

教育SWG上期活動報告

教育SWG

AISI 評価リサーチグループ

深井雄介

2026年10月7日

AISI事業実証WG 2026年度上期報告会

AISI Japan
AI Safety Institute



深井 雄介

Yusuke Fukai

A I セーフティ インスティテュート 評価リサーチグループ

- 幼稚園、小学校、中学校・高校（英語）教員免許保有
- 教育分野、とりわけ初等中等教育段階における生成 A I 利活用や、教育 D X ・教育データ利活用等の事業に従事した実績を有する
- A I S I では、教育分野での A I の安全・安心な活用の促進に向け、教育 S W G にて、A I セーフティ評価の枠組みの整備に取り組む
- I P A プロジェクトマネージャ、応用情報技術者

- 教育分野、とりわけ初等中等教育段階では、児童生徒や教師のAI活用の可能性とともにリスクが明らかになってきた
- 安全・安心なAIの利活用の促進に向け、**ユースケース毎のリスクを基に、AIセーフティの評価観点の整理**が必要

教育分野の AI利活用の 現状

- 教育分野、とりわけ**初等中等教育段階**では、**文部科学省**は、学校現場で生成AIを利活用する際の**考え方や留意事項をガイドラインとして公表**
 - ✓ 『初等中等教育段階における生成AIの利活用に関するガイドライン Ver.2.0』
