

令和6年度 最終評価様式

子どもからの聴取に関するAI訓練ツールの開発

令和7年5月
警察庁

- 実施する重点課題（特に該当するものには◎、そのほかで該当するものには○（複数可）を記載）

業務プロセス転換・ 政策転換に向けた取組	次期SIP/FSより 抽出された取組	SIP成果の社会実装 に向けた取組	スタートアップの事業創出 に向けた取組	若手人材の育成 に向けた取組	研究者や研究活動が 不足解消の取組	国際標準戦略の促進 に向けた取組
○					○	

- 関連するSIP課題（該当するものには○を記載）

持続可能な フードチェーン	ヘルスケア	包摂的 コミュニティ	学び方・ 働き方	海洋 安全保障	スマート エネルギー	サーキュラー エコノミー	防災ネット ワーク	インフラ マネジメント	モビリティ プラットフォーム	人協調型 ロボティクス	バーチャル エコノミー	先進的量子 技術基盤	マテリアル 事業化・ 育成エコ
			○										

1. 社会実装に向けた施策・取組等の全体俯瞰の中での成果（進捗の説明）

① 全体概要

1 解決すべき社会課題

- 現状
児童虐待等の事案において、警察、検察、児童相談所では、児童の心身の負担軽減と供述の信用性確保を図るため、事情聴取に先立って、児童から聴取する方法・内容等を関係機関で協議し、代表者が聴取を行う「代表者聴取」という施策が運用されている。代表者聴取においては、児童から正確な情報を多く得るため、児童の特性を踏まえた構造化された聴取技法が用いられている。
- 課題
代表者聴取を適切に運用し、児童の心身の負担軽減と供述の信用性確保を図るために、聴取を担当する者は、構造化された聴取技法に基づく聴取技術を身につける必要がある。
警察においても、講義等の教養を行い聴取技術の習得に努めているが、知見を有する教養担当者が限られていることもあり、教養の日程調整を図る上で時間的・場所的制約が生じている。また、聴取技術の習得や維持・向上には、実際に聴取の経験を積むことが有効であり、ロールプレイング等による実践的な訓練も取り入れているが、そのような機会を定期的に設けるには、人的・経済的負担も大きい。これらは、教養・研修を必要とする分野において共通の課題と考えられる。

