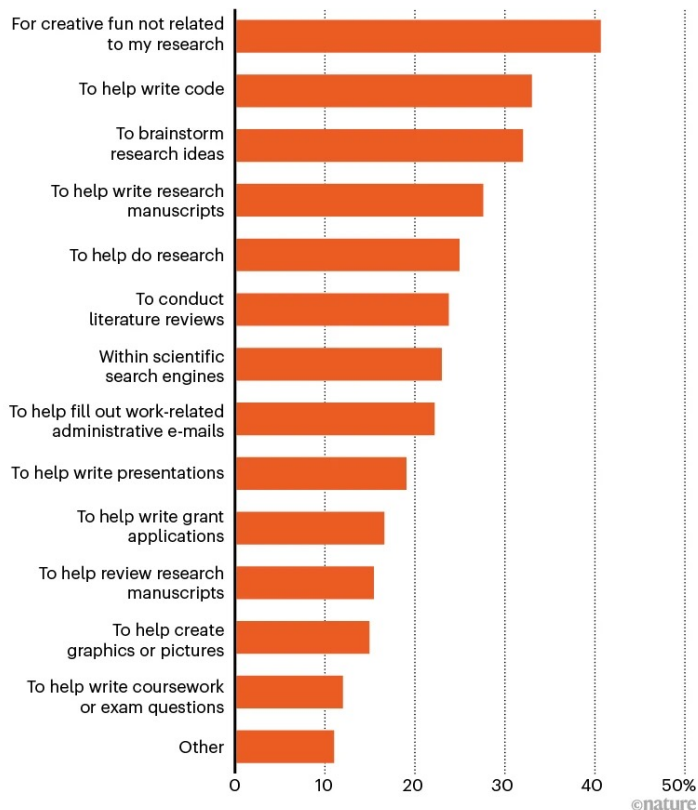


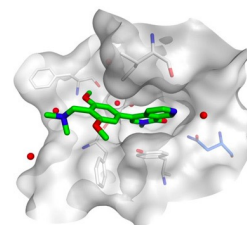
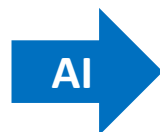
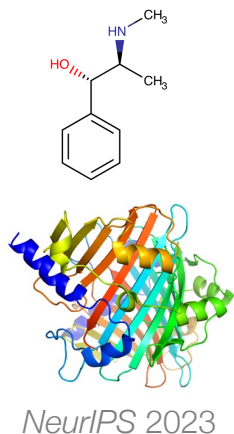
# Medicine in the age of generative AI

## HOW RESEARCHERS USE LARGE LANGUAGE MODELS

Q: What do you use generative AI tools (such as ChatGPT and other large language models) for? (Choose all that apply.)

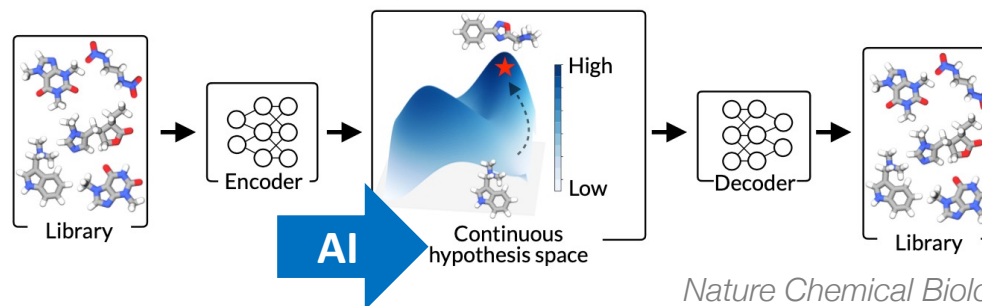


**Generative AI is changing the way science is done**



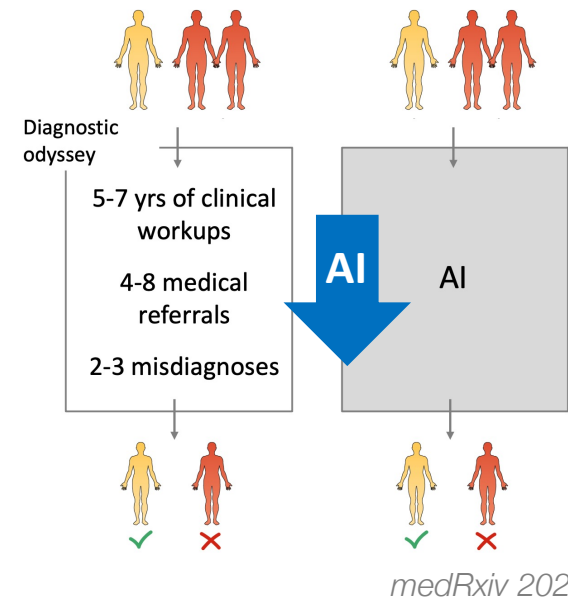
Binding of novel drugs to therapeutic targets

## Generative AI to design biomolecules



**AI is used to augment research, providing insights that might not have been possible using traditional methods alone**

## Therapeutic matching



# The era of AI-generated content

**You**

Please give me some suggestions on a good presentation

**ChatGPT**

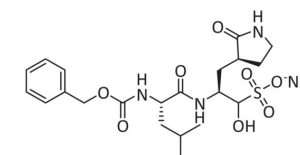
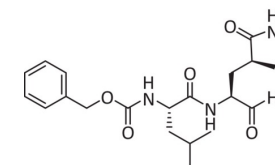
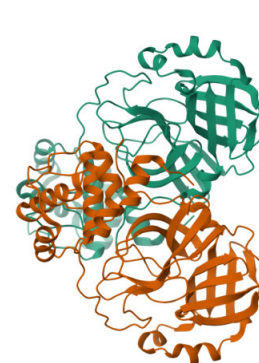
Certainly! Here are some specific suggestions for creating a good presentation:

1. Start with a Strong Opening:
  - Begin your presentation with a compelling hook or an intriguing question to grab your audience's attention from the start.
  - Share a relevant quote, anecdote, or surprising fact to pique curiosity.
2. Clearly Define Your Main Message:
  - State your key message or takeaway early in your presentation, so your audience knows what to expect.

“ A cat and a female character in a spaceship exploring a hidden galaxy. With detailed backgrounds, expressive characters, including magical elements, illustration made by hand. ”



***Can we transform drug discovery with Generative AI?***



⋮

“Generate drug candidates for the SARS-CoV-2 3CL protease.”