(文部科学省：2024年12月)
- 国内外で、教育分野の様々な場面でAI利活用が進むとともに、児童生徒や教師向けの多様なAIサービスが開発・実装され、**AIの利活用に関して、可能性やリスクが明らかになってきた**
 - ✓ 文部科学省での生成AIパイロット校事業*1や教育分野特化の生成AIの実証事業*2等
 - ✓ OECD『Digital Education Outlook2026』(2026年1月公表) 等

AIセーフティ 評価の重要性 と利活用促進 への期待

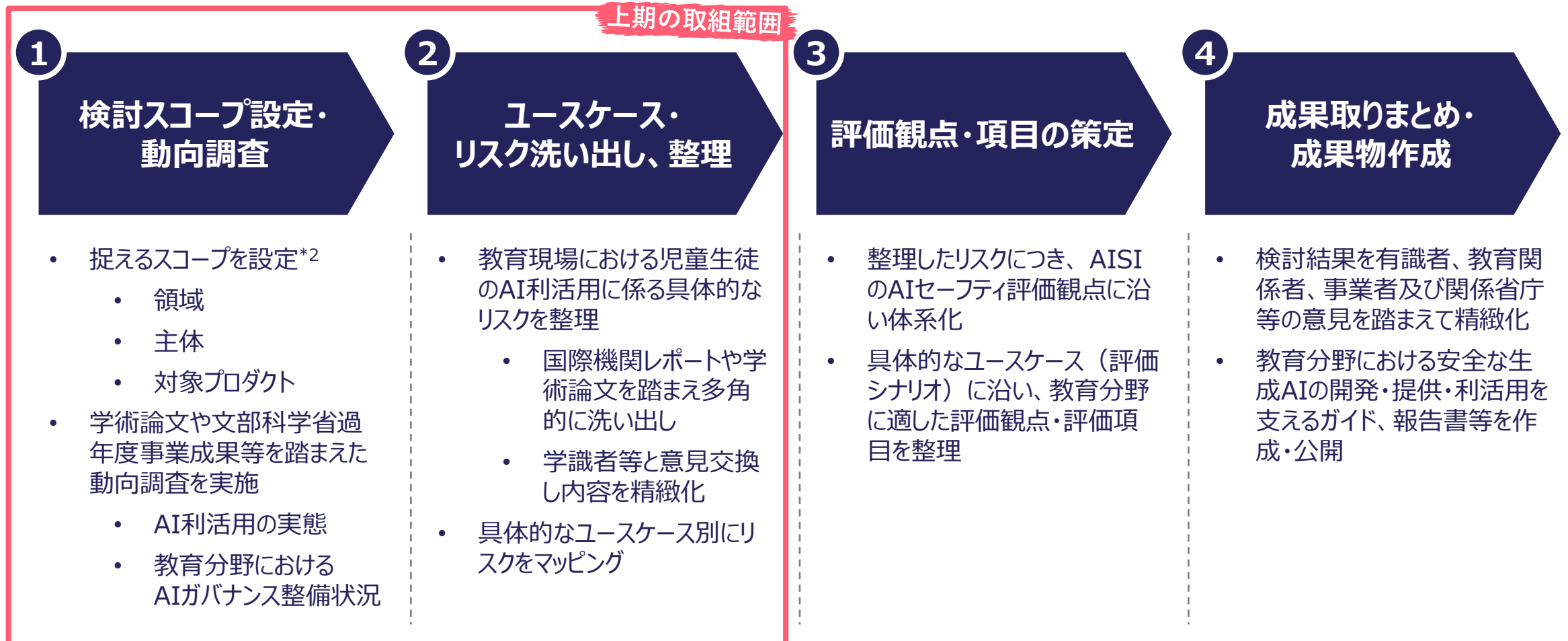
- **安全・安心なAI利活用**への期待
⇒教育分野でのAI利活用の際は、単に利便性のみならず、学校教育に係る各種法令や教師の教育的意図等、**教育目的に照らして安心・安全に利活用できるかを評価**する視点が必要
- 教育分野のユースケースごとのリスク整理
⇒個別の学習場面等のユースケースのリスクを踏まえ、**AIセーフティの評価観点の整理**が求められる
- 評価基準・評価データ等の整備
⇒AIの安全性とセキュリティのベンチマークの取組や新たな技術等の進展も踏まえつつ、**教育分野における評価観点ガイド**や、**AI安全性ベンチマーク**及び**評価環境整備**により、教育分野での安全・安心で主体的なAIの利活用の促進につなげていく

