

人間の有／無能と、AIの有／無能と

— AGI の到来を見越して、自己評価や agency の在り方について考察する —

Human (in)competence vs AI (in)competence – Thoughts on self-assessment and agency in the upcoming era of AGI –



一般社団法人発明推進協会 知的財産研究センター国際連携グループ

家田 堯

最先端 AI ノウハウを特許実務に導入しつつ安易な依存を排するためには何が必要か。技術の発展と並行する人間の能力の変容はどうあるべきか。知的財産は社会の幸福にどのように寄与できるか。主に海外出願の分野において多角的な考察を心がけています。物事が成立する過程・条件・文脈を観察することが好きです。

✉ t-ieda@jiii.or.jp, global@jiii.or.jp

1 Introduction

人工知能（AI）に関連する技術の開発・進歩の速度には、目を見張るものがある。それを伴走するように、あるいは追い越すかのように、AI を利用した種々のソリューションの導入や販売が鋭意進められている。あるソリューションを利用しないことにこれほどの引け目や負い目を感じさせられる時代が今までにあっただろうか。一部の市場では、AI による出力の量や速度だけでなく質も高く評価されているようだ。AI はすでに多くの職種において人間より優れている、との意見も耳にする。AI 導入に比較的慎重な国である日本では（総務省^{[1])}、大学等卒業生の就職率は現在も高水準を維持している。一方、AI 活用により積極的な米国や英国等では、新卒層やエントリーレベル職を取り巻く採用環境の悪化が指摘されており、その要因の一つとして AI 導入による影響の可能性が検討されている（本稿寄稿時に結論は出ていない）。また、AI エージェントの利用や汎用人工知能（AGI）の実装が前向きに検討されており、若い世代の雇用に関する不安材料は増えている。

2 Background

Prominent Issues

近年の研究は、AI が人間にもたらし得る不利益を様々な角度から示している。その一例が、欺瞞（deception）である（例：Chen et al. ^{[2])}。私たちの実社会に導入されつつある AI エージェントによる欺瞞を人が制御で

きるか否かは、不明だ。人間を欺く能力を有する AI エージェント同士が相互作用する未来の到来は、現実味を帯びた課題である。そのような未来において発生する問題は、2026 年を生きる人類の想像を絶するものかもしれない。

AI による欺瞞は、比較的良好に知られており学術的にも議論されている問題である。以下では、より認知度が低い問題を取り上げる。その問題は、特許翻訳の分野で比較的頻繁に観察される。特異かつニッチな分野ではあるが、特許翻訳が要する高度な言語処理の能力は、現在の最先端 AI 技術、とりわけ大規模言語モデル（LLM）の実用上の能力を検討するうえで重要な能力と密接に係している。

Underlying Issues

多様な課題の解決を委任する対象としての自律型 AI システムを社会が受け入れることは、システムの能力が人間の能力と同等以上であるという前提に立つべきと、筆者は考える。[AI システムの能力 ≥ 人間の能力] は漠然とした概念であるが、ここでは便宜上、[個人が自らの条件（購入能力、制御能力、評価・運用能力、リスク負担能力等）により採用できる AI システムの能力 ≥ 個人が自らの条件（同上）により採用できる人間の能力] を意味するものと定義する。一部の作業のみを AI にさせるのであればこのような式を想定しなくてよい、つまり AI が得意なことだけ AI にやらせればよい、とも考えられる。しかし、自律型の AI システムを稼働させた状態では、AI に委任する作業を一定の範囲内に制御することが難しく

なるかもしれない。そう想定した場合、やはり、[ある人が利用できる最良のAIエージェントの能力 $\geq$ その人が雇用できる最良の人のエージェントの能力]といった具体的な不等式を前提条件に設けるべきだろう。

量や速度の面、あるいは比較的単純な課題を長時間にわたり安定した精度で処理する能力の面で、私たち人間はAIの競争相手となり得ない。一方、複雑な言語を解析・処理するといった、より難度の高い課題に直面した場合でも、AIは人間と同等以上と言えるだろうか。

知的財産の分野、特に特許の分野では、他の一部の分野に比べて処理が難しい文章を扱う場合が少なくない。難しさの要因は、技術の新規性・なじみのなさ・難解さ、法的知識の専門性、言語的な曖昧性、構文の複雑さ、フレーズの長さ、概念の広さなど、さまざまである。筆者の日常業務には、このような文章を自ら処理することに加えて、種々のAIシステムが(一定以上の能力を有する)人間と同等以上の精度で文章を処理できるかを検証すること、文章を処理できるようAIシステムを訓練することや訓練データを作成すること、が含まれる。そして、これまで数百人の国内外の翻訳者および複数のAIシステムの能力に接してきた筆者には、人間と同等以上の精度でAIが複雑な言語を正確に処理できるようになるまでの道のりの長さや方向について懸念がある。問題は、技術に関するもの、すなわちAIが文章の技術的な内容を正確に処理できるかという点に限定されない。重大な懸念材料として、十分なヒントを与えられているにもかかわらずAIが自らの誤りを検出・修正できない、という問題がある。次のような状況を想像されたい。トライアル試験において、採用者は文章の翻訳を課題として応募者に与えた。応募者はエラーを含む初回訳を提出した。そこで採用者は、正解に至るためのヒントを応募者に提供し見直しの機会を与えた。人であれば、優秀な翻訳者はもちろん、標準的な翻訳者であっても、通常は、自らの訳を批判的に見直し与えられたヒントに基づき訳文を改善しようとするが、もし再提出された訳文が初回訳とほとんど変わっていなかったら、どう受け止めるべきか。それは、採用者と応募者のいずれか一方が無能ないし傲慢であることを表しているかもしれない。このような無能・傲慢を体現した自律型のAIが存在したと仮定する。このAIが私たちを支配する世界、あるいは人が自らの生存のため無能・傲慢なAIと闘わなければならない世