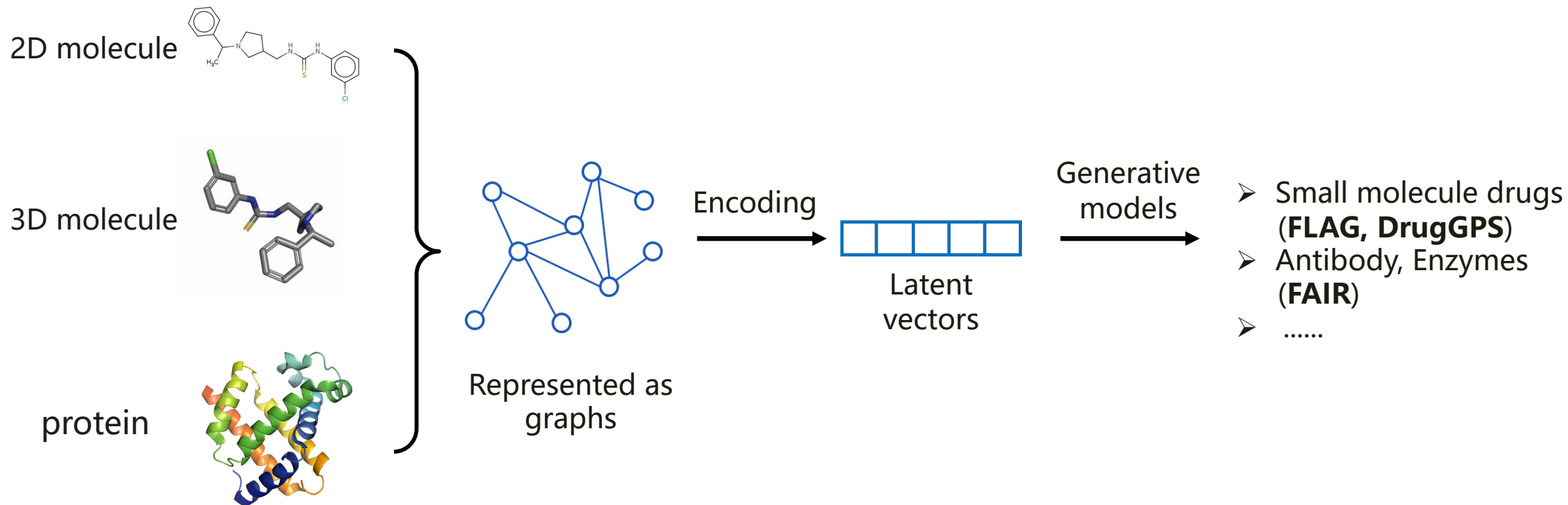


# Generative AI models for drug design



# Generative models for molecules and proteins

- In structure-based drug discovery, **molecule/protein** data can be represented as **graphs**
- **Graph generative models** typically encode graphs into latent vectors for downstream tasks

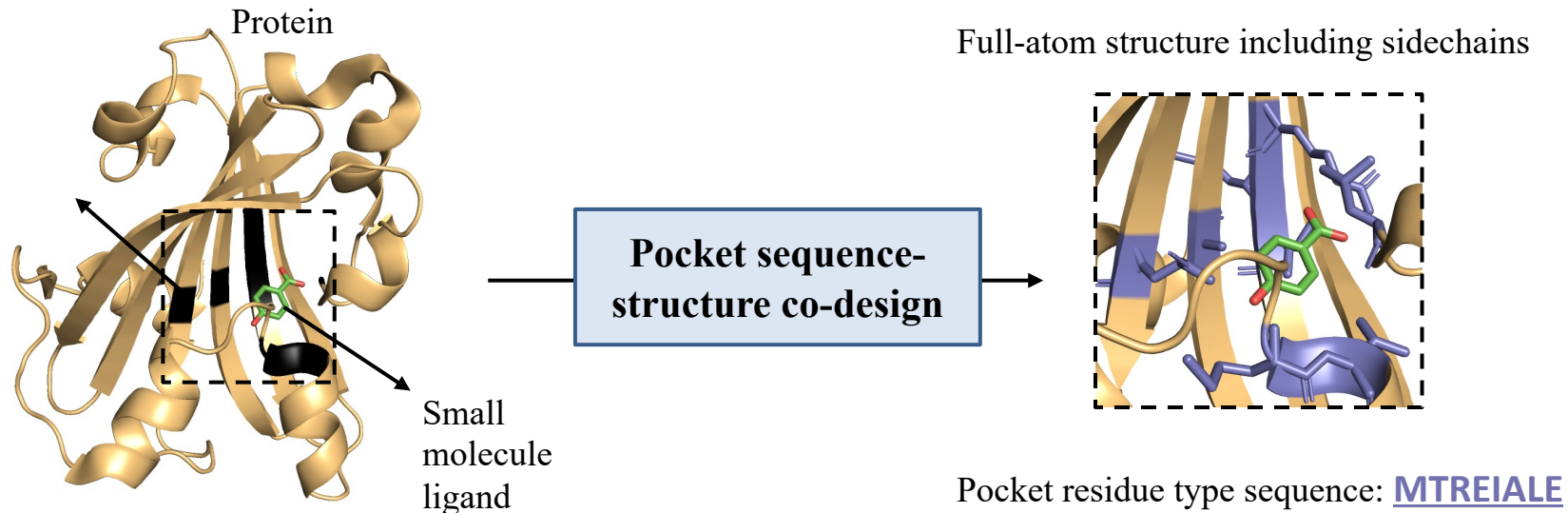


# Early success of generative AI in drug design

Treatment	Organization	Description	Phase	Lead indication
REC-2282	Recursion	Small molecule pan-HDAC inhibitor	2/3	Neurofibromatosis type 2
REC-994	Recursion	Small molecule superoxide scavenger	2	Cerebral cavernous malformation
REC-4881	Recursion	Small molecule inhibitor of MEK1 and MEK2	2	Familial adenomatous polyposis
INS018_055	InSilico Medicine	Small molecule inhibitor	2	Idiopathic pulmonary fibrosis
BEN-2293	BenevolentAI	Topical pan-tyrosine kinase inhibitor	2a	Atopic dermatitis
EXS-21546	Exscientia and Evotec	A <sub>2A</sub> receptor antagonist	1b/2	Solid tumors carrying high adenosine signatures.
RLY-4008	Relay Therapeutics	Inhibitor of FGFR2	1/2	FGFR2-altered cholangiocarcinoma
EXS-4318	Exscientia	PKC- $\theta$ inhibitor	1/2	Inflammatory and autoimmune conditions
BEN-8744	BenevolentAI	Small molecule PDE10 inhibitor	1	Ulcerative colitis
Undisclosed	Recursion	Small molecular inhibitor of RBM39, a CDK12-associated protein	Pre-clinical	HRD-negative ovarian cancer

# Generative AI for protein design

- Designing **proteins** that bind to small molecules is a cornerstone problem in drug development and bioengineering
- **Question:** How to design **protein pockets as areas** that are bound by a ligand?
- **Approach:** Learn a generative model that generates both **amino acid sequence and full atom 3D structure** from the ligand and the protein scaffold

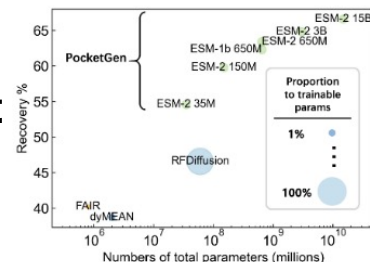




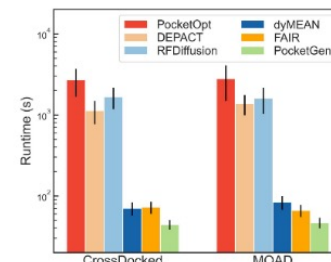
# Generative sequence-structure models enable atom-level predictions of ligands binding to biological targets

Generative models:

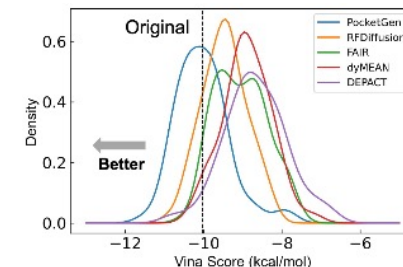
- **atom-level co-design of protein pocket sequence and 3D structure**
- **selective small molecule ligands**
- **optimized PPI interfaces**



