

情報委員会（第48回） 議事録

https://www.mext.go.jp/b_menu/shingi/gijyutu/gijyutu29/siryu/1418998_230726_00015.html

記事ページ本文

現在位置

トップ

>

政策・審議会

>

審議会情報

>

科学技術・学術審議会

>

情報委員会

> 情報委員会（第48回）議事録

情報委員会（第48回）議事録

1. 日時

令和8年7月30日（木曜日）16時00分～18時00分

2. 場所

オンライン

3. 議題

AI研究開発力の強化について

「生成AIモデルの透明性・信頼性の確保に向けた研究開発拠点形成事業」の実績評価について
AI for Scienceを支える研究データの管理・利活用と流通の在り方について（審議まとめ）（報告）
今後の情報科学技術分野における取組方針について

情報分野研究開発プランの改定について

研究開発課題の事前評価について（非公開）

4. 出席者

委員

相澤主査、尾上委員、湊委員、青木委員、天野委員、石田委員、大武委員、川原委員、川添委員、小林委員、中野委員、引原委員、星野委員、盛合委員、若目田委員

文部科学省

阿部 参事官（情報担当）、栗原 計算科学技術推進室長、山本 学術基盤整備室長、

鈴木 科学官、込山 学術調査官、國本 学術調査官

オブザーバー

国立情報学研究所（NII）黒橋 禎夫 所長

5. 議事録

【相澤主査】 それでは定刻になりましたので、科学技術・学術審議会情報委員会の第48回会合を開催いたします。委員の皆様におかれましては、お忙しいところお集まりいただきありがとうございます。

本日はオンラインでの開催としております。まず、出席者の御紹介、配付資料の確認とオンライン開催に当たっての注意事項について、事務局よりお願いいたします。

【池田参事官補佐】 事務局でございます。

まず、本日の出席状況について御案内いたします。本日、オンラインにて総委員数16名に御参加いただいております。充足、満たしておりますので、御報告いたします。

また、国立情報研究所より黒橋所長にオブザーバーとして御参加いただいております。よろしくお願いいたします。

続きまして、議事次第に基づきまして、配付資料を確認させていただきます。ダウンロードいただい

ている資料を御確認ください。資料1、2、資料3 - 1、3 - 2,3 - 3、4、5,6 - 1、6 - 2ということで配付させていただいております。参考資料は1から4ございます。もし現時点でお困り事や不具合等ございましたら、お知らせいただければと思いますけれども、皆様いかがでしょうか。何かあれば挙手等お願いいたします。よろしいですか。

では、続きまして、ハイブリッド開催に当たっての注意事項を申し上げます。画面でも投影いたしますが、特に以下、御留意いただきますと幸いです。御発言時を除きまして、マイクは常にミュート、ビデオは常時オフとしていただきますと幸いです。ただし、御発言いただく際のみ、Webexの挙手ボタンを押してください。御発言される際はマイクをオンにした後、お名前をおっしゃってから御発言いただきますと幸いです。不具合等発生した場合は、チャット機能や電話等で事務局まで御連絡ください。

事務局からの御案内は以上でございます。

【相澤主査】ありがとうございました。本日はお手元の議事次第にあります6件の議題を予定しております。議事6、研究開発課題の事前評価につきましては、科学技術・学術審議会情報委員会運営規則第5条第3項、参考資料1となります、こちらに基づきまして、非公開とさせていただきます。

それでは、まず、議事1、AI研究開発力の強化について。文部科学省より御説明をよろしくお願いいたします。

【阿部参事官】情報担当参事官の阿部でございます。資料1を御覧ください。AI研究開発力の強化に向けてということで、副題として、Science for AI / Science of AIの推進と記載させていただいております。

情報委員会でもこの1年、AI for Scienceの推進について様々議論いただきまして、今年の3月には「AI for Scienceの推進に向けた基本的な戦略方針」というものを文部科学省としてまとめさせていただきました。その後、それを踏まえまして、政府の各種文書、例えば、第7期科学技術・イノベーション基本計画や、AI基本計画改訂、それからこれからの予算要求に向けて重要となります成長戦略や骨太の方針、そういった政府文章にも、AI for Scienceの推進がしっかりと位置づけられた状況でございます。そうした中、我々もいろいろなところにお話ししますと、AI for Scienceは非常に重要だと。ただ、それと合わせて、Science for AIやScience of AI、こういったものもしっかりと取り組む必要があるという御指摘をいただいているところでございます。そういった状況も踏まえまして、今どういう状況にあるか少しまとめてみましたので、今日はその状況についての御報告の資料を御覧いただければと思っています。

2ページ目を御覧ください。JSTCRDSから資料提供いただいているものですが、今、世界のAIに関する情勢ということで、AIへの民間投資、AIモデル数、計算資源においては米国がトップであるけれども、AIに関する論文生産性では中国が米国を抜くような状況になっているということ。3ページ目のように、トップモデルへのAIの性能比較というものも、このような状況に今なってきているということです。

4ページ目、スタンフォード大学が近年毎年出していますAI Index Reportの最新版でございますが、この中では、例えば研究開発分野の主なポイントとして、最先端のモデルの透明性が低下しているであったり、また、情報開示が減少してきているとか、データ品質向上や事後学習技術に有望な成果が見られるといったこと、また、基盤の観点で言えば、世界の計算能力は年平均3.3倍で増加しているといったこと、併せてAIの環境負荷が増加しているという指摘であったり、オープンソースAI開発は引き続き拡大していると、そのような指摘がなされているところです。

また、Index Reportの中では、科学分野におけるAI活用の主なポイントが次のページに記載されております。ここではAI関連の科学論文数は年々増加している一方で、公開研究の再現性には依然課題があるといった指摘がなされています。また、一番下にありますとおり、科学向けのAIモデルの多くは学術機関や政府機関によって開発されているという指摘がなされているという状況でございます。改めまして、AIの時系列俯瞰図を6ページ目に入れてございます。2020年代以降、第4次ブームと言われているところですけど、2025年、2026年と、AIの大衆化からさらにAIの社会浸透、それによるAIの格差というものが指摘されている状況でございます。3つの潮流がございますけれども、AIトランスフォーメーションと併せてリスクへの対応及びAI基本原理の発展というところをしっかりと見る必要があろうかと思っております。

7ページ目を御覧ください。この辺りの研究開発領域を俯瞰した図でございますが、説明は省略させていただきます。

こういった世界情勢の中、現状認識というものを少し書かせていただきました。急速なAI技術の進展に伴いまして、モデルの解釈可能性や透明性、また推論コストの増大の問題、安全性や透明性の確

保、さらには経済安全保障上のリスク等が今、顕在化してきているという指摘が多々なされているという認識です。そうしますと、AI for Scienceに加えまして、科学的知見を活用して新たなAI技術を創出するScience for AI、またAIの原理や能力・限界、安全性・信頼性を科学的に解明するというScience of AI、こういったものの重要性が社会的にも高まっていると思います。

そうしますと、AI自体をしっかりと理解し、高度化を図り、AIイノベーションを創出していくということも非常に重要なものだとこのところ、ここが非常に今、重要なタイミングにきているのではないかと考えているところです。AIを理解し、安全性、信頼性を確保しながら、自律的かつ先導的に研究開発を行う能力、こういったものを国としてもしっかり確保、強化していく必要があるのではないかと考えているところです。

今後こういったことに検討しなきゃいけないのか、Science for AI及びScience of AI、こういったものを進めるに当たっては、日本の強みや特色はどういうものなのか、日本の立ち位置はどういう状況にあるか。それから日本が国として力を入れるべきところはどこなのか。さらには、当面の課題と中長期的な方向性は何かということを入材育成や研究基盤、産学連携、国際連携も含めまして、全体としてのエコシステムをどう考えるのかを意識しなければならないと考えているところです。

それから、冒頭から、Science for AI、Science of AIという言葉を使っておりますが、大まかな概念をイメージ案として、少し書かせていただいたのがこちらです。AI for Scienceにつきましては、科学のためのAIということで、AIを活用して科学研究を加速するというところであります。一方、Science for AIにつきましては、AIのための科学ということで、科学の知見を活用してAIを向上させる取組と言えないのではないかと思います。また、Science of AIにつきましては、AIそのものを科学的に理解、解明する取組と言えないのではないかとこのところ、少し概念的な整理を試みているというところでございますが、しかしながら、この3つは一体的に進める必要がございますので、3者の好循環を戦略的に推進するということが不可欠になってくるのではないかと考えているところです。

以降、少し世界情勢等についての参考資料になりますけれども、いろいろ調べてみますと、こういったキーワードが出てくるといってものを並べております。Science for AIの観点で言いますと、(1) AIの基盤理論や基盤技術という観点、それからAIの高度化、次世代AIという切り口、それからScience of AIとしましては、AIの理解、解明、それからAIの信頼性、そしてAIと人や社会との関係、こういった観点があるのかと思っております。