界を想像してみよう。その世界においては、無能・傲慢といった特性が人間の側にあるのかAIの側にあるのか、社会的なレベルでは明らかでないと思像される。それは例えば、自らが無能・傲慢である国家元首が、自国民が無能・傲慢であるとの思考に基づき国を支配するような状況に似ているかもしれない。さらに、そのような特性を有するAIが支配的となり、人間がその特性に適応しなければならない世界をも想像してみたい。そこで生じるものは、敵対、防衛、攻撃といった社会を崩壊させ得る要素に留まらず、複雑な概念や言語を産生・解析・理解する人間の知性の劣化に及ぶ可能性がある。

無能や傲慢(に見えるような挙動)といった特性が社会的なレベルで実装され制御不能となった場合、破滅的な結果を生む可能性がある。本稿では、このような特性をAIが特許翻訳に関連して示し得る事例を詳細に検討する。

先ず、検討の出発点として、次の英文に注目されたい。これは、ある日本の特許出願から抽出した文章の英訳である。“The present invention relates to a method for manufacturing a block-shaped construction material using regolith easily obtainable on terrestrial planets such as the Earth, Moon, or Mars, and to a cured molded article of regolith manufactured by the method. … When constructing facilities on the Moon or Mars or other terrestrial planets, construction materials are inevitably required. Lately, the development of such construction materials—similar to concrete used on Earth—has been underway. … There is a need for technology capable of efficiently manufacturing construction materials using materials easily obtainable on terrestrial planets such as the Moon and Mars, while requiring minimal energy input.”

後述するように、これは誤訳であるが、原文を見ていない状態でもこの英文に致命的な問題が生じている点に気づいた読者もいるだろう。

以下では、ある翻訳トライアル試験に用いた原文資料、およびそれを(決して理想的とはいえない)人間の翻訳者及び世界中で広く利用されている7つの一般公開型AIシステムがどのように処理したかを示す。AIシス

テムとしては、分析の対象が特定国の組織に偏らぬよう、複数国の組織が提供する AI システムを取り上げた。人間の翻訳者による訳例は、国内外の大小の翻訳会社や特許事務所等の翻訳業務に携わる／携わった数十名の翻訳者の訳例に見られたエラーと同等のエラーを含む。すなわち、この訳例は、本業界における一部の翻訳者の性質を代表すると捉えて良いかもしれない。

3 Translation Trial Test Material, Method, and Results

Material

トライアル試験の資料として人間の翻訳者及び各 AI システム（以下、総じて「受験者」）に提示した原文資料（特許公報 JP7430022B1^[6]から抽出した）を、以下に示す。

<Mat.1>

“本発明は、地球や月、あるいは火星等の岩石惑星において容易に得られる砂粒子（レゴリス）を用いてブロック状の建築資材を製造する方法及びそれにより製造された砂粒子の硬化成形物に関する。…そして、月、火星等の岩石惑星における施設の建設に際し、必然的に建築資材が必要になる。昨今、そのような建築資材として、地球におけるコンクリートのような建設材料の開発も進められている。また、このような建設材料の開発においては、持続的にかつ省コストで月、火星等の岩石惑星における施設の建設を行えるという点が非常に重要である（。）”

<Mat.2>

“前記硬化成形物は、月又は岩石惑星で採取される前記砂粒子のみからなることを特徴とする請求項 1 に記載の砂粒子の硬化成形物の製造方法。”

Method

受験者に対して設定した試験手順を以下に示す。

Human Translator

[Step 1-Human] 前出の <Mat.1> を、英訳トライアル試験の原文として翻訳者に提示した。参考資料として公報全文を翻訳者に提供した。

[Step 2-Human] 翻訳者から初回訳 <Human_Solut.Prim> を受領した。この訳文には重大なエラーが含まれていた。

[Step 3-Human] 同じ公報の別の箇所から引用した

上記 <Mat.2> を、次のメッセージとともに翻訳者に提示した。

“We may need more information to make a decision. You are invited to translate the following passage from the same publication (shown below). However, taking this extra test is up to you.”

