#### トーゴーの日シンポジウム2025

# 「データベースなしじゃ語れない!~マルチモーダル×AIは科学にどんな変革をもたらすの?~」

パネルディスカッション

相澤彰子

国立情報学研究所 コンテンツ科学研究系 教授

# テキスト×AI、データ×AI

AI Transformer (テキスト/文字列)

#### 論文テキストを学習した大規模言語モデル(LLM)

• **読む**: 文献検索、推薦、要約、比較表作成

書く:翻訳、文章校正、論文全体の下書き、査読コメント

研究する:アイディア出し、新規性チェック、コーディング、図表作成、実験計画

#### 各種基盤モデルによる予測

AlphaFold 3, AlphaGenome, ...

# マルチモーダルデータ×AI

AI Transformer (テキスト/文字列) マルチモーダルデータ (データベース)

# マルチモーダルデータ×AI

#### AI Transformer (テキスト/文字列)

/ 小分子は SMILES 形式で表現 タンパク質・DNA・RNA は FASTA 形式で表現 結晶材料データは、シーケンスにフラット化 ✓

# 現実世界 (研究室)

#### テキストと「シーケンス」を交互に配置 (Interleaved Sequence)

A prospective, randomized clinical trial was performed to study the efficacy of povidone iodine \(\langle\mol\rangle\mol\rangle\mol\rangle\mol\rangle\mol\rangle\mol\rangle\mol\rangle\mol\rangle\mol\rangle\mol\rangle\mol\rangle\mod マルチモーダルデータ(データベース)

#### PubChem、UniProt、NCBI などのデータベースからエンティティと説明を抽出

https://pubchem.ncbi.nlm.nih.gov/compound/aspirin

Pub Chem Aspirin (Compound)

CgH8O4
CH5COOC6H4COOH
aspirin
ACETYLSALICYLIC ACID
50-78-2
2-Acetoxybenzoic acid
2-(Acetyloxybenzoic acid
2-(Acetyloxybenzoic acid

- Xia, Yingce, Peiran Jin, Shufang Xie, Liang He, Chuan Cao, Renqian Luo, Guoqing Liu, et al. 2025. "Nature Language Model: Deciphering the Language of Nature for Scientific Discovery." http://arxiv.org/abs/2502.07527.
- Wang, Eric, Samuel Schmidgall, Paul F. Jaeger, Fan Zhang, Rory Pilgrim, Yossi Matias, Joelle Barral, David Fleet, and Shekoofeh Azizi. 2025. "TxGemma: Efficient and Agentic LLMs for Therapeutics." http://arxiv.org/abs/2504.06196.
- Wang, Yizhou, Chen Tang, Han Deng, Jiabei Xiao, Jiaqi Liu, Jianyu Wu, Jun Yao, et al. 2025. "SciReasoner: Laying the Scientific Reasoning Ground across Disciplines." http://arxiv.org/abs/2509.21320.

# マルチモーダルデータ×AIエージェント

AI Transformer (テキスト/文字列)

現実世界 (研究室)

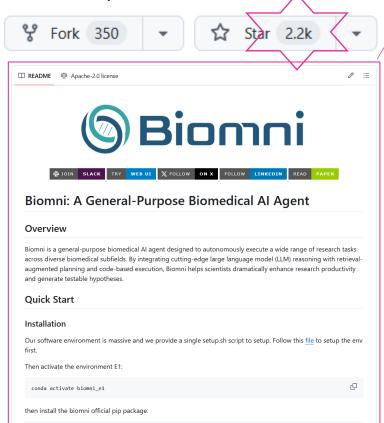
マルチモーダルデータ(データベース)

タスクデータ (ツール呼出し含む)

#### Biomni: A General-purpose biomedical AI Agent

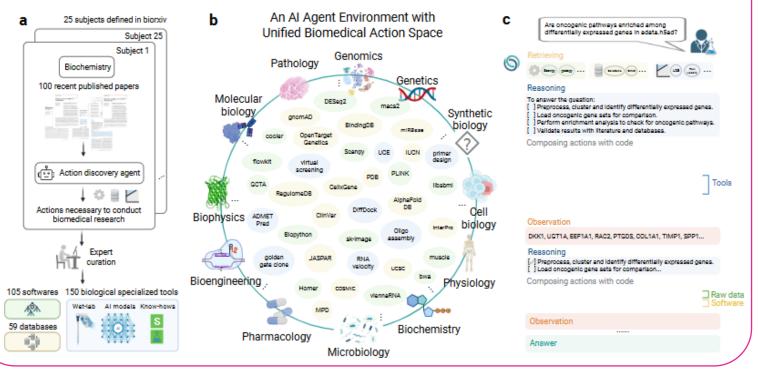
https://github.com/snapstanford/Biomni

Apache-2.0 license



Huang, Kexin, Serena Zhang, Hanchen Wang, Yuanhao Qu, Yingzhou Lu, Yusuf Roohani, Ryan Li, et al. 2025. "Biomni: A General-Purpose Biomedical AI Agent." Bioinformatics. bioRxiv. https://www.biorxiv.org/content/10.1101/2025.05.30.656746v1.

- 生命医学分野に特化した105個のツール、105個のソフトウェアパッケージ、59個のデータベース(専門家がキュレーション)
- これらのリソースを使いこなすためのAIエージェントと環境



# マルチモーダルデータ×AI:必要なもの

AI Transformer (テキスト/文字列)

AI-ready なデータ創出 メタデータ自動付与 自動アノテーション

現実世界 (研究室)

マルチモーダルデータ(データベース)

タスクデータ (ツール呼出し含む)