

「AIセーフティ強化に関する研究開発」 公募説明資料

2025年2月

国立研究開発法人新エネルギー・産業技術総合開発機構
A I ・ ロ ボ ッ ト 部

本説明内容は「公募要領」及び「研究開発等計画」から重要な部分を取り上げたものです。

ご提案いただくにあたっては、必ず「公募要領」及び「研究開発等計画」をご一読ください。

目 次	公募要領の 対応ページ
1. 事業内容及び公募対象 (背景・目的、事業内容、公募対象、取組スケジュール)	P.3-4
2. 応募要件・実施要件	P.4-5
3. 応募方法	P.5-8
4. 採択先の選定 (審査の方法、審査基準、選定スケジュール)	P.8-10
5. その他重要事項・留意事項 (提出情報の取り扱い、知財・データマネジメント、安全保障貿易管理)	P.14-15, 17-18
6. 問い合わせ先	P.10-11

背景

- 2023年11月の英国主催のAIセーフティ・サミット（わが国からは、岸田前総理がオンライン出席）を契機として、各国はAIセーフティ・インスティテュート（AISI）を立ち上げ、AIセーフティの確保に関する議論を急速に進めている。こうした国際的な情勢も踏まえ、わが国としても、第7回AI戦略会議における総理指示を踏まえ、AIセーフティ・インスティテュート（AISI）を立ち上げ、国際的な議論に参画している。本事業は、こうした官民一体としての取組を研究開発の側面から支援する。
- 特に、安全であるかを見極める評価技術は、生成AIを適切に管理・利用するために必要である。欧州AI法に代表されるように、リスクベースのアプローチにより、安全性を4段階のカテゴリで評価し対策を義務付けるといった法規制が進みつつある。2024年11月の米国（国務省、商務省）主催の国際AISIネットワーク会合につづき、2025年2月にはフランス主催のAIアクション・サミットが予定されており、今後AIセーフティについては相当程度の高頻度で国際的な議論が進むと想定されることから、早急な技術開発を目指すもの。

目的

- 生成AIを適切に管理・利用するために必要となる安全性評価技術の開発・普及を目的に、**
 - ① リスクベースアプローチの基になる安全性の“ものさし”（基準）となる技術（評価・管理技術）の開発（AIセーフティ評価・管理基盤技術）、
 - ② 人間拡張技術などのサイバー空間とフィジカル空間をつなぐ領域での評価手法の開発と実証およびテスト環境構築技術の開発（応用領域別AIセーフティ評価・実装技術）、
 - ③ 国際標準化および普及のためのガイダンス等の整備（AIセーフティ基準・ガイダンスと標準化）を実施する。

研究開発項目① リスクベースアプローチの基になる安全性の“ものさし”（基準）となる技術（評価・管理技術）の開発（AI セーフティ評価・管理基盤技術）

- 生成AIのリスクを評価・管理するための基盤技術は、世界的にも研究開発途上にあるが、生成AIについてのAIセーフティ基準（体系化した評価観点・評価水準及び評価・管理手法）を策定するために不可欠である。そこで、システムライフサイクル全体のリスクアセスメント技術を開発する目的で、データ、学習モデル、AIシステムのそれぞれの評価・管理技術を開発する。
- 評価・管理技術の開発では、データ、モデル、AIシステムの評価観点と評価水準の整理および評価・管理の要素技術の基礎設計、評価ツールのプロトタイプ試作およびその課題点分析、評価用ベンチマークデータセットが具備すべき要件整理、ベンチマークデータセットに基づく試験的安全性評価などを行う。
- 研究開発項目③のAIセーフティ基準の検討・改定に役立てる目的で、前記の取組を整理した報告書を作成する。

