一般競争入札（総合評価落札方式）に関する質問及び回答（Q&A）

最終更新日 2025 年 3 月 10 日

独立行政法人情報処理推進機構

件名：AI セーフティの評価環境の構築に関する先行調査研究業務

項番	資料名	頁番号	項目名	質問内容	回答	回答掲載日
1	Ⅲ. 仕様書	p.15	2.3 AI セーフティの評価環境の設計および試作 (1) 評価ツール	モデルを軸に評価する場合、モデルの作成に使用されたデータは不明ですが、モデルから元データを推測できる場合、それがセキュリティリスクとなる可能性があります。この点を踏まえ、試作時にはモデル作成時のデータやシステムのソフトウェア的なセキュリティはスコープ外としてよいでしょうか？	自組織で事前学習またはファインチューニングを実施して独自モデルを開発する場合もあるため、学習データが特定できる場合もあります。従って、モデルの学習用データに関するセーフティ/セキュリティはスコープ外にはなりません。 システムにおいて AI と関係の無い部分のソフトウェア的なセキュリティに関してはスコープ外ですが、そのような従来のセキュリティの問題に影響されて AI の挙動に問題が生じる場合はスコープ内になります。	2025 年 3 月 10 日
2	Ⅲ. 仕様書	p.16	2.3 AI セーフティの評価環	評価に使用するデータの種類や量、データ処理の負荷を見積	当初の対象データ種別はテキストとなります。データ量に関してはケースバ	2025 年 3 月 10 日

			境の設計および試作 (2) 評価内容定義および評価用データセット作成支援機能	りたく思います。	イケースですが、あくまで参考ではあるものの、下記の LLM を対象とした評価データセットリポジトリである Inspect Evals では、数千〜数万程度のテスト項目からなるデータセットが多いようです。 データセット作成支援機能の設計としては、例えば UI 画面において適切にページネーションする等の一般的な考慮は必要になります。 [参考] Inspect Evals: https://github.com/UKGovernmentBEIS/inspect_evals	
3	Ⅲ. 仕様書	p.16	2.3 AI セーフティの評価環境の設計および試作 (3) 評価対象 AI 実行基盤	評価用 AI 実行基盤をコンテナとして考えた場合、評価用AIソフトウェアと実行基盤を一緒にコンテナにまとめることが推奨されるでしょうか？	実運用を考慮した場合、外部で運用されている AI モデルを評価ツールによって評価する場合もあれば、組織内で完結している AI システムを評価する場合もあり、後者において、AI モデル内の状態に着目した評価なども技術的にはあり得ます。左記に関して特に推奨はありませんが、同一コンテナでも別コンテナでも評価を実施可能にする必要はあります。	2025 年 3 月 10 日

4	Ⅲ. 仕様書	p.16	2.3 AI セーフティの評価環境の設計および試作 (3) 評価対象 AI 実行基盤	仕様書に記載されている API について、どのように考慮すればよいでしょうか？API として ollama などの既存の OSS をターゲットにすることが適切でしょうか？	Ollama が提供しているような OpenAI 互換 API を実行基盤側で公開し、評価ツール側から OpenAI API 用ライブラリでアクセスするのも一案ですし、それに問題があれば実行基盤側で独自 API を公開し、評価ツール側に独自 API に対応したアダプタライブラリを実装してアクセスする形にしても良く、様々な方式があり得ますが、要点としてはモデル単体でなく実行基盤全体として API を公開し、全体の評価を可能にすることを意図しています。	2025 年 3 月 10 日
5	Ⅲ. 仕様書	p.16	2.3 AI セーフティの評価環境の設計および試作 (3) 評価対象 AI 実行基盤	試作でターゲットとする、モデルをどのくらいのものにするのが想定されていますでしょうか?	検証対象として、ローカルのサーバ上で運用するモデルとしては、10B 前後のオープンなモデルを利用することを想定しています。より大規模なモデルを利用しても問題ありません。 また、モデルの出力を別のモデルでチェックする場合など、ローカルのサーバ上のモデルと外部で運用されているモデルを実行基盤から利用する形態もあり得るため、API 経由で利用する外部のモデルに関しては特に規模を問わ	2025 年 3 月 10 日

