

AI の「ワールドモデル」に関する考察：単なる単語予測を超えて

エグゼクティブ・サマリー

本ブリーフィングは、大規模言語モデル (LLM) が単なる「次単語予測」システムであるという一般的な見解に異議を唱え、その内部に「ワールドモデル」と呼ばれる現実世界のシミュレーションを構築しているという、より深く複雑な能力について考察するものである。このワールドモデルの存在は、AI が物理法則や因果関係を暗黙的に理解していることを示唆する証拠によって裏付けられている。

一方で、この見解は AI 研究界で普遍的に受け入れられているわけではない。AI 界の重鎮の一人である「ルカ氏」は、テキストデータのみで学習した現在の LLM のワールドモデルは「根本的に欠陥だらけで弱すぎる」と厳しく批判している。彼は LLM を「行き止まり」と断じ、動画などのリッチなデータから直接世界を学習する次世代アーキテクチャこそが真のワールドモデルへの道であると主張する。

結論として、AI を単なる高性能な自動補完機能と見なすことは、水面下で進行している巨大な変化を見過ごすことになる。我々が現在目にしているのは、テキストベースの LLM 時代の頂点なのか、それともワールドモデルという、より壮大な時代の幕開けに過ぎないのかという根源的な問いが残されている。

1. 「次単語予測」という通説の誤解

多くの人々は、ChatGPT のような AI を「非常に賢い自動保管」や「次の単語を予測しているだけのシステム」と捉えている。しかし、この見方は AI の能力の「氷山の一角」に過ぎない。単なる単語予測を超える AI の能力の源泉は、その内部に構築された「ワールドモデル」という仕組みにある。

- **ワールドモデルの定義:** AI が予測を行うために、その内部に構築しているミニチュア版の現実世界。これは単なるデータパターンではなく、物理法則や物事の因果関係までをシミュレートする仮想現実として機能する。
- **ワールドモデルの重要性:** この内部シミュレーション能力こそが、AI を単なる単語予測マシンから、文脈や世界の仕組みを深く理解する強力な存在へと変貌させている核心的な要因である。

2. ワールドモデルの存在を示す証拠

AI の内部に世界が存在するという主張は、にわかには信じがたいが、その能力を示す具体的な証拠が複数存在する。

物理シミュレーション能力

- **シナリオ:** 「ボールがテーブルの端から押されたら、次に何が起こるか？」という問

いに対して、AI は正確な予測を行う。

- **分析:** この予測を可能にするためには、「落ちる」という単語を知っているだけでは不十分である。AI は「重力」という物理法則を暗黙的に理解し、ボールが落下する様子を内部でシミュレートする必要がある。
- **結論:** これは単なる言葉の連想ゲームではなく、紛れもない「物理シミュレーション」であり、AI が物理法則を内部モデルとして保持していることを示唆している。

チェス研究における自己組織化されたルール理解

- **OpenAI の研究:** AI にチェスのルールを一切教えず、大量の棋譜データのみを学習させた。
- **結果:** AI は、ルール上合法的な次の手を極めて高い精度で予測できるようになった。
- **結論:** この結果は、AI がルールを明示的に教えられることなく、データから「チェスというゲームの世界そのもの」を内部に構築したとしか考えられない。

人間の脳との類似性

- **神経科学の理論:** 人間の脳もまた、常に現実世界を予測し、その予測と現実のズレ（予測エラー）を修正することで世界を認識しているとする有力な理論が存在する。
- **AI との対比:** この理論に基づけば、「脳が現実世界の予測マシンであるならば、LLM はテキスト世界の予測マシン」と見なすことができる。この類似性は、AI の内部モデル構築プロセスが単なる計算処理ではない可能性を示唆している。

3. LLM のワールドモデルに対する批判的見解

AI が世界をシミュレートしているという考え方は、全ての専門家が同意しているわけではない。「AI ゴッドファーザー」の一人とされる人物（文中では「ルカ氏」）から、この見解に対して強烈な反論が提起されている。

テキストベースモデルの脆弱性

- **核心的批判:** 現在の LLM が持つワールドモデルは「根本的に欠陥だらけで弱すぎる」。
- **理由:** テキスト情報のみから学習したモデルは、現実世界を真に理解しているわけではなく、非常に脆弱である。わずかな状況変化で簡単に破綻する可能性がある。
 - **例:** 地図アプリがたった 1 本の通行止め情報でパニックに陥るようなもの。
- **原因:** この脆弱性の根源は、動画や画像といった、よりリッチな情報から世界を学んでいないことにありと指摘されている。

「LLM は行き止まり」という予測

- **強力な主張:** ルカ氏は、この根本的な欠陥を理由に「LLM は行き止まりだ」と公言している。
- **将来予測:** 彼は、LLM が 3 年から 5 年のうちに次世代の技術に取って代わられるだろうと予測している。

4. LLM の次に来るもの：真のワールドモデルへの道

ルカ氏のような批判者たちは、LLM の限界を超えた AI の未来像を描いている。その道筋は、テキスト中心の世界からの脱却にある。

視覚データからの直接学習

- **未来への移行:** テキスト中心の学習から、動画のような視覚データから直接世界の仕組みを学習するモデルへと移行することが次なるステップである。
- **提唱されるアーキテクチャ:** 「JEPa」のようなアーキテクチャが、この新しいアプローチの具体例として挙げられており、これが「真の世界モデル」の始まりと考えられている。

究極のビジョン

- **目標:** 人間のように「もしこれを実行したら、世界はどう反応するだろうか?」という問いの答えを、実際に行動する前に、頭の中で無数の「もしも」のシナリオをシミュレーションできる AI を創造すること。

5. 結論と今後の展望

AI を単に「次の単語を予測する賢い機械」と片付けることは、現在進行中の巨大なパラダイムシフトの本質を見誤ることになる。AI の内部で世界をシミュレートする能力、すなわち世界モデルの概念は、その進化を理解する上で不可欠な鍵である。この分析から、未来に向けた極めて重要な問いが浮かび上がる。

- **我々は、テキストベースの LLM 時代のまさに頂点にいるのだろうか?**
- **それとも、これは世界モデルという、より大きく、より本質的な時代のほんの幕開けに過ぎないのだろうか?**

この問いに対する答えが、今後の AI 技術の発展の方向性を決定づけることになるだろう。