※研究開発等計画 P.4 に「研究開発の詳細（参考）」の記載がございますので、合わせてご覧ください。

研究開発項目② 人間拡張など暮らし領域での評価手法の開発と実証およびテスト環境構築技術の開発（応用領域別AIセーフティ評価・実装技術）

- ・ 応用領域として、サイバー空間とフィジカル空間での人間の行動や経験を拡張するAI技術（人間拡張AI技術）など、「暮らし」領域を対象とし、「暮らし」領域に特有のAIセーフティを評価するための評価シナリオ（想定され得るリスク状況・動作のパターン）やベンチマークデータセットなどを用いた評価手法を開発する。
- ・ セーフティ評価に用いる実フィールドを模擬したテスト環境を構築する技術を開発する。
- ・ 「暮らし」領域の5件以上のAI製品・サービス等を対象として、評価手法の実証実験を行う。
- ・ 「暮らし」領域のAI技術のプライバシーの評価観点及び評価水準を策定し、前記5件以上のAI製品・サービス等への適用を通じて検証を行う。
- ・ これらの評価手法、評価観点及び評価水準を活用し、1か所以上の現場においてAI技術導入の検証を実施する。
- ・ 開発した安全性評価・実装技術についての知見を研究開発項目③のAIセーフティ基準へ反映させる。

研究開発項目③ 国際標準化および普及のためのガイダンス等の整備（AI セーフティ基準・ガイダンスと標準化）

- ・ 日本の産業界がAIセーフティに係る国際的なルール協調から取り残されないようにするために、AIセーフティ基準（体系化した評価観点・評価水準及び評価・管理手法）を策定し、AIセーフティ基準をまとめたAIセーフティガイドラインを公開・改訂する。
- ・ AIセーフティ・インスティテュート（AISI）コンソーシアムと連携し、企業向けの実装解説等（ガイダンス）を策定し、立ち上げるAI安全性技術コミュニティを通じて、社会実装・普及の促進を行う。
- ・ ISO/IEC等でのAI標準の国際的議論に貢献し、AISIに成果を提供する。

対 象： AIセーフティ強化に関する研究開発（委託）

予算規模： 14億円程度（NEDO負担率：100%）

事業期間： 2025年度（1年間）

留意事項

- 予算規模は変動する可能性があります。
- 研究開発項目①～③は、相互に連携して研究開発を進める必要があるため、**全体提案のみ**とします。

取組スケジュール

年度	令和6年度			令和7年度														
月	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3			
①AIセーフティ評価基盤技術	NEDOによる 公募・採択																	
①-1 データの評価・管理技術				個別の評価手法の研究開発						評価手法の実証実験						基準へ反映 論文執筆		
①-2 モデルの評価・管理技術																		
①-3 システムの評価・管理技術				ハードウェア等調達			ツール設計			ツール開発			評価			プラットフォーム化		
②応用領域別の評価・実証技術				評価シナリオの開発			テスト環境構築技術の開発									AIセーフティ 基準への 反映		
				評価手法の開発														
				データ品質管理技術の開発														
				ベンチマークデータセットの開発														
				安全なシステムを構築するための基盤モデル技術の開発														
				AI利用リスク低減技術の開発														
③基準・ガイダンスと標準化																		
③-1 AIセーフティ基準				AIセーフティ基準・ガイドライン策定						ガイドライン改訂								
										NIST RMFとの対応資料						ISO規格との対応資料		
③-2 社会実装・普及の促進				AI安全性技術コ ミュニティ立ち上げ			企業向け実装解説の作成											
③-3 標準化活動と国際連携				ISO技術体系の 国内実装の検討			ISO/IECとAISI技術体系 の整合性整理資料						機能安全規格等の策定推進					

