

生成AI時代の特許翻訳

—オープンな基盤モデルと専門知の未来—

Patent Translation in the Age of Generative AI



大学共同利用機関法人情報・システム研究機構 国立情報学研究所 所長

黒橋 禎夫

1994年京都大学大学院工学研究科電気工学第二専攻博士課程修了。博士（工学）。2006年4月より京都大学大学院情報学研究科教授、2023年4月より国立情報学研究所所長。2026年度から開始された我が国の AI for Science のフラグシップ事業 ARISE 戦略ターゲット型の推進統括（PO）を務める。

✉ kurohashi@nii.ac.jp

1 生成 AI は翻訳をどう変えたか

生成 AI の登場は、翻訳という営みに大きな転換点をもたらした。従来の機械翻訳は、ある言語を別の言語へ正確に変換することを主な目的として発展してきた。一方、現在の大規模言語モデル（Large Language Models: LLM）は、文章全体の意味を理解し、要約し、推論し、さらには利用者の目的に応じて生成する表現や文体を調整する能力を備えている。

この変化は、特許翻訳において特に大きな意味を持つ。特許明細書は単なる文章ではなく、技術的内容、法的要件、さらには権利範囲に関わる精緻な表現が織り込まれた専門文書である。その翻訳には、個々の文を翻訳するだけではなく、技術全体の理解、専門用語の一貫性、さらには発明者の意図や請求項の構造まで把握することが求められる。LLM はこうした高度な文書理解を得意とするため、翻訳支援にとどまらず、関連特許や論文との比較、専門用語の統一、技術内容の要約、さらには外国語文献の調査支援まで、知財業務全体を支える知識基盤となりつつある。翻訳は、生成 AI が担う多様な知的活動の一つへと位置付けが変わり始めている。

2 急速に進化する生成 AI

ChatGPT が公開された 2022 年末以降、生成 AI の進歩は予想をはるかに超える速度で続いている。2024 年には GPT-4o による画像・音声を含むマルチモーダル化、推論能力を強化した o1 の登場、DeepSeek に

よる高性能かつ低コストなモデルの公開など、基盤モデルそのものが大きく進化した。さらに 2025 年以降は、AI エージェントがソフトウェア開発や調査、文書作成など複数の作業を自律的に実行する段階に入った。

この流れの中で、翻訳は一つの独立した技術ではなく、AI が知識を理解し活用する能力の一部となった。例えば、外国語の特許を検索し、その内容を理解し、日本語で要約し、自社技術との差分を整理する一連の作業を、一つの AI システムが支援できるようになりつつある。

こうした能力は非常に魅力的である一方、海外企業が開発した巨大な基盤モデルに全面的に依存してよいのかという問題も浮き上がりつつある。学習データはどのようなものだったのか、なぜその回答に至ったのか、安全性はどのように評価されているのか、といった点は十分に公開されていない場合が多い。また、収益性の観点から、主要な対象は英語や世界市場で需要の大きい分野となり、日本語や日本固有の制度・文化への対応は必ずしも十分とはいえない。

知財分野においても、日本語の技術文書、日本の法制度、日本企業特有の表現を深く理解できる AI が求められる場面は少なくない。こうした背景から近年、「ソブリン AI」という考え方が世界的に注目されるようになってきている。

3 ソブリン AI とオープンな研究基盤

ソブリン AI とは、単に自国で AI を開発することを意味するものではない。自国の言語や文化、制度を深く理解し、その仕組みや性能を社会が検証できる形で維持・

発展させていく能力を持つことである。

国立情報学研究所（NII）が中心となって進めている LLM-jp および大規模言語モデル研究開発センター（LLMC）の活動も、この考え方に基づいている。我々が目指しているのは、世界最高性能だけを競うフロンティアモデルではない。実社会で十分利用できる高性能な日本語 LLM を継続的に開発するとともに、その学習データ、学習レシピ、評価手法などを可能な限り公開し、AI そのものを科学的に研究する Science of AI の基盤を提供することである。

現在、中国の高性能モデルがオープンウェイトとして公開されているが、学習データや構築方法までは公開されていない。そのため、利用はできても、「なぜその能力を獲得したのか」を科学的に検証することはできない。AI の透明性や信頼性を高めるためには、モデルだけではなく、その背後にある研究プロセスまで共有することが必要である。

LLM-jp/LLMC では日本語能力や日本文化への理解を高めることにも取り組んでいる。これは単に日本国内だけを対象とするという意味ではない。ソブリン AI は一国のみで完結するものではなく、多様な言語・文化を持つ国々がそれぞれの知見を持ち寄り、相互に連携することによって初めて発展するものである。現在、我々は韓国、インド、タイ、カンボジア、フランス、イギリス、アイスランド、ブラジルなどとの協力も進めている。オープンな研究基盤を通じて、多様なソブリン AI が相互に連携するエコシステムを構築することこそ、今後の生成 AI の健全な発展につながると考えている。

4 生成 AI 時代に 特許翻訳が直面する課題

生成 AI は人間の知的作業の効率や品質を飛躍的に向上させつつある。一方で、その普及は我々の知的営みそのものに新しい課題を投げかけている。特許翻訳もその例外ではない。その課題は大きく三つに整理できるだろう。

第一は「情報の爆発」である。生成 AI は知識を生み出すコストを劇的に下げた。論文、技術報告書、特許明細書、さらには調査レポートまで、AI を活用して短時間で作成できるようになりつつある。今後、世界中で技術文書が爆発的に増加することはほぼ間違いない。これは翻訳の需要がなくなることを意味しない。むしろ逆である。

翻訳し、理解し、比較しなければならない文書の量は飛躍的に増える。これからは「何を読むべきか」「何を翻訳する価値があるか」を選択することが重要になる。生成 AI は翻訳そのものだけでなく、膨大な情報の中から価値ある知識を見いだす役割も担うようになるだろう。

