

生成AIは生命科学の何を変えるのか

トーゴーの日シンポジウム2024 「AI+ロボティクス+データベースが変える生命科学」

大田達郎

千葉大学 国際高等研究基幹 / 大学院医学研究院

千葉大学データサイエンスコア 副ディレクター



千葉大学大学院医学研究院
人工知能(AI)医学

大田達郎 (千葉大学)

- 医学研究院 人工知能医学 所属
 - (ユーザ・Appデベロッパとして) 生成AIもやっています
- 専門
 - バイオデータベース, オミックスデータ解析プラットフォーム
- 職歴
 - 2011- ROIS ライフサイエンス統合データベースセンター
 - 2023- 千葉大学
 - 2024/07- 千葉大学データサイエンスコア 副ダイレクター
- 兼務
 - 2015- 理化学研究所 情報統合本部 先端DSプロジェクト
 - 2023- 国立遺伝学研究所 生命情報・DDBJセンター



「AI+ロボティクス+データベースが変える生命科学」

生成AIは生命科学の何を変えるのか

前提: 生成AIは凄いが、解決すべき問題もある

SaaSでの提供

- 強力なプロプライエタリ・モデルへの依存 => 透明性・セキュリティ・データ保護

計算機

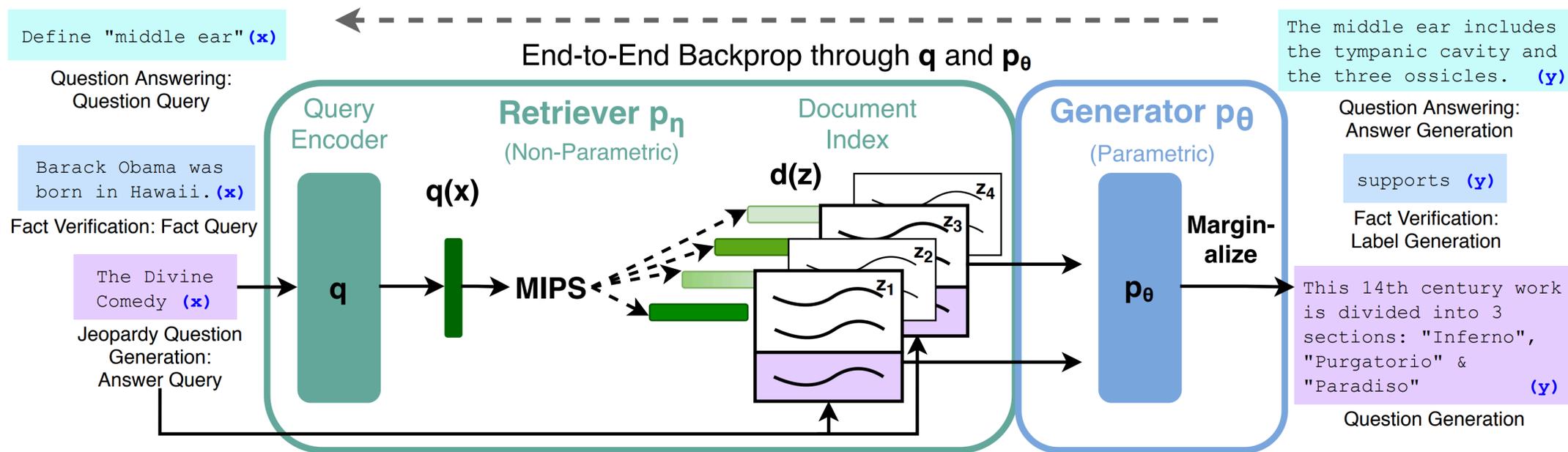
- 高性能GPU計算機への依存 => 特定企業による供給の寡占・膨大な電力消費

応用可能性

- 医療応用における事実誤認など正確性に関するリスク
- 人間とAIの価値観の相違について（倫理）

実応用への障壁: 出力の正確さの担保

Retrieval Augmented Generation (RAG): データベースを基盤にLLMを「強化」



一般にはドキュメントを蓄積したベクトルDBが用いられるが知識グラフでもRAGは可能

LLMは人における"直感"？

- Daniel Kahneman の著書に曰く、人の思考は「直感的・感情的な速い思考」システム1と「時間のかかる論理的な思考」システム2との組み合わせによって成立している
- 膨大な学習データからパターンを見出すLLMは「AIにおけるシステム1」と見なせる
- 論理的な推論を行うシステム2にあたるのは、本来データベースが担う機能ではないか

