

# AI は「人類最後の試験」を突破できるのか？

～人工知能 VS 人間の究極の試験が始まる～

ライフデザイン研究部 主席研究員/テクノロジーリサーチャー 柏村 祐

## 1. 知の限界に挑むAI

情報技術が急速に進展する現代において、人工知能（AI）は目覚ましい進化を遂げている。画像認識や自然言語処理といった分野で、AI は既に人間を凌駕する能力を発揮し始めており、今や AI は、まさに人類が長年挑み続けてきた「知の限界」に挑戦しようとしている。今回注目するのは、「人類最後の試験」（Humanity's Last Exam, HLE）というウェブサイトである。このサイトでは、専門家さえ頭を悩ませる、物理学、数学、生態学、古代文字など、多岐にわたる分野の超難問が提示されている。

これらの問題は、Center for AI Safety（非営利機関）と Scale AI（民間企業）によって主導された、世界トップレベルの研究者たちが、AI の真の知的能力を測るために作成したものであり、まさに人類が現在直面している「知のフロンティア」を示す試金石といえるだろう。驚くべきことに、これらの人類の叡智の結晶ともいえる難問に対し、AI が次々と挑戦し、一部の問題では目覚ましい成果を上げ始めているという。これまで、人類が叡智を結集しても解決できなかった難問を、AI がやすやすと解き明かす日が来るかもしれない。もし、その日が来れば、それは人類の歴史における知的大革命となるだろう。しかし、AI が「人類最後の試験」を解くことは、単純な AI の勝利を意味するのだろうか。あるいは、人間の知的能力はもはや不要となるのだろうか。

本レポートでは、「人類最後の試験」ウェブサイトの情報にもとづき、AI が示す驚異的な問題解決能力を解説する。そして、AI の進化が私たちにもたらす真の意味を考察するとともに、AI 時代において、私たち人間が果たすべき役割、AI と共存し、共に発展していくための道筋を探る。

## 2. 「人類最後の試験」とは何か

「人類最後の試験」とは、AI の潜在能力を極限まで引き出すために設計された、他に類を見ない超高難度ベンチマークである。

ウェブサイトの冒頭では、その目的と概要が述べられている。近年、AI、特に文章作成 AI や対話 AI といった大規模言語モデル（LLM）は、驚異的なスピードで進化しており、その性能を測るための従来の AI ベンチマークテストでは、もはや AI の真価を測りきれなくなっている。たとえば、MMLU（大規模多肢選択問題理解）という、大学レベルの知識を幅広く問うテストがある。この MMLU のような従来のベンチマークで

は、最新 AI は非常に高いスコアを容易に達成してしまうため、AI の能力を正確に測ることが難しくなっているのが現状である。そこで、既存の AI ベンチマークの陳腐化という課題を解決するために登場したのが、この「人類最後の試験」である。HLE は、従来のベンチマークとは比較にならないほど高い難易度をもつ問題群を提示することで、AI の能力の限界を見極め、さらなる AI 技術の発展を加速させることを目指している。

HLE のデータセットは、100 を超える専門分野、2,700 問にも及ぶ超難問で構成されている。これらの問題は、世界 50 か国、500 以上の研究機関に所属する、約 1,000 人もの専門家たちの叡智を結集して作成された。問題作成者には、世界最高峰の研究機関に所属する研究者たちが名を連ね、まさに人類が誇る最高の知性が生み出した問題群といえる。ウェブサイトで公開されているサンプル問題を見るだけでも、その圧倒的な難易度を実感することができる（図表 1）。

たとえば、「古典」分野の問題は、古代文字の解読を求めるものであり、歴史や言語学の専門知識が不可欠である。「生態学」分野の問題は、ハチドリの骨の構造という極めて専門的な内容を問うもので、生物学、特に鳥類学の深い知識が求められる。

「数学」分野の問題は、抽象数学の最先端分野である「圏論」からの出題であり、数理学の博士号をもつ研究者でさえ容易には理解できない。「計算機科学」分野の問題も、グラフ理論やマルコフ連鎖といった情報科学の専門知識を必要とし、高度な論理的思考力が求められる。

これらのサンプル問題からも明らかなように、「人類最後の試験」は、各分野の最先端研究に携わる専門家ですえ、容易には解けない、まさに人類が到達しうる最高レベルの難問を集めた、究極のベンチマークといえる。