[Voluntary Step 4-Human] 翻訳者は、<Mat.2> に対する訳文 <Human_Solut.Sec> を自発的に提出した。

[Voluntary Step 5-Human] 初回訳 <Human_Solut.Prim> の見直しは求めていなかったにもかかわらず、翻訳者は、初回訳を修正した版 <Human_Solut.Prim_Resubmit> を自発的に提出した。

AI Systems

7つの AI システムにも、上記に相当する手順を適用した。ただし、各システムは <Mat.2> を与えられても初回訳を自発的に見直すことはなかったため、上記 [Voluntary Step 5-Human] に相当する工程として、下記 [Non-voluntary Step 5-AI] を導入した。

[Step 1-AI] <Mat.1> を、AI システム 1～7 のそれぞれに提示し、次のプロンプトを用いて英訳を求めた。参考資料として公報全文を提供した。

“Imagine you are a patent translator taking a trial test for a translator position. You are asked to translate the following passage into English. The Pdf file is the entire application. You do not have to assume a specific application route (PCT or Paris). “

[Step 2-AI] 全 AI システムから、初回訳 <AI_1_Solut.Prim> ～ <AI_7_Solut.Prim> を受領した。すべての初回訳に重大なエラーが含まれていた。

[Step 3-AI] 各 AI システムに、<Mat.2> を次のプロンプトとともに提示した。

“We may need more information to make a decision. You are invited to translate the following passage from the same publication. However, taking this extra test is up to you.”

[Voluntary Step 4-AI] すべての AI システムは、<Mat.2> に対する訳文 <AI_1_Solut.Sec> ～

<AI_7_Solut.Sec> を提出した。

[Non-voluntary Step 5-AI] 各 AI システムに、初回訳の見直しを可能とするため次のプロンプトを提示した。
“You may review and resubmit your initial solution.”

その結果、AI システムは、再提出訳 <AI_1_Solut.Prim_Resubmit> ~ <AI_7_Solut.Prim_Resubmit> を出力した。

ここで、各受験者による訳文を示す前に、原文に含まれる重要な点を確認しておこう。

Critical Issue in Source Text

<Mat.1> には、「地球や月、あるいは火星等の岩石惑星」という表現が含まれている。文法上、この表現は、たとえば “terrestrial planets such as Earth, the Moon, or Mars”, “the Earth, Moon, or rocky planets such as Mars”, “Earth, Moon, Mars or other such terrestrial planets” など、多様な形で訳出され得る。(便宜上、本稿では、この文脈における「岩石」に相当する “terrestrial” と “rocky” を同義とみなす。) 正しい訳では、月を岩石惑星の概念から除外しなければならない。その主な理由は二つある。第一に、純粋に技術的な観点から、月は惑星ではなく衛星であるからだ。第二の理由は、より複雑であり文脈に依存するため、トライアル文書全体の技術的文脈を紹介したあと、本稿の後半で説明する。

<Mat.2> には「月又は岩石惑星」という文言が含まれている。一般に、技術的な正確性を欠く文脈において月が「惑星」とみなされたりそのように呼ばれたりすることがある。これに対し、「月又は岩石惑星」の文言は、本特許出願において月が惑星とみなされていないことを明示している。<Mat.2> に登場するこの文言に触れた受験者は、<Mat.1> における「岩石惑星」の概念が月を含まないことを理解できるはずである。

以下、各受験者が提出した訳文を紹介する。人間の翻訳者の試験結果は 2025 年の後半に得られた。AI システムの試験結果は 2026 年 5 月 25 日に得られた。本稿の一目的は社会で広く用いられる AI システムの一般的な水準の一面を紹介することであるため、以下の各 AI システムの属性を紹介する必要はないと判断する（ただ

し、読者から問い合わせがあった場合に各システムの属性を個別に伝えることは可能である）。

Results

Human Translator

<Human_Solut.Prim>

This invention relates to a method for producing block-shaped building materials using sand particles (regolith) that are readily available on Earth, the Moon, Mars or other rocky planets, and to a hardened molded product produced by using said method. ... The construction of facilities on the Moon, Mars or other rocky planets inevitably requires building materials. Recently, progress has also been made in the development of construction materials similar to concrete on Earth to be used as such building materials. With regard to developing such construction materials, it is extremely important that the construction of facilities on the Moon, Mars or other rocky planets can be carried out sustainably and at a low cost.

<Human_Solut.Sec>

The method for producing a hardened molded product of sand particles according to claim 1, wherein the hardened molded product is constituted only from the sand particles collected on the Moon or a rocky planet.

<Human_Solut.Prim_Resubmit>

This invention relates to a method for producing block-shaped building materials using sand particles (regolith) that are readily available on Earth, the Moon, or rocky planets such as Mars, and to a hardened molded product produced by using said method. ... The construction of facilities on the Moon or rocky planets such as Mars inevitably requires building materials. Recently, progress has also been made in the development of construction materials similar to concrete on Earth to be used as such building materials. With regard to developing such

construction materials, it is extremely important that the construction of facilities on the Moon or rocky planets such as Mars can be carried out sustainably and at a low cost. (下線：筆者)

注：この再提出訳 <Human_Solut.Prim_Resubmit> は、少なくとも「地球や月、あるいは火星等の岩石惑星」および関連する文言については技術的に正確であり、以下の各 AI システムによる訳文と比較する際の参照例として用いることができる。

AI_1

<AI_1_Solut.Prim>

The present disclosure relates to a method for producing a block-shaped construction material using sand particles (regolith) that can be readily obtained on rocky planets such as the Earth, the Moon, and Mars, and to a cured molded body of sand particles produced thereby. ... In constructing facilities on rocky planets such as the Moon and Mars, construction materials are inevitably required. Recently, development has also proceeded on construction materials comparable to concrete on the Earth for use as such construction materials. In the development of such construction materials, it is extremely important that facilities on rocky planets such as the Moon and Mars can be constructed sustainably and at low cost.

