

仕様書

I o T 推進部

1. 件名

生成 A I 開発加速に向けた新たなデータセットの構築に関する調査

2. 目的

生成 A I は、従来の A I では不可能だった、様々な創造的な作業を人間に代わって行える可能性があることから、産業活動・国民生活に大きなインパクトを与えると考えられており、ポスト 5 G 時代のキラーアプリケーションとしても位置づけられる。

生成 A I の開発には大量かつ良質なデータが必要であり、その確保が生成 A I の競争力を左右するが、生成 A I の開発にあたって必要なデータは限られており、そのデータセットの構築手法といったデータセットのあり方も模索が続いているところである。また、2023 年 5 月に A I 戦略会議で取りまとめられた「A I に関する暫定的な論点整理」においても、「A I 利用に大きな期待があるものの課題を抱える分野のデータを整備し、その分野に変革をもたらす A I 開発を促進すべき」と明記されている。

こうした背景から、例えば音声データセットの構築において効率的なアノテーションの手法が定まっていないといったデータセットの構築手法に課題を抱える分野のデータセットの構築と、それに対するユーザー（基盤モデル開発者）からのフィードバックを通じて、生成 A I の開発を加速するためのデータセットの構築のあり方を調査する。本調査事業では、経済産業省の定める「ポスト 5 G 情報通信システム基盤強化研究開発事業研究開発計画」に定められた「競争力ある生成 AI 基盤モデルの開発」に参画している事業者等と連携し、そのニーズの取り込みも図る。

3. 内容

生成 A I の開発に効果的なデータセットの構築の調査にあたっては、そのデータセットが基盤モデル開発にとって有用なものとなるよう、基盤モデル開発者の意見を最大限取り入れて進めることが重要である。本調査事業の成果を最大化する観点から、データセットの構築者に加えて、基盤モデル開発者も本調査事業に参加することとし、本調査事業の中ではデータセットの構築とその利活用方法の実証を同時並行的に行い、その結果をデータセットの構築のあり方の検討にフィードバックさせていく。調査に当たっては、以下の内容は必ず実施すること。

また、本調査事業を通じて構築したデータセットについては、生成 A I の開発に資するものであると考えられることから、適切な形で保管及び提供される必要があるため、データ管理者が、以下 4.（3）に示す一定の基準を満たす生成 A I 開発者に提供することと

する。

- (1) 基盤モデル開発者の開発に資する、データセットの構築手法に課題を抱える分野のデータセットについて、データセットの有効な構築手法に係る仮説を立てた上で実際にデータセットを構築し、構築する中で出た課題やその対応策を調査する。
- (2) 構築したデータセットの基盤モデル開発者による利活用方法の実証とそのフィードバックを通じて、基盤モデル開発者の開発加速に資するデータセットのあり方（課題、対応策、利点等）を調査する。
- (3) (1)で構築されたデータセットを実際にデータ管理者が保管し、生成A I開発者に提供する。データセットの提供にあたっては、管理体制を整備し、また、例えば、セキュリティが確保されたクラウドサービスを利用する等、技術的な観点から効率的かつセキュアな対策を講じ、その過程で発生した課題についても調査する。

4. 調査の進め方

調査を進めるにあたっては、経済産業省、N E D O及び成果普及等支援者と調整し、指示に従うこととする。

- (1) データセット構築者、基盤モデル開発者及びデータ管理者の3者で実施すること。ただし、必ずしも基盤モデル開発者が本調査事業の共同提案者もしくは再委託先等の実施体制に入る必要は無い。
- (2) データ管理者については、広くデータセットを提供する主体となることから、高い公益性を持ち、データセット管理や提供に当たっての審査を実施できる能力を持つ事業者とすること。なお、調査期間終了後の保管費用をN E D Oから支出することはできないことから、調査期間終了以降もデータセットの保管・提供にあたっての経費を継続的に支出できる事業者であること。
- (3) データ管理者は、データセット管理についての体制を整備の上、データセット提供のための申請方法、申請様式、基準について、経済産業省及びN E D Oの承諾を得ること。これらを変更する際も同様に承諾を得ること。また、申請状況等について経済産業省又はN E D Oからの照会に対し、速やかに対応すること。
なお、データセット提供の基準には以下の4点を必ず入れること。
 - ・ データセットの取得を希望する者（以下、「申請者」）は、日本国内で登記され、かつ、日本国内に開発拠点を有する企業・大学等であること。
 - ・ 申請者は、従事者（データセットを扱う者）のリストをデータ管理者に届け出た上で、データ管理者の許可を得ること。そのリスト以外の人物はデータセットを扱わないこと。また申請者における従事者は、外為法・個人情報保護法・著作権法等の法令等に反しないこと。
 - ・ 申請者は、データセットを生成A I開発のみに用いることとし、データ管理者の

許可無く他の企業・大学等にデータセットを再配布しないこと。データ管理にあたっては、十分な体制を整備していること。

- ・ 申請者におけるデータセット利用の結果として得られた基盤モデル等の成果の公開や販売等にあたっては、事前にデータ管理者の許可を得ること。

5. 調査期間

NEDOが指定する日（2024年度）から原則1年。

6. 予算規模

15億円以内

7. 報告書

提出期限：

事業終了時には調査報告書を所定の期日までに提出。調査報告書では、本調査事業で発生した事項や過程（課題や対応の方向性を含む）についても、詳細に記載すること。

提出方法：

NEDOプロジェクトマネジメントシステムによる提出。

記載内容：

「成果報告書・中間年報の電子ファイル提出の手引き」に従って、作成の上、提出のこと。

<https://www.nedo.go.jp/itaku-gyomu/manual.html>

8. 報告会等の開催

委託期間中又は委託期間終了後に、成果報告会における報告を依頼することがある。

9. その他

- ・ 本調査事業の実施にあたっては、別途募集する、実証・調査やその成果普及の支援等を一元的かつ効率的に行う者（成果普及等支援者）と連携することとする。
- ・ 生成 AI 開発者に対して、構築するデータセットについて、ヒアリングを広く実施した上で提案すること。
- ・ 実施にあたっては、各分野に精通する専門性を持った企業等と可能な限り連携すること。
- ・ 実施事項の内容や進め方及び本仕様書に定めなき事項等については、経済産業省、NEDO及び成果普及等支援者と実施事業者が協議の上で決定するものとする。