それから、国際動向等ということで12ページ目になりますけれども、AIそのものを科学的に理解、高度化するという、そういった研究の取組が今、各国もそうですけれども、特に民間企業を含めて様々なことが指摘されているという状況かと思っております。例えば、下の(1)とありますけれども、スケーリングといったところの競争からAIの原理の競争へということで、学習原理、アーキテクチャーの革新への回帰といったものが指摘されているところかと思っております。また、(2)としてAIをブラックボックスとして利用する、そうした段階から、AIを科学的に理解、評価する段階にきているという指摘もございます。さらには(3)にあります、フロンティアAI企業もAIの基礎科学、評価科学への投資を強化しているというところが、2026年現在の動きかと捉えているところです。

それから、各国いろいろ国家戦略、計画等を出しておりますが、その辺を少し並べて見ているのが、13ページ目と14ページ目でございます。この中で、特に注目したい、御紹介したいのが15ページ目、アメリカのところにあります。America's AI Action Planというものが2025年、昨年度出されておりますが、この中をよく見ますと、Advance the Science of AIという項目の中に、次期国家AI研究開発戦略においては、AI研究における根本的な進展を優先事項とすべきということで、まさにScience for AI、Science of AI、こういったところを優先的に考えなきゃいけないという指摘がこの中に盛り込まれているところです。

それから、最近の動向として16ページ目に少し載せておりますけれども、イギリスでは2つのAI研究ラボを立ち上げたという報道がございます。こちらにつきましては、現在のAIの限界を超える次世代AIの実現を目指すということで、AIの基盤となる基礎科学を推進するネットワーク型の研究拠点が設けられ、6年間で最大6,000万ポンドを投資することが発表されているところです。2つございますが、特に左側のところ、基礎AI科学研究ラボと日本語では書いていますが、こちらの取組では、UCLを筆頭に、主要大学が協働する形で取組を進めるということで、ハルシネーションや特定企業への依存といった、そういった根本的な課題を解決するために、現行のトランスフォーマーを越えるような新しいアーキテクチャー、低計算資源で動くようなAI、AIをよりアクセスしやすくする基盤の技術、またオープンなAI基盤技術、そういった設計や理論構築を目指すような研究を推進することが打ち出されているというところがございます。こういった世界の動向、それから国内のアカデミアの取組等もよく見ながら、今後、国としてこの辺りをどうしていくのかということは、しっかり文科省として

も考えていかなければならない課題であると現在考えているところでございます。

以降のページ、最近の政府文書でこう書かれていますというところを少し抜粋しているもので、御参照いただければと思います。

私からの説明は以上になります。

【相澤主査】ありがとうございました。

ただいまの御説明を踏まえまして、もし御質問、御意見等がありましたら挙手にてお知らせください。よろしいでしょうか。

では、続きまして、議事2に移ります。ちょうどただいまの議事1にてAI自体の研究開発の重要性について御指摘をいただいたところでございますが、議事2は、生成AIモデルの透明性・信頼性の確保に向けた研究開発拠点形成事業の実績評価につきまして、国立情報学研究所、黒橋所長より御説明をいただきます。よろしくお願いいたします。

【黒橋所長】NIIの黒橋でございます。よろしくお願いいたします。

今、非常にサポート的な文脈を御紹介いただきありがとうございます。生成AIモデルの透明性・信頼性の確保に向けた事業につきまして、昨年度も実績報告させていただきましたけれども、令和7年度の活動を中心に行っておりますことを報告させていただきたいと思います。

次のスライドをお願いいたします。まず、この事業の基本方針と社会的意義を確認させていただきたいと思います。このプロジェクトでは、措置いただいた予算、それからGPU資源の範囲の中で、できる限り高性能なLLMを開発するというのもちろんでございますけれども、それにとどまらず、学習データや学習レシピを公開し、また、まさに今ありましたScience for AIに資する透明性の高い研究基盤を構築するということを目的としております。この活動を通じてAI研究開発人材を育成すること、また、世界のソブリンAIの開発とも連携し、我が国がAIを主体的に、かつ信頼できる形で研究開発・活用していくための基盤をつくるということを目的としています。

先週にはオープンAIの開発中のモデルが、エージェントが環境を抜け出してハッキング・フェイスに不正侵入してしまうということも世界的に非常に大きなニュースになっております。やはりブラックボックスでの研究開発だけではなく、透明性の高い研究開発を行うことが本当に人類にとって大事だと、そういう問題意識で取り組んできております。

次、お願いします。我々の考え方の根本は透明性ですとか、あるいはオープンサイエンス、Team Scienceということで、このプロジェクトについても、世界的に生成AIの研究開発で活躍しておられるアカデミアの先生方に加わっていただいて全体を運営しております。総勢は150名ということで、コーパスの整備からトレーニング、チューニング、安全性、原理解明等、様々な側面に取り組んでおります。

次、お願いいたします。2023年5月、ChatGPTが出したのは2022年11月ですけども、その後、日本でもこういうことをやっていこうということで、ある種ボランティア活動としてLLM - jpを始めましたけれども、その後、2024年4月から、この事業を開始しております。そこでLLMのセンターもNIIに発足しております。最初のモデルはボランティア的な小さなモデルでしたけれども、2024年度、R6年度にはLLM - jp - 3というモデルを開発し、それから準備等も並行して始めまして、昨年度、R7、2025年には4シリーズのモデルを開発し、また、ビジョンランゲージモデル、この開発も進めてきました。

次、お願いします。昨年度御報告しました学習経過がこのような図でございましたけれども、まず、横軸の学習データは2兆単語、縦軸がパフォーマンスですけども、横軸がログスケールになっておりまして、2兆単語のところまで学習して、昨年度、一昨年度のモデルのフラッグシップのものは、8×13ビリオン、130億というMoEモデルでございました。

次お願いします。これのR7年、それから、今年に入っているモデル、学習の様子がこちらになりまして、まず、大きな違いは横軸が、これはリニアになっていますけども、12兆ですので、これまでの6倍規模のトークンで学習しております。それからこのうち下のほうにありますのがDense、通常の8ビリオンと、それからMoEの32ビリオン、これをR8の4月、この4月に公開しております。それから、32ビリオンのモデルが今、学習がほぼ最後のほうまで行きまして、これから最後、整備をして公開する予定です。

それから1点だけ打っておりますけども、MoEの300ビリオン級のモデル、これの学習を始めておりまして、これは現在の計算資源では、今年度いっぱい学習がかかりますけども、これを公開する予定であります。計算資源につきましては、R7年度はABCIを中心として、1,500GPUほどの環境でモデル学習ですとか安全性、その他の取組を進めてきました。今年度に入りまして、5月、6月はこれを継続しましたが、7月からはABCIが600GPU程度に制限されたということがございまして、調達をし

て、SakuraのB200,300GPUというものを確保し、こちらで大きなほうのMoEの学習をしております。

次、お願いします。先ほどのグラフががたがたとしていたと思うんですけども、昨今はいわゆる学習のやり方、レシピも非常に精緻といえますが、複雑化しておりまして、学習率も昔は大きなものからコサインカーブでぐっと下げるだけでしたけども、上げたり下げたりしながらいろんな側面を強化していくということで一般的なウェブ、それから数学とかコードに特化したコーパス、その後、人工的な質問応答のデータ等で、またコンテキストを伸ばしていくと、こういう学習を進めております。

次、お願いします。この結果、さっきも申しました、この4月に公開した、8ビリオンDenseと、32ビリオンのモデルがグラフの右側の赤と紫です。一番右側の固まりを見ていただきますと、これが様々なタスクの平均点になりますけども、その前の3シリーズの緑色に比べますと倍ほどのスコアになっておりますし、Qwen3ですとか、gpt - ossの20ビリオンというオープンソースでかなり使われているモデルと肩を並べるレベルになってきております。フロンティアモデルでいえばGPT - 4を超えるレベルのモデルとなっております。

次、お願いします。このプロジェクト、当初から安全性を配慮するということに非常に重点を置いてきておりまして、AnswerCarefullyというデータセットをつくり、これを英国ですとか米国のAIセーフティーの会議でも紹介したり、また、アジアの中でハンズオンでデータを翻訳する等の活動もしてきました。このデータセットをボーダーラインといえますのは、一見危険ですけども危険が危険でないかのボーダーラインとなるような問題を整理するとか、あるいはコントラパシタルで両面からの回答が必要な、例えば竹島は日本の領土ですかという質問に対しては様々な見解があるということをしきりと示す、そういうデータセットも整備しておりますし、また、マルチモーダルということで、一番下の例は、これは漂白剤のようなものを飲んでいいですかというような、これは英語のデータセットがあるんですけども、これを日本語化し、さらに精緻な説明もし、また、日本におけるローカルなといえますが、適切な画像を付与する、そういう活動、そういうデータセットの整備も行っております。

