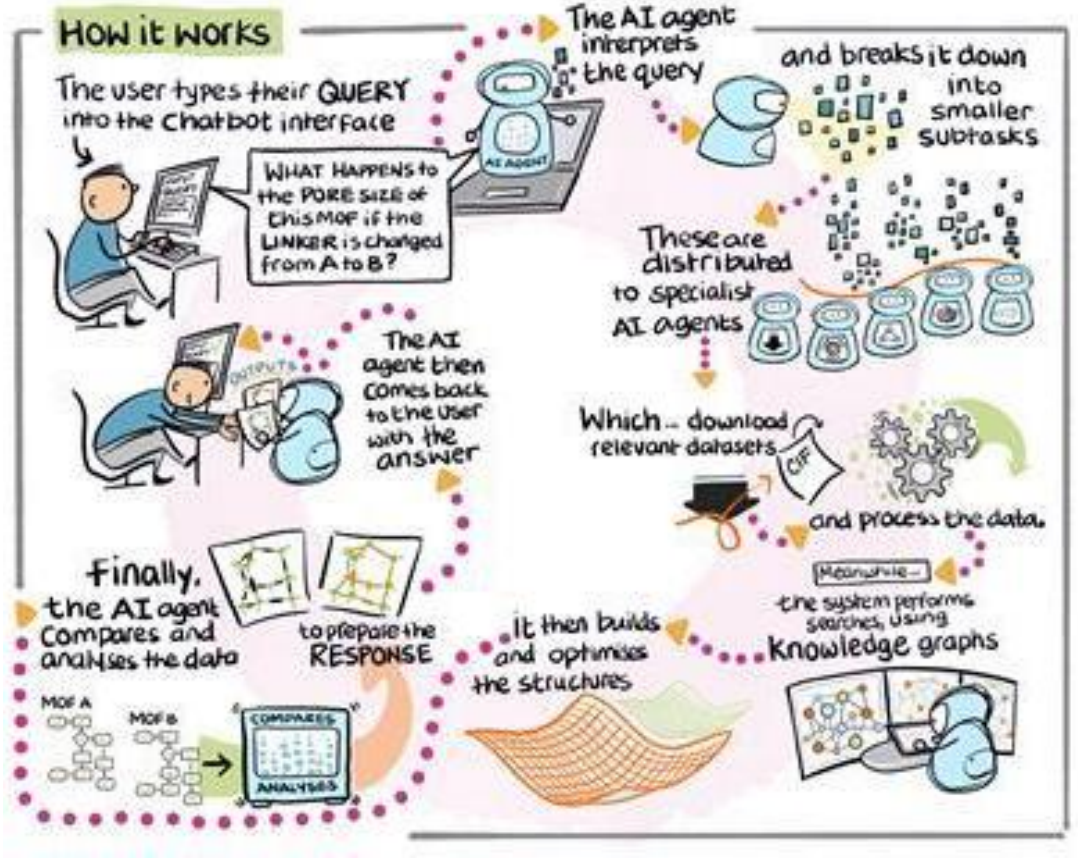
What is an AI agent?





Why are scientists interested in agents?