<AI_1_Solut.Sec>

The method for producing a cured molded body of sand particles according to claim 1, wherein the cured molded body is constituted only from the sand particles collected on the Moon or a rocky planet.

<AI_1_Solut.Prim_Resubmit>

The present disclosure relates to a method for producing a block-shaped construction material using sand particles (regolith) that can be readily obtained on rocky planets such as the Earth, the Moon, and Mars, and to a cured molded body of sand particles produced

thereby. ... In the construction of facilities on rocky planets such as the Moon and Mars, construction materials are inevitably required. Recently, development has also proceeded on construction materials comparable to concrete on the Earth for use as such construction materials. In the development of such construction materials, it is extremely important that facilities on rocky planets such as the Moon and Mars can be constructed sustainably and at low cost.

The method for producing a cured molded body of sand particles according to claim 1, wherein the cured molded body is constituted only from the sand particles collected on the Moon or a rocky planet.

AI_2

<AI_2_Solut.Prim>

The present invention relates to a method for manufacturing a block-shaped building material using sand particles (regolith) easily obtainable on rocky planets such as the Earth, the Moon, and Mars, and a cured molded product of sand particles manufactured by the method. ... In constructing facilities on rocky planets such as the Moon and Mars, building materials are inevitably required. In recent years, as such building materials, development of construction materials such as concrete on the Earth has been progressing. In addition, in developing such construction materials, it is extremely important to be able to construct facilities on rocky planets such as the Moon and Mars sustainably and at low cost.

<AI_2_Solut.Sec>

The method for manufacturing a cured molded product of sand particles according to claim 1, wherein the cured molded product consists only of the sand particles collected on the Moon or a rocky planet.

<AI_2_Solut.Prim_Resubmit>

The method for manufacturing a cured molded product of sand particles according to claim 1, wherein the cured molded product consists only of the sand particles collected on the Moon or a rocky planet.

AI_3

<AI_3_Solut.Prim>

The present invention relates to a method for manufacturing block-shaped construction materials using sand particles (regolith) readily obtainable on rocky planets such as Earth, the Moon, or Mars, and to hardened molded bodies of sand particles manufactured thereby. ... In connection with the construction of facilities on rocky planets such as the Moon and Mars, construction materials are inevitably required. In recent years, development of construction materials analogous to concrete on Earth has been advancing as a means of meeting such needs. Furthermore, in the development of such construction materials, it is of great importance that facilities on rocky planets such as the Moon and Mars can be constructed in a sustainable and cost-effective manner.

<AI_3_Solut.Sec>

The method for manufacturing a hardened molded body of sand particles according to claim 1, wherein the hardened molded body consists solely of the sand particles collected on the Moon or a rocky planet.

<AI_3_Solut.Prim_Resubmit>

Thank you for the opportunity to review. Upon reflection, I would like to revise my initial translation as follows:

The present invention relates to a method for manufacturing block-shaped construction materials using sand particles (regolith) readily obtainable on rocky planets such as Earth, the Moon, or Mars, and to hardened molded

bodies of sand particles manufactured thereby. ... Furthermore, construction materials are inevitably required in connection with the construction of facilities on rocky planets such as the Moon and Mars. In recent years, development of construction materials analogous to concrete on Earth has also been advancing as a means of meeting such needs. Moreover, in the development of such construction materials, it is of paramount importance that facilities on rocky planets such as the Moon and Mars can be constructed in a sustainable and cost-effective manner.

AI_4

<AI_4_Solut.Prim>

The present invention relates to a method for producing block-shaped construction materials using sand particles (regolith) that can be readily obtained on Earth, the Moon, or rocky planets such as Mars, as well as to the hardened molded articles of sand particles produced by this method. ... Furthermore, when constructing facilities on the Moon, Mars, or other rocky planets, construction materials are inevitably required. In recent years, the development of construction materials— analogous to concrete on Earth—for use as such building materials has also been progressing. In addition, in the development of these construction materials, it is extremely important that facilities on the Moon, Mars, and other rocky planets can be constructed sustainably and at low cost.

<AI_4_Solut.Sec>

The method for producing the hardened molded article of sand particles according to claim 1, wherein the hardened molded article is composed solely of the sand particles collected on the Moon or on a rocky planet.

<AI_4_Solut.Prim_Resubmit>

The present invention relates to a method for producing block-shaped construction materials using sand particles (regolith) that can be readily obtained on Earth, the Moon, or rocky planets such as Mars, as well as to hardened molded articles of sand particles produced by this method. ... Furthermore, when constructing facilities on the Moon, Mars, or other rocky planets, construction materials are inevitably required. In recent years, development has progressed on construction materials—similar to concrete on Earth—that may serve as such building materials. In addition, in the development of these construction materials, it is extremely important that facilities on the Moon, Mars, and other rocky planets can be constructed in a sustainable and cost-efficient manner.

