

改訂にあたり

1980年のノーベル化学賞に輝いたFrederick SangerとWalter GilbertのDNAシーケンシング法は、生命の設計図を「読む」技術の礎を築き、次世代型DNAシーケンサー（NGS）へと進化を遂げました。そして2024年、膨大なデータを「理解する」技術として、タンパク質立体構造予測AIのAlphaFoldを開発したDemis HassabisとJohn Jumperが化学賞を、ニューラルネットワークの基礎を築いたJohn J. HopfieldとGeoffrey Hintonが物理学賞を受賞したのです。これは、生命科学における「読む」技術から「理解する」技術へという歴史的な転換点を象徴する出来事でした。

一方で、これらの技術を誰もが簡単に利用できる時代だからこそ、新たな課題も浮上しています。先日、自由研究のため私たちの研究室を訪れた高校生グループが、温泉水中の微生物を対象としたメタゲノム解析に挑戦しました。彼らはPythonプログラミングはおろか、メタゲノム解析に関しても全くの初心者でした。われわれが指導したのは、ChatGPTとGoogle Colaboratoryの基本的な使い方のみ。しかし驚くべきことに、彼らはこれらのツールを巧みに操り、データを解析し、結果を可視化することに成功したのです。この時点では、生成AIの可能性に改めて驚かされると同時に、科学技術の民主化の力を実感しました。

しかし、ここで私たちは科学における本質的な課題に直面することになります。プログラムにバグが発生した際、彼らはその原因を特定することも修正することもできず、さらに重要なことに、可視化されたグラフの解釈や結果の生物学的な意味を理解することができなかつたのです。これは、生成AI時代の科学における象徴的な出来事かもしれません。

AlphaFoldが示したように、AIは確かに人間の直感を超越する発見をもたらします。一方で、ゲノム科学における代表的な推論方法であるアブダクション（仮説形成）において、AIが提示する解析結果や仮説が生物学的文脈で本当に意味をなすのか、実験的検証に値するのか、を判断するには、依然として深いアルゴリズムの理解と生物学的洞察が不可欠です。特に、マルチオミクスと称する膨大かつ多様なデータ群の解析では、使用するアルゴリズムの本質を理解していないと、AIが提案した解析結果が誤った結論へと導く危険性すらあります。

本改訂では、このような歴史的転換点を背景としつつも初版を踏襲し、単なるツールの使い方にとどまらず、アルゴリズムとその理論的基盤を深く理解することを重視しました。読者が生成AIや新しい解析技術を活用しつつ、それらの背景にある科学的根拠を理解し、独自の視点で問題を解決する力を育むことをめざしています。初版で掲げたD. E. Knuth著『The Art of Computer Programming』の理念をさらに発展させ、現代の科学者に求められるスキルを提供する一助となり、本書が次なる科学の躍進を支える基盤となることを願っています。

最後に、国立遺伝学研究所 生命情報・DDBJセンターの皆様、ライフサイエンス統合データベースセンター (DBCLS) の皆様、先進ゲノム解析研究推進プラットフォームの皆様、さらには私どもと哲学を共有し忍耐強く改訂版の出版に導いていただいた羊土社の皆様に心より深く感謝申し上げます。

2024年11月

編集を代表して
黒川 顕