

新規研究開発課題に係る基本計画書概要 【研究推進室】

安全なデータ連携による最適化AI技術の研究開発

1

研究開発の背景・目標

○背景

我が国自身の社会課題の克服や産業競争力の向上を目的として、AIの更なる社会実装が期待されており、その実現のため、AIの適用領域の拡大や、我が国産業が有する分野毎のデータ活用や他分野との連携、日本が強みを有する分野とAIの融合が求められている。一方、プライバシーデータや機密データ等について、分野を越えて活用することが難しく、分野横断的にデータを活用したAIによる課題解決が困難となっている。

○政策目標(アウトカム目標)

本研究開発で確立した技術を活用して、多様な者が有するデータを安全に連携させることを可能とする「データ連携AIプラットフォーム」が創出され、分野を横断したデータ活用が進められることにより、我が国の社会課題の解決や産業競争力の向上に貢献する。

○研究開発目標(アウトプット目標)

「マルチモーダルAI技術」、「エッジAI技術」、「連合学習技術」を確立し、更にこれらを組み合わせることで、実空間に存在する多様なデータを安全に連携し分野横断的な課題解決を可能とする分散型機械学習技術を確立する。

技術課題

○課題ア マルチモーダルAI技術

多様なデータを組み合わせ複雑な予測を可能とする大規模マルチモーダル深層学習モデルの構築技術、及び、実空間から収集するデータの差異を吸収可能な、ロバストなマルチモーダルAI技術についての研究開発を行う。

○課題イ エッジAI技術

マルチモーダル深層学習モデルを対象に、エッジ環境の限られた計算資源の規模に応じて、効率的に学習を行う技術の研究開発を行う。

○課題ウ 連合学習技術

マルチモーダル深層学習モデルを対象に、多数のエッジ環境間におけるデータの偏りを前提とした高精度な連合学習技術の研究開発を行う。

※研究開発の実施に当たっては、上記3つの要素技術を組み合わせ分散型機械学習システムを試作し、具体的な社会実装シーンを想定した技術実証を行う。

到達目標

○課題ア

3種類以上のマルチモーダルデータを組み合わせ複雑な予測を可能とする1億5000万パラメータ規模のマルチモーダル深層学習モデルの構築技術の確立、及び、収集された場所や期間等によってデータの量や粒度が1/10程度の少量になってしまう場合(あるエッジ環境にて収集されるデータの量や粒度が、他のエッジ環境の1/10程度の少量になってしまう場合)でも、予測性能を概ね維持できるようなロバストなマルチモーダルAI技術の確立を目標とする。