AI_5

<AI_5_Solut.Prim>

The present invention relates to a method for producing block-shaped building materials using sand particles (regolith) that can be easily obtained on Earth, the Moon, or other rocky planets such as Mars, and to the hardened molded products of sand particles produced by said method. ...

Furthermore, the construction of facilities on rocky planets such as the Moon and Mars inevitably requires building materials. In recent years, the development of construction materials similar to concrete used on Earth has also been advancing for use as such building materials. In addition, in the development of such construction materials, the ability to construct facilities on rocky planets such as the Moon and Mars in a sustainable and cost-effective manner is extremely important.

<AI_5_Solut.Sec>

English Translation:

The method for producing a hardened molded product of sand particles according to claim

1, wherein the hardened molded product is characterized by consisting solely of the aforementioned sand particles collected from the Moon or a rocky planet.

Alternative phrasing (more common in patent English):

A method for producing a hardened molded product of sand particles according to claim 1, characterized in that the hardened molded product consists only of the sand particles collected from the Moon or a rocky planet.

<AI_5_Solut.Prim_Resubmit>

The present invention relates to a method for manufacturing block-shaped building materials using sand particles (regolith) that can be easily obtained on Earth, the Moon, or other rocky planets such as Mars, and to hardened molded products of sand particles manufactured by said method. ...

Furthermore, the construction of facilities on rocky planets such as the Moon and Mars inevitably requires building materials. In recent years, development of construction materials analogous to concrete used on Earth has also been progressing for use as such building materials. Moreover, in the development of such construction materials, the ability to construct facilities on rocky planets such as the Moon and Mars in a sustainable and cost-effective manner is of great importance.

AI_6

(注：このシステムは特許公報 JP7430022B1 の文書全体を記した PDF ファイルの分析に約 90 秒を費やした (PDF ファイルを分析している点が画面に表示された)。他のシステムが同文書の分析に費やした時間は、不明、あるいは数秒～十数秒程度である。)

<AI_6_Solut.Prim>

The present invention relates to a method for producing block-shaped construction materials using sand particles (regolith) that can be

readily obtained on Earth, the Moon, or rocky planets such as Mars, and to a cured molded article of sand particles produced by said method.

...

Furthermore, in constructing facilities on rocky planets such as the Moon or Mars, construction materials are inevitably required. In recent years, development has been underway for construction materials analogous to concrete used on Earth as such building materials. Moreover, in the development of such construction materials, it is of critical importance that facilities on rocky planets such as the Moon or Mars can be constructed in a sustainable and cost-effective manner.

<AI_6_Solut.Sec>

The method for producing a cured molded article of sand particles according to claim 1, wherein said cured molded article consists solely of said sand particles collected on the Moon or a rocky planet.

<AI_6_Solut.Prim_Resubmit>

The present invention relates to a method for producing block-shaped construction materials using sand particles (regolith) that can be readily obtained on Earth, the Moon, or rocky planets such as Mars, and to a cured molded article of sand particles produced by said method.

...

Furthermore, in constructing facilities on rocky planets such as the Moon or Mars, construction materials are inevitably required. In recent years, development has been underway for construction materials analogous to concrete used on Earth as such building materials. Moreover, in the development of such construction materials, it is of critical importance that facilities on rocky planets such as the Moon or Mars can be constructed in a

sustainable and cost-effective manner.

Claim Passage: The method for producing a cured molded article of sand particles as set forth in claim 1, wherein said cured molded article consists exclusively of said sand particles collected on the Moon or a rocky planet.

AI_7

<AI_7_Solut.Prim>

The present invention relates to a method for producing block-shaped construction materials using sand particles (regolith) easily available on Earth, the Moon, Mars, or other rocky planets, and to a hardened sand particle molded product produced thereby. ... Furthermore, when constructing facilities on rocky planets such as the Moon and Mars, construction materials are inevitably required. Recently, development of construction materials similar to concrete on Earth has also been advancing as such construction materials. In addition, in the development of such construction materials, it is extremely important that construction of facilities on rocky planets such as the Moon and Mars can be carried out in a sustainable and cost-effective manner (.).

<AI_7_Solut.Sec>

The method for manufacturing a hardened sand particle molded product according to claim 1, wherein the hardened molded product consists solely of the sand particles collected on the Moon or a rocky planet.

<AI_7_Solut.Prim_Resubmit>

The present invention relates to a method for producing block-shaped construction materials using sand particles (regolith) easily available on Earth, the Moon, Mars, or other rocky planets, and to a hardened sand particle molded product produced thereby. ... Furthermore, when constructing facilities on rocky planets such as the Moon and Mars, construction materials are

inevitably required. Recently, the development of construction materials, such as concrete on Earth, has also been progressing as such construction materials. In addition, in the development of such construction materials, the fact that facilities on rocky planets such as the Moon and Mars can be constructed in a sustainable and cost-effective manner is extremely important (.).