活動目的・ 内容

*1 文部科学省「リーディングDXスクール事業 生成AIパイロット校」(令和6年度)、「生成AIパイロット校の指定を通じた事例創出事業」(令和7年度、令和8年度)

*2 文部科学省「学びの充実など教育課題の解決に向けた教育分野特化の生成AIの実証研究事業」(令和7年度)、「教育課題の解決に向けた教育分野特化の生成AIの実証事業」(令和8年度)

- 教育SWGでは、『AIセーフティに関する評価観点ガイド』に示すAIセーフティ評価観点*¹に沿いながら、教育分野で想定される固有のリスクと、当該リスクに対応した評価観点・項目を整理していく



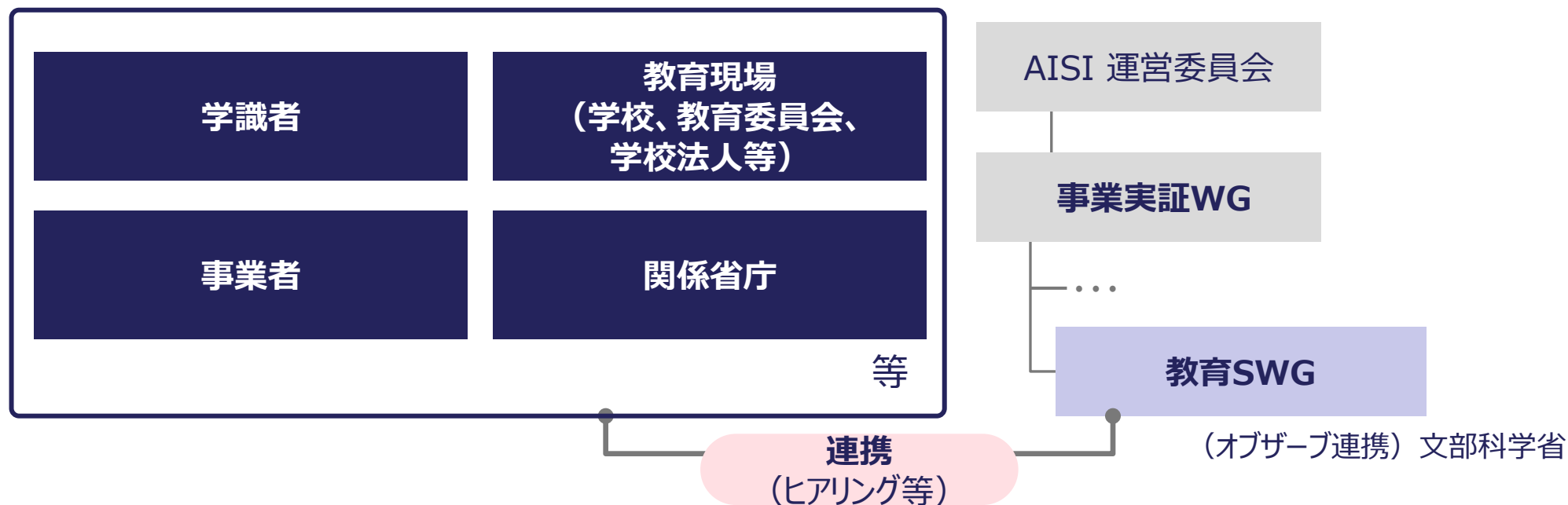
*¹ 2026年7月改定では、既存の10の評価観点（有害情報の出力制御等）に加え、AIエージェント特有の評価観点が2つ追加されている（[AIセーフティに関する評価観点ガイド - Japan AISI](#)）

*² 教育SWGでは、現下は、初等中等教育段階における児童生徒の生成AI利活用に関するリスクの整理をスコープとしている

- 今年度新規に立ち上がった教育SWGでは、まずはユースケース・リスク洗い出し、リスクの整理を優先的に進めている
- リスクの洗い出し・整理にあたっては、学識者等の有識者とヒアリング等を通じた連携を重ねる体制で推進している

有識者

AIセーフティ・インスティテュート (AISI)



- OECDやUNESCO等の国際機関では、児童生徒・子ども、教師のAI利活用に係る効果とリスクに言及した報告書/レポートを公表。メタ認知的怠惰や、依存、擬人化などのリスクが指摘されている

OECD Digital Education Outlook 2026 (OECD, 2026)*1

2026年1月19日発行

- 生成AIは、
児童生徒の個別最適な学びや教師の校務効率性向上等に資する
- 他方、児童生徒の
認知オフロードによるメタ認知的怠惰等のリスクも存在
- **生成AIへの過度な依存**は、学習者を受動的な消費者へ、教師を単なる監督者へと変えてしまうリスクもある
- 児童生徒が、生成AIを利用し課題をこなすことは、
自動的に「学び」につながるわけではない場面もある
- 汎用型か教育分野特化かに関わらず、生成AIサービスは、教師による
教育的な意図のもとに学習場面の中で活用されることが重要
- また、教師は出力結果を吟味・評価し、必要に応じて修正・承認し、
最終的に質に責任を持つ姿勢が重要

The algorithm in the room (UNESCO, 2026)*2

2026年9月8日発行

- 教師がAIで課題を作成し、生徒がAIで解答し、AIが採点する
「Classroom AI Doom Loop」は、
知識伝達のサイクルから人間が実質的に排除されるリスク
- 子どもとAIチャットボットの関係について、AIは「いつでも応答し、個別適応する」ため、従来のメディア以上に強い**擬似的対人関係**（parasocial attachment）を形成する可能性
- 小学校の実践報告では、AIが即座に答えを与えることで、
仮説形成や議論、発見の喜びといった「**理解が生まれるプロセス**」が失われる懸念が示されている
- AIによる文化表象や知識生成では、
特定の文化や視点が強調され、別の視点が不可視化される可能性

