

資料5

科学技術・学術審議会 情報委員会
情報科学技術分野における戦略的重要
研究開発領域に関する検討会（第2回）
令和6年5月14日

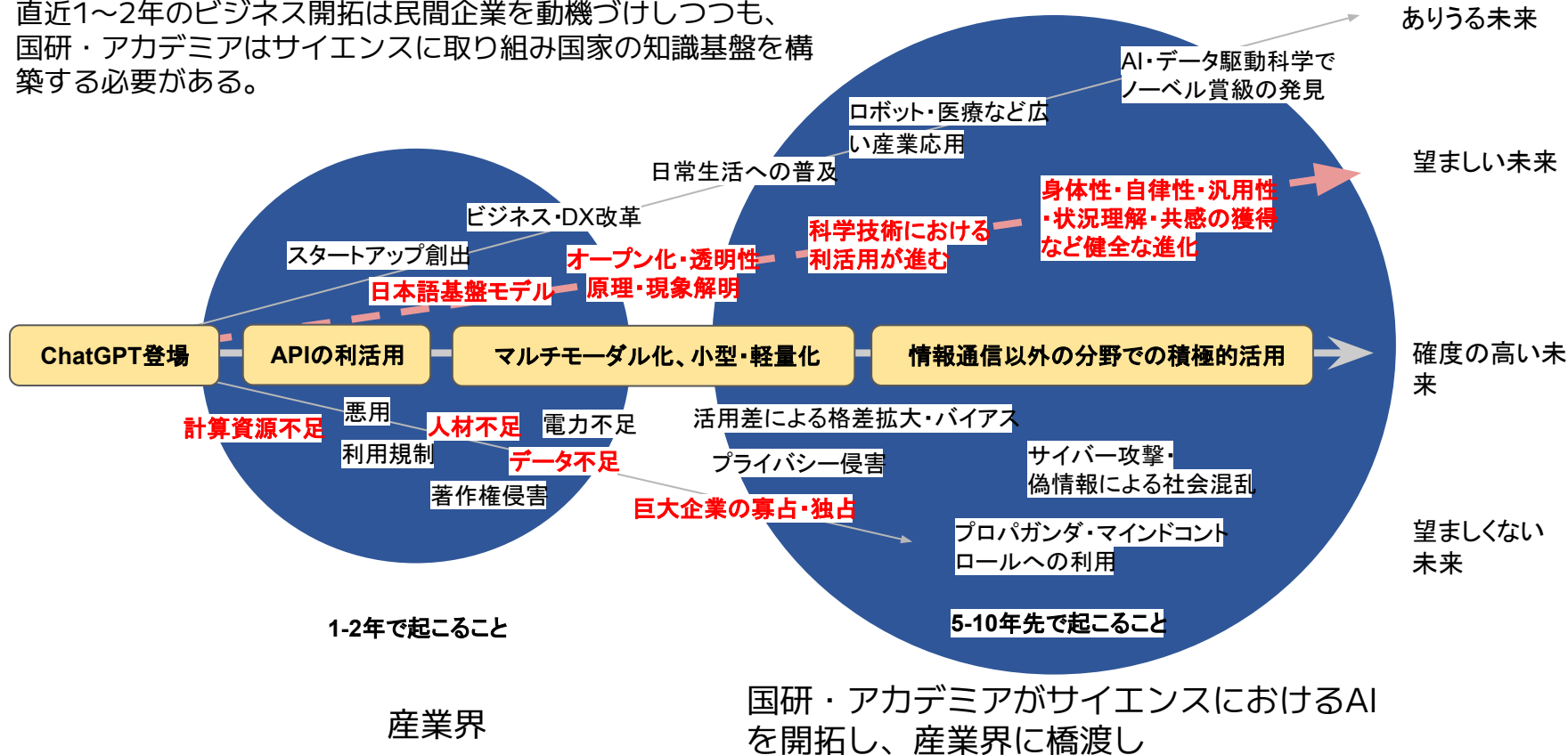
情報科学技術分野の最新動向等を調査する検討会

生成AIの登場と情報学の諸分野の役割

東京大学 工学系研究科 教授
川原圭博

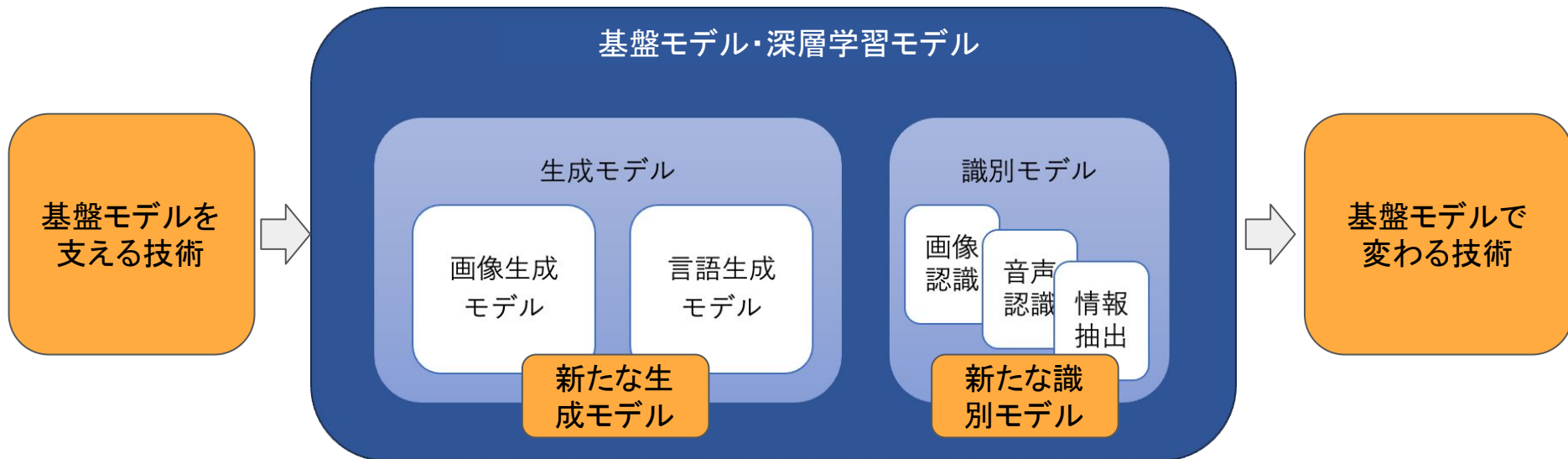
2023年に調査した、5年先を見据えた産業界とアカデミアの役割分担と連携

直近1~2年のビジネス開拓は民間企業を動機づけしつつも、
国研・アカデミアはサイエンスに取り組み国家の知識基盤を構築する必要がある。



謝辞: 2023年に本資料の作成に当たりご助言を頂いた多くの方々に感謝いたします

2024年現在、生成AIの登場により情報学の諸分野も活性化している



主にオレンジの領域が拡大している

生成AIの登場とそれを支えた情報技術

- 大規模言語モデルのような生成AIは、情報学の諸技術が組み合わさった技術であり、自然言語処理分野だけの問題ではない。
 - 核となる技術
 - 深層学習技術や自然言語処理技術・コンピュータビジョンなど
 - 生成AIに不可欠な周辺技術
 - データサイエンス・統計
 - Web・データベース技術
 - 高性能計算、GPUといった半導体技術、高速ネットワーク技術（計算基盤）

大規模言語モデルを更に良くするための研究(1)