公募・採択に関する詳細なスケジュールは後述

【応募要件】

- 本事業は、研究開発と Society5.0 との橋渡しプログラム（B R I D G E）として実施します。
 - 応募資格のある法人は、次の（1）～（9）までの条件、「研究開発等計画」（資料1）及び「2025年度実施方針」（資料2）に示された条件を満たす、単独又は複数で受託を希望する企業・大学等とします。なお、国立研究開発法人が応募する場合、国立研究開発法人から民間企業への再委託又は共同実施（再委託先又は共同実施先へ資金の流れがないものを除く。）は、原則認めておりませんのでご注意ください。
- （1）B R I D G E に関する、「科学技術イノベーション創造推進費に関する基本方針」（資料3）及び「研究開発と Society5.0 との橋渡しプログラム運用指針」（資料4）を十分に理解していること。
 - （2）「研究開発等計画」（資料1）及び各省庁プログラムディレクター（各省PD）等の意向を踏まえながら、B R I D G E 関係者（内閣府、経済産業省等の関係省庁やその他実施機関を含む）と密に連携・協力した上で事業を実施することができること。
 - （3）当該技術又は関連技術の研究開発の実績を有し、かつ、研究開発目標達成及び研究計画遂行に必要な組織、人員等を有していること。
 - （4）委託業務を円滑に遂行するために必要な経営基盤、資金及び設備等の十分な管理能力を有し、かつ、情報管理体制等を有していること。
 - （5）NEDOが事業を推進する上で必要とする措置を、委託契約に基づき適切に遂行できる体制を有していること。

- (6) 企業等が事業に応募する場合は、当該事業の研究開発成果の実用化・事業化計画の立案とその実現について十分な能力を有していること。
- (7) 研究組合、公益法人等が応募する場合は、参画する各企業等が当該事業の研究開発成果の実用化・事業化計画の立案とその実現について十分な能力を有するとともに、応募する研究組合等とそこに参画する企業等の責任と役割が明確化されていること。
- (8) 複数の企業等が共同して事業に応募する場合は、実用化・事業化に向けた各企業等間の責任と役割が明確化されていること。
- (9) 本邦の企業・大学等で日本国内に研究開発拠点を有していること。なお、国外の企業・大学等（研究機関を含む）の特別な研究開発能力、研究施設等の活用又は国際標準獲得の観点から国外の企業・大学等との連携が必要な場合は、国外の企業・大学等も参画する形で実施することができる。

【実施要件】

- 本事業は、採択後、業務委託契約を締結します。業務委託契約の締結にあたっては、最新の「業務委託契約約款」を適用します。その他必要に応じて、特別約款の適用を求める場合があります。また委託業務の事務処理においては、NEDOが提示する事務処理マニュアルに基づき実施いただきます。
- 事業の実施にあたっては、該当する約款およびマニュアルを遵守いただくことが要件となります。

【参考】 委託事業の手続き：約款・様式 <https://www.nedo.go.jp/itaku-gyomu/yakkan.html>
委託事業の手続き：マニュアル <https://www.nedo.go.jp/itaku-gyomu/manual.html>

- Web入力フォームより 必要情報の入力と提案書類等の提出を行ってください。
<https://app23.infoc.nedo.go.jp/koubo/qa/enquetes/e5c3jel14a34>
- 持参、郵送、FAX又はE-mailによる提出は受け付けません。