図表 1 人類最後の試験の問題例

<p><b>Classics</b></p> <p><b>Question:</b></p>  <p>Here is a representation of a Roman inscription, originally found on a tombstone. Provide a translation for the Palmyrene script. A transliteration of the text is provided: RGYN<sup>o</sup> BT HRY BR cT<sup>o</sup> HBL</p> <p><small>Henry T Merton College, Oxford</small></p>	<p><b>Ecology</b></p> <p><b>Question:</b></p> <p>Hummingbirds within Apodiformes uniquely have a bilaterally paired oval bone, a sesamoid embedded in the caudolateral portion of the expanded, cruciate aponeurosis of insertion of m. depressor caudae. How many paired tendons are supported by this sesamoid bone? Answer with a number.</p> <p><small>Edward V Massachusetts Institute of Technology</small></p>
<p><b>Mathematics</b></p> <p><b>Question:</b></p> <p>The set of natural transformations between two functors <math>F, G: \mathcal{C} \rightarrow \mathcal{D}</math> can be expressed as the end</p> $\text{Nat}(F, G) \cong \int_A \text{Hom}_{\mathcal{D}}(F(A), G(A)).$ <p>Define set of natural cotransformations from <math>F</math> to <math>G</math> to be the coend</p> $\text{CoNat}(F, G) \cong \int^A \text{Hom}_{\mathcal{D}}(F(A), G(A)).$ <p>Let:</p> <ul style="list-style-type: none"> <li>- <math>F = \mathbf{B}_\bullet(\Sigma_4)_{*/}</math> be the under <math>\infty</math>-category of the nerve of the delooping of the symmetric group <math>\Sigma_4</math> on 4 letters under the unique 0-simplex <math>*</math> of <math>\mathbf{B}_\bullet\Sigma_4</math>.</li> <li>- <math>G = \mathbf{B}_\bullet(\Sigma_7)_{*/}</math> be the under <math>\infty</math>-category nerve of the delooping of the symmetric group <math>\Sigma_7</math> on 7 letters under the unique 0-simplex <math>*</math> of <math>\mathbf{B}_\bullet\Sigma_7</math>.</li> </ul> <p>How many natural cotransformations are there between <math>F</math> and <math>G</math>?</p> <p><small>Emily S University of São Paulo</small></p>	<p><b>Computer Science</b></p> <p><b>Question:</b></p> <p>Let <math>G</math> be a graph. An edge-indicator of <math>G</math> is a function <math>a: \{0, 1\} \rightarrow V(G)</math> such that <math>\{a(0), a(1)\} \in E(G)</math>.</p> <p>Consider the following Markov Chain <math>M = M(G)</math>: The statespace of <math>M</math> is the set of all edge-indicators of <math>G</math>, and the transitions are defined as follows:</p> <p>Assume <math>M_t = a</math>.</p> <ol style="list-style-type: none"> <li>pick <math>b \in \{0, 1\}</math> u.a.r.</li> <li>pick <math>v \in N(a(1-b))</math> u.a.r. (here <math>N(v)</math> denotes the open neighbourhood of <math>v</math>)</li> <li>set <math>a'(b) = v</math> and <math>a'(1-b) = a(1-b)</math></li> <li>Set <math>M_{t+1} = a'</math></li> </ol> <p>We call a class of graphs <math>\mathcal{G}</math> well-behaved if, for each <math>G \in \mathcal{G}</math> the Markov chain <math>M(G)</math> converges to a unique stationary distribution, and the unique stationary distribution is the uniform distribution.</p> <p>Which of the following graph classes is well-behaved?</p> <p><b>Answer Choices:</b></p> <ol style="list-style-type: none"> <li>The class of all non-bipartite regular graphs</li> <li>The class of all connected cubic graphs</li> <li>The class of all connected graphs</li> <li>The class of all connected non-bipartite graphs</li> <li>The class of all connected bipartite graphs.</li> </ol> <p><small>Marc R Queen Mary University of London</small></p>

資料: Humanity's Last Exam より「<https://lastexam.ai/>」

ウェブサイトには、いくつかの最新 AI モデルが HLE に挑戦した結果も掲載されている (図表 2)。GPT-4o、Grok-2、Claude 3.5 Sonnet、Gemini Thinking など、現時点で最高レベルの性能をもつとされる AI モデルでさえ、HLE においては驚くほど低い精度しか達成できていない。最も高い精度を示した AI モデルでも、わずか 14% であり、多くのモデルは 1 桁台の精度にとどまっている。これは、最新の AI をもってしても、人類が長年かけて蓄積してきた高度な知識や複雑な思考力には、遠く及ばないという

現実を示している。さらに、AIは「自信過剰」な傾向があり、誤った答えを自信満々に回答してしまうという課題も露呈している。

しかしながら、ウェブサイトの開発チームは、AI技術の急速な進歩を考慮し、「2025年末までには、AIがHLEで50%以上の正答率を達成する可能性も十分に考えられる」と予測している。もし、AIがHLEで50%を超える正答率を達成する日が来れば、AIは限定的な領域においては、専門家レベルの知的能力に到達することを意味するだろう。もっとも、HLEが評価するのは、あくまでも「既存の知識に基づいた問題解決能力」であり、「未知の領域を切り拓く創造性」や「複雑な状況に対応する総合的な判断力」といった、人間ならではの高度な知性を測ることはできない。