2 取組施策の内容

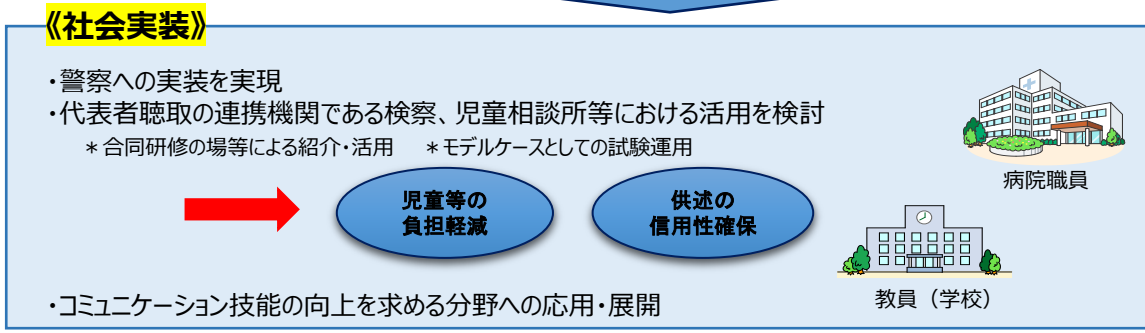
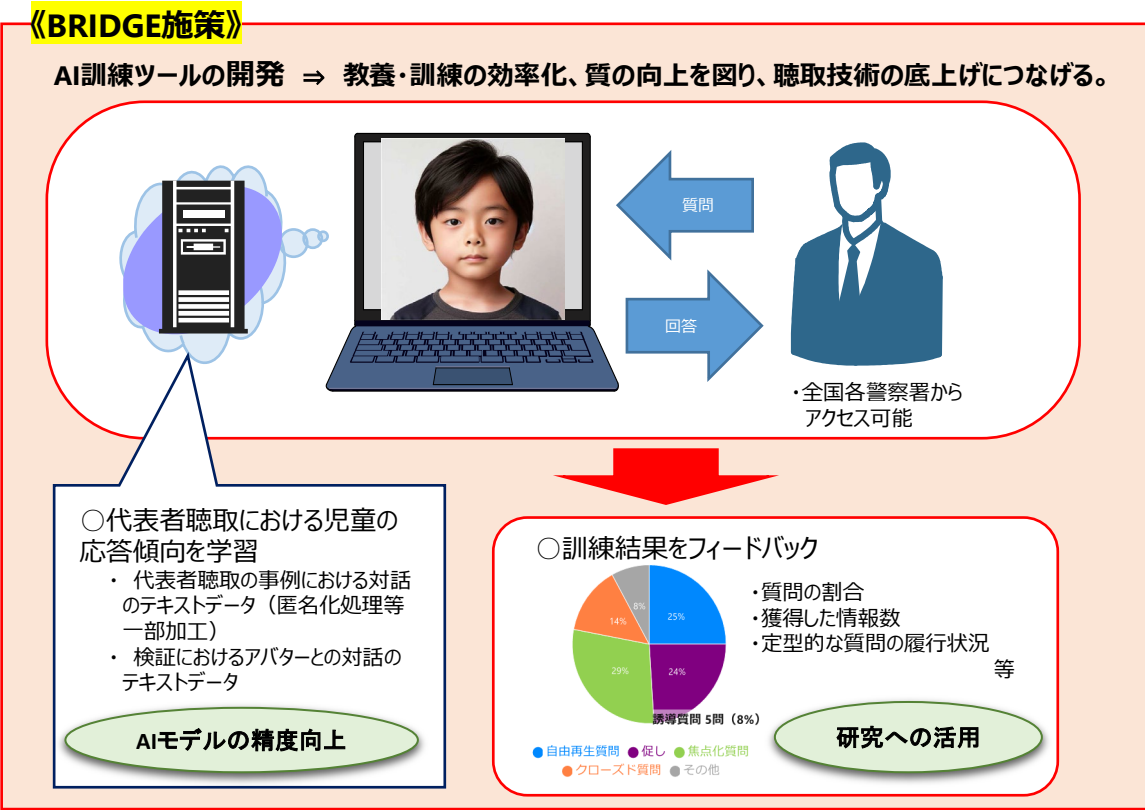
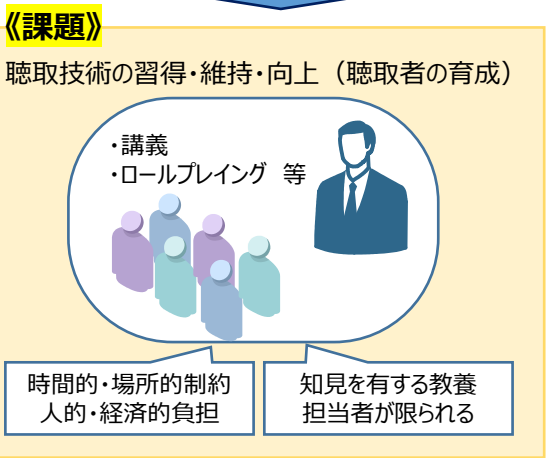
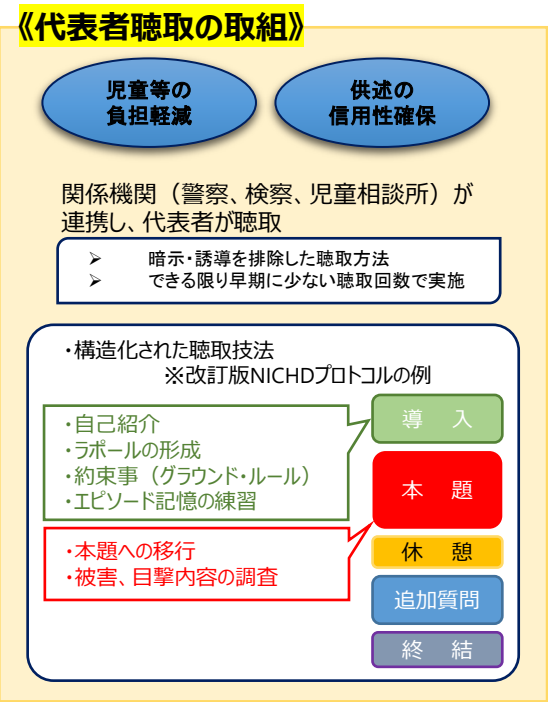
- 本施策は、AI技術を活用し、児童のアバターを相手に聴取の訓練ができ、その結果についても評価を得られる訓練ツールを開発するもので、AI訓練ツールの実装により、教養・訓練の効率化、質の向上を図り、聴取技術の底上げにつなげるものである。
- 業務プロセスの転換
AI訓練ツールの開発・実装が実現すれば、教養・訓練において必ずしも受講者を一箇所に集める必要はなく、教養担当者を必要とせずに関心のある人が必要なタイミングで訓練を行うことができるため、時間的・場所的制約が低減し、業務プロセスの転換が期待できる。また、受け答えを含め児童のアバターの精度が高いほど、実践的な訓練が可能となり、成人同士のロールプレイングに比べて聴取経験の質が向上することが期待される。
 - 研究活動の促進
対話によるコミュニケーション能力が求められる分野では、アバターを用いた訓練のニーズがあると考えられるが、アバターを用いた聴取訓練に関する研究知見は少なく、研究活動が不足している。AI訓練ツールの開発・実装を実現することで、研究活動への機運を高め、効果的なアバターの活用策や聴取技術を高める訓練手法の研究、AI技術への投資促進につながることが期待される。