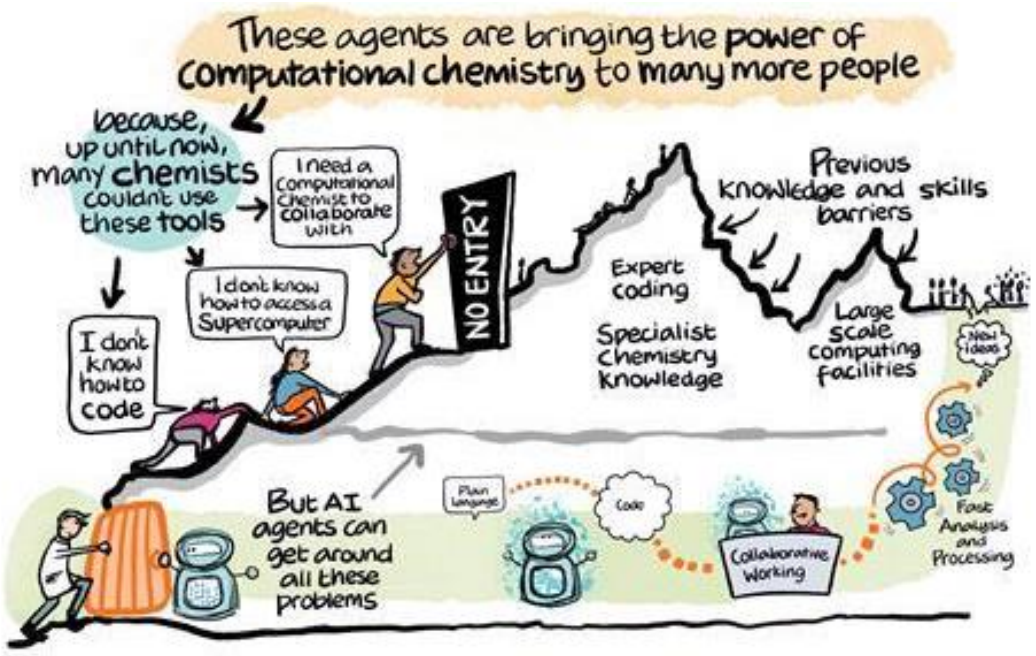
Automation of complex workflows



https://www.chemistryworld.com/news/ai-agents-set-to-democratise-computational-chemistry/4022465.article

Why are scientists interested in agents?

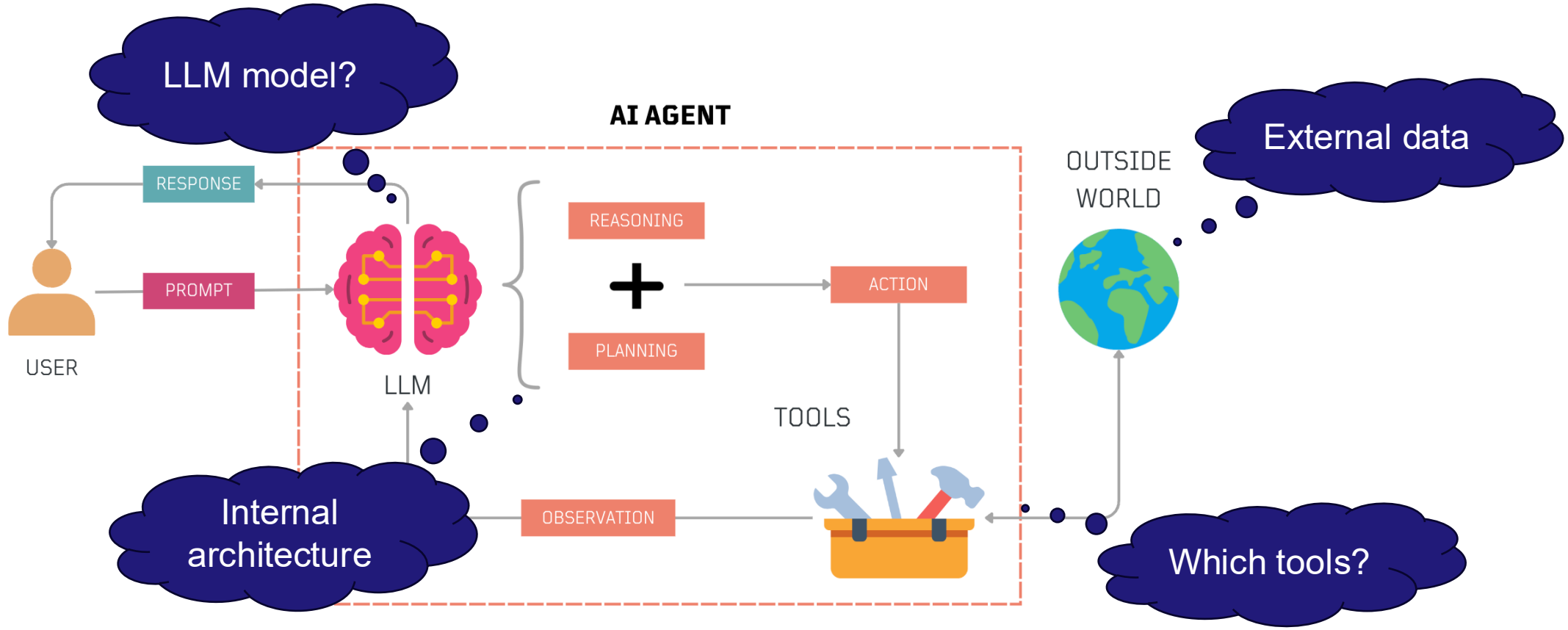
Accessibility



Democratization of scientific tools

https://www.chemistryworld.com/news/ai-agents-set-to-democratise-computational-chemistry/4022465.article

