別紙　事業範囲等詳細

大学等と連携した行政特化型国産 AI モデルの構築・実証事業

別紙　事業範囲等詳細

目次

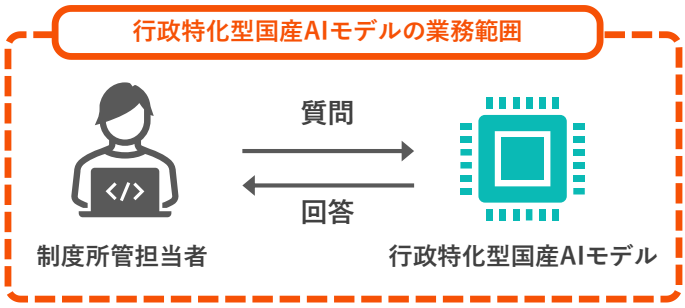
1　業務要件.....3
2　行政特化型国産 AI モデルの構成等.....3
3　機能要件.....4
4　非機能要件.....6
5　構築・実証スケジュール.....8

1　業務要件

業務と行政特化型国産 AI モデルの範囲は以下のとおり。

（1）令和 8 年度の業務要件

以下の図のとおり、制度所管部署の職員（担当者）が行政特化型国産 AI モデルを利用することを想定（利用対象は全庁職員ではなく、特定の業務を所管する担当者に限定する）。



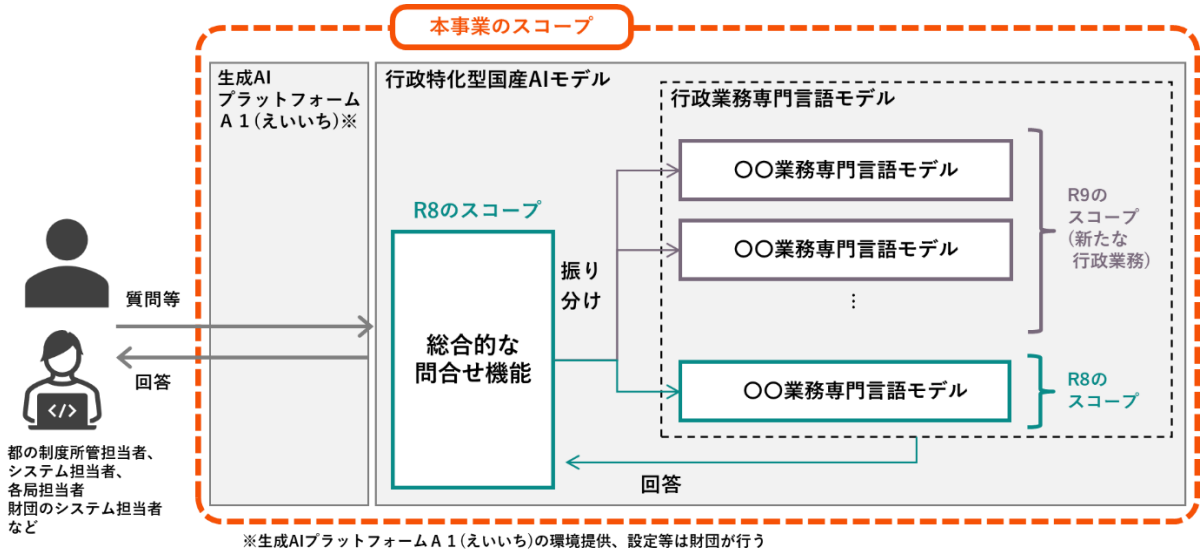
（2）令和 9 年度の業務要件（仮）

令和 8 年度の構築・実証成果を踏まえ、東京都（以下「都」という。）及び一般財団法人 GovTech 東京（以下「財団」という。）と協議の上、未着手の行政業務を新たに 2 業務程度対象とし構築する。業務の選定にあたっては、令和 8 年度の実証実績及び行政的なニーズに基づき、都及び財団と協議の上決定する。