3 成果の社会実装

- 代表者聴取における児童の応答傾向をAIモデルに学習させ、AI訓練ツールの開発を実現するとともに、警察組織で運用可能な環境を構築し実装した。検証においては、特に初心者に対して聴取技術を高める効果があることを示す結果が得られたことから、今後、検察、児童相談所等との合同研修の場など、機会を捉えてAI訓練ツールの紹介を行い、同機関における活用についても検討していく。
- 運用によって明らかとなる課題もあると考えられ、警察における運用実績を踏まえつつ他組織への実装を検討していくこととなるが、まずは、開発企業において、特定の自治体の児童相談所をモデルケースとして試験的に導入・運用していくことを検討している。
- また、本施策の成果を活用することで、コミュニケーション技能の向上を求める分野への展開も期待できる。

1. 社会実装に向けた施策・取組等の全体俯瞰の中での成果（進捗の説明）

② 全体俯瞰図



2. 研究成果及び出口戦略、達成状況（取組全期間）

テーマ

① AI訓練ツールの試作版の課題や基礎研究結果を踏まえた改訂版の開発と訓練・検証

① 研究成果及び達成状況

- 効果的な訓練を実現するため、アバターの回答精度の向上や利用者が訓練後に受けるフィードバックの工夫等が課題としてあげられたことを踏まえ、改訂版を開発した。
 - ・ アバターの回答精度については、導入する大規模言語モデルの検証、聴取事例や検証時の対話データを活用した追加学習、プロンプトの体系化等を行うことで改善が見られ、明らかに設定情報と矛盾する回答を生成する割合が低減した。
 - ・ フィードバックについては、「質問の割合と発話文字数の割合」「獲得した（事件に関する）情報」「定型的な質問の履行の有無」の3つを主要な指標とし、グラフ等を用いて視覚的にも分かりやすい構成とした。
上記3つのフィードバックの指標の判定にも大規模言語モデルを活用しており、その精度に課題があったところ、ファインチューニングを行うことで改善が図られた（訓練するに当たって、著しい支障はない程度であると考える。）。
- 警察組織のネットワーク上での動作確認や訓練効果の検証等を目的として、全国警察協力のもと、計3回の検証を実施した。
検証の結果、特に、代表者聴取に関する研修歴がない者において、適切な質問の割合が増加するなど聴取技術が高まることが示され、本ツールによる訓練が聴取技術の習得に有効であることを示す結果が得られた。

② 出口戦略・研究成果の波及

- 警察組織において、AI訓練ツール改訂版を既存の教養・研修と組み合わせながら運用していくことで、聴取技術の底上げを図り、ひいては、実務における児童の心身の負担軽減・供述の信用性の確保につなげる。

③ 目標達成状況等の特記事項

- 実戦的な訓練を可能とするためには、アバターの回答精度が重要であるところ、不自然で現実と乖離した回答では、訓練効果が見込めないことはもとより、ツールの利用意欲も低下させることとなる。限られたリソース（機器・モデルの性能、データ量、期間等）の中で試行錯誤を重ね、結果として、検証参加者の多数から「また使いたい」との評価を得るに至る程度に回答精度を向上させることができた。
- 教養担当者を必要とせず、個々人において訓練を実施することを想定したツールであることから、フィードバックの正確性が低ければ利用者（特に聴取に関する知識の浅い利用者）に誤った知識を与えてしまうことが懸念された。
この点について、一定程度の精度の改善は図られたが、誤りを完全になくすことは困難であるため、ツールに組み込んだ基礎知識を解説する研修動画や具体的な正誤を示した選択式の訓練を活用しつつ、運用において注意を促していくことで対応していく。

2. 研究成果及び出口戦略、達成状況（取組全期間）

テーマ

② AI訓練ツール改訂版の実装

① 研究成果及び達成状況

- 本施策内においては、警察組織への実装を実現した。
 - ・ 聴取の訓練ツールとして運用していくに当たり、警察における利用場面を考慮したペルソナを設定して必要な機能を洗い出し、聴取の経験や知識の浅い者、スキルアップを臨む者等、利用者の状況や目的に応じてツールを活用できるように、利便性を考慮しつつシステム構築した。
 - ・ また、組織として聴取技術の底上げを図っていくためには、本ツールによる訓練状況を把握・管理していく必要があるところ、訓練に必要な機能だけでなく、ツールの利用状況や訓練結果を把握できる管理者機能も設けることで、組織としての指導・教養に資する形で実装を実現した。
- より広範な社会実装に向け、代表者聴取の関係機関である法務省（検察）、こども家庭庁（児童相談所）に対するヒアリングを実施した。
 - ツールのフィードバック機能（質問の割合等）について有用である旨の意見を得られるとともに、それぞれの業務性により即した機能面や精度面のニーズがあることを把握した。

② 出口戦略・研究成果の波及

- 警察組織において、AI訓練ツール改訂版を既存の教養・研修と組み合わせながら運用していくことで、聴取技術の底上げを図る。
- 検証では、特に初心者に対して聴取技術を高める効果があることを示す結果が得られたことから、今後、検察、児童相談所等との合同研修の場など、機会を捉えてAI訓練ツールの紹介を行い、幅広い活用につなげる。
- 実運用によって明らかになる課題もあると考えられることから、警察における運用実績を踏まえつつ、さらなる社会実装の方法を検討していくことが必要であり、開発企業においては、本施策の成果を活用し、特定の自治体をモデルケースとして試験的にツールを導入・運用していくことも検討している。