○課題イ

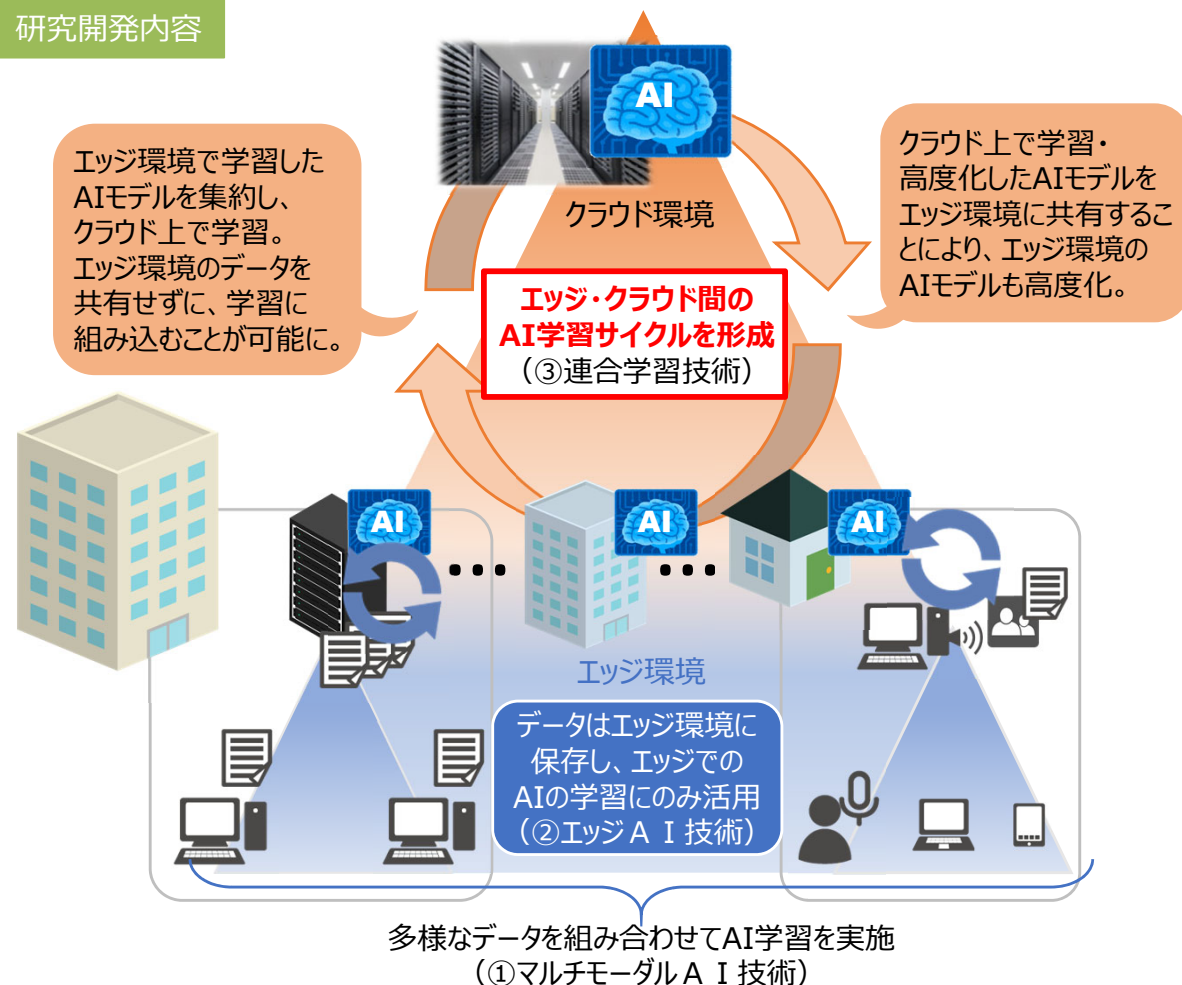
マルチモーダル深層学習モデルを対象に、1つの建物内や1つの地区内等の、小規模な範囲をカバーするエッジ環境の限られた計算資源の規模に応じて、効率的に学習を行う技術の確立を目標とする。

○課題ウ

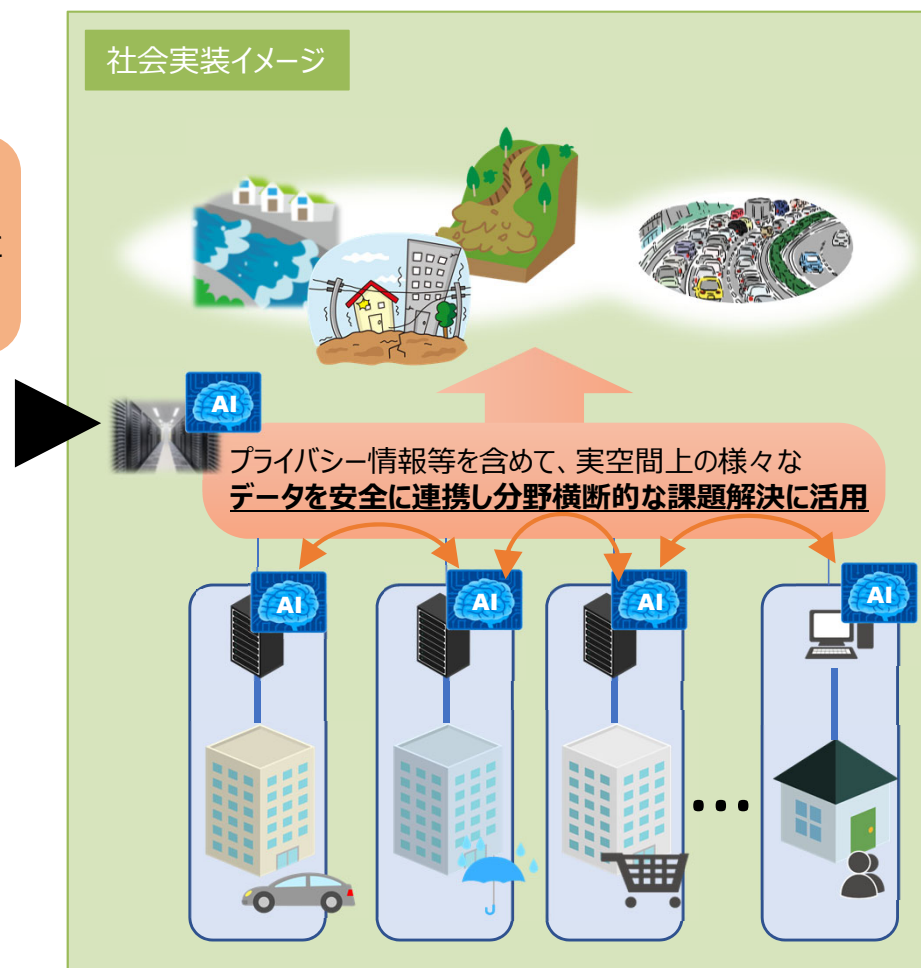
マルチモーダル深層学習モデルを対象に、数百程度のエッジ環境間におけるデータの偏りを前提に、データ集中型AIに近い予測性能を実現可能な連合学習技術の確立を目標とする。