なお、将来的な行政業務専門言語モデルの追加に際し、システム全体の改修が最小限で済むような拡張性の高い構成であることを前提とする 。

2　行政特化型国産 AI モデルの構成等

行政特化型国産 AI モデルの論理構成及び本事業のスコープは以下の図のとおり。



3　機能要件

行政特化型国産 AI モデルの機能要件は以下のとおり。

項番	分類	機能等	機能要件	機能実装年度
1	全般	生成 AI プラットフォームＡ１（えいいち）からの利用	Ａ１から都の職員が利用可能なこと。Ａ１との接続にあたっては、OpenAI Chat Completions API 互換のインターフェースを基本とする。 なお、Ａ１の環境提供、設定等は財団が行う。	R8 年度
2		行政データの学習	モデルに組み込む知識及び参照する外部文書等のデータについて、追加・変更を可能とすること。	R8 年度
3		会話内容の学習	AI の回答に対し、職員が入力した評価（Good/Bad 等）及び正しい回答への修正情報を蓄積し、継続的にモデルの回答精度を向上させることが可能なこと。	R8 年度
4		会話履歴による精度向上	Ａ１のメモリ機能により送信される会話履歴を適切に処理し、対話の文脈を踏まえた一貫性のある回答を行うこと。	R8 年度から R9 年度
5		ハルシネーション対策	回答にあたり、モデル内在の知識及び外部文書の参照のいずれからも十分な根拠が得られない場合、又は回答の信頼度が低いと判断される場合は、その旨を利用者に明示した上で、関連する情報の提示や確認先の案内等、利用者の次の行動を支援する回答を行うこと。また、根拠が不十分な情報を確定的な表現で回答しないこと。	R8 年度
6	業務	総合的な問合せ機能	入力内容を行政の用語や考え方も含めて適切に解釈し、利用者が意識することなく、最適な行政業務専門言語モデルへ自動的にルーティングすること。本機能は、単一の API エンドポイントとして提供し、Ａ１から利用可能とすること。ルーティングの実現方式は提案に含め、行政業務専門言語モデルの追加時における拡張性を考慮すること。	R8 年度

項番	分類	機能等	機能要件	機能実装年度
7	業務	行政業務専門言語モデル（1 業務）	特定の行政業務に関連する法令、条例、要綱、マニュアル等の知識をモデルに組み込み、当該業務における回答精度を担保すること。加えて、外部文書の検索・参照（RAG 等）により回答の正確性を補強すること。 また、回答にあたっては、モデル内部の知識に基づく場合及び外部文書を参照した場合のいずれにおいても、回答根拠となる引用元情報（法令名、条項、文書名、該当箇所等）を API レスポンスに含めること。	R8 年度
			外部文書の検索・参照（RAG 等）の実現にあたって、文書及びユーザーの質問を意味的なベクトルデータに変換するエンベディングモデルを用いること。当該モデルは日本語テキスト間の類似性を適切に捉えられるものであること。	R8 年度から R9 年度
8		行政業務専門言語モデル（複数業務）	複数の行政業務の専門性を備えた言語モデルにより、当該業務における回答精度を担保すること。 ※機能要件は項番 7 と同じとする ※複数の行政業務に関する詳細は「1（2）令和 9 年度の業務要件（仮）」のとおり	R9 年度
9	モデル検証	モデル検証用フロント画面	都のシステム部門、業務所管部門及び財団が、行政特化型国産 AI モデルの検証・評価をＡ１から実施可能にすること。ただし、効率等の観点から、検証において専用の検証用フロント画面を用意する方が望ましい場合は、当該画面を介した実施も可能とする。	R8 年度
10	ガードレール	入力チェック	利用者の入力に扱ってはならない情報が含まれていないか検知し、遮断すること。検知対象のカテゴリ及び検知の感度等を財団が設定・変更できること。	R8 年度
11		プロンプトインジェクション対策	利用者の入力により、システムの動作や権限を意図せず操作されるプロンプトインジェクション攻撃への対策が可能なこと。	R9 年度
12		出力チェック	回答に出力してはならない情報が含まれていないか検知し、遮断すること。検知対象のカテゴリ及び検知の感度等を財団が設定・変更できること。 また、財団が回答ログの確認及び修正を行い、修正内容を再学習データとして活用できる仕組みを提供すること。	R8 年度