③ 目標達成状況等の特記事項

- アバターを用いた聴取訓練についての研究知見はいまだに少ないことから、社会実装については、運用実績や効果のエビデンス等を踏まえ段階的な展開を検討していくことが、施策内のリソース的にも現実的であると考えられた。また、サーバーの保管管理や情報管理の観点から、本施策では、警察庁にサーバーを設置して実装を実現した。
- GPU等の処理性能上、同時に訓練可能な人数が限られるところ、予約機能を設けることでGPUへの接続を制御し、アバターのレスポンス速度を維持しつつ支障のない訓練を提供できる最大数を確保した（同時接続20名まで可）。

2. 研究成果及び出口戦略、達成状況（取組全期間）

テーマ

③ 事案内容や難易度に応じたアバターの研究開発等

① 研究成果及び達成状況

- 多様な事案設定（シナリオ）に対応できるよう、年齢、性別、被害態様の異なる8体のアバターを開発した。
試作版において開発した2体のアバター（未就学児想定）に加え、小学生（低学年、高学年）、中学生を想定した男女のアバターを開発し、計8つのシナリオを作成した。
- 様々な事案内容に基づく訓練を可能とするため、シナリオの編集や新規作成ができる仕組みを構築した。
聴取技術の維持・向上には、継続した訓練が必要であるところ、既存シナリオの繰り返しだけではなく、新たなシナリオが追加されることによって、ツールの利用意欲にもつながることが期待される。
- アバターの回答を生成するモデル（回答選択モデル）は、オープンソースの中でも大規模で高性能なモデルに学習させて構築し、アバターが被害を開示する確率を、実際の代表者聴取事例における分析結果や海外の先行研究の知見に基づいて設定した。

② 出口戦略・研究成果の波及

- 既存の8つのシナリオの利用状況も踏まえつつ、編集や新規作成できる仕組みを活用し、開示抵抗の高い児童を想定したシナリオを作成する等、実戦的な訓練に資するツールとなるようブラッシュアップを図っていく。
- 運用を通じて蓄積する訓練データを加工する等して学習データに活用することで、ツールの各精度の向上が期待できる。
- 年齢に応じた話し言葉や発話量などは、ますます進歩していくであろうAI技術を踏まえれば、最新モデルへの変更やさらなる追加学習をすることで、よりリアルな応答を実現できる可能性があり、コミュニケーション技能を要する様々な業種への応用が期待される。

③ 目標達成状況等の特記事項

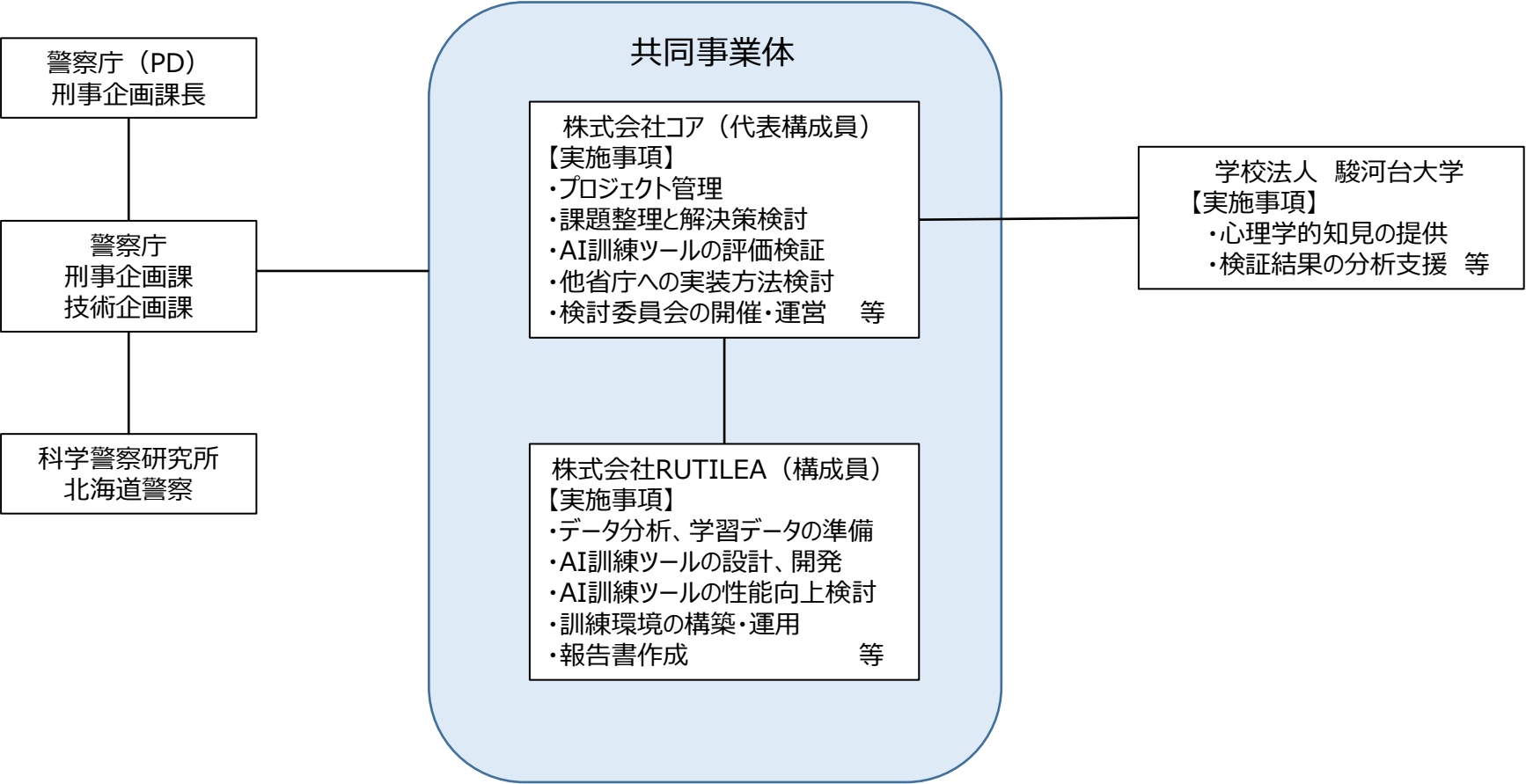
- 児童からの聴取においては、ラポール形成（話しやすい関係性の構築）の発話が重要であるところ、ラポール形成の程度を定量的に示すのは困難であった。訓練中の「定型的な質問」の履行状況からラポール形成が図られたかを判定するなどして、アバターの回答傾向に反映させる仕組みを構築したものの、アバターの回答傾向に顕著な差異は見られなかったため、この点はさらなる研究知見が求められる。