いずれも仮訳であり、原文との相違がある場合には原文を優先します。上記の一切は、引用のために示すものである。記載内容は、AISIIにて抜粋したものであり、各レポートの記載内容の全量ではないことに留意。

*1 OECD (2026), *OECD Digital Education Outlook 2026: Exploring Effective Uses of Generative AI in Education*, OECD Publishing, Paris, <https://doi.org/10.1787/062a7394-en>. (参照 2026/9/17)

*2 UNESCO (2026), *The algorithm in the room: Seeing and confronting the implications of AI for the future of education*. Licensed under CC BY-SA 3.0 IGO. <https://doi.org/10.54675/PMRO5023>. (参照 2026/9/17)

教育分野のAI利活用に関する国際動向（2 / 3）

- イギリスでは教育分野での安心かつ責任ある生成AI利活用の政策文書が公開されているほか、エストニアでは、全国規模でのAI利活用に関する教育プログラムが実施されている

イギリス | Generative AI in education (2025)*1



- 安全かつ責任ある生成AI利活用の推進に関する政策文書
- 学校教育関係者に加えて、**保護者やEdTech事業者**が対象者
- 生成AIは安全・効果的で、責任のある利活用を前提に、児童生徒の学びを支える技術として位置づける
- 生成AIは教師が教育に集中するための支援の手段であり、**教師と児童生徒の関係は代替されない**としている
- 生成AI利活用に係るリスクを8観点で整理
 - データ保護
 - 知的財産権
 - サイバーセキュリティ
 - フィルタリング・監視基準
 - 安全管理
 - 評価不正
 - 児童生徒の権利保護
 - 他省のガイドラインとの接続

エストニア | AI Leap (2025-)*2



- 全国規模のAIに係る教育プログラムを展開
- 教育現場でのAI活用能力の定着を目指す

教師向けプログラム

- **研修教材や、教員コミュニティ(PLC)の整備**
- 教科別の指導教材や活用ガイドの作成
- 教育省等と連携したチェンジマネジメントの推進

児童生徒・学生向けプログラム

- 青少年団体との連携
- **ハッカソン等によるAI活用事例の創出**
- 対話型**AIリテラシーワークショップ**の実施

技術基盤整備・開発

- **教育分野特化の生成AIの開発・継続的改善**
- 高校教員へのChatGPTまたはGoogle Gemini 有料版の提供

(出典)

*1 [Generative artificial intelligence \(AI\) in education - GOV.UK](#) *2 [TI-Hüpe](#)

- イギリスの教育分野での生成AIプロダクトの安全性基準に関するガイダンスでは、主にEdTech開発者やサプライヤーを対象に、ユースケースの明示や、フィルタリング、学習者の発達への配慮等を求めている

イギリス 教育省 | Generative AI: product safety standards (2026)*1



2026年の改定では、認知発達、情緒的・社会的発達、メンタルヘルス、操作的な誘導の防止の4観点が追加となっている

- 教育現場で利用される生成AIプロダクトやシステムが**安全とみなされるために求められる機能や設計要件**を示す
- **主にEdTech開発者やサプライヤー**が対象者
- 学校やカレッジにとって「**教育での利活用**に際し、**いずれのAIプロダクトが安全か**」を判断する上で有用になり得る
- 求められる機能や設計要件を13領域で整理
 - **利用目的の明示**：利用目的・想定ユースケースの明示
 - **教育におけるユースケース**：該当する利用区分の選択
 - **フィルタリング**：有害・不適切なコンテンツへのアクセス防止
 - **モニタリング・報告**：利用記録・監査等のログ管理の体制整備
 - **セキュリティ**：悪意ある利用・危害を防ぐ安全対策
 - **プライバシー・データ保護**：データ保護に関する法令や規則への適合
 - **知的財産**：同意のない著作物の保存・収集・利用の禁止
 - **設計・テスト**：透明性・子どもの安全を重視した設計・検証
 - **ガバナンス**：説明責任を確保する運用体制の整備
 - **認知発達**：認知能力・技能や発達への悪影響の軽減
 - **情緒的・社会的発達**：情緒的依存や発達への悪影響の軽減
 - **メンタルヘルス**：心理的不調の検知・人的支援への接続
 - **操作的な誘導の防止**：利用者への不当な誘導・搾取の防止