項番	分類	機能等	機能要件	機能実装年度
13	運用保守	ログによる透明性担保	回答結果の出力プロセスのログを記録・可視化し、出力の根拠となった学習データや参照ドキュメントを特定できることで透明性を確保すること。	R8 年度
14		モデルのバージョン管理	行政特化型国産 AI モデル更新後に回答品質が劣化した場合、以前のバージョンへ復旧できる機能又はその手順を設けること。	R8 年度
15		行政業務専門言語モデルの追加	行政業務専門言語モデルの追加が可能なこと。 その際、総合的な問合せ機能の改修が発生しない若しくは最小限で対応できる構成等にすること。	R9 年度

4　非機能要件

（1）行政特化型国産 AI モデル全般

ア　行政特化型国産 AI モデルは国内で構築すること。

イ　ベースとする言語モデルを採用する場合は、透明性が確保され、信頼性が認められるモデルを採用し、提案時にモデル名及び選定理由を明示すること。

ウ　行政特化型国産 AI モデルの学習及び推論過程を把握可能とし、ブラックボックス化を回避することで透明性を確保すること。

（2）性能・拡張性

ア　業務処理量

・令和 8 年度

ユーザー数　30 名程度

同時アクセス数　5 名程度

・令和 9 年度

ユーザー数　数百名程度

同時アクセス数　数十名程度

イ　令和 10 年度以降の実運用に向けた要件時の想定

令和 10 年度以降の都及び財団における実運用に支障がないよう、以下の要件を満たす構成とすること。

・業務処理量

ユーザー数　最大 2 万名

同時アクセス数　数百名程度

・業務継続性

同時アクセス数が 1 時間継続しても稼働を継続できること。

・拡張性

GPU 等の拡張により性能向上が可能なこと。

なお、提案にあたっては、推論に必要なマシンスペック（GPU 構成等）と応答性能（レイテンシ、スループット）の関係を示すこと。また、同時アクセス数の増加に対する性能特性及

びスケーリング方針を含めること。

（3）移行性

協定期間終了後、構築したAIモデルが財団の環境で動作可能であること。

協定期間中に財団の環境への移行及び検証を実施し、正常に動作することを担保すること。

（4）セキュリティ

ア　前提条件

東京都サイバーセキュリティ基本方針及び東京都サイバーセキュリティ対策基準を遵守すること。

イ　不正追跡・監視

ログを取得し、不正アクセス時等に追跡可能なこと。ログの保管期間は1か月程度とする。

ウ　ネットワーク

以下の要件を満たすこと。

・閉域（IP-VPNレベル以上）

・ネットワーク通信の暗号化はTLS 1.2以上を必須

・外部ネットワークからの入り口にはWAFを設置し、不要なポートはアクセス不可

・クラウドサービス間の通信は、プライベート通信を原則

・技術的制約上、プライベート通信が難しい場合は、アクセス元制限を行う

エ　ウィルス・マルウェア対策

対策を必須とする。

オ　クラウド環境

クラウド環境を利用する場合は、法律準拠は日本国法が適用される閉域の環境とすること。

カ　データ

データの保存及び処理を行う環境について、法律準拠は日本国法が適用される環境とすること。

また、提供するデータを外部AIサービスの学習に利用しないこと。

キ　開発工程等における生成AI利用

設計、開発、テスト等の工程で生成AIを利用する場合は、使用するAIツール等を明示すること。生成AIで作成された成果物は人間によるレビューを必須とすること。

5　構築・実証スケジュール

構築・実証スケジュールは以下の想定。