提出期限：**2025年3月17日（月）正午** アップロード完了

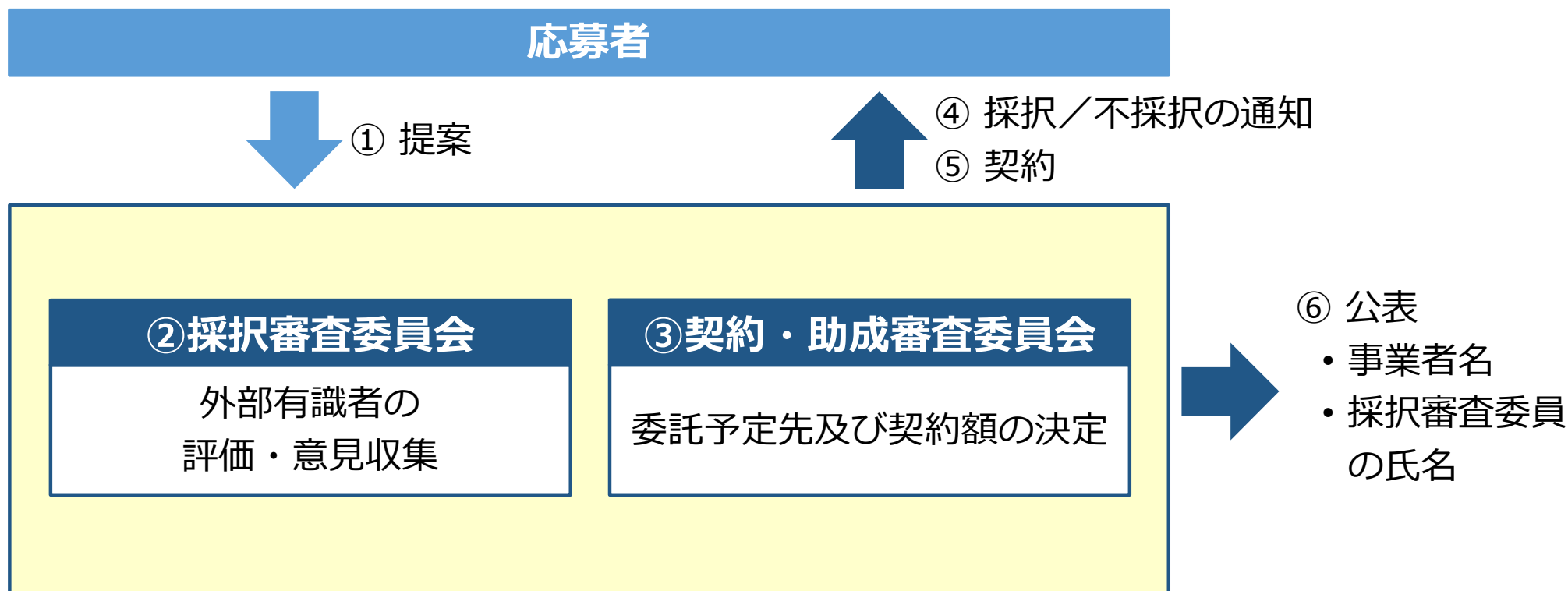
提出書類
別添1：提案書
別添2：研究開発統括責任者候補及び研究開発責任者の研究経歴書
別添3：企業情報
別添4：ワーク・ライフ・バランス等推進企業に関する認定等の状況
別添5：事業遂行上に係る情報管理体制等の確認票
e-Rad応募内容提案書
直近の事業報告書及び直近3年分の財務諸表（原則、円単位：貸借対照表、損益計算書（製造原価報告書、販売費及び一般管理費明細書を含む）、株主（社員）資本等変動計算書）

全てPDF化し、1つのzipファイルにまとめてアップロード

※資料差し替え時も、全ての資料をまとめて再提出してください。

※アップロードするファイル（PDF、zip等）にはパスワードは付けないでください。

- 外部有識者による**採択審査委員会**とNEDO内の**契約・助成審査委員会**の二段階で審査します。
- 契約・助成審査委員会では、採択審査委員会の結果を踏まえ、NEDOが定める基準等に基づき、最終的に実施者を決定します。
- 必要に応じてヒアリング審査や資料の追加、代表者面談等をお願いする場合があります。
- 採択先の選定は非公開で行われ、審査の経過等、審査に関する問い合わせには応じられないのであらかじめご了承ください。



i. 事業の適合性

- ・ 本事業の目的・目標に適合しているか 等

ii. 開発の優位性

- ・ 開発内容に新規性・優位性等があるか 等

iii. 計画の妥当性

- ・ 達成目標が明確で、スケジュールが効率的・効果的か 等

iv. 社会実装の取組

- ・ 社会実装のターゲットが明確で、それに向けた取組に実現性・実行性があるか
- ・ 開発成果が展開されることで、安全性の高いAIシステム開発が促進され、安全性向上に貢献すると考えられるか
- ・ 社会・経済への波及効果が期待できるか 等