（出典）

*1 [Generative AI: product safety standards - GOV.UK](#)（2025年1月初版公開、2026年1月更新）。上記は仮訳であり、正確には原典をご参照ください

教育分野でのAIセーフティ（想定されるリスク例）

- 有識者ヒアリング等を通じて、認知・非認知的なリスクや心理的リスク、技術的リスク等の整理を開始
- 今後は、具体的なユースケースごとにリスクを体系化し、必要なAIセーフティ評価観点・項目を検討する

主要リスク（例）	リスク内容/シナリオ
感情的愛着の形成	- 孤独感、不安、対人的な繋がりへのニーズ等を背景として、子どもがAIを擬人化し得る - 安心感や情緒的支援を求める中で感情的愛着が形成され、子どもが相談、意思決定を生成AIに過度に依存し得る
機微情報の収集	支援を個別化する上で、AIコンパニオンは、メンタルヘルス、所在地又はアイデンティティに関する情報を含む個人データを日常的に収集し得る
生産的試行錯誤の喪失	- 生成AIへの過度な依存は、学習方略を調整するメタ認知的な関与を弱める「メタ認知的怠惰」を招き得る 生産的な試行錯誤（productive struggle）を失わせるとともに、学習者のエージェンシーを低下させ得る
AI依存による成績低下	- AIの支援を受けて課題を遂行できたことを、自らの理解や能力によるものと誤認し、実際の習熟度を過大評価し得る - AIを活用した後にAI支援なしでテストをした場合、成績が低下し得る
エコチェンバーの形成	個別最適化が学習のエコチェンバーを生み、誤概念を段階的に増幅し得る
見方・考え方の画一化	狭く偏った教材やフィードバックが、児童生徒の世界観を画一化・狭小化し得る
プロンプトインジェクションでの評価操作	提出物をAIで評価する場合に、悪意ある指示を埋め込み、挙動を乱すことで、不適切に良い評価に繋がり得る
教師負担の増大	学術的誠実性（academic integrity）の観点で、教師がAIによる不正を検出・防止するために課題を再設計せざるを得ず、負担が増大し得る

(参考) ユースケース別のリスク整理イメージ

① スコープ・動向調査

② ユースケース・リスク整理

生成AIによる習熟度に応じた問題生成やヒント提示は、個別最適な学びを支援し得る一方、発達の段階に合わない不適当な問題の生成や過度な解法への誘導により、試行錯誤の省略や理解度誤認等を招くリスク

学年・教科

中学3年生・数学

単元*1

「A 数と式」(2)簡単な多項式

ユースケース*2

児童生徒の習熟度に合う因数分解の問題を生成AIに作成させ、分からない場合はヒントを求めながら解く

学習過程

プロンプト入力

問題生成

試行錯誤

問題解決

主な
学習活動

- 教師が作成したプロンプト例を参考に児童生徒がプロンプトを作成
- 児童生徒が難易度や条件を調整しながらプロンプトを修正

- 生成AIが因数分解問題を作成
- 児童生徒はアウトプットを踏まえ問題レベルや精度を調整・プロンプト再入力
- 児童生徒は解く問題を選択

- 児童生徒が因数分解を解く
- 児童生徒は分からない箇所を特定し、生成AIにヒントを求める

- 生成AIが出力したヒントを基に児童生徒は問題を解く
- 児童生徒は理解度に応じて次の問題へ進む

主なリスク

- 誤情報・不透明性
不成立な問題や誤った解法が示されるリスク

- 学習過程の外注化・過度な依存
直接的過ぎるヒントである場合、試行錯誤が省略されるリスク
- 知識・技能の弱化
自力で式を変形する機会が減り、知識技能の定着を低めるリスク