- プライバシーデータや機密データ等を含め、我が国に存在するデータを分野の垣根を越えてAI学習に活用することを可能とし、分野横断的な我が国の社会課題の解決や産業競争力の向上に貢献するため、実空間に存在する多様なデータを安全に連携させることを可能とする分散型機械学習技術を確立。

研究開発内容



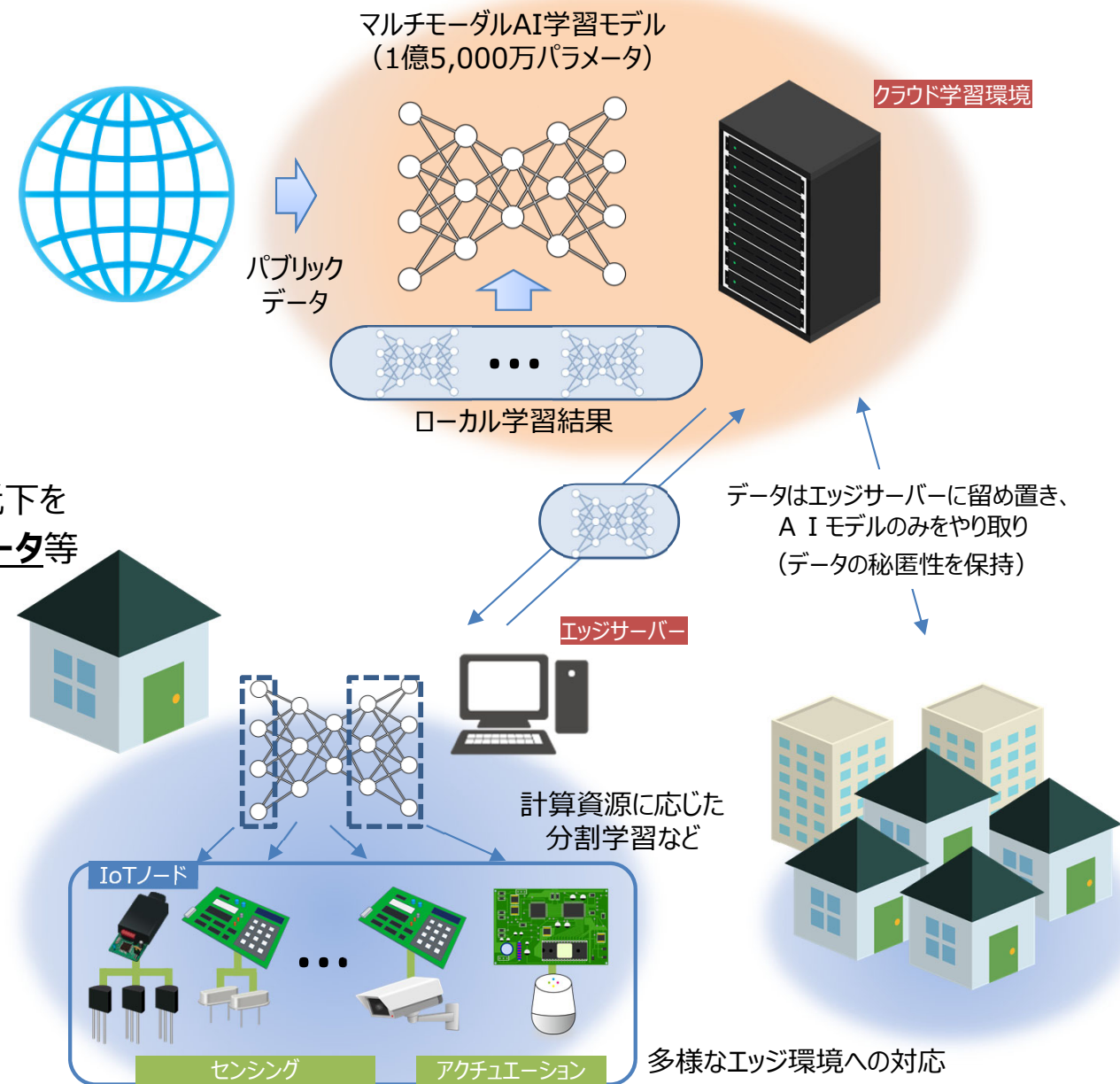
社会実装イメージ



【前提条件・課題サマリ】

- 3種類以上のマルチモーダルデータを組み合わせた
1億5000万パラメータ規模の
マルチモーダル深層学習モデルを想定
- 実空間の多様なエッジ環境に存在する様々なIoTデバイスが収集するデータの活用を想定し、エッジ環境ごとの
 - データの量・精度・粒度等のバラつき
 - 計算資源のバラつき

を考慮し、これらに起因する学習効率・予測精度の低下を可能な限り抑えながら、プライバシーデータや機密データ等を含めたデータについて AI 学習に活用し、
連合学習により安全なデータ連携を行う。



研究開発内容