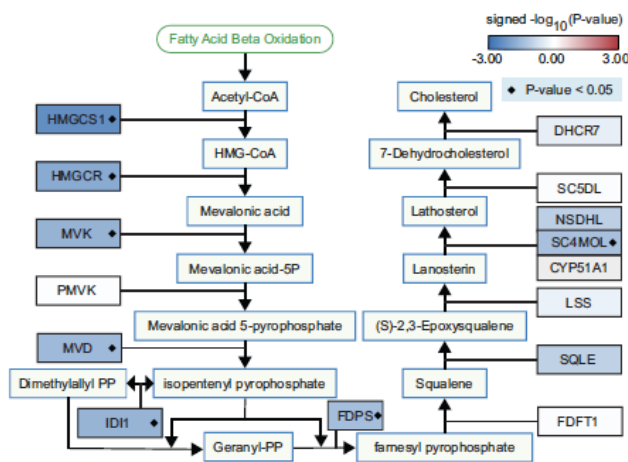
Iterative refinement based on side-chain effects, ligand flexibility, and sequence-structure consistency



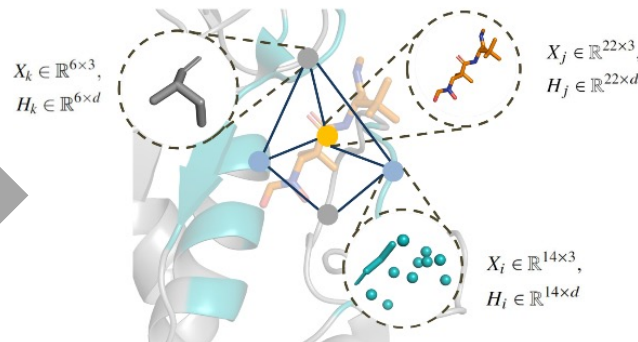
10x faster than current AI, 15% better accuracy (AAR, RMSE, docking score)



45% better hit rate than current AI, need to generate fewer molecules to find a hit



Key disease mechanisms, shared effects, interaction effects



Unified Representation

Generative sequence-structure models

Molecular property predictors

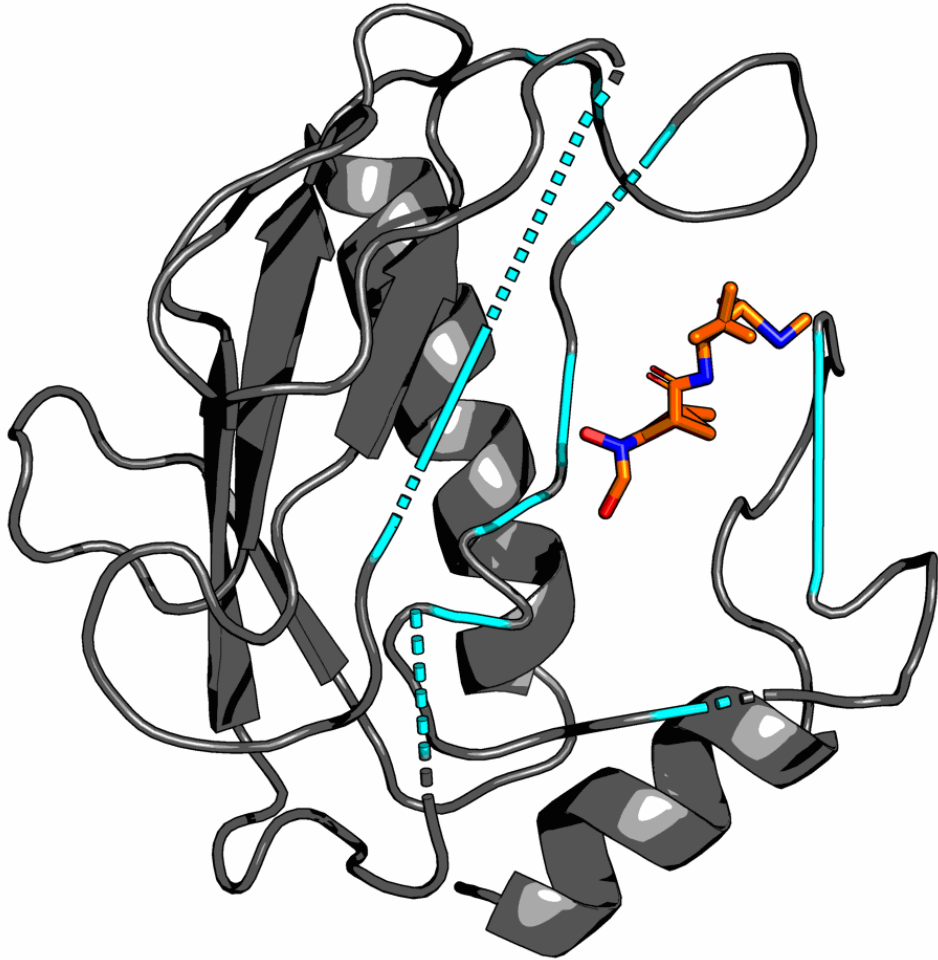
- LSS-MSMO1 binding
- HMGCR-targeted
- PROTAC comprising a VHL ligand

Predictions



Priority lists of generated molecular structures

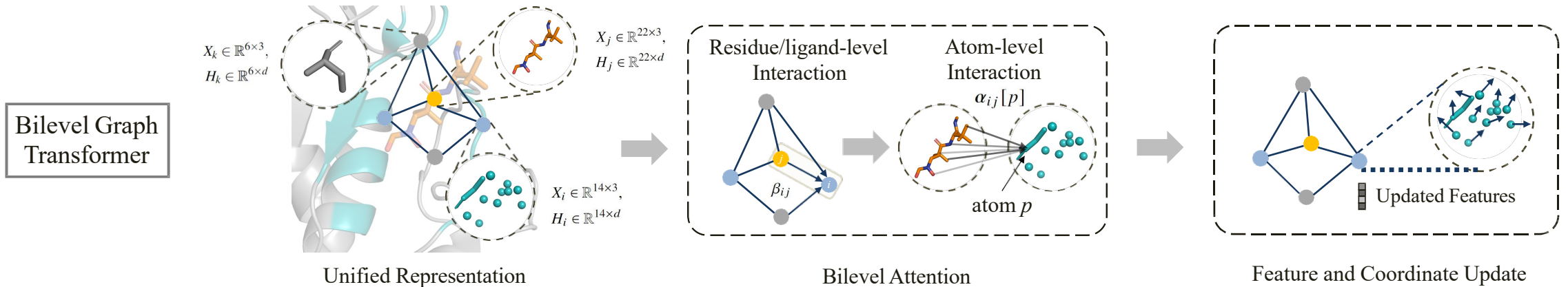
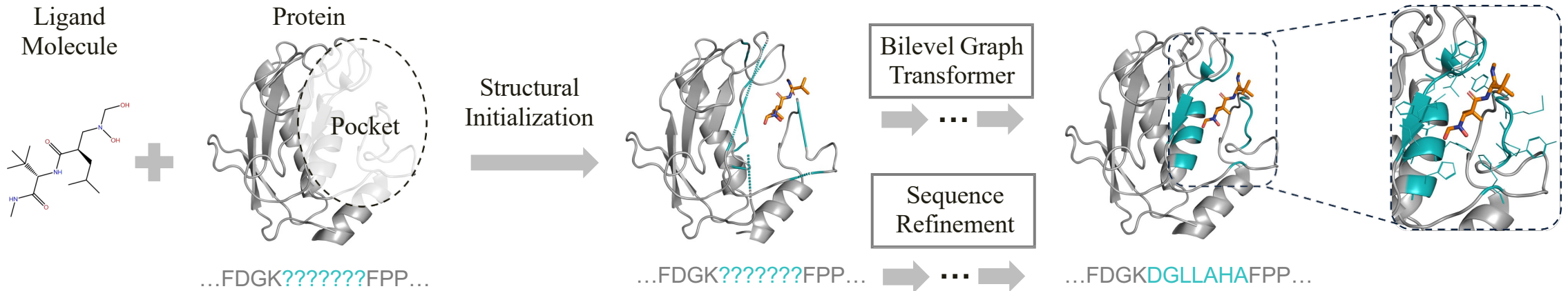
# Sequence-structure co-generation of protein pockets