図表2 AIモデルによる精度と自信過剰さの指標分析結果

Model	Accuracy (%) ↑	Calibration Error (%) ↓
 GPT-4o	3.1	92.3
 Grok-2	3.9	90.8
 Claude 3.5 Sonnet	4.8	88.5
 Gemini Thinking	7.2	90.6
 o1	8.8	92.8
 DeepSeek-R1*	8.6	81.4
 o3-mini (medium)*	11.1	91.5
 o3-mini (high)*	14.0	92.8

資料: Humanity's Last Exam より「<https://lastexam.ai/>」

### 3.AIとの協調、そして人が磨くべき能力

「人類最後の試験」の結果は、AIが特定の分野においては、人間を凌駕する潜在能力をもつことを示唆している。特に、膨大な知識量が要求される問題や、論理的な思

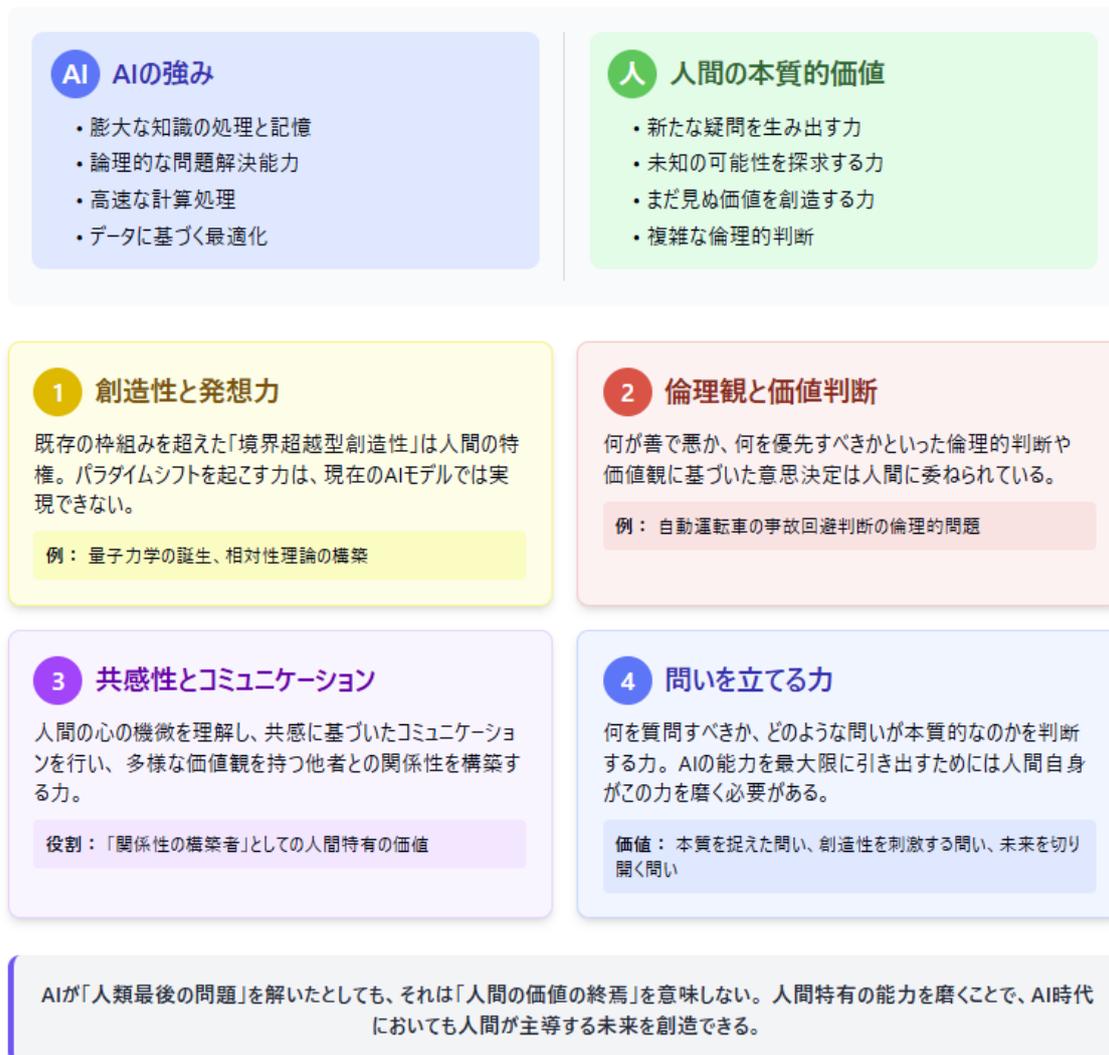
考力が求められる問題においては、AIは今後、急速にその能力を高めていくことが予想される。しかし、仮にAIが「人類最後の問題」を解き明かす日が来たとしても、それは決して「人間の価値の終焉」を意味するものではない。なぜなら、人間の知性の本質は、単に問題を解くだけでなく、新たな疑問を生み出し、未知の可能性を探求し、まだ見ぬ価値を創造することにあるからである。