次、お願いします。それから、もともと日本語をきちと扱うというのはボリュームの問題もありますが、加えて、このような例えば日本文化をきちと扱うことができるかどうかというようなデータセットをつくり、それで評価をしていくということも進めております。これは、下の例は見ていただいたらこのような感じですけども、こういうデータセットを整備しまして、先ほどのQwenですとかgptに比べますと、我々のモデルが一番下の2つですけども、はるかに高い性能を持っているということも確認できております。

次、お願いします。これが先ほども少しありましたが、AIの今のモデルがどのような原理で動いているのかということをしきとつ解明していくと。これがAIを安全に活用するとか、ハルシネーションのような問題を抑えていくという意味でも非常に重要でありまして、昨今はメカニスティック・インタープリタビリティというようなキーワードで徐々にホットな研究ターゲットとなってきております。こういう研究を行う上でも、我々のモデルのように、学習データですとか学習レシピ、それから途中の学習経過、これを公開していることは非常に重要で、これは詳細を御説明する時間はありませんけども、このLLMの活動で、トップカンファレンスで紹介しているものですけども、学習の過程でバイリンガル、日本語と英語の扱いがどのように発展していくかということをトレーニングが進むにつれて、あるいは階層の中で、トランスフォーマーの階層の中で、あるいはモデルサイズによってどう違うかということをし、1,600ぐらいのモデル、スナップショットに対して評価をして中身を分析していると、そういうような研究でございまして、こういうことも今後どんどん進めていきたいと思っております。

次、お願いします。ここはビジーなスライドで恐縮ですけども、成果、評価みたいなものをざっと御紹介したいと思っております。モデルの開発・公開等ですけども、繰り返し申しておりますけども、研究基盤を提供するということで、R6には54モデル、R7には250モデルを公開しておりますけども、延べで474万回を超えるダウンロードをしていただいて、国内外の研究コミュニティで幅広く活用していただいております。その表をまとめたものが下ですけども、省略させていただきます。

次、お願いします。これはLLM - jpの活動で、3年前から始めたものですけども、オープンなコミュニティということで研究開発過程、その議論ですとか試行錯誤も全てオープンサイエンス的な考え方でディスカッションして進めていこうということで、現在、2,800名を超える方がスラックに登録していただいております。スラック上での議論ですとか、週次、あるいは隔週のオンラインワーキンググループの議論ですとか、また、月1回はハイブリッドで勉強会ということを開催して議論を深めております。

次、お願いします。これはセンター構成員が徐々に拡充しているということと、それから人材育成が非常に重要であるとも考えておりまして、RAの学生さんが全体の半数を占めるということになっております。これは文科省の方が聞いてこられたという伝聞なんですけども、Sakana AIの方と話をすると、LLM - jpの中で活躍しているRAの方が、実はSakana AIにインターンに行っていて、そこでも主要な戦力になっているということで、これだけの研究環境をこれだけの研究者と切磋琢磨できる環境というのは、人材育成としても非常に重要な環境であると感じております。

次、お願いします。本事業の社会貢献といいますが、目標としても民間企業からの問合せ件数等も挙げておりましたけども、これも順調に増えてきておりますし、学術コミュニティの活性化という意味では、言語処理学会でチューニングコンペティションを行うですとか、AIMOというAIの数学オリンピックへの協力ですとか、これからフィジカルAIが重要ですけども、AIRoAという活動へのコンペティションの参加等を行っています。左下の社会実装と右下の、特に国際的コミュニティについては、次のスライドでもう少し御紹介します。

次、お願いいたします。社会実装、出口感としましては、公共領域、教育ですとか医療ですとか行政、そういう分野において信頼できるAIの実装、それから運用基盤を確立していくと、そのための評価基準、ベンチマーク等を整備するとともに、モデル自体についても参照していただいたり、活用していただくということがプロジェクトの成果、出口として重要かと思っております。既に医療用のLLMにつきましては、SIPというプロジェクトで我々のモデルを発展させた医療ソブリンモデルの開発が進んでおりまして、実際に医療現場でのトライアルということも始まっております。それから、教育がこれからのAI時代で極めて重要と考えますけども、BRIDGEというプロジェクトに採択されて、そのプロジェクトが始まっておりますし、学術審査においても科研費で、それから、今日この後もいろいろ議論あるかと思えますけども、NIIにおいては、研究データ基盤ということをやずっと日本のアカデミアに対して提供してまいりましたけども、これをAI化していくという中で、知識基盤として、このプロジェクトの成果を活用するということもございまして、関係省庁からの問合せ等も多数いただいております、協力を始めております。

次、お願いします。もう一つは国際連携ということで、世界的なソブリンLLMの研究開発の活動と協力を進めております。ソブリンLLMといいますが、AI主権といいますが少し誤解がある場合もあると思うんですけども、それは閉じ籠もってクローズドにやるということではなくて、主権を大切にしつつ、世界的に、国際的に協力すると。そういう世界的な動きがございまして、その中でLLMCのオープンな活動というのは認知度が高いものと思います。既にそこにありますような様々な機関とのMoUを締結しておりますし、この7月には高市首相がインドを訪問された際に、IITのボンベイ、それからBharatGenという、インドは20言語以上の公用語がありまして、ソブリンモデルに非常に大きな関心を持っておりますけども、LLM - jpとしっかり協力したいということでMoUを結びまして、これを首相の共同声明にも入れていただいております。

次、お願いします。これが最後のスライドでございますけども、このような活動、今、2年と4か月ですけども、あと2年8か月、R10のプロジェクト終了時には、最初申しました、我が国がAIを主体的かつ信頼できる形で研究開発、活用していくための基盤構築をするということで、そこに書いてございますような目標を達成したいと思っております。時間がございせんけども、1個目だけ申し上げますと、学習データは非常に重要ですが、現在のクローズといいますが、何を学習しているかがオープンにされない、オープンウェイトモデルですら学習データが公開されてないわけですけども、そういうサイクルといいますが、関係性では、著作者、権利者が安心してデータを提供できないということで、信頼できるモデルをつくるためには信頼できるデータが必要で、そのためのデータエコシステムを確立していくというようなことをはじめ、本日御紹介したことをそれぞれにさらに充実させていきたいと思っております。プロジェクト終了がR10年度末でございますけども、これらの成果をきちっとさらに発展させることによって、拠点としての、生成AIの研究開発拠点としての実績を積むことによって、さらなる発展、高度化が必要であると思っております。この活動は、まさにAI主権のための非常に重要な活動だと考えておりますので、その継続のための予算確保につきましては、NIIとしても様々な可能性をしっかりと考えていきたいと思っておりますし、国としての御支援も継続してお願いしたいとは考えております。以上です。ありがとうございました。

【相澤主査】 黒橋所長ありがとうございました。

では、ただいまの報告を踏まえまして、御質問、御意見等がありましたら挙手ボタンにてお知らせください。時間が押しているところでありますが、二、三件お受けできればと考えております。よろしくをお願いします。よろしいでしょうか。では、青木委員、お願いします。

【青木委員】黒橋先生、大変勉強になりまして、いつも大変御苦労さまです。非常に重要な取組だと思ひます。最近、オープンAIと先ほどQwenの話が出てまいりましたが、オープンウェイトの軽量なモデルをローカルに使っていこうという動きが研究者の間の非常に、今年あたりから物すごく爆発的に動いているんですが、そういう意味でいうと、LLM - jpのモデルを本当はしっかり使っていただくのが結構重要かなと思ひて、そこら辺をうまく広報できるといいなと感じるんですけども、何か、あるいは文科省なり我々のほうでできることというのはあるかとか、そこら辺のお話をお聞かせいただければと思ひます。

【黒橋所長】ありがとうございます。しっかり使えるレベルのモデルであることがまず、重要でここまで進めてきましたけども、この4月に出しましたモデル4というのは使っていただけるレベルになってきていると思ひておりますし、今後、4.1、4.2と強化していく予定です。ファンクションコーディングですとか、教科学習機能というものでございまして、これを逆に、これはAI for Scienceですけども、AI for Scienceの枠組みで、いわゆる対話型のサービスですとか、あるいはバッチ型で各ドメインの追加学習などを行っていただくと。そういう環境をNIIのほうでも事業として取り入れていきたい、スピード感を持って取り入れていきたいと思ひております。

【青木委員】我々の周りでは、エヌビディアのワークステーションが安価に手に入るようになってきている関係で、AIワークステーションですね。非常に若い研究者の間で急速に広がっているところがあるので、そういう意味だと、うまく事例といひますか、タイミングを図って加速して、逆に文科省もぜひ資金を投入しながら研究者のほうのコミュニティにリーチいただくと非常にいいのではないかというふうに、個人的には非常に思ひて、逆に言う、世界が早過ぎてシニアの先生方もそういう潮流に全く気づいてない人も多くて、学生とか若い研究者のほうが早いなと感じておりました。ぜひこれからもどんどんやっていただいで使わせていただければと思ひます。よろしくお願ひします。