多様なデータを組み合わせ複雑な予測を可能とする大規模マルチモーダル深層学習モデルの構築技術、実空間から収集するデータの差異を吸収可能な、ロバストなマルチモーダルAI技術についての研究開発を行う

技術的ポイント

■ 課題

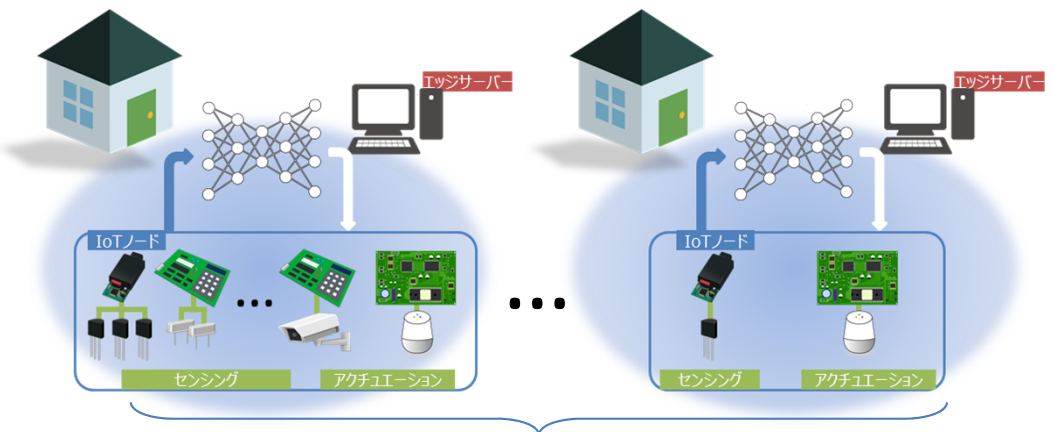
- マルチモーダルAIにおける大規模深層学習は発展途上
- 連合学習において、実空間の様々なIoTデバイスが収集するデータの活用を考える際、IoTデバイスの性能の差異や設置環境等により、十分なデータを収集できないエッジ環境においては、AIの予測性能が低下

■ 本研究開発における取組

- 近年の深層学習モデルの大規模化傾向を考慮に入れ、研究開発終了時点（R7年度末）でのパラメータ規模を想定した、大規模マルチモーダル深層学習モデルの構築技術確立。
- 複数のエッジ環境において、データの量や粒度に一桁程度の差異がある場合でも、予測性能を概ね維持できるよう、データに対するロバストネスを持たせることで、環境全体の予測性能の向上を図る。