- アラインメント(倫理・安全性のため)
 - 好ましい受け答えをさせるための技術。学術的には強化学習や最適化。
 - 人手によるデータセットの構築も重要。倫理学・社会学連携テーマとして有望。
 - GPTは学習アルゴリズムの開発よりもデータ生成の方がずっと人数が多いとも言われている。
- 品質を保った軽量化(経済効率改善のため)
 - データの巨大化とモデルの巨大化は(先行グループの中では)ひと段落している。
 - データの品質向上: 良質なデータを使い、小さなモデルで高い性能を出すこと
 - モデルのQuantize(量子化): 重み表現のビット幅を減らす。[BitNet](#) (Microsoft)
 - Pruning(枝刈り): 影響の小さいノードを削除する。
 - Distillation(蒸留): 大きいモデルやアンサンブルモデルを教師モデルとして、その知識を小さいモデル(生徒モデル)の学習に利用する。
 - Mixture-Of-Expert: 巨大モデルでなく小さなモデルの集合を使う技術。[Mixtral](#) (Mistral)

→これらの工夫・改良はLLM・生成AIのみならずAI一般の改良でもあり影響が大きい

大規模言語モデルを更に良くするための研究(2)

- 効率化 (Attentionの代替機構の探索)
 - State-Space-Model
 - トランスフォーマーのAttentionはトークン長に対して二乗の計算量がかかるのでこれをいかに線形オーダーに近くさせるかという研究が加速 (例: Hyena, Mamba, RetNet, TransformerFAM)
- LLMの理解
 - LLM内部の機構を実験的・理論的に理解する研究も進む。 (実験的アプローチ、理論的アプローチ)
- 分散化
 - 並列処理 (スパコン)、エッジAI (5G, 自動運転)、連合学習 (プライバシー保護) と異なる文脈で分散化の方法が様々に検討されている

基盤モデル・深層学習をうまく使うことで進展する他の研究分野

- **ライフ・マテリアル**:タンパク質の構造解析、様々なマテリアルの発見の応用は数多くある。論文を読み込んで実験を行う、ロボット駆動科学も盛んに。
- **ロボット・IoT**:ロボット自体の行動計画、制御、状況認識
- **設計・証明**:半導体や回路設計、数学の定理証明
- **通信**:通信環境のモデル化、通信制御、管理
- **HCI・教育・医療**:対話型インタフェース、長期かつ大規模なデータ収集

→ とにかくすべての情報分野に急速にLLM・生成AI・深層学習の波及効果が出初めている状況。マテリアル、ライフ、通信、ロボティクス、エンターテインメントなどで主役交代もありうるのでは(私見)。

今必要なこと

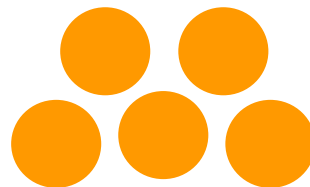
情報学のすべての分野で、どのように生成 AI に貢献できるか、生成 AI をどのように利用するか の検討が進んでいる。

生成 AI 関係で国内にも特徴ある研究拠点が形成されつつあり、連携が期待される。

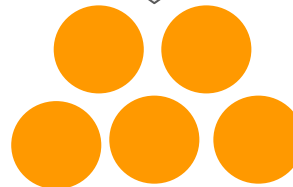
- 理研AIPでは様々な理論研究が進んでいる
- LLM開発の直接的な開発はLLM-jpのNIIを中心としたグループが健闘している
- LLMに関する"研究"となるとLLM-jpの活動と別に、企業を含めて広く行われている
- サイエンス分野への応用については理研R-CCSが行っている

LLMを支える研究、LLMで伸びる研究にも投資するタイミングになっている。

LLMを支える情報技術

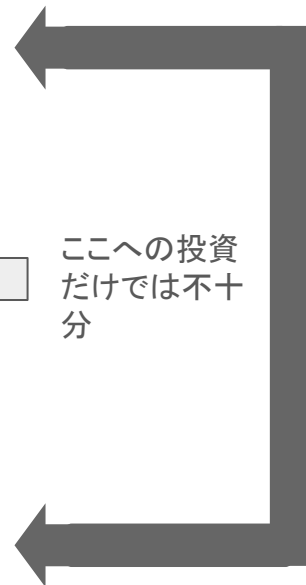


LLM・生成AIを含む
基盤モデル・深層学習



LLMで進展する
様々な技術

LLM・生成AIから見た、川上・川下への投資がより重要になる



ここへの投資
だけでは不十分