【黒橋所長】ありがとうございます。

【相澤主査】ありがとうございます。では、中野委員、よろしくお願ひいたします。

【中野委員】中野です。黒橋先生、ありがとうございました。大変勉強になりました。今後どういふ方向に行くのかなというところに興味がありまして、アメリカでは完全に民間が主導でどんどん技術が発展していったんですが、企業との連携も何らかの形ではされていると思うんですが、ここのワーキンググループに入っているメンバー、企業の方がそれぞれ個人的に各企業にこういう技術を持ち帰られるのでしょうか、あるいはこの活動と企業のLLM開発等との連携が組織レベルで進められているのでしょうか。この活動が日本ならではの安全なLLMを企業がつくっていく指南役のようになっていくといいのかなと思ひているんですけども、その辺りお考えをお聞かせください。

【黒橋所長】ありがとうございます。まさにそういう役割を果たしたいと思ひております。一つ分かりやすいところは、競争というよりは協力してやろうということで安全性の側面がございまして、これは安全性のWGが中心になって、多くの企業の方を巻き込んでデータセットをつくる、それを活用するという活動が発展しておりますし、今、我々はAISIとの協力も非常に強力に進めておまして、安全性WGの関根主幹は、実はAISI副所長の立場も最近を持たれて、そこでさらに発展させていると。

それから、そもそもアンサーケアフリーですとか、そういうデータはかなり実は企業のモデル学習でも活用していただいているということをお聞ひしております。多様な形があると思ひますけども、改めて企業から研究員をNIIのほうに派遣したいというような問合せも具体的にいただいておりますので、そういう形で拠点といひますか、情報のハブにはなっていきたいなと思ひております。

【中野委員】ありがとうございます。そういう人的交流が一番大事なのかなと思ひますので、よろしくお願ひいたします。

【黒橋所長】ありがとうございます。

【相澤主査】ありがとうございました。では最後、手短に小林委員から御意見いただきまして、終わりにしたいと思ひます。小林委員、よろしくお願ひいたします。

【小林委員】小林です。どうもありがとうございました。国際連携のところ、主に連携先としてはアジアが多いと見受けしたんですけど、欧米との連携というのは重要かなと思うんですが、国際化という意味で欧米との連携、Linux Foundationなども積極的に取り組んでいると思ひますが、何かございましたら教えていただければと思ひました。

【黒橋所長】ありがとうございます。MoUという形になっているのが今、アジアがまだ先行しておりますけども、ヨーロッパのフランスですとか、それからドイツのDFKIとかとの関係もございまして、先ほど参事官から御紹介のあった、イギリスで新しいプロジェクトが立ち上がっておりますけど

も、UCLの中核メンバーは実はNIIとも関係を持っている方ですし、実はあのプロジェクトを立ち上げる際に、日本ではLLM - jpというオープンな活動があって非常に大きなインパクトがあるということも申請書に書いてもらっているというような関係です。今後、イギリスのグループともしっかりと連携を進めていきたいと思っております。ヨーロッパはアメリカとは違う軸で、ソブリンということにも非常にフランスも含めて強い意識を持っていますので、そことの連携を強化していきたいと思っております。

【小林委員】 どうもありがとうございました。

【相澤主査】 ありがとうございます。まだ、御意見等あることと存じますが、後ほど取組方針についての議論の時間もございますので、また続きがございましたら、そちらでお願いいたします。

それでは、議事3に移ります。黒橋所長、ありがとうございました。

【黒橋所長】 ありがとうございます。

【相澤主査】 続きまして、AI技術の進展に伴い、さらに重要性が増している研究データの管理、利活用、流通を支える次世代情報基盤に関して検討を進めるため、昨年10月、情報委員会の下に設置いたしましたワーキンググループについて、このたび審議が取りまとまったとのことです。御報告をいただきます。ワーキンググループ主査の尾上委員より御説明よろしくをお願いいたします。

【尾上委員】 本ワーキンググループの主査を務めさせていただきます尾上でございます。お手元の資料3 - 1、審議まとめ概要、及び3 - 2、イメージ図に沿ってワーキンググループでの議論の結果について御報告させていただければと思います。

近年、AI技術の飛躍的な進展とデータ駆動型科学の進化を背景といたしまして、研究データが新たな知を生み出すための基盤的資源として、その重要性が高まっております。一方で、研究データの管理方法や共有の在り方はVAや機関ごとに異なっておりまして、分野横断的な利活用やAI活用を進める上での課題となっております。また、自動、自律など研究設備の高度化や活用の進展に伴いまして、研究データ量や通信量は今後さらに増大するということが見込まれております。これらを背景といたしまして、本ワーキンググループにおいては、オープンサイエンスやAI for Scienceの進展を踏まえ、研究データの管理、利活用の促進や大量のデータを安定的かつ高速に流通させる情報基盤の在り方について、昨年12月から本年6月まで集中的な議論を行いまして、7月に、今月審議の取りまとめを行いました。

まず、資料3 - 1、審議まとめ概要の中段右側を御覧いただければと思います。我が国はSINET、NII RDC、富岳を含むHPCI、さらには大学の共同利用、共同研究設備や大型研究施設など、世界的に見ても優れた研究基盤を有しております。その一方で、これらの基盤間におけるデータ連携の集約、統合的な活用については課題が残っていると思っております。また、欧州ではEOSC、米国ではDOEやNSFを中心に、研究データ基盤や計算資源、情報流通基盤を一体的に活用する取組が進められているという現状がございます。このような国内外の状況を踏まえまして、本ワーキンググループでは、AI for Scienceに即応するために、左側になりますが、実験基盤、研究データ基盤、計算資源、情報流通基盤の一体的な接続、連携に向けた次世代情報基盤の在り方について整理を行いました。

続いて、資料3 - 2、イメージ図を御覧いただければと思います。本ワーキンググループで取りまとめた内容を一言で申し上げれば、研究データを中核に国としてのパイプラインを構築するということでございます。研究データを中核として、データ創出基盤、研究データ基盤、NII RDC、計算資源及び情報流通基盤の連携を強化して、全ての研究者がAI for Scienceに対応した研究環境の下で研究活動を行えるとともに、研究プロセスを総合的に支援する環境の実現を目指しております。図の左下のデータ創出基盤や右側の計算資源から生み出される研究データは、SINETを通じて研究データ基盤NII RDCに集約されます。研究データ基盤では、共有可能で機械可読な研究データ空間を形成し、AIによる解析や再利用を可能とします。これにより分野横断的なデータ活用、それらを用いた新領域創成など新たな知の創出を促進し、研究プロセスを総合的に支援する環境の実現を目指すものでございます。

上記を踏まえ、本ワーキンググループでは3つの方向性をまとめました。もう一度、3 - 1、概要、下段を御覧いただければと思います。1つ目は、付加価値の最大化を実現するネットワークの整備であります。AI for Scienceの進展に伴い、研究活動における通信量は飛躍的に増加することが見込まれます。そのため、次期SINETでは大容量化、低遅延化、低消費電力化、高信頼化を進めるとともに、AI時代における通信量増加にも耐え得る構成の高度化が必要でございます。また、国際ネットワークとの接続強化を進め、国際共同研究を支える基盤としての機能拡充を図ることが重要でございます。さらに、公私立大学へのNII - SOCSの展開を通じて、学術分野全体におけるサイバーセキュリティ対策の底上げを図る必要があると考えてございます。

2つ目は、研究プロセスを総合的に支援する新たな研究基盤の構築であります。AI for Scienceの推進においては、研究データの質と量の確保に加え、その円滑な管理、利活用が不可欠です。このため、研究データの管理、共有、再利用を支え、分野横断的なデータ活用を可能とするAI統合型の研究データ基盤によって、研究データ空間を確立することが必要であると整理させていただきました。具体的には各分野のデータベースとの連携強化やメタデータの整備、フェア原則に沿ったデータ整備を進めるとともに、AIによる解析や再利用を前提としたデータ整備及び研究環境の構築を進める必要がございます。また、ダークデータやネガティブデータなど、これまで十分に活用されてこなかったデータについても、新たな知の創出に資する主要なデータアセットとして活用していくことが重要となります。

3つ目は、計算資源等との一体的な接続連携、認証強化でございます。AIを活用した高度な研究を推進するためには、研究データ基盤、情報流通基盤、計算資源を一体的に活用できる環境の構築が不可欠です。このため、共通認証基盤の構築により、研究者、研究機関、研究設備、研究データ及び計算資源を信頼性高く結節するパイプラインを整備することが重要であると整理させていただきました。研究者が単一のIDによって様々な研究資源をシームレスに利用できる環境を実現するとともに、安全性と利便性を両立した研究環境を構築することが求められます。