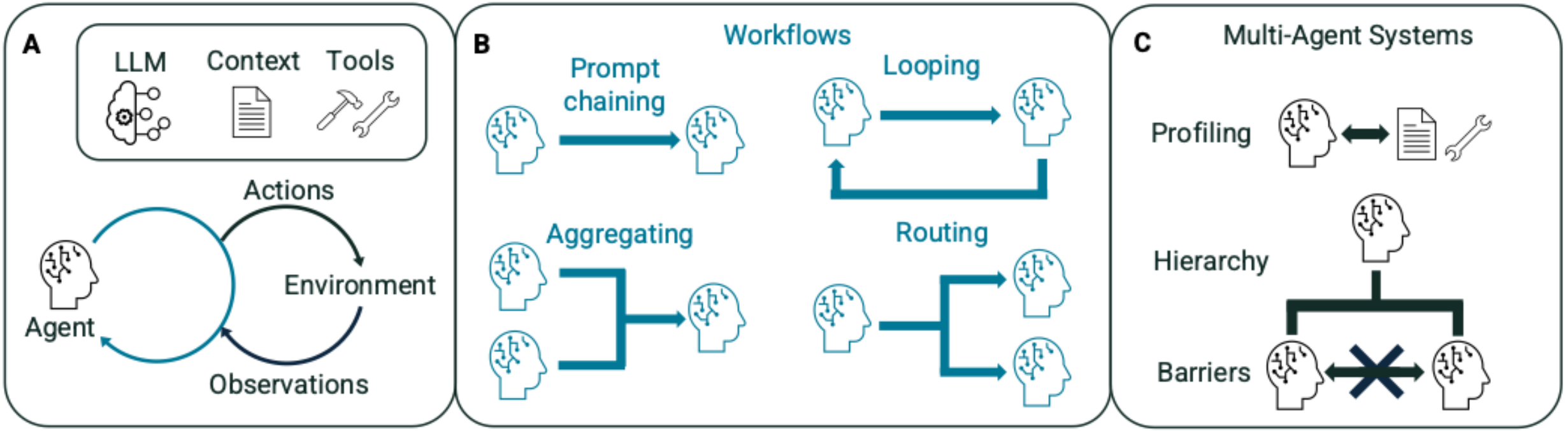
How to build a simple LLM from scratch?



LLM agent frameworks for agent design ➔ *smolagents*, *LangChain*, …

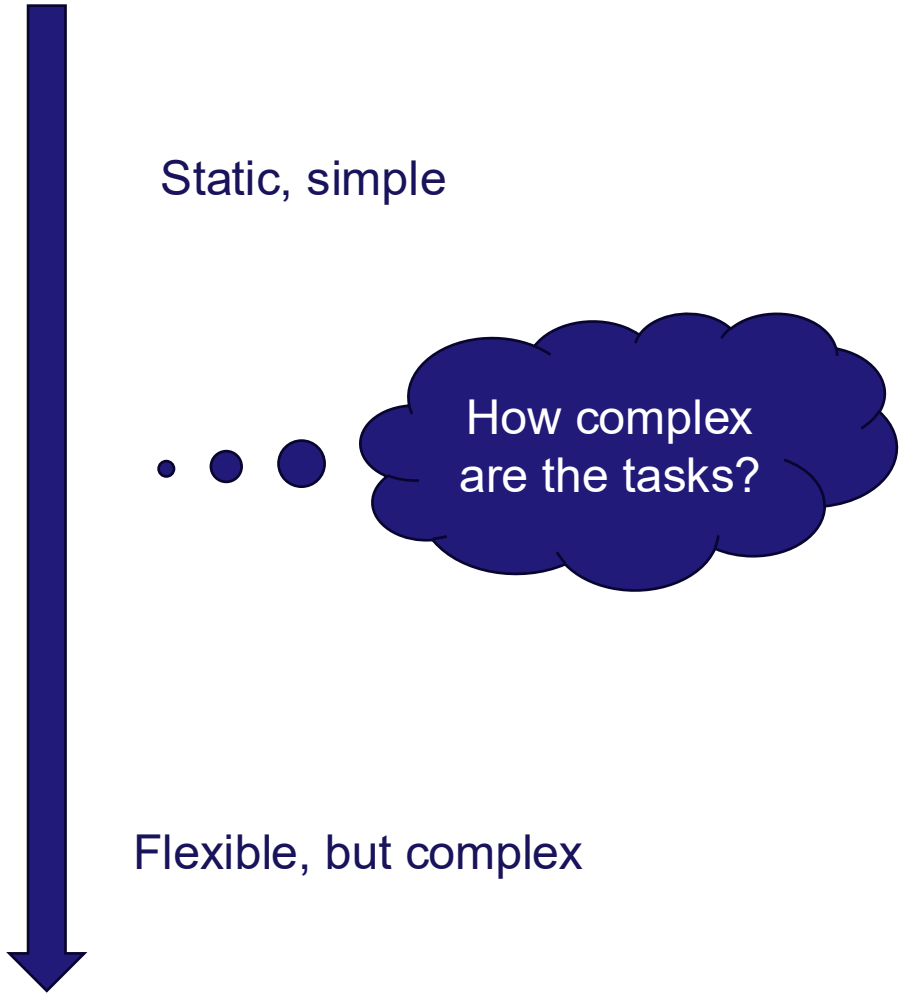
https://blog.openthreatresearch.com/content/images/2024/05/image-29.png