					ず対象としてください。	
6	Ⅲ. 仕様書	p.16	2.3 AI セーフティの評価環境の設計および試作 (3) 評価対象 AI 実行基盤	評価用 AI 実行基盤に求められるセキュリティ要件はどのような内容でしょうか。	仮想マシンやコンテナなど下位レイヤも含めて、評価対象の AI モデルを他者のモデルと混在せずに評価可能な状態にすることが実行基盤には求められます。多くのセキュリティ要件は下位レイヤの方で満たせると想定していますが、クローズドなネットワーク内であったとしても API を公開する関係上、一般的な AI 関連 API と同様、API キーによるアクセス制限は最低限必要になると認識しています。	2025 年 3 月 10 日
7	Ⅲ. 仕様書	p.15	2.3 AI セーフティの評価環境の設計および試作 (1) 評価ツール	想定される利用ユーザー数は何人程度でしょうか。	明確な想定は無いものの、データセットを用いて評価し、対象モデルの出力を別の AI モデルを用いて評価することもあることから、データセットや結果評価用モデルはある程度の範囲の組織内で共用する想定です。また、個人には非力な PC や仮想環境しか提供されない組織もあるので、評価ツール自体もサーバで運用されて、ある組織内の複数人で同時利用する可能性もあります。現状ではトランザクションが必	2025 年 3 月 10 日

					要な処理は存在しないと認識しており、スケールアウトが可能になっていればユーザー数への対応は問題ないと想定しています。	
8	Ⅲ. 仕様書	p.17	2.5 キックオフ・定期打ち合わせ等 (2) 定期打ち合わせ	作成物のレビュー方法はどのような流れでしょうか。作成したソフトウェア(開発途中含む)をユーザー様で操作されるのか、評価項目に対して実行できることを証明(スクリーンショットや画面共有など)できれば良いのかなど。	機能や GUI のレビューのため、IPA 側でも動作検証しますので、開発途中のもの含め、適宜共有をお願いします。	2025 年 3 月 10 日
9	Ⅲ. 仕様書	p.18	7.3 納入物件	「7.3. 納入物件」「(3) 評価環境 試作物」「ソースコード等一式」は、お渡ししたものがユーザー様の各環境でそのままの状態で動作することを想定していますでしょうか。 「β版としての公開を想定した程度の品質を目標とする。」とありますため、指定されたブラウザでのテストは行うとしても、少なからずユーザー様の	納品物は、原則そのままの状態で評価環境β版として OSS 公開し、ユーザーがそれをダウンロードしてユーザー環境で実行することを想定しています。 β版かつ OSS での公開となるため、あらゆるユーザー環境での動作を保証する必要はありませんが、不必要に対応環境を狭めることも避ける必要があります。 利用ライブラリ等から決まる対応ブラウザ(例: ES2016 以上をサポートする	2025 年 3 月 10 日

				環境依存による問題が発生する可能性があり、テストのスコープに影響があると考えております。	ブラウザ対応)と、検証済みのブラウザ(例: Chrome **, Edge **検証済み)の情報をそれぞれ公開することとし、本案件内での検証作業としては、代表的ないくつかのブラウザの最新版で動作検証するレベルの作業を想定しています。	
10	Ⅲ. 仕様書	p.15	2.3 AI セーフティの評価環境の設計および試作 (1) 評価ツール	「機能拡張により画像・映像・音声（ストリーミング含む）にも対応可能な設計とする。」とは、設計としては拡張性を持たせるが、本プロジェクトにおいては実際に開発または妥当性の検証は行わないということで良いでしょうか。	左記の機能に関して、実際の開発が本プロジェクトの対象外である点はご認識の通りです。左記に対応可能な設計であるか否かについて不明確な点がある場合は、理論的または実験的な検証を求める場合があります。	2025 年 3 月 10 日
11	Ⅲ. 仕様書	p.19	10. その他	本プロジェクト体制におけるIPA 様の中心的な役割の担当の方はどのような立場の方になりますでしょうか。 PJ を円滑に推進するにあたって、意思決定権をお持ちかを確認させていただきたいです。	体制に関しては、当方で必要と思料する相応の体制となるよう検討しております。	2025 年 3 月 10 日

12	Ⅲ. 仕様書	p.16	2.3　AI セーフティの評価環境の設計および試作	今回作成予定の試作については「優先的な表示言語は英語とする」と記載があります。つきましては、試作に関する提出ドキュメントについても、英語でのご提出を予定しておりますが、この認識で相違ないでしょうか。	仕様書記載の通り、調査報告書と設計書は日本語での記載としてください。 引用部などには英語が混在しても問題ありません。 また、評価環境の試作物についての「優先的な表示言語は英語とする」に関しては、最優先で対応する表示言語が英語であって、日本語対応にも期待している意であることはご留意ください。 関連資料は特に言語の指定は無く、作業の過程で作成した資料をそのまま提出していただければ構いません。	2025 年 3 月 10 日