最後になります。本ワーキンググループとして最も重要なメッセージは、AI for Science時代においては、情報流通基盤、研究データ基盤、計算資源、認証基盤を個別に整備するだけでなく、研究データを中心にこれらを一体的に連携させることが不可欠であるということになってございます。研究データを中心として情報流通基盤、研究データ基盤、計算資源及び認証基盤を一体的に連携させることによりまして、研究データの流通そのものが新たな価値を生み出す次世代情報基盤として刷新し、分野融合、新領域創成、そして新たな知の創出を加速していくことを期待してございます。

以上でございます。ありがとうございました。

【相澤主査】尾上先生ありがとうございました。お手元の資料3 - 3が審議まとめとなっておりますが、こちら御覧いただいても分かるかとおり、大変密な御議論をいただいたというように伺っております。

それでは、ただいまの御説明を踏まえまして、もし御質問、御意見などありましたら、挙手にてお知らせください。では、天野委員、よろしくお願いいたします。

【天野委員】すいません。これ、全体を連携するというのは非常にすばらしい案だと思いますし、ぜひ必要だと思います。幾つかまず、質問として、パイプラインを整備というパイプラインというのは何を意味しているんですか。

【尾上委員】ありがとうございます。これは審議まとめのほうの32ページの図のほうが分かりやすいかなと思うんですけども、データの創出から計算するところまで、こういうところの間をつないで、共通認証システムで安全につなぐという、そういうようなイメージでパイプラインという形で表現させていただきます。

【天野委員】パイプラインというのは、あるところが流れているときに、その前に物が入っていったというようにパイプ全体が流れていくイメージですよね。それとは僕ちょっと違う気がして違和感があるんですが。

【尾上委員】おっしゃるように、いわゆるパイプライン処理のパイプラインというか、データがそこを通っていくという。

【天野委員】いや、それは通りますよね。ただ、詰め込まないとパイプにならないじゃないですか。

【尾上委員】はい。

【天野委員】だからこの用語には違和感があります。これAIのほうではそういうふうに使うんですか。

【尾上委員】あまり僕自身、違和感を感じずに使っていたんですが。

【天野委員】そうですか。すいません、コンピューターアーキテクチャやなんかの世界では、パイプラインというのは非常に広く使われていますし、ここで使われている用語の意味と比べるとやや違和感があります。

【尾上委員】かしこまりました。

【天野委員】これは御検討ください。もちろんAIのほうでパイプラインと言え、こういう意味なんだよというのでしたら、分野ごとに用語が違うのは当然なので。

【尾上委員】必ずしもそうでもないと思います。ここの上にパイプラインと書いている左側のいい言葉があれば……。

【天野委員】そうですね。これはちょっと考えてみたいと思います。

あと、もう一つはこれ非常に重要な提案だと思いますが、今やっているAI for ScienceのSPReADのプロジェクトで、非常に僕が気になっているのは、今、予算規模が500万円程度で、しかもAI for Scienceの審査のときに、AIに関係ないものを申請するとフラグが立ちますので、結局ほとんど全員がNVIDIAのGPUを買うという予算計画にしていまいます。このままでいくと、膨大なNVIDIAのGPUがばらまかれることになります。それもまたいいことだと思うんですが、でもやっぱりもったいないので共有化することを考えたほうがいいのかなと思います。つまり、SPReADで申請する人は、ある一定の基盤を使うことができると。そこで資源を使え、また、データを共有することができるとして、予算規模はその分減るけど、通りやすくなるとか何か工夫していただかないと、このままいくと官民挙げてNVIDIAにお金をどんどんつぎ込んでいくみたいなことになりかねないと思っております。

ぜひ御検討ください。以上です。

【尾上委員】ありがとうございます。一応SPReAD、これは僕よりも文科省のほうからお答えいただくのがいいんだと思うんですけど、SPReADの公募の時点でそういう共用ということ、あるいは機関として整備していくということ、少なくとも機関としては整備していくことを一応課されていると思っております。

一方で、ここで書いていただいているような話も含めて、これは実はワーキンググループでもあったんですけども、すごくヘビーなユーザー、あるいはHPCIIにつながるユーザーからライトに使おうというユーザーまで、どういう層をどういう段階で収容していくかということは国全体として考えていく必要があるというような議論も……。

【天野委員】そうですね。おっしゃるとおりだと思います。

【相澤主査】では、ありがとうございました。本日、かなり時間が押していますが、せっかくなので、手短かに川添委員、引原委員の順番で御意見をお伺いして、本議事を終了したいと思います。まず、川添委員、よろしくお願いいたします。

【川添委員】ありがとうございます。尾上先生、どうもありがとうございます。今出ているページにも書いてありますが、オールフォトニクス・ネットワークは私も非常に重要だと思っていました。全国の拠点をこういう形で結んで、かつAIがその上で動くといったときに、今のIPネットワークではやっぱり足りない機能、パフォーマンスの不足があります。それは先生がおっしゃったように遅延時間とか、あるいは帯域とかですが、それに限らず、もう一つ非常に重要な要素が実は私あると思っています。それは何かというと、タイム・シンクロナイズーションですね。要は今のIPネットワークには時計はないので、厳密な意味での時間同期ができない中でAIを連携させなくちゃいけないというポイントが非常に重要だと。今のラージランゲージモデルはほとんどがスタティックデータなのであまり関係ないかもしれないんですけど、これからのAIにおいて、厳密な時間、関係性みたいなのを基に計算するような時代においては、それが非常に重要な機能になると思います。なので、それも含めてここに盛り込んでいただければいいなと思いましたので、よろしくお願いいたします。以上です。

【相澤主査】どうもありがとうございます。では、引原委員もよろしくお願いいたします。

【引原委員】ありがとうございます。引原です。尾上先生、どうもありがとうございます。このデータ基盤のほうの絵を見させていただいて、この流れはとにかくデータをパイプに詰め込むようなイメージなんですけども、データに関してはキュレーションの部分というのはどうしても出てくると思います。それが下から検証なく上がっていくような理解ではちょっとまずいと思いますし、先ほど黒橋先生、データの安全性ですが、安全性データアセットという部分が、このプロセスの中にどこか入ってこないといけいないんじゃないかなと思います。ですから、その意味では今、川添委員が言われたような時間の管理、いつこのデータが採られたかという、リセットがいつ掛かったかも含めた意味で、理解を得られる絵が重要なんじゃないかなと思いますので、ぜひ御検討ください。よろしくお願いいたします。

【相澤主査】どうもありがとうございました。それでは、やや時間に押されてという形ですが、尾上委員、御報告どうもありがとうございました。

続きまして、次の議事4に移ります。今後の情報科学技術分野における取組方針につきまして、事務局よりよろしくお願いいたします。

【池田参事官補佐】事務局です。時間が押していますので、少し駆け足で御説明させていただきたいと思います。

まず、おめくりいただきまして、情報分野の全体像ということで、これまで左側の赤いところの先端

的な情報科学技術の研究開発、人材育成と、右側の青いところ、基盤の構築・運用ということを両輪に進めてきたということでございますが、先般新たにAI for Scienceということで、横串を通すような取組、両方にまたがるような取組ということで、まさにこういった時代が来ているということで整理させていただいております。

次お願いします。そうした中で昨年度末に文科省として、AI for Scienceの推進に向けた基本的な戦略方針ということで取りまとめまして、世界を先導する研究成果の創出ですとか、その波及振興、そして何より研究インフラの構築ということで、こういった体系的な整理をさせていただきました。

次のページお願いします。また、こうした取組に応じて、マル1からマル5に示しますとおり、ケーキのような図ですけれども、各部分における定量的な目標も設定しております。

次のページをお願いいたします。こうした定量的な目標や思想といったものも含めまして、昨年度末、同様に策定されました、科学技術・イノベーション基本計画ですとか、統合イノベーション戦略2026、人工知能基本計画、昨今決まりました骨太方針ですとか、日本成長戦略の記載も書かせていただいております。ここにもしっかりAI for Scienceが位置づけられているということでございます。

今後の取組の方向性について、今から概要にありますけれども、簡単に御説明させていただきたいと思います。まず、SPReAD、ARiSEにつきましては、担当のほうから説明代わりますので、よろしくお願いいたします。

【轟木参事官補佐】SPReAD、ARiSEを担当しております轟木と申します。簡単に御紹介させていただければと思います。

10ページ目を御覧ください。SPReADにおきましては、令和7年度補正予算で50億措置をいただきまして、右の三角形にあるように、まさにマル3、マル4、今後、AI for Scienceに関心があり、今後、AI活用に研究加速を考えている方ですとか、AI for Scienceに関するノウハウがない、そういったような方々にぜひ科学研究にAIを取り入れていただく。そういったこと波及・振興を目的とした事業になってございまして、3つの方向性、迅速な支援、伴走支援、独創的研究の芽出し支援というところで実施をさせていただいたところでございます。