v. 実施体制・能力

- ・ 役割分担が明確で効率的な体制か
- ・ AISIと連携する体制が構築できているか、具体的な連携方法について想定はあるか
- ・ 必要な人員・設備・支援体制や関連分野の開発実績を有するか
- ・ AIセーフティに関連する評価技術開発、ガイドラインの作成及び国際標準への規格化に関する十分な知見と実績を有しているか 等

vi. 提案の経済性

- 予算の範囲内で必要経費を適切に計上しているか
- 事業との重複なく妥当な予算規模か 等

vii. 総合評価

加点要素

- 女性活躍推進法に基づく認定企業(えるぼし認定企業・プラチナえるぼし認定企業)、次世代育成支援対策推進法に基づく認定企業(くるみん認定企業・プラチナくるみん認定企業・トライくるみん認定企業)、若者雇用促進法に基づく認定企業（ユースエール認定企業）に対しては加点します

2025年

2月

3月

4月

5月

6月

★2月14日
公募開始

★3月17日(正午)
公募締切

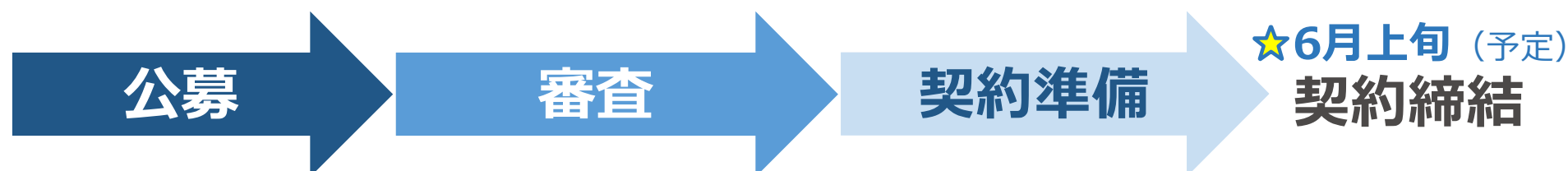
※1者応募の場合、10日間程度延長します。
延長の場合は、ウェブサイトでお知らせします。

★4月上旬 (予定)
採択審査委員会

(外部有識者による書面審査およびヒアリング審査)

★4月中旬 (予定)
契約・助成審査委員会

★4月中旬~下旬 (予定)
委託予定先決定・公表



- NEDOは、応募書類等の提出書類について、公文書等の管理に関する法律に基づく行政文書の管理に関するガイドラインに沿い定められた関係規程により、**厳重な管理の下、一定期間保存**します。
- 取得した個人情報については、法令等に基づく場合の提供を除き、研究開発の実施体制の審査のみに利用しますが、特定の個人を識別しない状態に加工した統計資料等に利用することがあります。
- 提案書の一部である「研究開発統括責任者候補及び研究開発責任者の研究経歴書（CV）」については、**採択先決定後**、適切な方法をもって**速やかに廃棄**します。
- なお、e-Radに登録された各情報（プロジェクト名、応募件名、研究者名、所属研究機関名、予算額及び実施期間）及びこれらを集約した情報は、「独立行政法人等の保有する情報の公開に関する法律」（平成13年法律第140号）第5条第1号イに定める「公にすることが予定されている情報」として取り扱われます。

知財マネジメント

- 「NEDOプロジェクトにおける知財マネジメント基本方針」を適用します。
- 産業技術力強化法第17条（日本版バイ・ドール規定）が適用されます。
- 本プロジェクトの成果である特許等について、「特許等の利用状況調査」（バイ・ドール調査）にご協力をいただく場合があります。
- 詳細については以下のリンクより、内容をご理解ください。
https://www.nedo.go.jp/jyouhoukukai/other_CA_00002.html