LLM Agent design patterns



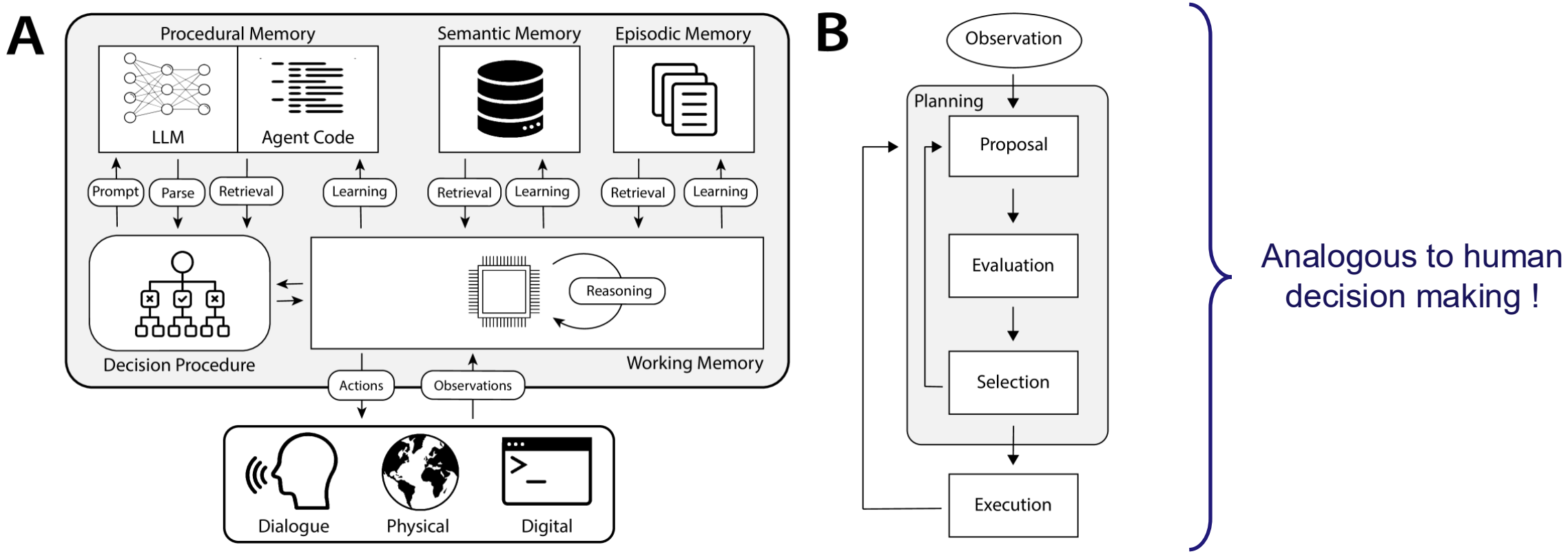
Agent Complexity

Agency Level	Description	Short name
☆☆☆	LLM output has no impact on program flow	Simple processor
★☆☆	LLM output controls an if/else switch	Router
★★☆	LLM output controls function execution	Tool call
★★☆	LLM output controls iteration and program continuation	Multi-step Agent
★★★	One agentic workflow can start another agentic workflow	Multi-Agent
★★★	LLM acts in code, can define its own tools / start other agents	Code Agents



LLM Agent Architectural Framework

Cognitive Architectures for Language Agents (CoALA)



Analogous to human decision making !