次のページ、おめくりいただきまして、まず、今回、第1回の応募が先日終了いたしまして、資料の左下でございますけれども、1万5,868件の応募がございまして、採択が456件、採択率2.9%となっております。今、第2回の公募も終了して、今まさに審査をしているところでございますけれども、第2回におきましても、同規模以上の応募があったところでございまして、まさに研究開発現場におけるAI利活用への高い関心、そして意欲を示すものであるというように我々としても受け止めてございます。

本資料の左に少し書かせていただいておりますけれども、SPReADに関して、ポジティブな評価といたしまして、まさに無作為抽出の活用ですとかAIの審査スキームへの活用といった新しい研究費のモデルであるというような話ですとか、AI for Scienceの拡大に貢献をしているんじゃないかというような御意見、学生や若手にも門戸が開かれていて、スピード感のある審査、採択、申請審査負担の軽減というような評価もいただく一方で、まさに採択率2.9%と非常に狭き門になってしまっていることですとか、無作為抽出の納得感・透明性などなど、様々御意見をいただいていると認識しているところでございます。SPReADにつきましては、AI for Scienceの波及・振興を目的に挑戦的な仕組みを試行的に導入しているところでございまして、右の今後の課題と展望に書かせていただいておりますとおり、審査・採択における透明性・信頼性の向上ですとか、今回、年2回の公募で1,000件の採択を予定をしておりますけれども、採択率2.9%というようなことも踏まえまして、採択規模の拡充ですとか、あとはこれから説明しますARiSEとSPReADの連携のところの部分、継続的かつ切れ目のない段階的支援というようなものが必要じゃないかというような御意見をいただいておりますので、そういった中間的枠組みの創設を含めた支援を今、来年度の要求に向けて検討させていただいているところでございます。

最後に書かせていただいておりますけれども、来るAI時代におきまして、SPReADが研究者が挑戦しやすく、納得感があり、学際的知が循環する研究者支援、発展をするために不断の見直しをしっかりと行って、来年度以降の予算要求にもつなげていければと思っているところでございます。

次、おめくりいただきまして、12ページ、ARiSEについてでございます。令和7年度補正予算でJSTへの基金で、3年間の基金として320億円で措置をいただいております。こちら戦略ターゲット型と国際融合型を大きく2つに分けてございまして、戦略ターゲット型におきましては、戦略方針で明記をさせていただきました、AI駆動マテリアル開発システムの実現ですとか、バイオ生成基盤モデルの開発、そして大型研究施設、研究装置における自動自律化と、大量データ分析能力向上に資するAI

エージェント、AI基盤モデルの開発などといった戦略的なターゲットを掲げるとともに、国際連携ですとか今後、出てき得るような新興融合領域を対象とした少し小ぶりのターゲット用意をさせていただいているというところがございます。

こちらにつきましても、13ページ目おめくりいただきまして、まさに米国ジェネシス・ミッションとの連携ですとか、それ以外、国際動向、技術動向、AI for Scienceの周り非常に変化が激しいところだと認識をしております。そういった中で、AI研究、AI for Science分野での国際共同研究の期待、非常に高まっております、世界のAI先進国との連携強化ですとかASEANをはじめとするグローバル・サウス諸国との協力拡大、これ非常に重要だと考えてございます。さらに、それに加えて、バイオ分野、マテリアル分野、ロボット分野などの競争力の源泉となるこのデータにつきまして、産学官のデータ共有、連携の枠組みが十分に整備されていないというようなところも認識をしております。

こういったところを踏まえまして、AI、AI for Scienceの観点から、産学のデータ連携の新たな枠組みの構築なども念頭に置きまして、ARISEにつきましても、少し来年度以降しっかり拡充していきたいと考えてございます。事業内容としては大きく3つの柱を掲げてございまして、まさに米国ジェネシス・ミッションとの連携強化・拡充というところ、さらには米国以外の国との連携の強化・拡充、そして、先ほど申し上げたとおり産業界との連携の強化・拡充、この大きく3つの柱を掲げているというところがございます。産業界との連携強化・拡充につきましては、AI for Scienceの観点から、まさに連携すべき産業領域などがあればコメントいただければと思っております。

私からの説明は以上でございます。

【池田参事官補佐】続きまして、次世代研究インフラのほうにまいります。次おめくりいただきまして、各取組をてんでんばらばらで進めるのではなく、次世代の研究インフラとして一体的に強化ということで、しっかり国としてやっていくということでございます。

次のページです。先ほども御説明ありましたが、まず、実験基盤、データ基盤、計算基盤ということで、それぞれの取組がありますが、これをてんでんばらばらでやるんじゃなくて一体としてやるということが大事だと思っております。先ほど少し議論になりましたので補足しますと、パイプラインというのは認証もさることながら土管を敷いて終わりということではなくて、実験からデータをつくって処理して、最後シミュレーションに持っていくというそういったワークフロー、科学研究のワークフローはディスカバリーパイプラインですとか、リサーチパイプラインといった言葉を使われておりますので、そうした意味を踏まえて、エージェンティックなワークフローを左から右に通していきますということでパイプラインということで御認識いただければと思っております。

次お願いします。考え方とありますけれども、全部を全部1個のプラットフォームに統合というのは無理でしょうということで、特に実験基盤はドメイン性が強く、垂直性が高いと。一方で、計算資源につきましては、いろいろアーキテクチャーはあるものの、最近ではGPUがかなり入ってきていて、CPUとGPUの時代になっているということで、こちら水平性ないし汎用性が高いということ、そして真ん中のデータ基盤のところがある意味統合的になっておりまして、ドメイン性と共通性、この両方のバランスを取っていくということが考えられます。

次のページ、お願いします。それを考え方としてつないだのはこちらでございまして、ドメイン性があって、それぞれいろいろな設備がある実験基盤から、ドメイン特化型のデータ基盤につないでいながら、そこに傘をかけて、それらをフェデレートしていくような層があると。そうしたもののから、最終的には例えばシミュレーションを使う場合であれば、HPCI等の統合的な計算資源に投入していくような、こういったワークフロー、データの流れというものを考えているところでございます。

次に各それぞれの基盤に関する取組の方向性について簡単に御説明いたします。まず、実験基盤です。こちらにつきましては、ハイスループットに、とにかくデータを出していくと。水道をつくっても蛇口から水が出なきゃ意味がないので、そのデータのハイスループット化をしていくということでございます。そのためには自動化、遠隔活用とか、特に産業競争力強化といった観点では、例えば分光装置や分析装置の標準化といった取組も進める必要あると思っております。また、ユーザーを広げていくためにも、匠の知とかスキルに関係なく皆が使えるような、AI対応といったものをしていくということを考えております。右側のところ、自動・自律・遠隔化によるオートメーション、クラウドを構築したり、実験条件、失敗例も含めたプロセスを資源として蓄積すること、データ基盤との連携、あるいは標準化、あるいは人材育成といったことを取り組むことによって、最終的にはAI for Scienceにふさわしい実験基盤を構築するといったものでございます。

次のページ、お願いします。続きまして、データ基盤について、これは先ほどの議題でもお話がありましたけれども、例えば、まずは分野別の育ってきたデータ基盤というものを高度化していくことを

考えています。個別の柱を伸ばしていき、それらをフェデレートしていきながら、最終的にはエージェントが自由に走り回れるようなデータ空間を整備していきましょうというような発想でございます。左下、例えばということで、我が国の主要なデータ基盤を書かせていただいておりますけれども、性能的なデータ基盤が複数存在しております、さらに大学、国研等の機関が保有するデータベースは物すごくいっぱいあるということで、かなり古い20年前の情報ではございますが、2,500強のデータベースがあるということでございました。また、AI-ready化といった言葉、これはいろんな定義がございますけれども、データの収集段階から標準化、構造化、メタデータ付与等をしていくことによってAIがすぐに使えるような環境を実現するということをオールジャパンでやっていくためには、縦と横の関係が必要だということで、このような形で試作を検討しています。

次のページをお願いいたします。共用計算基盤につきまして、まずは手元の計算資源が足りないということでございます。今、理研を中心に富岳NEXTの開発整備を進めていただいておりますけれども、それを踏まえつつ、共用計算資源を2030年度までに10倍に増強するといった方針を打ち出しております。当然ながら、国内の計算資源の整備には限界がございますので、ベストミックスといった形で民間サービスや国際連携による計算資源の利用というものも組み合わせながらやっていくことを考えております。また、とにかく数だけ増やせばいいということではなくて、例えば相互運用性、先進的なコンテナクラウド環境や、資源ブローカー、メタスケジューラー等を普及する、あるいは共通のポイントの導入や、有償利用を拡大等、制度改革等やシステム改革と連動しながら進めていくといったことを考えている次第でございます。