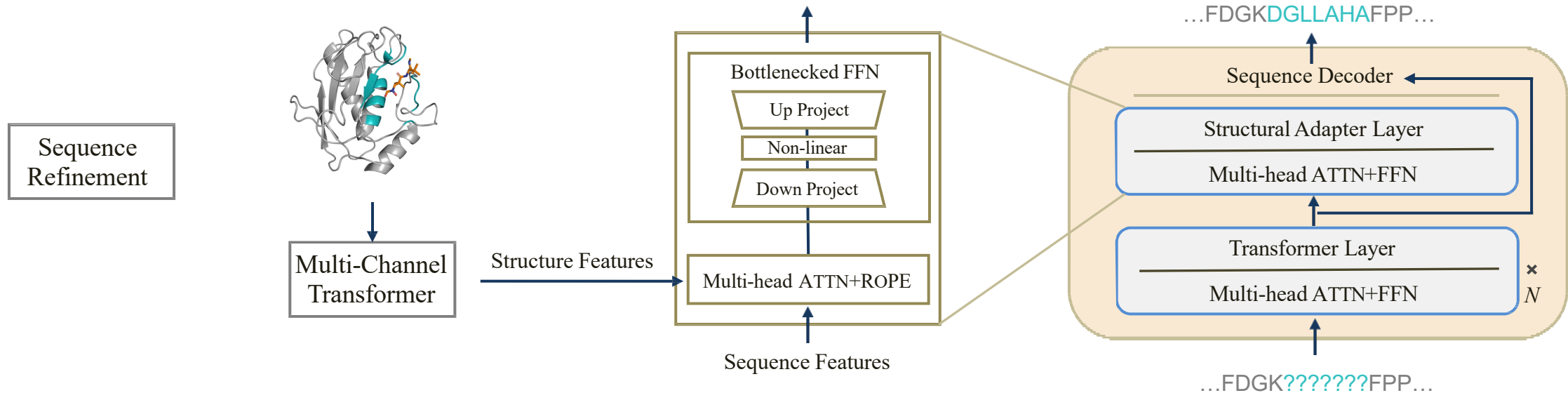
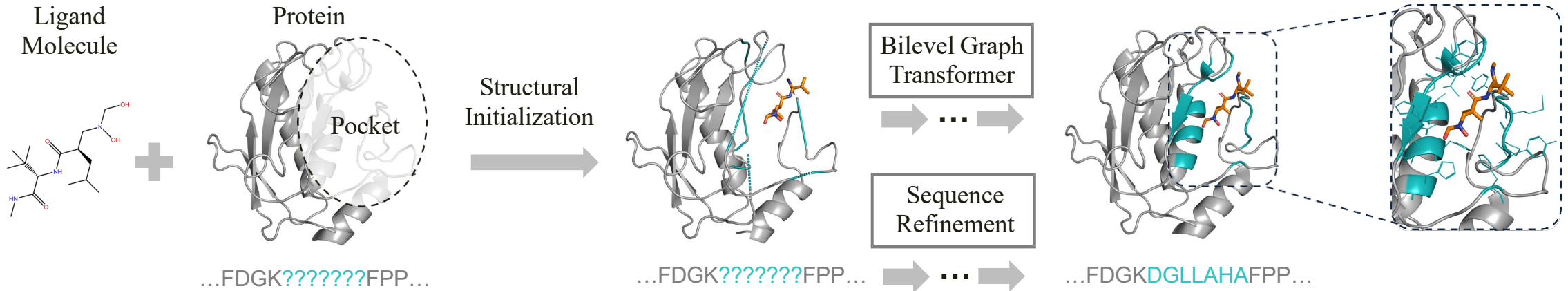
- Generating **high-fidelity protein pockets** — an area where a protein interacts with a ligand molecule
  - Complex interactions between ligand molecules and proteins
  - Flexibility of ligands and AA side chains
  - Complex sequence-structure dependencies
- PocketGen generates residue sequence and full-atom structure within protein pocket region



# Iterative refinement of both sequence and structure in the protein pocket to maximize binding affinity with small molecule ligand



# Iterative refinement of both sequence and structure in the protein pocket to maximize binding affinity with small molecule ligand

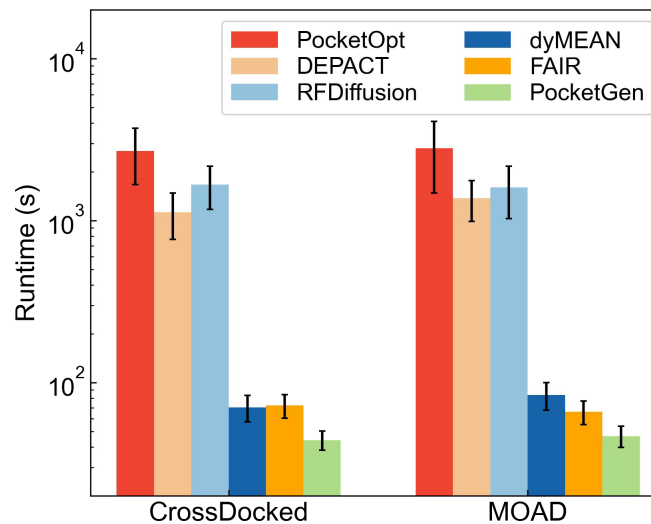


# PocketGen generates protein pockets with higher binding affinity and structural validity than existing models

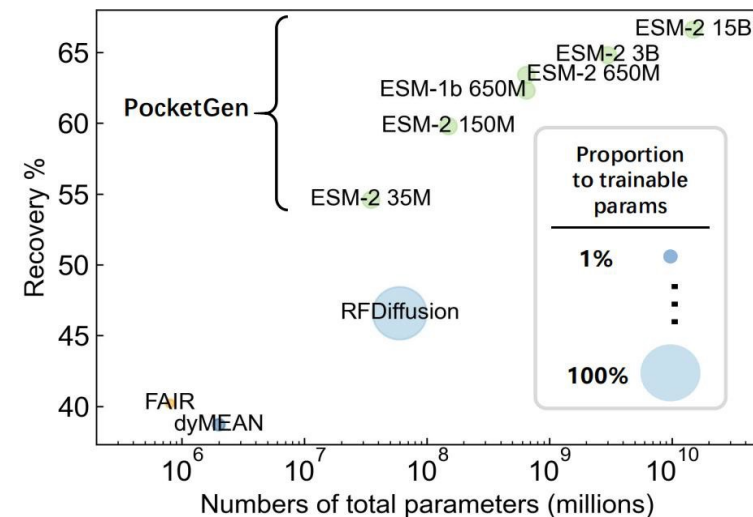
	PocketOpt	DEPACT	dyMEAN	FAIR	RFDiffusion	PocketGen
Top-1 generated protein pocket						
Vina score (↓)	-9.216	-8.527	-8.540	-8.792	-9.037	<b>-9.655</b>
Success Rate (↑)	0.92	0.75	0.76	0.80	0.89	<b>0.97</b>
RMSD (↓)	-	1.47	1.44	1.39	<b>1.13</b>	1.21
pLDDT (↑)	-	82.1	83.3	83.2	84.5	<b>86.7</b>
scTM (↑)	-	0.901	0.906	0.899	0.924	<b>0.937</b>
Top-3 generated protein pockets						
Vina score (↓)	-8.878	-8.131	-8.196	-8.321	-8.876	<b>-9.353</b>
RMSD (↓)	-	1.45	1.43	1.40	<b>1.18</b>	1.24
pLDDT (↑)	-	81.9	82.8	83.1	84.6	<b>86.2</b>
scTM (↑)	-	0.896	0.892	0.897	0.929	<b>0.934</b>
Top-5 generated protein pockets						
Vina score (↓)	-8.702	-7.786	-7.974	-7.943	-8.510	<b>-9.239</b>
RMSD (↓)	-	1.46	1.45	1.42	1.25	<b>1.22</b>
pLDDT (↑)	-	82.2	82.9	83.3	84.3	<b>86.1</b>
scTM (↑)	-	0.892	0.903	0.886	0.926	<b>0.935</b>
Top-10 generated protein pockets						
Vina score (↓)	-8.556	-7.681	-7.690	-7.785	-8.352	<b>-9.065</b>
RMSD (↓)	-	1.53	1.44	1.41	<b>1.26</b>	1.28
pLDDT (↑)	-	81.5	82.7	83.0	84.2	<b>85.9</b>
scTM (↑)	-	0.895	0.896	0.884	0.924	<b>0.931</b>