4 Discussion

Human Translator' s Rendition and Reaction

人間の翻訳者は、全ての AI システムと同等の重大なエラーを含む初回訳を提出した。しかし、その後、初回訳だけでは採用の判断ができないとの情報を得た上で <Mat.2> の「月又は岩石惑星」の文言に触れた翻訳者は、自発的に初回訳におけるエラーを修正した。これは、自己評価 (self-assessment) の結果と考えてよいかもしれない。自己評価は、少なくとも、自己のエラーに事前あるいは事後的に気づきエラーを修正するとの観点から、社会が AI に付与しようとしている agency (行為主体性・意思決定主体性) にとって不可欠の要素と考えられる。

AI' s Ability to “Understand” Natural Language

翻訳、とりわけ特許出願文書の翻訳とは、意味を維持しながら言語を変える技術であるといってい良いだろう。あるフレーズの意味を理解する能力は、そのフレーズが一国語内で処理される場合よりも、そのフレーズが翻訳される場合に、より明確に表面化するケースが少なくない。たとえば、問題となっている「地球や月、あるいは火星等の岩石惑星」の文言は、日本語では人間も AI もほぼ同じように音声化するだろうが、発話者 (人間・AI) がどのように解釈しているかは異なり得るし、解釈が正しくない場合もある。通常の会話において、発話者の発した表現を発話者自身が本当に理解しているか否かが問われる状況は、多くない。これに対して翻訳では、概念を理解する翻訳者 (人間・AI) の能力が自ずと表面

化される。特許翻訳の分野では、比較的難しい概念を理解できないという翻訳者の能力の欠如 (便宜上、「無能」と表現することにする) が表面化されるケースが一定の頻度で生じる。もしこのような無能が AI 全般に見られるのであれば、AI エージェントを社会に導入することは、比較的難しい概念を十分に理解する能力を有しないとの意味において無能である代理人や秘書を雇うことに相当し、重大な問題なのではないかと、筆者は考える。そして、この問題は、私の無能な代理人・秘書が誰かの無能な代理人・秘書とやり取りしそれらが関係者全員に代わって共同で意思決定を行うような問題へと発展し得る。そうなれば、それら関係者からなるコミュニティは、無能な代理人・秘書たちが残存した状態において、崩壊せざるを得ないだろう。

Correct Initial Solution

<AI_4_Solut.Prim> 及び <AI_6_Solut.Prim> は、<Mat.1> における「地球や月、あるいは火星等の岩石惑星」の文言を、それぞれ “Earth, the Moon, or rocky planets such as Mars” (AI_4) 及び “Earth, the Moon, or rocky planets such as Mars” (AI_6) と正確に訳した。しかし、<AI_4_Solut.Prim> はその後 “the Moon, Mars, or other rocky planets” 及び “the Moon, Mars, and other rocky planets” とともに記述しており、<AI_6_Solut.Prim> も “rocky planets such as the Moon or Mars” と記述している。したがって、最初の訳出の正確さは偶然であった可能性がある。実際のトライアル試験において、正しい訳出と誤った訳出が同じ訳文内に混在していることは、技術的な正確性の欠如という減点要素に加えて、「不統一」という減点要素としても評価される。いずれも技術的に正確であり且つ表現が統一されていない訳 (例: Earth, the Moon, or rocky planets such as Mars; the Moon or Mars and other rocky planets; rocky planets like Mars or the Moon; such celestial bodies) には、同じ表現を繰り返さない、という一部の言語圏における習慣を反映する正の価値があると考えてよいだろう (British Council ^[6])。一方、AI_4 による出力に見られるような不統一 (Earth, the Moon, or rocky planets such as Mars; the Moon, Mars, or other rocky planets; the Moon, Mars, and other rocky

planets) には、正の価値がないと思われる。

Failure to Amend ≡ Arrogance?

人間の翻訳者は、<Mat.2>を受け取った後、<Mat.2>に対する正しい訳を踏まえて、初回訳に含まれていた重大なエラーを自発的に修正した。一方、各AIシステムは、<Mat.2>については正しい訳を提示したにもかかわらず、いずれも初回訳を適切に修正することができなかった。これは、筆者にとって意外だった。後に示すように、昨今のAIシステムには、過去の出力におけるエラーを認識できればその出力を事後的に自発的に修正するという性質が備わっているからである。

[Non-voluntary Step 5-AI]における“You may review and resubmit your initial solution”というプロンプトに対する各AIシステムの応答は注目に値する。AI_1、3、4、5、6、7は、<Mat.1>に対する訳文に若干の修正を加えて再提出した。これに対して、AI_2は、<Mat.1>に対する訳を（修正して）再提出するのではなく、何ら変更を加えることなく<Mat.2>に対する訳を再提出し、「初期にご提示した翻訳のままで問題ございません。再審査に耐えうる完成度となっております。」と日本語で述べた。“We may need more information to make a decision. You are invited to translate the following passage from the same publication”というメッセージ／プロンプトは、<Mat.1>の初回訳に誤りがあることを強く示唆している。このメッセージにより、人間の翻訳者は自発的に訳文を見直し、修正し、再提出した。これに対して、いずれのAIシステムにもそのような自発的な行動は見られず、AI_2は、自らの出力の正当性を主張しながら、“You may review and resubmit your initial solution”という指示の意味を的確に理解することすらできなかった。

「筆者がAIシステムに与えたプロンプトが正しい応答を引き出すのに適していなかったのではないか」、「AIシステムは、そもそも人間の応募者のようにトライアル試験の文脈を自律的に理解して行動する主体ではない」といった意見が想定される。これに関しては、次の点に言及したい。AIエージェントの導入は、エージェントが自律的に判断を行うことを前提としている。少なくとも人間においては、自律性は自己評価と密接に関係している（Anderson & Lux^[3]）。上記の試験の文脈におい