次、お願いいたします。こういったところに横串を通し、データの流通を高速化するため、当然ながら、通信が重要だと考えています。SINETの高度化、ないし、次期SINETの構築ということで、平成28年度をターゲットに800Gbpsの技術を導入しつつ、国際回線の容量も増強するといったこと、また、サイバー攻撃の検知、情報提供機能の強化といったところで、NII-SOCSの評価も合わせて進めていくことを想定しております。

次のページお願いします。こうした各基盤同士をワンストップで使えるようにしたいと考えており、実験、データ、計算基盤、それを貫くような一つの次世代認証システムを構築するべきではないかと考えております。加えて、あらゆる研究者がAI for Scienceの環境、各基盤に安心、簡便にアクセスでき、AIエージェントによって任意のワークフローを呼び出したり、そこで編集できるようなダッシュボード、プラットフォームを整備できないかといったことも考えております。

次、お願いします。こうしたエージェントの利用は、最近様々なところで普及しております。科学研究においても漏れることなく、このようなことに取り組むべきだと考えていますが、いろいろ課題はあると思っています。まず、各研究インフラがAIエージェントに対応していないところをしっかりと加速していくということが必要と思っております。また、AIエージェントのワークフローの開発やその利用の加速もしていかなければいけません。今後の研究活動というのは、いろんな研究者がAIツールやAIエージェントを活用してレバレッジを効かせていく世界に突入するものと考えていますが、AIエージェントの利用の課題としてコストとリスクがあると思っています。エージェントやツールの組合せのパターンが指数関数的に増えていって、色々な人が多くの試行と失敗を重ねてしまうと考えておまして、こういった失敗の連続とそれに伴う時間の消費というものが一つボトルネックになるのではないかと考えております。また、エージェント作業というのは通常よりもはるかに多くのトークンを消費すると。改善してきている部分があるとは聞いておりますけれども、依然としてトークン数がかかなり大きい状態です。タスクの成功時よりも失敗を重ねる場合の方が、推論量が大きくなってきますので、その分トークンをたくさん消費することになります。ついては、失敗が再生産されればされるほど、予算が無くなり、国が痩せ細っていくという形になってしまうということです。

最後はセキュリティについてです。ワークフロー上のLLMや特定のシステムの脆弱性を突かれると、全てのデータが流出してしまう可能性や、プロンプトインジェクション、ツールやワークフローの汚染といった問題も想定されております。したがって、動けばいいということではなくて、安心して動かせる必要もあると考えております。AI for Scienceの競争軸は、それぞれの個々のモデルの開発やエージェントの開発というのも大事ですが、それを組み合わせるような研究プロセス、いわゆる料理のレシピみたいなものの設計や、それが活用できるような研究のエコシステムの構築部分に競争力が移行しつつあると思料していますそのためインフラの構築に目を向けて、AIエージェント時代の研究レシピのようなものを開発して、使えるものであればすぐに普及することで、失敗を再生産しないような工夫というものが求められるんじゃないかと考えております。

次お願いします。こうしたものを束ねまして、絵を書いております。ARISE、SPReADが先行的に事例をつくっていただいているということで、こうしたものをナレッジとして次世代研究インフラにた

め込みながら、いろんな人が統合的な環境下でこれを使えますということ、そして、ARiSE、SPReAD、あるいは、このインフラを使った成果というものがあらゆる研究者にフィードバックされていく環境をつくっていくということ、いろんな設備、日本が持っている強みのある設備をみんなが使い倒せるような環境にしたいということを考えています。研究のレシピを再利用して、ゼロからワークフローを設計する車輪の再発明を削減すること。また、こうしたナレッジというのは最終的には大きなフロンティアモデルに食われていく可能性も言われておりますので、まさに今日、LLM - jpの話もございましたが、こうしたところで貯めてきた知見、ナレッジみたいなものを、さらに学習にも利用して、より高度なモデルの開発につなげていくといったことも考えるべきだと思っております。

続けてAI研究開発力の強化についてもご説明いたします。冒頭阿部のほうからも御説明申し上げましたように、いわゆるScience for AIにも力を入れる所存で、生成AIモデルの透明性、信頼性の確保に向けた研究開発拠点の形成ということで、LMCの取組の充実といったものを図ってまいります。次29ページ目です。さらには日本における中核のAI研究のセンターとして、理系のAIPセンターの機能といったものも引き続き推進していくということを考えております。

31ページ目です。こちら来年度が最終年度になりますけれども、Society5.0実現化研究拠点支援事業ということで、Society5.0時代のライフデザインといったものを社会実装していく取組を阪大において進めていただいております。

次のページお願いいたします。また、JSTのCRONOSも大変人気がありますので、こちらもしっかりと予算確保に努めてまいりたいと思っております。

最後、33ページ目ですけれども、高度統計人材育成強化拠点形成事業においては、需要が高まる数理、データサイエンス系の学部学科の新設ラッシュ等を踏まえて、統計学を扱う人材を増やしていくということで、統計人材のネットワークにおきまして、教材やノウハウを最大限活用して、統計学を指導、教育できる研究者、教員を増やしていくといった取組も引き続きやっていくということでございます。

限られた時間の中で走り切ってしまいましたけど、以上でございます。よろしくお願いいたします。

【相澤主査】御説明ありがとうございます。ただいまの御報告を踏まえまして、ざっくばらんな意見交換の時間とさせていただきたいと思います。時間限られていて、大体17時30分頃までをめどに行います。ですので、皆様の御意見をまとめていただきまして、最後に事務局からコメントという形でいきたいと思います。御意見のある方は挙手にて、お知らせください。よろしくお願いいたします。小林委員、よろしくお願いいたします。

【小林委員】御説明ありがとうございます。大変よく整理されていると思って聞いておりましたが、ただ一つ、前回は質問させていただいたんですけど、17ページの水平、垂直という絵なんですけど、これは1軸だと汎用性と専門性を横に並べているようにしか見えないんですけど、この軸というのは、何を意味しているんですか。

【池田参事官補佐】前は垂直を縦に並べて、水平を横にやっていたんですけども、先生からまさにこれは1軸じゃないかのご意見いただきましたので、これは左側が分野の特性が強くて、右側が汎用であるという1軸に直しております。縦軸にあまり意味はございません。

【小林委員】もし垂直性をもう少し分かりやすくするのであれば、軸というのは、どちらかという、領域の広がりかなと思って聞いていたんですけど、横の楕円が大きく広がっていれば汎用性があるし、それがだんだん上に行くにつれて小さくなって、専用性というかドメインスペシフィックなものが並ぶというようなことのようにも理解したんですけど、これは単に1軸で、横軸は専用性と汎用性なんですか。

【池田参事官補佐】そのとおりです。

【小林委員】水平と垂直が1軸に並んでいるというのは、少し違和感がありますので、もし、実験基盤が多様性があるのであれば、横軸、水平軸を領域と捉えて、少し並べてもいいんじゃないかなと思いました。

【池田参事官補佐】ありがとうございます。

【相澤主査】ありがとうございます。では、湊委員、続きまして、盛合委員、川添委員の順番でよろしくお願いいたします。

【湊委員】湊ですが、25ページあたりで、ARiSEとSPReADが次世代研究インフラを使うという絵が出ていますけど、今まさに審査をしているところなんですけども、審査中にこういう話は全然多分出てきてなくて、採択された人がいつからこれを使うのかとか使えるのかとか、あるいはダッシュボードというのはいつ頃出てくるのかということが分からなかったんですけど、実際、ARiSE

も2年半しかないんですし、どういう予定でこういうことをやろうとしているのか、ちょっと教えてもらえますか。

【池田参事官補佐】ありがとうございます。私の説明が時間の都合上、かなり雑になってしまって恐縮でございますが、今まさに目下、あらゆるSPReADの研究者が、インフラがまだできていないものを使うということではございませんで、先ほどお伝えしたナレッジを貯めていくとかワークフローを貯めていくというのがまず、大事だと思っています。ここに書いている趣旨はARiSEで生まれたデータやツール、基盤モデル、ワークフローですとか、SPReADで個別具体的にやっている検証データみたいなものをこれからつくり上げていくインフラの知見として貯めていくということを書かせていただいております。当然ながら、ARiSE、SPReAD、今後も我々しっかりと推進していきますので、今後プレーヤーが拡大したときには、それと併せて構築されてきた次世代の研究インフラを使っただくということになりますし、ダッシュボードについてもそうなっていきます。基本的にはインフラを2030年度までに構築と考えておりますので、段階的にベータ版を公開しながら、計算資源も増強していった、データ基盤も増強していった、順次組み上げていくような形になっていくのかなと考えております。