3. 到達目標（KPI）に対する実績

テーマ名	実施内容の概要と到達目標（KPI）	到達目標（KPI）に対する実績	最終年度（単年度）の実施内容と実績
① AI訓練ツール試作版の課題や基礎研究結果を踏まえた改訂版の開発と訓練・検証	開発したAI訓練ツール試作版による訓練・検証結果から明らかとなった課題から実装に向けたAI訓練ツール改訂版の開発を行い、訓練・検証を繰り返し、精度を高めていく。 （TRL 4 から 7 へ、BRL 5 から 7 へ）	- 試作版における課題等を整理し、アジャイル型の開発を行うとともに、検証を繰り返し精度を向上させ、運用に耐える改訂版を完成させた。 【①～③共通】 * 生産技術を確立（TRL 7） * 検証を経て有効性が実証され、事業モデルの確立が見込める（BRL 7）	- アバターの回答精度の向上やフィードバックの工夫等の課題を踏まえ、改訂版を開発した。 - AIの精度検証を繰り返し行い、課題となっていた精度の改善を図った。 - 全国警察の協力を得てツールによる訓練効果の検証を行い、聴取技術の習得に有効であることを示す結果が得られた。
② AI訓練ツール改訂版の実装	社会実装に向けた調査・研究等の結果を基に、令和 7 年度からの運用に向けた取組を行う。 （TRL 4 から 7 へ、BRL 5 から 7 へ）	令和 7 年度から警察において運用すべく実装方法を検討し、警察組織のネットワークを利用することで、各都道府県警察の拠点（警察署等）で訓練可能な環境を構築した。	- 警察における利用場面を考慮したペルソナを基に、運用上必要な機能を洗い出し、利用者の立場や目的に応じて、利便性のあるシステムを構築し、警察組織に実装した。 - ツールの利用状況や訓練結果を把握するための管理者機能を設け、組織的な指導・教養に資するものとした。 - 法務省やこども家庭庁等にヒアリングを行い、より広範な社会実装に向けた課題を整理した。
③ 事案内容や難易度に応じたアバターの研究・開発等	調査・研究等から得られた知見により、より実戦的な訓練を行うための事案内容や難易度（なかなか話さない等）を設定したアバターの研究・開発を行う。 （TRL 4 から 7 へ、BRL 5 から 7 へ）	実際の代表者聴取事例の分析結果や海外の先行研究の知見の一部をアバターの回答傾向に反映させるとともに、シナリオを編集・新規作成できる機能を構築することで、開示抵抗の高い児童の想定等、ニーズを考慮したシナリオの自由な作成を可能にした。	- 年齢、性別の異なる 8 体のアバター（シナリオ）を開発した。 - 様々な事案内容に基づく訓練を可能とするため、シナリオの編集や新規作成ができる仕組みを構築した。 - アバターの被害開示率について、代表者聴取事例の分析結果や海外の先行研究の知見に基づいて設定した。

4. 実施体制及び実施者の役割分担

実施体制



5. 民間研究開発投資誘発効果及びマッチングファンド

① 民間研究開発投資誘発効果（財政支出の効率化）

AI訓練ツールの開発・実装を実現したことにより、既存の教養・訓練において生じていた各種制約にとらわれることなく訓練を実施できる手段が確立され、聴取技術の底上げを図ることが期待できる。