Common issues when designing agents

- Reproducibility: probabilistic LLMs ➔ non-reproducible results!
- Robustness: complex agentic systems ➔ complex failure modes
- Debugging: iterations of agent design and prompt engineering
- Cost: good LLMs are expensive – small LLMs require more time

➔ Benchmarking an agentic system is generally quite complex!

Let's build a simple comp. chem. agent!

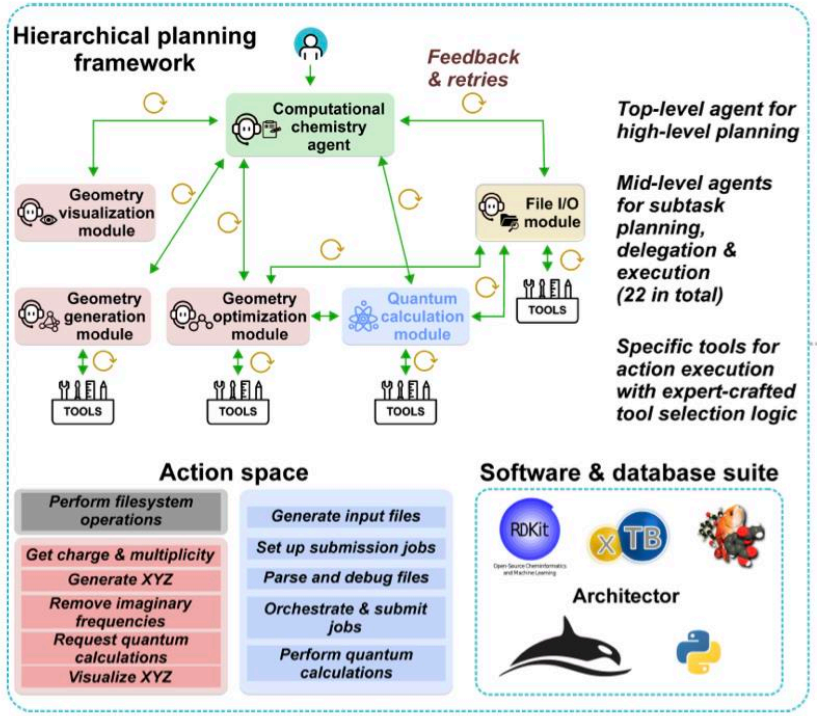
1. Task
2. How would a human solve this?
3. How complex is the task?
4. Which tools are required to solve this?
5. Choose architecture:
   - 1 vs n agents
   - Memory / external knowledge
   - Planning
   - base model
6. Testing + debugging

Done! But what is the advantage?

- accessibility
- automation

➔ more obvious for more complex tasks:

Ring strain of cycloalkanes	**Task:** Compute ΔG and ΔH of ring strain due to cyclization: Cyclo(C_nH_{2n}) → Cyclo($C_{n-1}H_{2n-3}$)-CH_3. **Systems:** The test set includes $n = 3 - 8$. **Anticipated challenges:** Correctly optimizing the geometry, keeping track of multiple files, and analyzing correctly.
pKa of common acids	**Task:** Compute the pKa of carboxylic acids, $H-A \longrightarrow H^+ + A^-$, with Gibbs energies computed using an implicit solvent. The proton solvation energy may be calibrated against literature pKa values of related carboxylic acids. **Systems:** This is tested for either acetic or chlorofluoroacetic acid. **Anticipated challenges:** Keeping track of multiple files, long-term planning of multiple steps.
Absorption spectrum of organic molecules	**Task:** Calculate the singlet (including the oscillator strength) and the singlet-triplet gap of vertical excitation energy of organic molecules with time-dependent density functional theory. The geometries are provided for the agent. **Systems:** The test set includes three organic molecules. **Anticipated challenges:** Reporting the summary of the results in a table.



Practical Part