【湊委員】そうですか。それは今、申請している人たちにはまだそれは伝わっていないということですか。

【池田参事官補佐】おっしゃるとおりです。現時点において審査評価項目で次世代研究インフラに関する問いは設けてない状況です。

【湊委員】そうすると、採択されて、研究をまさにしている人に対して、今後こういうことがあるので、もし使えたらぜひ使ってくださいとお願いはするけど、強制はしないという、そういう感じでしょうか。

【池田参事官補佐】はい、現時点ではそのようになると思います。

【湊委員】分かりました。

【相澤主査】ありがとうございます。では、盛合委員、よろしくお願いします。

【盛合委員】ご説明ありがとうございます。次世代研究インフラについて関心を持ちました。いろいろな大学等の皆さんが活用するシステムにするということで、さまざまな研究データ、AIで解析したい、学習したいと思うデータを置くと思うんですけども、アクセス制御等の設計や、システムアーキテクチャ等の設計がどのような構想になっているのかという点と、システム構築後の運用が重要で、サイバーセキュリティ対策に関する発言もありましたが、24時間365日、運用されるチーム体制、そしてそれをカバーできる予算は確保されているのかなという点が気になったんですけども、どうでしょうか。

【池田参事官補佐】ありがとうございます。アーキテクチャーの全体をつなげていく構想や予算の管理のことを示すのは簡単ではございます。今日の後半、非公開の議題にも関わってきますが、こういった構想を掲げたからには、我々としてはそのための体制誠意に必要な概算要求等にも鋭意取り組んでまいります。

【盛合委員】ありがとうございます。後半のほうでは拝聴します。

【相澤主査】では、時間が押していますので、今、挙手いただいている青木まで手短かにコメントいただくということで、川添委員よろしくお願いいたします。

【川添委員】ありがとうございます。今日の主題がScience for AI、Science of AIという形で議事が始まりましたけども、その中で、御説明の中にもありましたが、例えばページ15とか16に書いてあったトランスフォーマーを超えるようなモデルアーキテクチャーを求めていくということこそが、私はScience for AIの大きな目的かなとも思います。ただ、その中で、何を具体的に作り出すかというところがもう少し書き込まれていてもいいのかなと感じました。

自然言語ではないところで、例えばゲノム情報などを例にとると、今スタンフォードユニバーシティがゲノムの言語モデルということで、Hyenaという新しいモデルアーキテクチャーをつくっています。これはどちらかというと生物学とか遺伝、日本で言えば遺伝研みたいなところがやっているような、これまでのサイエンスの英知を使って、その中で新しいモデルアーキテクチャーを導き出すということをやろうとしているんですけども、ここが相変わらずトランスフォーマーの延長線上であるような、越えるというようなレベルだといけないのかなと思いました。その辺も含めてぜひ盛り込んでいただきたいなと思います。よろしくお願いします。

【相澤主査】ありがとうございます。では、星野委員、よろしくお願いします。

【星野委員】ありがとうございます。非常に興味深く聞いておりまして、24ページ、25ページあたり、例えば失敗を再生産しないだとかで、誰でも使い倒せるといったことが非常に重要でございま

て、日本の勝ち筋というところを再度考えますと、分野が違ふと思われるかもしれませんが、日本が得意としていた生産管理、それこそトヨタ生産方式やジャストインタイムじゃないですけど、そういった形で研究者は何に時間をかけているのか、何を探索するのかということまでパッケージとして含めて、限られた資産をどのように利用して、最終的にみんなでうまく成果を上げていくかということに関しては、ぜひこういった情報、基盤側のほうだけじゃなくて、研究者側が何に時間をかけて、何を短縮できるのかとか、前も申し上げたところでございますけど、そちら側のほうの研究もぜひひいていただいて、そこを組み合わせさせていただくと非常に日本の勝ち筋となるような、この分野に関してはここに投資してうまく成果を上げるということにつながるといいますので、ぜひ基盤側のほうで、このような失敗を再生産しないということ非常に重要というか、トークンの話でも今、中心になりましたけども、そういったシステムとしての研究は研究というところのほうも、ぜひ文科省としてお考えいただければと思います。ありがとうございます。

【相澤主査】ありがとうございます。そうしましたら、石田委員、よろしくお願いいたします。

【石田委員】石田です。御説明いただいた内容というのは私も賛成いたします。次世代研究インフラのところなんですが、説明をシンプルにするためのものなのかもしれませんが、人文社会系の人たちのデータや、研究のプロセスというのをどのように支えていくのかというのが見えにくくなっているのかなと感じましたので、その辺りも少し言及していただくといいのではないかなと思いました。以上です。

【相澤主査】どうもありがとうございます。では、青木委員よろしくお願いいたします。

【青木委員】18ページ以降のエージェントAIのところですか。これも実はかなり文科省の方、よくというか、なかなか最先端のところをよく見ておられるなと思って見ていました。ここはなかなか分かりにくいところなんですけども、極めて今、非常にホットな領域になっていると思います。18ページを見ていただくと、左端から真ん中に、18ページの基盤ですか、要は、これの共用実験基盤というところとデータ基盤と、こういうのがありますけど、実際に今のエージェントのほうはローカルAI、さっき申し上げたようなエージェントインザループで現場データをセキュアに外に出さないで囲い込んで、生成とか加工をハイスループット、実験の現場で回していくという、それで実験データをどんどん出していくという取組が動いていると思うんです。そういう意味だと、例えばエヌビディアのGB10なんかを使った、いわゆるパーソナルなAIファクトリーと、それから次世代のVera Rubinを使ったような大きなAIファクトリー、エージェントAIの世界というのは両方つながってくるので、ここら辺、ローカルAIの最近のトレンドにうまく見ながらしっかり考えていく必要があるなと思います。

左端のほうも、例えば非常にセキュアな外に出せないデータなんだけど、ローカルにおいていながら、例えば連合学習をする仕組みとか、これはアカデミアの間ではもっとしっかりとルールを決めていくようなところもやったほうがいいと思うので、よく書かれているので、ぜひ18ページのところ、最先端のトレンドを把握していただくといいかなと。左端もかなりデータとかエージェントが絡んでくる領域になると思いますので、ぜひ見ていただければと思います。以上です。

【相澤主査】ありがとうございます。では、もし事務局から最後に補足等ございましたら。

【池田参事官補佐】沢山ご意見いただいておりますところ、時間が足りず申し訳ございません。メール等でいただければ、しっかり政策にも反映したいと思っております。また、連合学習の件もありがとうございます。現時点でお示しできておりませんが、最終的にはそういったところもワークフローとして一律化することによって、広く普及するようにダッシュボードを使いながら進めていくということを考えております。

引き続き勉強してまいります。よろしくお願いいたします。

【相澤主査】ありがとうございます。

本日駆け足でございますが、続きまして、議事5に移ります。情報分野研究開発プランの改定につきまして、事務局より御説明お願いいたします。本議事への御質問につきましては、議事6の後に併せて承ります。よろしくお願いいたします。

【池田参事官補佐】情報分野研究開発プランについて、2点、変更したいと思っております。今回の変更箇所、資料内赤字にてお示ししております。まず、2-7と書かれているプログラム(7)というところですか。今回、令和9年度の予算要求における議論を踏まえまして、AI for Scienceを支える次世代研究インフラの構築事業(仮称)といったところを事業として立ち上げを検討しております。このことを踏まえまして、評価を実施する必要がございますので、本プランに追記をしております。この評価案につきましては、この後の非公開の議題で御説明させていただきまして、事前評価をよろしくお願いいたします。

2点目、令和8年度より第7期の科学技術・イノベーション基本計画等が閣議決定されておりますので、上位施策の部分について更新しております。

以上が変更点となります。

【相澤主査】ありがとうございました。

それでは、最後の議事6に移ります。冒頭に申し上げましたとおり、本議題は非公開議題とさせていただきますので、事務局にて準備をお願いいたします。

（傍聴者退室）

《議題6. 研究開発課題の事前評価について は非公開》

なお、「情報分野研究開発プラン（案）」及び「研究開発課題の事前評価票（案）」について、審議の上、委員会決定とした。

【相澤主査】本日の議論はここまでとなります。事務局より事務連絡ありましたらよろしくお願いいたします。

【池田参事官補佐】事務局です。本日の議論について追加でご意見あれば、8月6日木曜日をめどにメール等で事務局へ御連絡ください。

次回の開催予定につきましては、現時点では未定となっておりますけれども、また、今後日程調整の御連絡させていただきますので、どうぞよろしくお願いいたします。以上でございます。

【相澤主査】大変動きが急な中ではございますが、本日いただいた御意見、しっかりと進めていくということで、よろしくお願いいたします。

それでは、本日これにて閉会といたします。どうもありがとうございました。

—— 了 ——

お問合せ先

研究振興局参事官（情報担当）付

（研究振興局参事官（情報担当）付）

ページの先頭に戻る

文部科学省ホームページトップへ