また、AI訓練ツールを活用することで、アバターを用いた聴取訓練に関する研究が活性化することが期待される。

本施策の実施によって、アバターを活用した訓練により聴取技術が高まることを示す結果も得られており、技術的成果を活用・発展させることで、代表者聴取における分野のみならず、コミュニケーション技能の向上を求める様々な分野への展開が期待される。

② 民間からの貢献度（マッチングファンド）

合計 75,610千円 ※マッチングファンド率 29.7%（要求額：255,000千円）

1 人件費 60,000千円

- ～ ・ 自治体、民間展開に向けた調査・ヒアリング対応に係る人件費
- ・ 社会実装に係る研究開発、システム開発

2 設備費 15,610千円

- ～ ・ 検証用サーバー、施設使用料
- ・ 言語モデル学習のGPUコスト

【今後見込まれる民間側の貢献額】

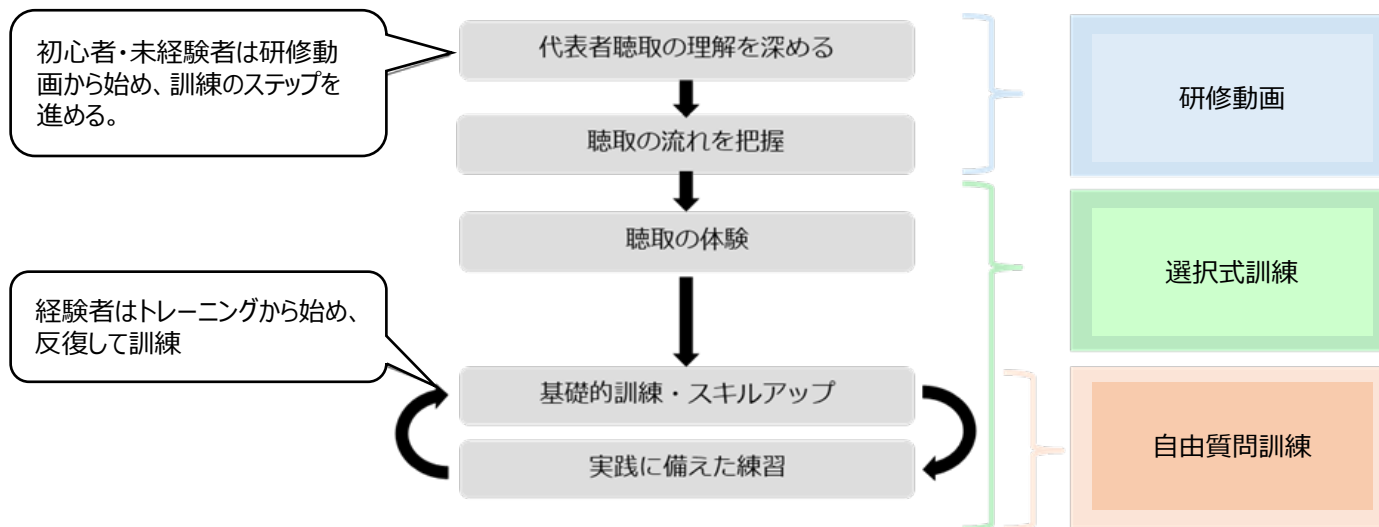
民間や医療業界への社会実装に関する研究開発予算 約10,000千円

実用性を重視した設計①

【被訓練者(ツール利用者)】

目的: 代表者聴取において必要な聴取技法(NICHDプロトコル)を習得する
→ 被訓練者のスキルレベルに応じた訓練の実施が望ましい

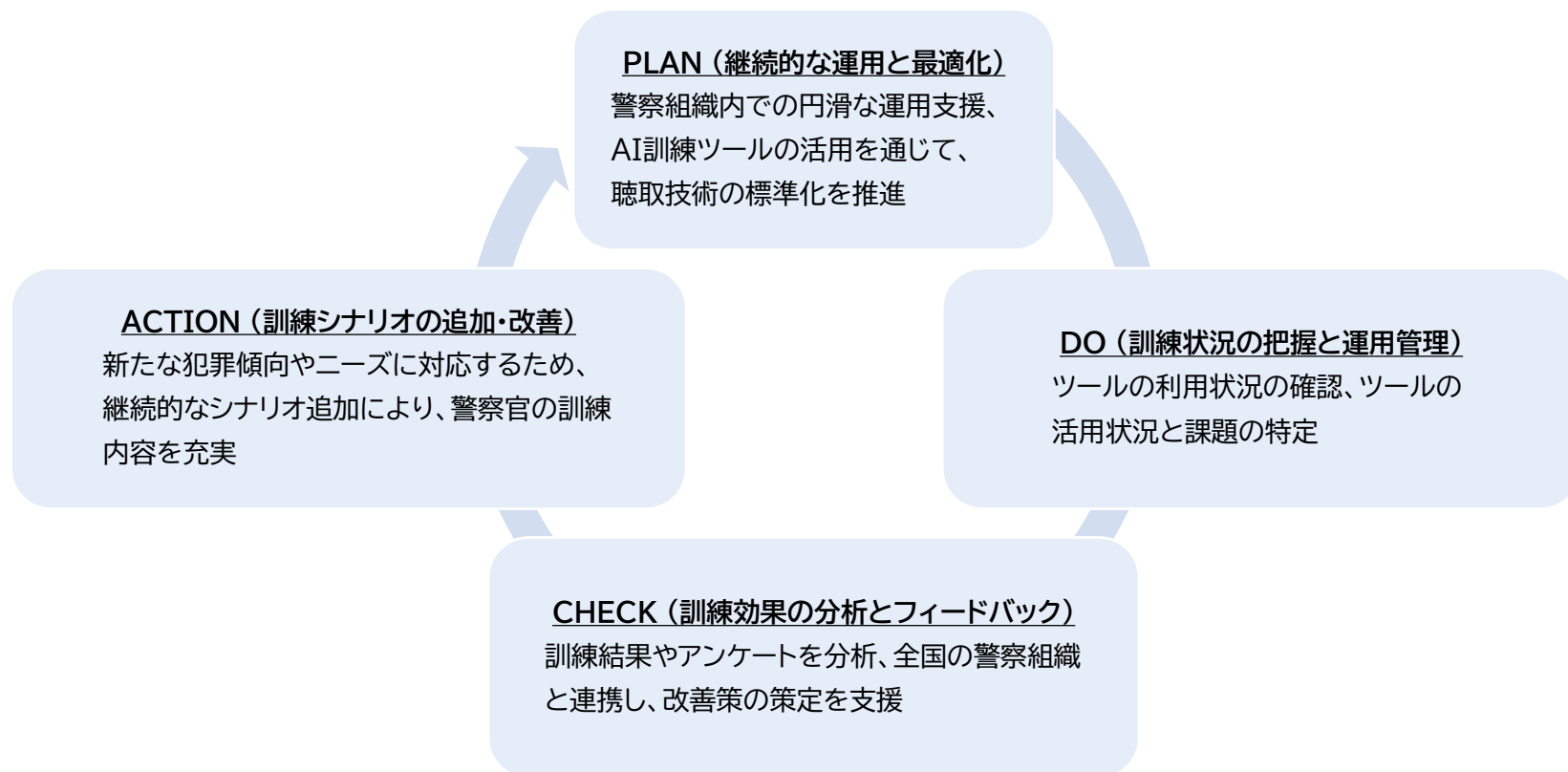
1. 研修動画視聴: 初心者や未経験者向けの訓練で、代表者聴取の理解を深め、聴取の流れを学ぶ