第二は「情報の信頼性」である。生成 AI は流暢な文章を生成する。しかし、その文章が事実として正しいかどうかは別問題である。さらに今後は、人間が書いた文書だけではなく、AI が生成した文書を AI が翻訳し、それを別の AI が要約するという状況も珍しくなくなる。そのとき我々は、どこまでその情報を信頼できるだろうか。

特許文書は、権利や技術を正確に伝達することが求められる文書である。一つの訳語や一つの表現の違いが、技術的解釈や法的解釈に影響を及ぼす可能性もある。その意味では、AI による翻訳結果を無条件に受け入れることはできない。

今後ますます重要になるのは、情報の来歴（プロビナンス）である。どの文書を学習し、どの AI が生成し、どのような過程を経て現在の文章になったのか。その履歴を適切に管理し、必要に応じて説明できる仕組みが求められる。生成 AI 時代には、「何を翻訳したか」だけではなく、「どのような知識を基に翻訳したのか」が問われるようになるだろう。

そして、第三の課題が「専門知の継承」である。翻訳とは、単なる言語変換ではない。翻訳者は対象分野の技術を理解し、背景知識を学び、用語の揺れを判断し、読み手にとって最も適切な表現を選択している。その過程では、多くの経験や暗黙知が蓄積される。生成 AI が高性能になればなるほど、人間はその過程を経験しなくなる可能性がある。最初から AI がかなり完成度の高い翻訳を提示してくれるようになると、若手の翻訳者がまず自ら訳し、試行錯誤し、徐々に専門性を身につけていく機会そのものが減ってしまう。

これは翻訳だけの問題ではない。研究における論文執筆や査読、医療の診断支援や治験計画、ソフトウェア開発など、多くの専門職で同じ問題が起き始めている。AI が優秀になるほど、人間が専門性を獲得する機会が失われる、いわば、「社会が専門知を再生産できなくなる」という問題が生じている。翻訳もまた、長年にわたり培われてきた専門知の一つである。その知識を AI が利用することは望ましいが、その知識を次世代へどのように

継承していくかという視点を忘れてはならない。

5 AI 時代に本当に守るべきもの

生成 AI の能力は、これからも着実に向上していくだろう。翻訳品質もさらに改善され、特許調査や文書作成、技術分析など、多くの知財業務が AI によって支援されるようになる。そのとき、我々は何を目指すべきか。最も重要なことは前節で第三の課題として指摘した「専門知の継承」、すなわち、人間の知性と専門知を社会の中で持続的に育てていくことであろう。

人類はこれまで、専門家を育てることによって社会を発展させてきた。研究者は研究者を育て、医師は医師を育て、技術者は技術者を育ててきた。その過程では、教科書だけでは伝えられない暗黙知や経験が受け継がれてきた。翻訳もまた同じである。優れた翻訳者は、辞書だけを見て訳しているわけではない。技術の背景を理解し、文脈を読み、読み手を意識しながら、一つ一つの表現を選択している。その判断は長年の経験の積み重ねによって支えられている。

AI がこのような専門知を利用することは、社会全体にとって大きな価値がある。しかし同時に、その専門知を生み出す人間が育たなくなれば、さらにいえば AI への過度な依存により人間の知性が低下するようなことがあれば、AI 自身も長期的には発展できなくなる。AI は既存の知識を学習することはできる。しかし、新しい知識を創造し、社会の変化に応じて専門知を更新していく主体は、依然として人間であろう。AI が優秀になるほど、人間はより高度な問いを立て、より本質的な判断を担うことが期待される。その意味で、AI は人間の代替ではなく、人間の知的活動を支える基盤として位置付けるべきである。

この考え方は、我々が LLMC でオープンな研究開発を進めている理由とも深く関係している。我々が公開しているのはモデルだけではない。学習データ、学習方法、評価方法、さらには失敗や改善の過程まで含めて共有しようとしている。その理由は、AI を社会全体の知的基盤として育てたいからである。ブラックボックス化された AI は便利だが、便利さだけでは専門家は育たない。なぜその性能が得られたのか、どのようなデータで学習したのか、どのような限界があるのか、それらを研究者、企業、利用者がともに理解し、改善していくこと

によって初めて、AI は社会の知識基盤として成熟していく。このような営みこそが Science of AI の本質である。AI を単なる利用対象ではなく、科学として理解し、社会全体で育てていく。そのためには、モデルの透明性や信頼性を高めることはもちろん、それを支える人材を継続的に育成することが欠かせない。

特許翻訳の世界も例外ではない。AI が翻訳を支援する時代だからこそ、翻訳者はより高度な技術理解、より深い法的理解、そしてより豊かな言語感覚を身につけることが期待される。AI が翻訳を行い、人間が最終判断を行うという役割分担だけでは十分ではない。AI を適切に評価し、改善し、新しい知識を与え、社会にふさわしい形で利用していくことも、これからの専門家の重要な役割になる。翻訳という営みは、異なる言語の橋渡しであると同時に、異なる知識体系や文化を結び付ける営みでもある。生成 AI は、その橋をより太く、より速く渡れるようにしてくれるだろうが、橋の向こうへ何を運び、何を未来へ残すのかを決めるのは、これからも人間である。

生成 AI 時代に本当に守るべきものは、人間の知性であり、その中でも長い時間をかけて育まれてきた専門知である。AI は、その専門知を置き換える存在ではなく、人類が培ってきた知の営みを次の世代へ受け渡していくための新しい知的基盤となることを期待したい。そして特許翻訳という営みもまた、その知の継承を支える重要な橋渡しであり続けるだろう。