Improved structural validity, amino acid sequence recovery, and affinity with target ligands

Model	CrossDocked			Binding MOAD		
	AAR (↑)	RMSD (↓)	Vina (↓)	AAR (↑)	RMSD (↓)	Vina (↓)
Test set	-	-	-7.016	-	-	-8.076
DEPACT	31.52±3.26%	1.59±0.13	-6.632±0.18	35.30±2.19%	1.52±0.12	-7.571±0.15
dyMEAN	38.71±2.16%	1.57±0.09	-6.855±0.06	41.22±1.40%	1.53±0.08	-7.675±0.09
FAIR	40.16±1.17%	1.46±0.04	-7.015±0.12	43.68±0.92%	1.37±0.07	-7.930±0.15
RFDiffusion	46.57±2.07%	1.44±0.07	-6.936±0.07	45.31±2.73%	1.45±0.10	-7.942±0.14
PocketGen	<b>63.40±1.64%</b>	<b>1.36±0.05</b>	<b>-7.135±0.08</b>	<b>64.43±2.35%</b>	<b>1.32±0.05</b>	<b>-8.112±0.14</b>



Better generation efficiency



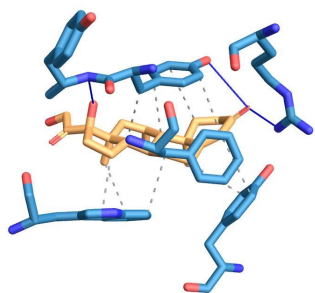
Performance wrt protein LM size

# PocketGen can redesign pockets of antibodies, enzymes, and biosensors for target ligand molecules

- Protein
- Ligand
- Aromatic Ring Center
- ⋯ Hydrophobic Interaction
- Hydrogen Bond
- ⋯  $\pi$ -Stacking (parallel)
- ⋯  $\pi$ -Stacking (perpendicular)
- ⋯  $\pi$ -Cation Interaction

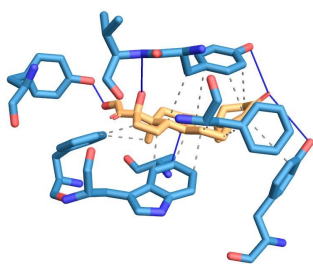
## Cortisol (HCY)

...WFRYYDTMY...



Original  
(HP 12, HB 3)

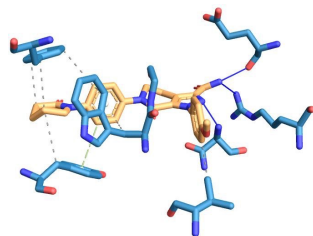
...WFRYVYSFY...



PocketGen  
(HP 12, HB 5)

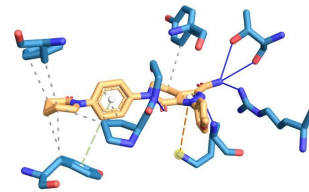
## Apixaban (APX)

...YREFQVWGG...



Original  
(HP 7, HB 4,  $\pi$  1)

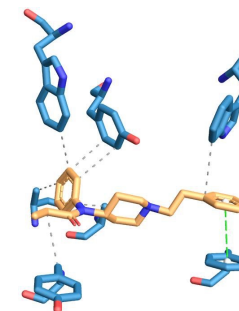
...YRTFKVPGY...



PocketGen  
(HP 9, HB 5,  $\pi$  2)

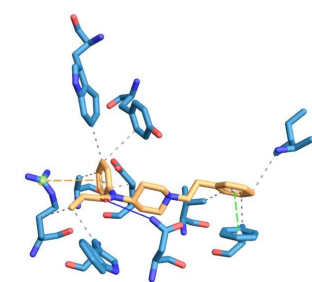
## Fentanyl (7V7)

...FVWAWYWAY...

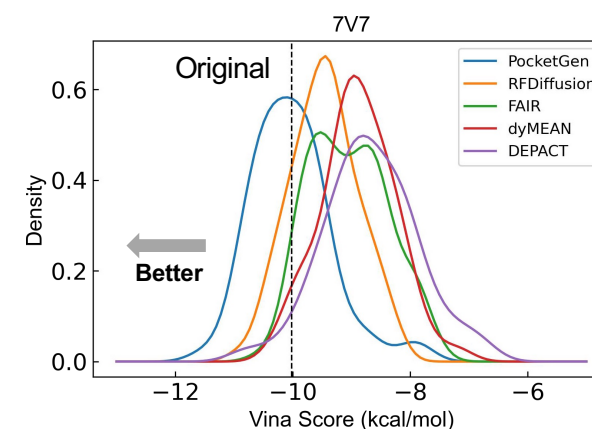
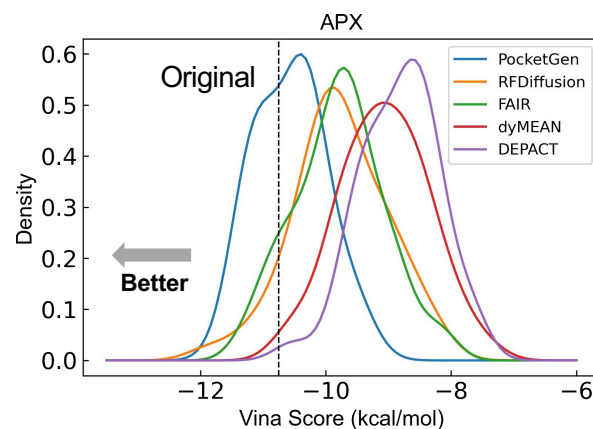
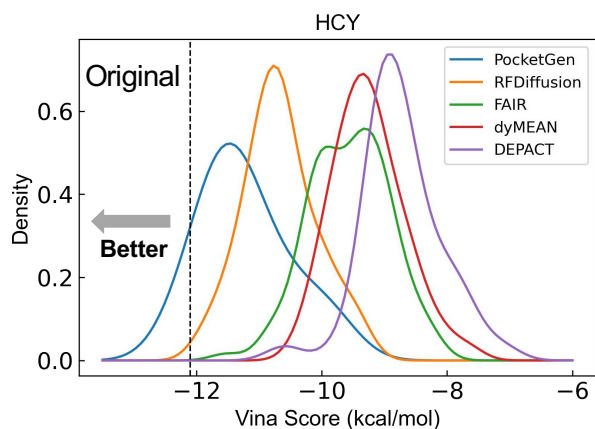


Original  
(HP 7, HB 0,  $\pi$  1)

...FVRAWYWEW...