到達目標

- 3種類以上のマルチモーダルデータを組み合わせ、複雑な予測を可能とする1億5000万パラメータ規模のマルチモーダル深層学習モデルの構築技術の確立
- 収集された場所や期間等によってデータの量や粒度が1/10程度の少量になってしまう場合でも、予測性能を概ね維持出来るようなロバストなマルチモーダルAI技術の確立



- 実空間においては、各エッジ環境で収集可能なデータの量や粒度・精度はセンシングデバイスの性能や環境に依存
- 多様な状況において予測精度が維持可能となるよう、データの補完等も行いながら、ロバストネスを持たせることが必要

研究開発内容

マルチモーダル深層学習モデルを対象に、エッジ環境の限られた計算資源の規模に応じて、効率的に学習を行う技術の研究開発を行う

到達目標

1つの建物内や1つの地区内等の、小規模な範囲をカバーするエッジ環境の限られた計算資源の規模に応じて、効率的に学習を行う技術の確立を目標とする。

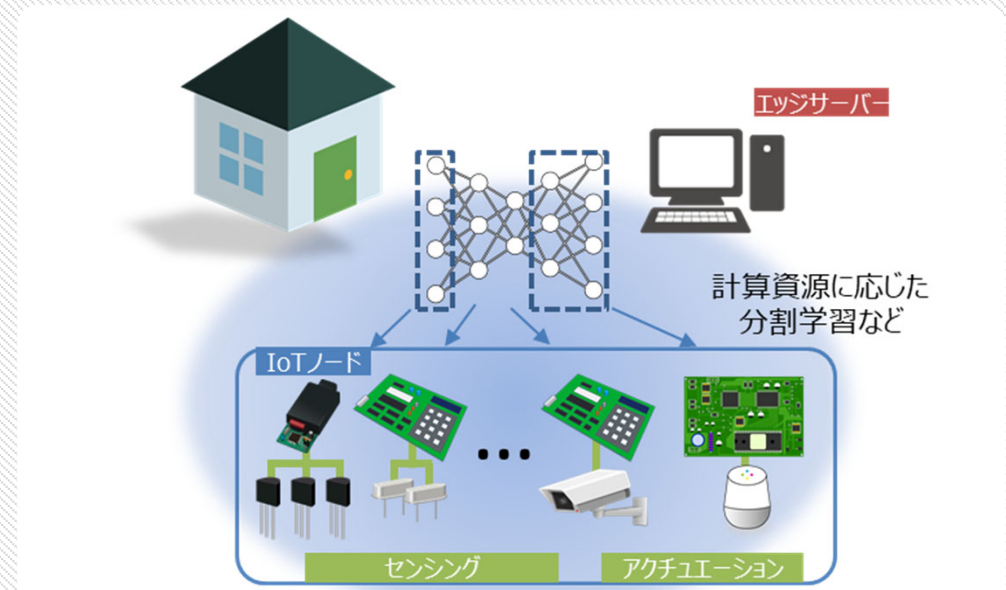
技術的ポイント

■ 課題

- エッジ環境では計算資源が限定的（今回想定する大規模マルチモーダル深層学習モデルに対しては、学習時は分散処理が現実的）
- 多数のエッジ環境を含めての実運用を想定した場合、各エッジに存在するコンピューティングデバイスの性能・配置が一定ではない

■ 本研究開発における取組

- 各エッジ環境における計算資源の多様な状況に応じ、今回想定する大規模マルチモーダル深層学習モデルに対してスケーラブルな学習が可能となるよう、分割学習などの手法により効率的に学習可能な技術の確立を図る。