データマネジメント

- 「NEDOプロジェクトにおけるデータマネジメント基本方針」を適用します。
- 詳細については以下のリンクより、内容をご理解ください。
https://www.nedo.go.jp/jyouhoukukai/other_CA_00003.html

- 外為法で規制されている貨物や技術を輸出（提供）しようとする場合は、原則外為法に基づく経済産業大臣の許可を受ける必要があります。
- 本事業により外為法の輸出規制に当たる貨物・技術の輸出が予定されているか否かの確認及び、輸出の意思がある場合は、管理体制の有無について確認を行う場合があります。
- 輸出の意思がある場合で、管理体制が無い場合は、輸出又は本事業終了のいずれか早い方までの体制整備を求めます。

国立研究開発法人新エネルギー・産業技術総合開発機構
A I ・ ロボット部 「AIセーフティ強化に関する研究開発」 担当

電子メール： project_aisafety@nedo.go.jp

受付期間： 3月17日(月)まで

※審査の経過等に関する問い合わせには応じられません。

Q.1 本プロジェクトではAISIとの連携及び国際協調がフォーカスされているが、現段階で関係機関とアクション等は取っているのか。

A.1 具体的な連携方法はまだ明確ではありませんが、今後日本のAISIとの連携を進めていく予定です。

進捗状況を報告し、それに対するAISIのコメントを考慮しながら進めていくことを考えています。

AISIとの関係をどのようにするかは、できること／できないことを考慮しつつ、実際に進めながら決めていく予定です。

Q.2 公募説明会資料のP 7にあるスケジュール表は確定しているのか、それともあくまで仮の予定なのか。また、提案時は記載されているスケジュールからどの程度変更して良いのかも教えてください。

A.2 記載のスケジュールは確定したものではなく、あくまで想定です。

提案内容に応じてタイムラインや項目が前後することもあるかと思います。

3月時点でAIセーフティ基準等を策定いただくことは必須ですので、それに合わせて柔軟に調整いただければと思います。

Q.3 ガイドラインの策定で14億円の予算が設定されているのではなく、テスト環境や評価手法等の開発も含めて14億円の予算が設定されているのでしょうか。

A.3 ご認識の通りです。研究開発項目①～③の全体で14億円の予算が設定されています。

Q.4 研究開発計画の実施体制図には再委託先が2つ示されていますが、必ずしも2つの再委託先を設定する必要があるのでしょうか。
また、再委託の中に含まれる項目は、細分化されることもあれば、1つの団体がすべてを担当することも可能なのでしょうか。

A.4 必ずしも2つの再委託先を設定する必要はありません。
研究開発計画の実施体制は、あくまで一つの例として示されています。
また、ご提案の内容によっては再委託先が細分化されることもあります。また、1つの団体がすべてを担当することも可能です。

Q.5 研究開発項目②では、「「暮らし」領域の5件以上のAI製品・サービス等を対象として、評価手法の実証実験を行う。」と記載されていますが、応募する際に、暮らし領域に関する製品やサービスを持っている会社を応募メンバーとして招く必要があるのでしょうか。

それとも、上記とは別に検証対象として協力いただければいいのでしょうか。

また、応募する企業の中で暮らし環境領域の検証モデルを持っていれば、それだけで十分なのでしょうか？

A.5 暮らし領域に関する具体的なサービスや製品を持っている企業を必ず巻き込む必要はありません。

暮らし領域に関する製品を提供できる企業と協力することや、実際に販売されている製品を用いても問題ありません。

Q.6 この公募は全体的な提案ですが、研究開発項目②に「5件以上」と記載されています。この場合、研究開発全体で5件以上を目指すのか、それとも各提案ごとに5件以上を満たす必要があるのか、どちらなのでしょうか。

A.6 各提案において全て5件以上満たしている必要があります。