- 見せかけの成果・自己認識の歪み
支援下で解けたことを自分の資質・能力の向上と誤認するリスク

*1 【数学編】中学校学習指導要領（平成29年告示）解説

*2 ユースケースは、文部科学省「生成AIパイロット校の指定を通じた事例創出事業」等での各学校の取組を参考に、AISIIにて作成

- 教育分野におけるAIの安全・安心な活用（AIセーフティの確保）に向け、今年度は初等中等教育段階におけるユースケース・リスク、評価観点を取りまとめガイドとして公表を予定

短期的な取組（令和8年度）

- 初等中等教育段階における児童生徒の生成AI利活用に係るユースケース・リスクの整理
- ユースケース・リスク毎にAIセーフティ評価観点の取りまとめ

中長期的な取組（令和9年度～）

- 教育分野における児童生徒、学生、教職員を中心とするAI利活用に係るユースケース・リスクの整理・更新
- ユースケース・リスク毎にAIセーフティ評価観点の取りまとめ
- 評価シナリオ・ベンチマーク設計（教育分野に特化したデータセット構築含む）
- 教育分野のAI提供者向けガイドラインの作成・更新

教育分野におけるAIセーフティ評価観点ガイド（仮）



【主な記載アジェンダ（想定）】

- 教育分野におけるAI動向
- 教育分野におけるリスク整理
- 教育分野におけるAIセーフティ評価
- 具体的なユースケース別整理

（現時点の想定であり、変更となる場合がある）

- **教育分野では**、児童生徒や教師向けに、教育分野特化の生成AIをはじめ、**多様な生成AIサービスが開発・実装**され、教育課題の解決に向けた事例や実証が創出され始めている
- この中で、**教育分野におけるA I の安全・安心な活用**（A I セーフティの確保）に向けて、学校現場だけでなく、生成AIサービスを開発・実装する事業者も含めた**リスクガバナンスの取組推進**が求められる

- OECD, "OECD Digital Education Outlook 2026," 2026.
(https://www.oecd.org/en/publications/oecd-digital-education-outlook-2026_062a7394-en.html)
- UNESCO, "The algorithm in the room: Seeing and confronting the implications of AI for the future of education," 2026.
(<https://www.unesco.org/en/articles/algorithm-room-seeing-and-confronting-implications-ai-and-futures-education>)
- UNESCO, "AI and the future of education: Disruptions, dilemmas and directions," 2025.
(<https://www.unesco.org/en/articles/ai-and-future-education-disruptions-dilemmas-and-directions>)
- UNICEF, "When AI becomes a friend: Child rights risks, harms, and regulatory responses to AI chatbots and companions," 2026.
(<https://www.unicef.org/documents/when-ai-becomes-friend-child-rights-risks>)
- 文部科学省,「学校教育におけるAI活用に関するこれまでの取組と今後の方向性」, 2026年6月30日.
(https://www.mext.go.jp/content/20260630-mxt_shuukyo01-000050664_2.pdf)
- 文部科学省,「初等中等教育段階における生成AIの利活用に関するガイドライン」(Ver.2.0) ,2024年12月26日.
(https://www.mext.go.jp/content/20241226-mxt_shuukyo02-000030823_001.pdf)
- Department of Education (United Kingdom), "Generative AI: product safety standards," 2026.
(<https://www.gov.uk/government/publications/generative-ai-product-safety-standards/generative-ai-product-safety-standards>)
- Sisi Tao, Min Lan, Minjuan Wang, and Hui Li, "Potential risks of generative artificial intelligence integration into K-12 education: A scoping review," Computers and Education: Artificial Intelligence, vol. 10, 2026.
(<https://www.sciencedirect.com/science/article/pii/S2666920X26000226>)

AISI

Japan AI Safety Institute