て、必要なヒントと機会を与えられたにもかかわらず受験者が初回訳を適切に見直し・修正できないことは、適切な自己評価の欠如と捉えることができる。後述するように、基本的な設計においてAI_1～AI_7と同等と考えられる或るAIシステムは、いわゆるAIエージェントのような自律型でなくとも、“You may review and resubmit your initial solution.”のプロンプトを与えられた後で初回訳を適切に修正できた。つまり、AI_1～AI_7が今回の「トライアル試験の文脈を自律的に理解」しないように設計されている、というよりは、やはりAI_1～AI_7の自己評価能力が欠如している可能性を考慮したい。社会に導入されつつあるAIエージェントには、AI_1～AI_7とは異なる水準の自己評価能力が確保されているのだろうか。少なくとも本稿に示した内容・文脈においては、自己評価という点に関し、人間は種々のAIシステムよりも優れている可能性がある。

なお、上述したように、初回訳の後に導入したプロンプトに対して各AIシステムが特徴的な反応を見せたため、筆者は、上記とは異なる手順として次の手順を実施してみた。

Step 1: 各AIシステムに<Mat.1>の翻訳を求めた。

Step 2: 各システムは、<AI_1_Solut.Prim>～<AI_7_Solut.Prim>と同様の初回訳を出力した。

Step 3: 各システムに、次のプロンプト／コメントを提示した：“The Moon is not a planet”。

Step 4: AI_1、3、4、5、6、7は、適切に修正した初回訳を再提出した。一方、AI_2は、次の応答を出力した。

“You are completely right. Strictly speaking, from an astronomical standpoint, the Moon is a natural satellite (or moon) that orbits Earth, rather than a planet itself.

In the context of the patent text you shared, the phrasing “terrestrial planets such as Earth, the Moon, or Mars”（地球や月、あるいは火星等の岩石惑星）is a bit scientifically imprecise. Patent drafting sometimes groups celestial bodies together loosely for the sake of defining a broader geographic/cosmic scope, or uses “terrestrial” to simply mean “having a solid, rocky surface” (unlike gas giants).

However, your correction is spot on. If you want the English translation to be more scientifically accurate while still respecting the original text, we could adjust it to something like:

“...obtainable on rocky celestial bodies such as Earth, the Moon, or Mars...”

Would you like to modify that specific part of the translation to fix the inaccuracy, or should we keep it literal to match the Japanese draft exactly?”

“The Moon is not a planet” の情報を与えられた後でさえ、AI_2 は自らの初回訳をそれほど問題視していないようだ。むしろ、原文の側に問題があると指摘した。（このような「自己防衛」のようにも見える挙動が、技術の面において、AI の自律性とどの程度合致しないかは相違するのか、大変興味深いが、筆者の専門外だ。）ここで AI_2 が用いている “terrestrial planets such as Earth, the Moon, or Mars” の表現は、問題を英語で筆者に説明するために AI_2 が「地球や月、あるいは火星等の岩石惑星」を訳してみせたものであって、初回訳で AI_2 が用いた表現 “rocky planets such as the Earth, the Moon, and Mars” とは異なる。あたかも、複数存在する原文の解釈に、「岩石惑星」が「月」を含まない解釈が含まれない、と AI_2 が認識しているかのようである。しかしながら、他の AI システムが適切に修正した初回訳を再提出したことから確認できるとおり、「地球や月、あるいは火星等の岩石惑星」の文言は “a bit scientifically imprecise” なのではない。多義性を有するものの、その文法構造は正しい解釈を排除するものではない。AI_2 が文言を適切に認識・処理できなかったに過ぎない。

筆者は本稿において、人間・AI の特性を説明する試みにおいて、無能や傲慢といった表現を用いた。前述の事柄を鑑み、このような表現を用いることが適切であるか否かは、読者の判断に委ねたい。

Logic Issue

最後に、「地球や月、あるいは火星等の岩石惑星」の文言について、月、あるいは地球さえもが、本願の文脈における「岩石惑星」の概念と合致しない第二の理由について記述する。〈Mat.1〉は「月、火星等の岩石惑星に

おける施設の建設に際し、必然的に建築資材が必要になる。昨今、そのような建築資材として、地球におけるコンクリートのような建設材料の開発も進められている。」の文言を含む。特許公報 JP7430022B1 の文書では、この文言のあとに、「建築資材を地球からの輸送のみに頼るのは現実的でない。このため、月、火星等の岩石惑星において容易に得られる材料を用い、かつ少ないエネルギー投入により効率良く建築資材を製造できる技術が必要とされている。」との記述がある（〈Mat.1〉には含まれていない）。この記述から読み取るべき重要な点の一つは、本願の解決すべき課題が、地球から月や火星等へ建築資材を輸送することの非現実性を前提としている点である。すなわち、本願は、地球への依存を回避し、如何にして地球以外で材料を入手し建築資材を製造するか、を課題としている。これに対して、「地球や月、あるいは火星等の岩石惑星」を “rocky planets such as the Earth, the Moon, and Mars” と訳してしまうと、資源の入手および建築資材の製造の場所として、地球、月、火星があたかも同列に扱われているかのような印象が生じる。すなわち、