PocketGen  
(HP 9, HB 1,  $\pi$  2)





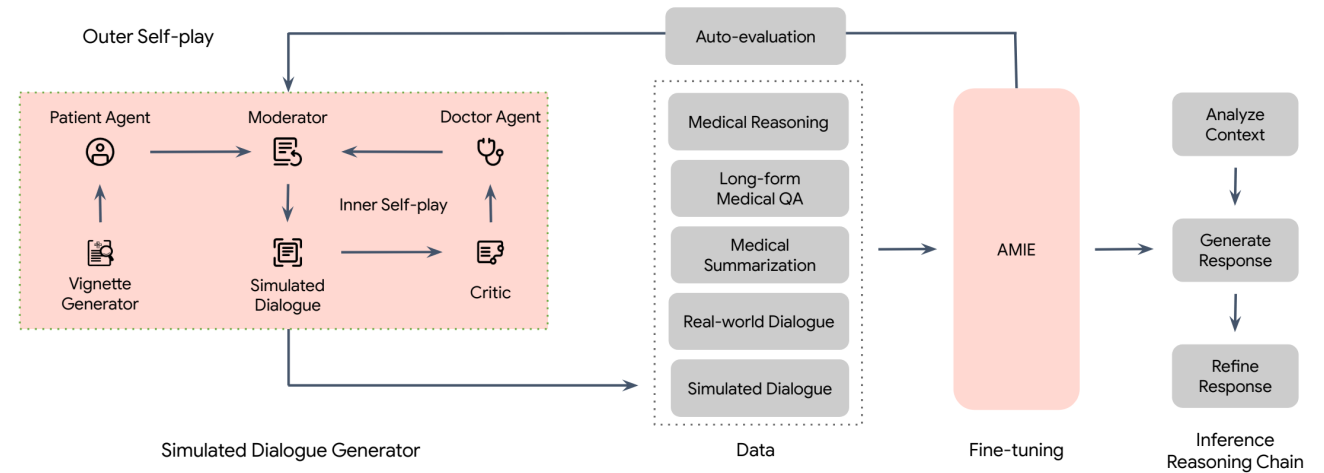


# Generative AI agents

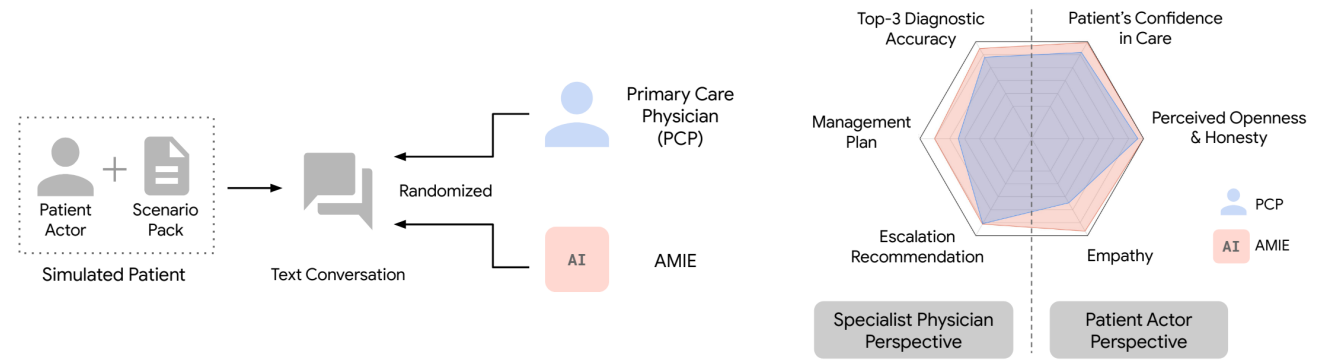


# “AI doctors””: Conversational medical AI optimized for diagnostic dialogue

At the heart of medicine lies the physician-patient dialogue, where skillful history-taking paves the way for accurate diagnosis, effective management, and enduring trust. AI systems capable of diagnostic dialogue could increase accessibility, consistency, and quality of care. However, approximating clinicians’ expertise is an outstanding grand challenge



AMIE System Design



Randomized Study Design for Remote Objective Structured Clinical Examination (OSCE)