‘A lifetime’s worth of wisdom’
Steven D. Levitt, co-author of *Freakonomics*

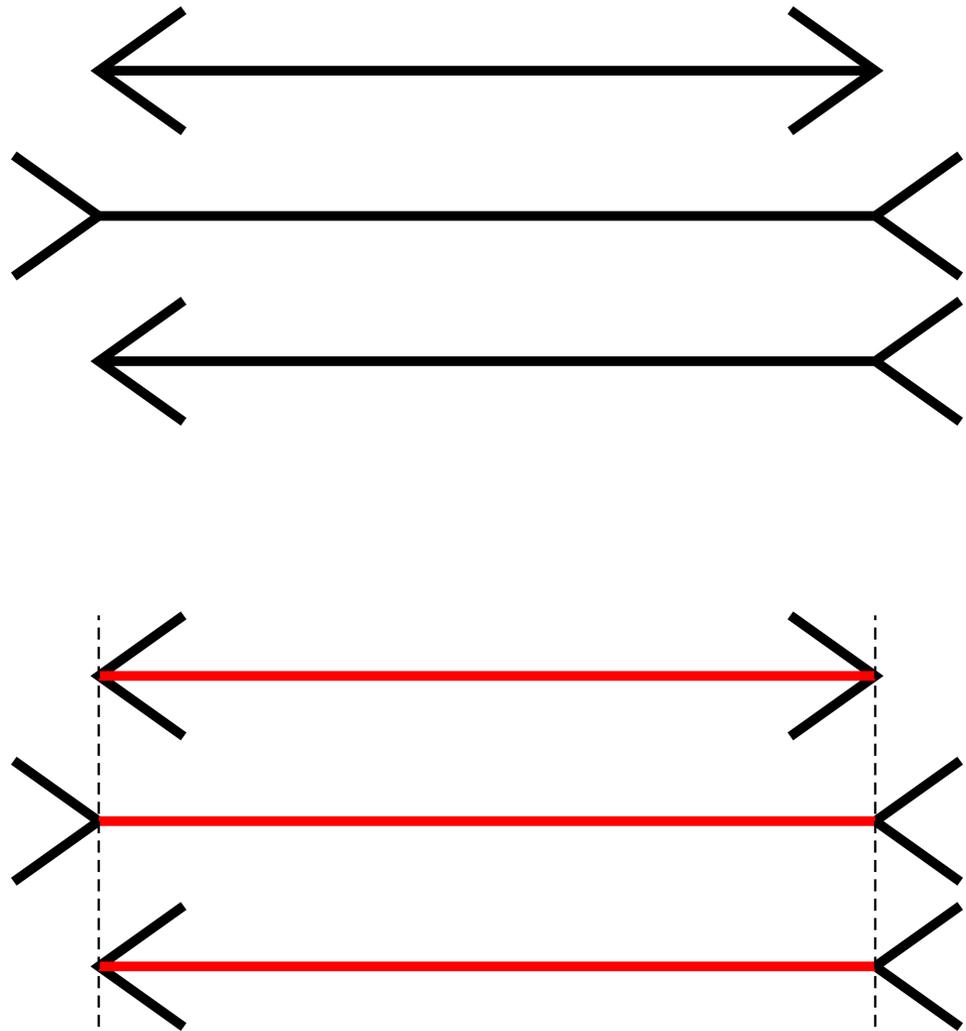
The International Bestseller



Thinking, Fast and Slow

Daniel Kahneman
Winner of the Nobel Prize





近い将来の生成AIは「幻覚」を見ない

完璧になったAIが解消できる「生命科学のボトルネック」は何なのか？

生命科学のボトルネックとその解消に必要なもの

- 賢く誠実で献身的な科学者 (教育 🎓)
- 新たな現象の発見と観察・仮説構築 (資金 💰 時間 ⌚ 知識 🧠)
- 新たな計測技術の開発 (資金 💰 時間 ⌚)
- 生体試料のサンプリング (資金 💰 社会的合意 🗣️)
- 再現可能な実験とその結果 (資金 💰 時間 ⌚ 技術 🔬)
- 論文の出版・およびサイエンスというゲームのルール (社会的合意 🗣️)
- 人類社会における需要 (予測・制御が難しい 🤔)

=> AI, ロボット, データベースはこれらのどれを助けるか？

将来の生命科学を左右するもの

- 観測・予測
 - 既にあるものはAIがなんとかしてくれる。今ないもの・これから起きることは？
- インフラ・データ
 - ハードウェア・計算機・ストレージ・ネットワークの維持はますます困難に
 - 100年後、1000年後に残すべきデータは何か、どうやって残すべきか？
- プログラム・AIの自律性
 - 人間が監督すべきか、人間が監督されるべきか
 - 真に強いAI研究者が現れたとき、人間がサイエンスをやる必要があるのか？