2. 選択式訓練: 基礎的な聴取技法を学ぶ訓練で、事前に定義されたシナリオに沿って、事前に定義された選択肢から適切な質問を選ぶ訓練をする
3. 自由質問訓練: AIアバターとの対話による実践的な訓練で、事前に定義された犯罪内容についての聴取を行う訓練をする



実用性を重視した設計②

【運用担当者(警察庁)】

目的:被訓練者の訓練状況や訓練効果をチェックしながら、継続的な運用と最適化を図ることで全国の警察官が聴取スキルを持続的に向上できる環境を確立する



研修動画

AI訓練ツールの研修資料

科学警察研究所 犯罪行動科学部
少年研究室作成資料

33分

研修動画

質問の種類

回答の自由度		誘導のリスク	
自由再生質問	焦点化質問		誘導質問
	非特定	特定	
聴取者から一切情報を出さずに聞く	聴取者から一切情報を出さずに聞く	子どもが話したことを追加的に尋ねる	子どもが話していないことを尋ねる
①誘いかけ ②それから質問 ③手がかり質問 ④時間の分割	「何」 「どうやって」 「どんな」	「誰」 「どこ」 「いつ」 「何」	①選択式質問 ②はい・いいえ質問
			「～したんだよね」 ※文脈によって、質問の種類に関係なく誘導になる

改訂版NICHDプロトコルの構造

- ・自己紹介
- ・ラポールの形成
- ・約束事(グラウンド・ルール)
- ・エピソード記憶の練習

- ・本題への移行
- ・被害、目撃内容の調査

- 導入
- 本題
- 休憩
- 追加質問
- 終結

【参考資料】

選択式訓練

- * 選択した質問が誤りだった場合
(表示される誤りの理由を参考にもう一度選択する)