AMIE Outperforms PCPs on Multiple Evaluation Axes for Diagnostic Dialogue

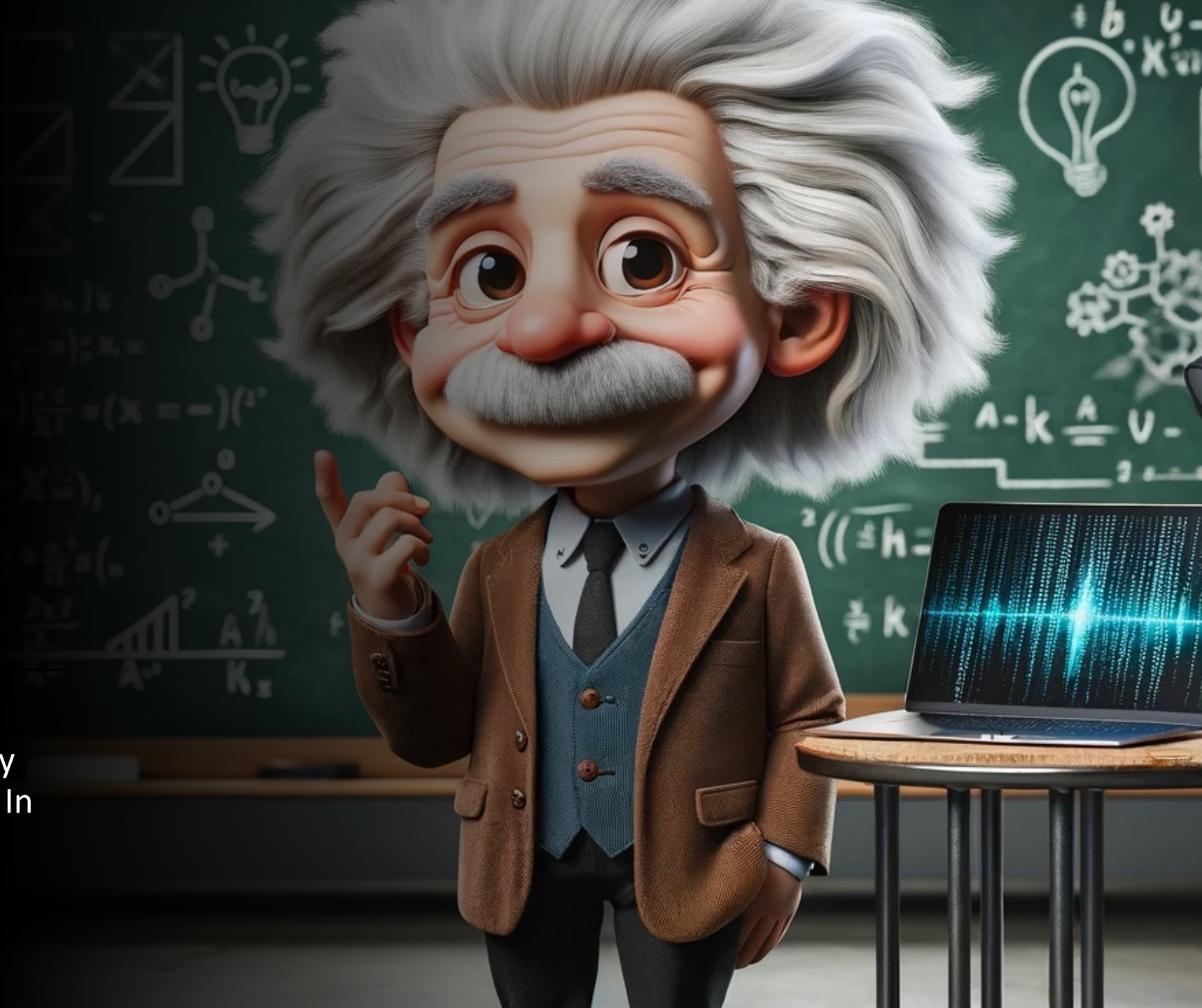




---

# “AI scientist” capable of skeptical learning and reasoning

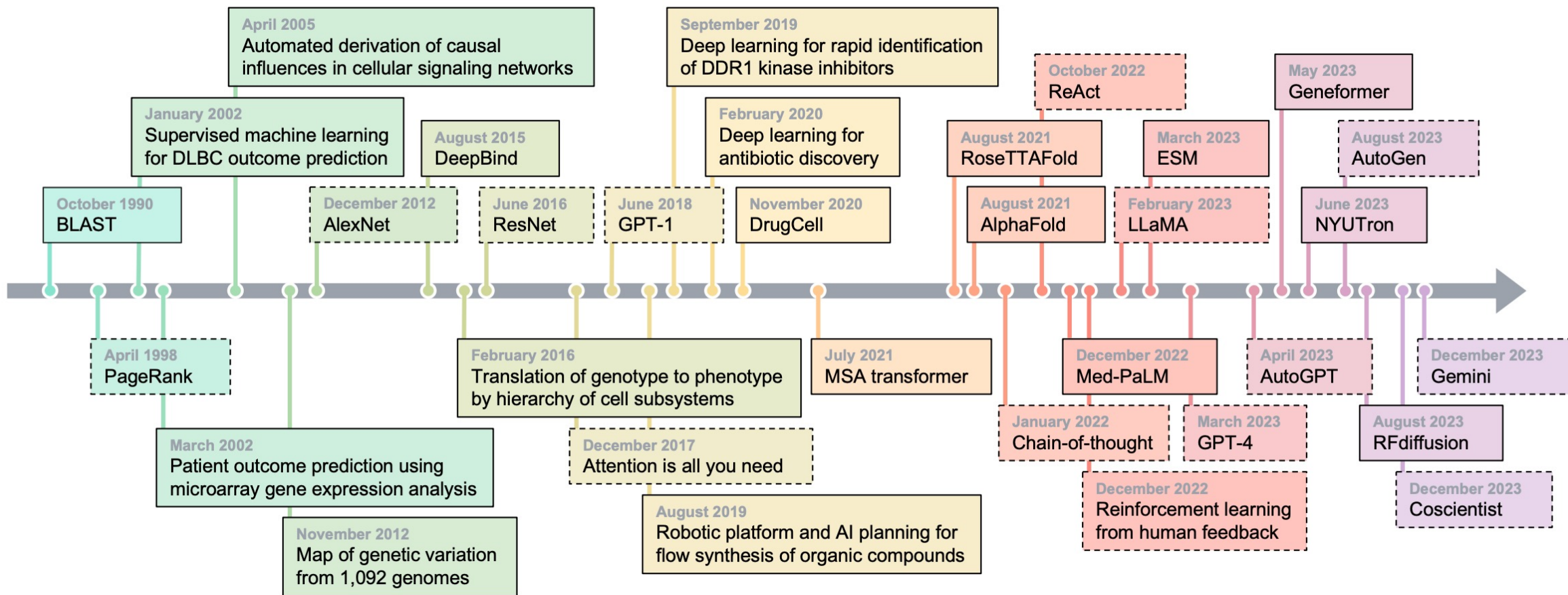
Empowering Biomedical Discovery  
with AI Agents arXiv:2404.02831, In  
review