- エッジの環境ごとに計算能力・計算資源の配置が多様
- 今回想定する大規模マルチモーダル深層学習モデルに対して学習時の分散処理を想定

研究開発内容

マルチモーダル深層学習モデルを対象に、多数のエッジ環境間におけるデータの偏りを前提とした高精度な連合学習技術の研究開発を行う

到達目標

数百程度のエッジ環境間におけるデータの偏りを前提に、データ集中型 AI に近い予測性能を実現可能な連合学習技術の確立を目標とする。

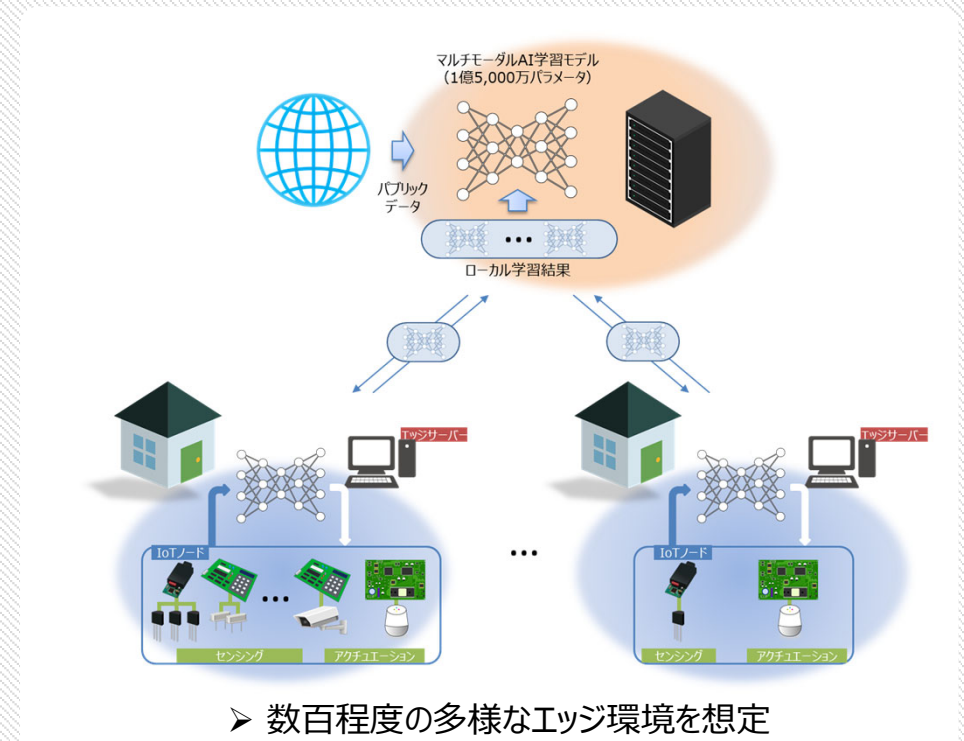
技術的ポイント

■ 課題

- 多数のエッジ環境を含めての実運用を想定した場合、エッジごとのIoTデバイスの性能の差異や設置環境の違い等により、データの“偏り”が生じることで、連合学習時の性能低下を懸念

■ 本研究開発における取組

- 数百程度の多様なエッジ環境を想定し、データの偏在を前提として、課題（ア）及び課題（イ）と合わせて連合学習における精度の向上を図り、効率的に、かつ、安全にデータを連携させ、AI学習を可能とする分散型機械学習の確立を目指す。



政策目標の達成に向けた取組方針

○研究開発期間中

- ・ 受託者が設置する研究開発運営委員会等において、政策意図を適切に反映させるとともに、学識経験者や有識者による幅広い観点からの助言を基に研究開発全体の方針を調整する。
- ・ 最新の技術動向や研究開発を進める中で新たに明らかになった課題等を踏まえて、各技術の更なる高度化等のために必要な追加予算の獲得を柔軟に検討する。
- ・ SIP、PRISM等の他省庁施策との連携や実際に各種データを収集している他省庁、民間企業等と連携し、本研究開発成果の適用範囲の拡大に取り組む。
- ・ 受託者等と連携しながら、本研究開発で確立する分散型機械学習技術の社会実装スキームの具体化を進める。

○研究開発期間終了後

- ・ 本研究開発で確立した技術を活用したデータ連携AIプラットフォームの創出に向けて、民間企業等に対し、受託者等と連携して、情報提供・技術移転等の各種支援を行う。
- ・ PRISM等の更なる予算の確保も検討しつつ、本研究開発で確立した技術の高度化、社会実装の加速化を図る。
- ・ 追跡調査・評価において、受託者等に製品化・事業化等の成果展開状況を確認するとともに、本研究開発成果やその後の成果展開状況についての情報発信等の活動を推進する。