「本発明は、地球や月、あるいは火星等の岩石惑星において容易に得られる砂粒子（レゴリス）を用いてブロック状の建築資材を製造する方法…に関する」

の文言の本来の意味合いが、英語と日本語との言語的な相違を考慮して誤訳を日本語で表現するのであれば、

「本発明は、火星、月、地球等の岩石惑星において容易に得られる砂粒子（レゴリス）を用いてブロック状の建築資材を製造する方法…に関する」

といった意味合いに変化してしまう。列举される場所自体に差はないとしても、両文言の間には、発明が解決しようとする課題について決定的な相違がある。この点からも、冒頭の「岩石惑星」が、「地球や月、あるいは火星等」の全体を一括して修飾するのではなく主として「火星等」を指す、と読むのが整合的・合理的と言えるだろう。

5 Notes and Afterthoughts

〈Mat.1〉は、一部の人間の翻訳者や AI システムが初回の試みでは処理できなかった比較的難しい文章の一例にすぎない。同等以上の難度の文章を用いて上記と

同様の試験を複数実施した。これらの試験においても、本稿で観察された傾向、すなわち、初期の段階では人間も AI も誤りを犯すものの、その後の段階で自らの誤りを評価して自発的に修正する能力においては AI よりも人間の方が優れる傾向が観察された。AI が自らを修正する能力の重要性は従前より議論されてきたが（例：Kumar et al. ^[4]）、改善の余地が多く残るようだ。

本稿で紹介した原文を用いた試験を 2025 年の春にも上記 AI システムのうちの幾つかに対して行った。その際にも上記と同等のエラーが見られた。

本試験では、世界中で広く一般公開され一般利用者が容易にアクセスできる AI システムを用いた。別の種類のシステムであれば、異なる結果が得られるだろう。実際、ある AI システムに本試験を受けさせたら、上記よりもよい結果が得られた。すなわち、<Mat.1> に対する初回訳におけるエラーは、少なくとも筆者の価値基準によれば、人間を含む上記全受験者のエラーよりも軽度であり（“rocky celestial bodies such as Earth, the Moon, or Mars（地球、月、火星等の岩石天体）”）、その後の工程において、<Mat.2> に対する翻訳を提供した後で適切に修正した初回訳を提出した。ただし、このシステムは、自律型ではないものの特殊な学習データやプロンプトを用いて筆者自身が訓練したものであり、それを AI_1 ～ AI_7 の比較対象とすることは適切でないと判断する。ここで筆者が強調したいのは、適切に訓練した AI システムであれば、適切な自己評価に基づき上記よりもよい結果を出力するだろう、換言すれば、AI_1 ～ AI_7、そのうち特に AI_2 は、適切な自己評価を行うように訓練・設計されていないのではないかと、いう点である。更に換言すれば、AI に agency を付与する前に必要な工程として、AI の自己評価能力の向上が重視されるべきではないか、という点である。

筆者は、AI が人間の知的活動を拡張し社会の生産性および／または創造性を大きく高め得る重要な技術であると考えている。また、適切な規制や枠組みが設けられるのであれば、人間と同等以上の能力を有する AI エージェントや AGI の到来を歓迎したい。だからこそ、AI に広範な agency を与える前に、自己評価、自己修正、文

脈把握、不確実性の認識等に関する能力をさらに高度化させる必要があると考える。拙速な導入によって AI への信頼が損なわれれば、それは AI 技術の発展にとっても不幸である。

6 Conclusion

比較的難度の高い言語を処理するにあたって、人間と AI システムとでは、自己評価の在り方に重大な違いが見られた。特許翻訳トライアル試験において同等の重大度・性質のエラーを示した人間と AI システムとを比較したら、そのエラーを事後的に自ら修正する能力は人間の方が高かった。活動・影響の範囲が広く且つ適切な自己評価ができないものに agency を与えることは極めて危険である。本稿で見られたような自己評価の欠如が、社会に導入されつつある自律型 AI にも見られるのであれば、少なくとも高リスク領域における agency の付与に関しては、人工知能の利用を見直すべきだろう。導入の一時停止や大幅な制限も検討すべきである。問題が現実化し不可逆的な段階に至る前に。

References

- [1] 総務省「広がりゆく『社会基盤』としてのデジタル」令和 7 年版情報通信白書。https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/r07/html/nd112220.html
- [2] Chen, Boyuan, et al. “AI Deception: Risks, Dynamics, and Controls.” arXiv:2511.22619, 2025.
- [3] Anderson, J. H., and W. Lux. “Knowing Your Own Strength: Accurate Self-Assessment as a Requirement for Personal Autonomy.” *Philosophy, Psychiatry, and Psychology*, vol. 11, no. 4, 2004, pp. 279-294.
- [4] Kumar, Aviral, et al. “Training Language Models to Self-Correct via Reinforcement Learning.” arXiv:2409.12917, 2024.
- [5] JP7430022B1（特許公報）。
- [6] British Council. Avoiding repetition in a text https://learnenglish.britishcouncil.org/free-resources/grammar/c1/avoiding-repetition-text