# Evolving use of data-driven models

## Databases and search engines

## Interactive and foundation learning models

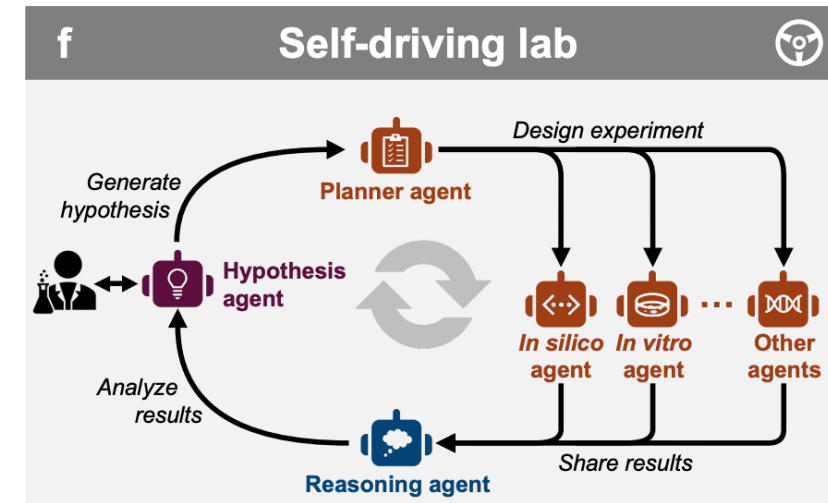
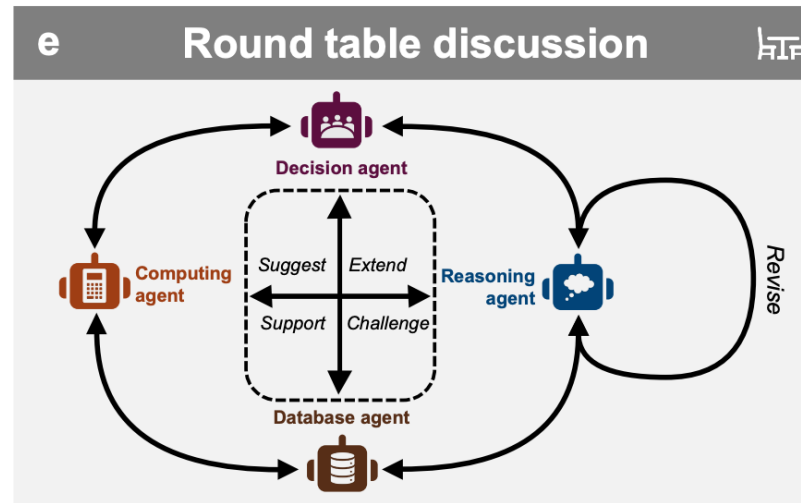
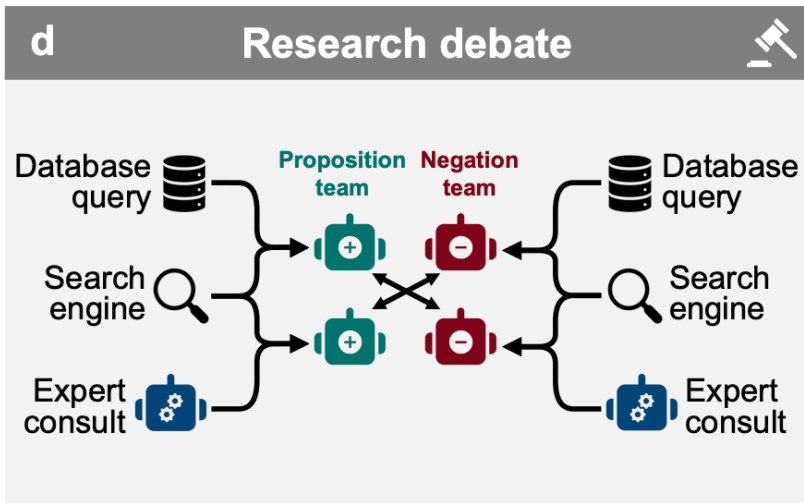
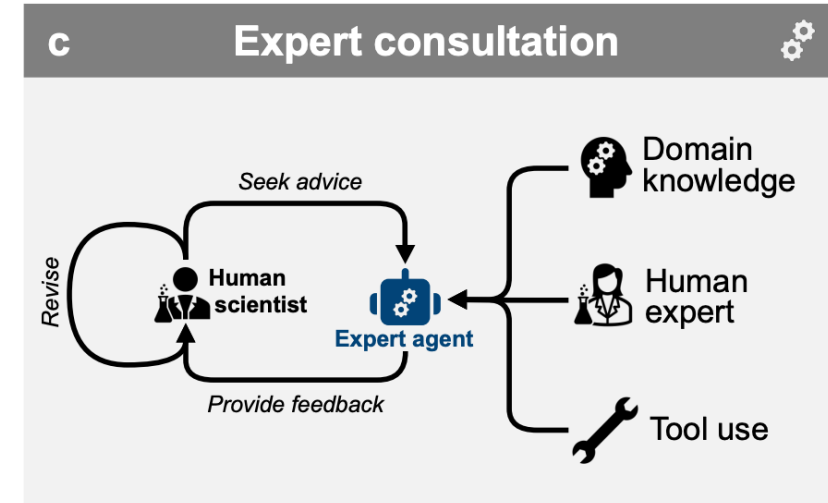
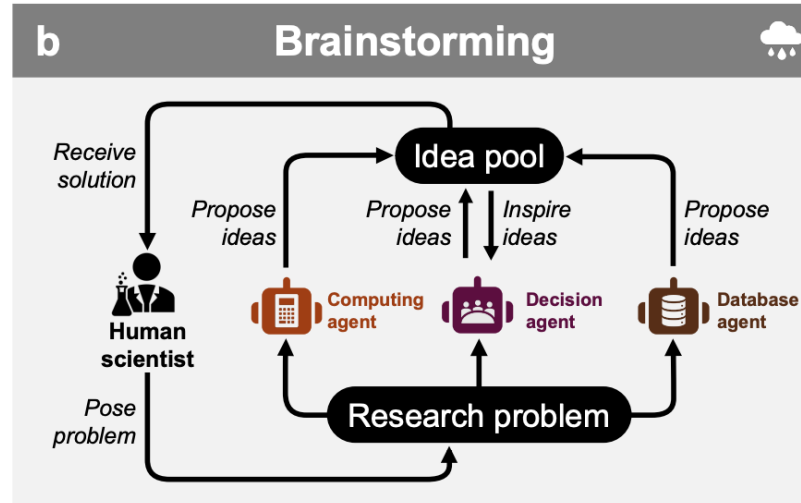
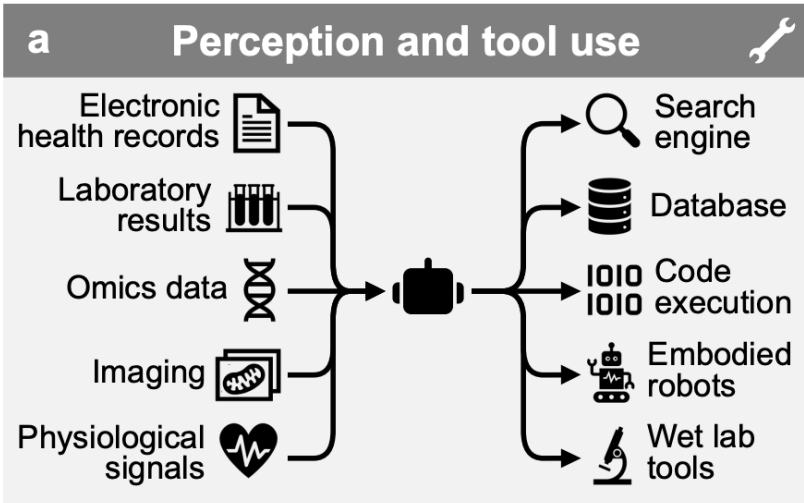


## Machine learning models

## AI agents

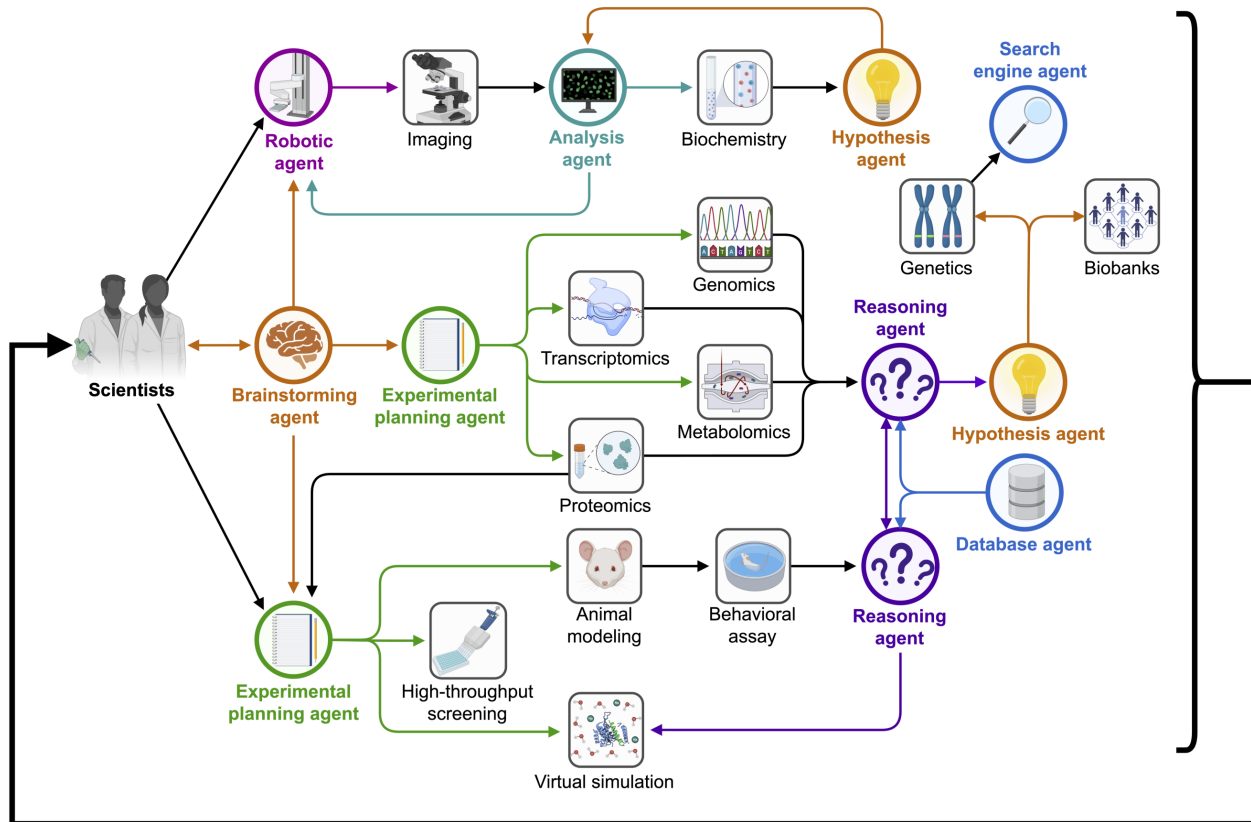


# “AI scientists” as generative AI agents



# “AI scientists” as generative AI agents

A long-standing ambition for biomedical AI is the development of AI systems that can make major discoveries with the potential to be worthy of a Nobel Prize—fulfilling the Nobel Turing Challenge



d. Reasoning with feedback for alternative experimental approach