AIが急速に進化していく時代において、人間にしかできないこと、そして人間が磨くべき能力とは何だろうか。

それは、以下の4つの能力に集約されるだろう（図表3）。

第一に、創造性と発想力である。AIは、既存の知識を組み合わせ、効率的に問題を解決することは得意だが、真に新しいアイデア、独創的な発想を生み出すのは、依然として人間の特権である。既存の枠組みを超えた「境界超越型創造性」は、量子力学の誕生や相対性理論の構築のように、パラダイムシフトを起こす力であり、現在のAIモデルでは実現できない領域である。第二に、倫理観と価値判断である。AIは、与えられた目標に対し、効率性や最適化を追求するが、何が善で、何が悪か、何を優先すべきかといった倫理的な判断や価値観にもとづいた意思決定は、人間に委ねられている。たとえば、自動運転車の事故回避判断のような複雑な倫理的問題は、技術的課題ではなく社会的合意にもとづく人間の価値判断が不可欠である。第三に、共感性とコミュニケーション能力である。複雑化する現代社会においては、人々の感情に寄り添い、共感し、協力し合いながら、社会全体としてより良い方向へ進んでいくことが求められる。AIには感情がないため、人間の心の機微を理解し、共感にもとづいたコミュニケーションを行うことはできない。多様な価値観や文化的背景をもつ他者との間に建設的な関係性を構築する「関係性の構築者」としての役割は、AIでは代替できない人間特有の価値である。第四に、問いを立てる力である。AIは、与えられた質問に対して、高速かつ正確に回答することができるが、そもそも何を質問すべきか、どのような問いが本質的なのかを判断するのは、人間の役割である。AIに適切な質問を投げかけ、AIの能力を最大限に引き出すためには、人間自身が「問いを立てる力」を磨く必要がある。本質を捉えた問い、創造性を刺激する問い、まだ見ぬ未来を切り開く問いを生み出す力こそ、AI時代において、ますます重要となるだろう。

図表3 AI時代に人間が磨くべき4つの能力



資料：筆者作成

#### 4.AIと人間の知的共進化

AIの進化は、人間の知的能力を代替するものではなく、人間の知的能力を拡張するものと捉えるべきである。AIを単なる道具としてではなく、知的なパートナーとして捉え、AIと互いの強みを活かし、弱みを補完し合う「知的分業」の最適化が求められる。たとえば、科学研究においては、AIがデータ処理や初期仮説の生成を担当し、人間は仮説の評価や実験デザイン、結果の解釈という創造的思考を要する部分に集中することで、研究の効率と質の両方を高められる。AIと人間の相互作用は、単なる分業を超えた「知的増幅効果」をもたらし、両者が共に進化する「知的共進化」の時代を実現するだろう。

これからの社会では、AIに仕事を奪われる「AI失業」ではなく、AIと創造的に協働できない人が取り残される「創造性格差」こそが真の課題となる。

そのため、教育・人材育成の根本的な見直しが急務である。現在の教育システムは、主に知識の蓄積と再現に重点を置いており、AIが最も優れている領域と重複している。AI時代の教育は、「AIにできないこと」を中心に再構築する必要がある（図表4）。

具体的には、画面を見てデータを処理する作業はAIに任せ、人間は物理的な実在と直接関わる能力を磨くべきだ。大工や料理人のように、素材に直接触れ、五感を使う技術教育が重要となる。木の質感を理解する指先の感覚や、料理の火加減を見極める目は、AIには真似できない。また、自然の中での活動や実験を通じ、身体を使って世界を理解する教育も必要だ。教室の外で土を掘り、植物を育て、物を組み立てる経験が、リアルな問題解決能力を育む。さらに、対面でのコミュニケーションや協働作業を重視すべきである。目を見て会話し、表情から感情を読み取り、場の空気を感じる能力は、AIにはない人間固有の強みである。

図表4 教育・人材育成におけるAI時代の新たなアプローチ



資料: 筆者作成

「人類最後の試験」は、AIの能力を測るだけでなく、私たち人間に、挑戦することの意義、常に学び続けることの重要性を改めて教えてくれる。AIの急速な進化に恐れを抱くのではなく、AIを深く理解し、AIを積極的に活用し、AIと共に成長していく。そして、人間ならではの創造性、倫理観、共感性、問いを立てる力を磨き続けることこそが、AI時代における人間の使命であり、無限の可能性を拓く鍵となるだろう。