

AI for Science推進委員会（第1回）での主なコメント等

<重点領域>

- AIは世界中で資金が流れ込んでくる領域であり、同じ領域で10倍の資金を使っているところでは日本は勝てない。日本がパワープレイできる領域の設定が必要（例えば、海洋生物学や食）。一方、創薬やマテリアルなどのパワープレイできないがやらないといけない領域では、どう戦略を立てるのが重要。※一見ハイインパクトな応用がなさそうに見える場合でも、どのように理解を得るかの戦略が必要
- 極めて先駆的な観測・実験を行う研究や、大量の高精度データを創出するノウハウや方法を有する研究等もプロジェクトとして考慮すべき。

<研究データ>

データの共有

- データの共有をどのくらいするかしないか、どのタイミングにするか、は大きな意思決定になる。
- （技術的に）実験装置から取得したデータは、チャンピオンデータだけでなくネガティブデータも、メタデータと一緒に全て自動でクラウドに保存し、機械学習等ができるようになるとうい。
- （研究者の意識的に）研究データの公開・共有に対する研究者のモチベーション向上が不可欠。そのためにデータの公開・共有が論文執筆・投稿と同等の実績として認められるような文化醸成や、研究者のデータ共有に対してインセンティブを与えるような人事評価を含めた制度設計が必要。
- AI for Scienceという枠組みの中で、民間企業も含めたデータ共有に関するエコシステムの設計が必要（データ提供の対価等を含めたエコシステムの設計が重要）。
- データ共有のためのリポジトリを国として整備することが非常に重要。
- 研究者にとってクローズなデータを出すことは非常にストレスなので、まずは公知データでAIを学習させ、有用なモデルが構築できれば、クローズな環境で独自のデータを学習させるというスキームが考えられる。

データの取得

- AI向けデータ（網羅性のあるデータ、コンテキストや装置マニュアル等のメタデータ含む）が不足している。
- AIのためにどういうデータを取得できるかを重視することが重要（特定目的のために取得されたデータはバイアスを含むことがある）。
- 研究データの整理・キュレーションをする者に対する評価体制や、研究費支援等のインセンティブ設計が必要。
- データセットの構築をどのようにAI駆動で行うか、研究コストを減らせるかの観点が重要。

高価値なデータ

- 極めてトップレベルの科学（今までにないような観測、シミュレーション、計算、実験等を組合せ）による質の良いデータを、世界に先駆けて創出することが重要。

<人材育成>

- AIに関するノウハウはないが、危機感とやる気を有している若手の層を拾い上げる必要がある。
- AIを前提とした知識や手法をいかに教育し、どうやって使いこなしていけるようにするか。

<AIの影響>

- 生成AIによる論文作成補助が受け入れられ始めている（AIによる論文の生成自体は禁止だが、リライトはOK等）。
- 生成AIによって、他分野の知識を容易に利用することが可能になり（異分野融合を通じて）新たな分野が創出される展開が予想される。そのような状況において、日本にAIEージェントをホストできる環境がない点（海外にデータを提供せざるを得ない状況）が問題。

<体制等：産学連携>

- （アカデミアの）プレイヤーが足りない中で民間企業の力は絶対必要になる（AlphaFoldが民間企業から出てきたことが衝撃。特にバイオインフォマティクスの主要なAIアプリは民間から出てくることが多い）。
- AGI含めてAIの能力を上げるためには、全産業領域に影響を与え得る、サイエンスができるAIが必要（海外ビッグテックも、AIを賢くする上でAI for Scienceが一番のキーであると認識し注目している）。
- 科学とビジネスの近接化があらゆる産業で起こっており、科学そのものがビジネスになりつつある。トップ科学者のためのAIというより、トップ層の科学者が実践していることをミドル層やボトム層でも実践できるようなAI for Scienceを重要視すべき。

<その他>

- AI for Scienceはいくつかのレイヤーがあり、その一つに科学者が使うAIツール（例:AlphaFold）を開発するレイヤーがある。これが発展すると、AI自体が科学者になり、ロボティクスも活用しつつ、人間の限界やバイアスを超えて自律的な研究が可能になる。