◀ 戻る

選択式訓練 自己紹介 ラポール形成 グラウンド・ルールの説明と練習 エピソード記憶の練習 本題 終結

✗

これはエピソード記憶の練習で聞く質問です。ここではラポール形成で聞く質問（何をするのが好きかお話して）をするのが適切です。

もう一度選択

1. 朝起きてからここに来るまでのことについてお話しして。

2. もっとタロウ君のことを知りたいので、タロウ君が何をするのが好きか、お話しして。

3. 嫌なことがあったときいたんだけど、それについてお話しして。

セッション終了までの時間: 14:50

- * 選択した質問が正しかった場合

◀ 戻る

選択式訓練 自己紹介 ラポール形成 グラウンド・ルールの説明と練習 エピソード記憶の練習 本題 終結

○

ラポール形成で最初に尋ねる質問としてふさわしい質問です。

OK

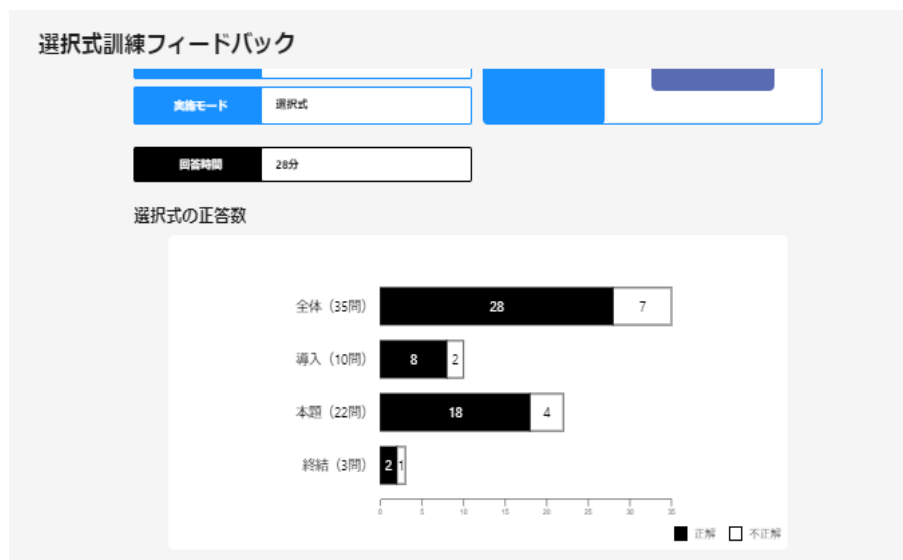
1. 朝起きてからここに来るまでのことについてお話しして。

2. もっとタロウ君のことを知りたいので、タロウ君が何をするのが好きか、お話しして。

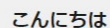
3. 嫌なことがあったときいたんだけど、それについてお話しして。

セッション終了までの時間: 14:51

- * フィードバック画面の例(一部分)



自由質問訓練

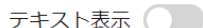


テキスト表示 ☒



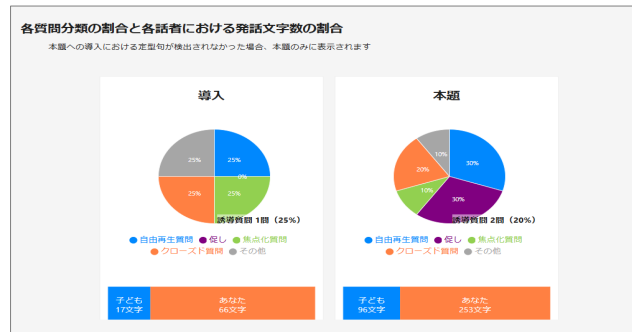
送信

* テキスト表示「オン」
マイクボタン、「送信」ボタンの操作あり



* テキスト表示「オフ」
会話の内容が表示されない。
より実戦的な訓練が可能
(ただし、マイクボタンの操作は必要)

* フィードバック画面の例(一部分)
《質問の割合と発話文字数の割合》



《獲得した事件情報》



《定型的な質問》



質問分類モデルの精度向上

多くの対話データを用いて、各質問分類の学習データ量ができる限り均一になるように準備し、学習を実施しました。

最終的に、一致度の指標を表すcohenの κ 係数において、0.918となり、開発した質問分類モデルは心理専門家の分類結果と「ほぼ完全な一致」であることが示されました。

※ 質問分類モデルによる分類結果と心理専門家による分類結果の対比

モデル	κ 係数
第二回効果検証時	0.918
第一回効果検証時	0.887
試作版	0.726

$\kappa \leq 0.2$: わずかな一致

$0.2 < \kappa \leq 0.4$: まずまずの一致

$0.4 < \kappa \leq 0.6$: 中程度の一致

$0.6 < \kappa \leq 0.8$: かなりの一致

$0.8 < \kappa$: ほぼ完全な一致

質問の分類	AIによる分類が心理専門家の分類結果と一致した割合（*1）	（AIによる分類の）心理専門家の分類結果に対する正解率（*2）
自由再生質問	96.6%	94.8%
促し	82.3%	95.3%
焦点化質問	94.5%	94.9%
クローズド質問	97.0%	95.8%
その他	91.1%	89.4%

（*1）割合は、各質問分類ごとに、[AIの分類と心理専門家の分類とが一致した数] / [AIが分類した数] で算出

（*2）正解率は、各質問分類ごとに、[AIの分類と心理専門家の分類とが一致した数] / [心理専門家が分類した数] で算出

評価検証の実施概要

全国の警察官を対象に大規模な評価検証を実施。

評価検証	実施期間	実施対象	目的	結果
AI訓練ツールの動作環境の事前確認(ネットワーク計測ツール)	2024年9月24日～ 2024年10月9日	全国都道府県警 1,242拠点	本プロジェクトにおいてAI訓練ツールの実行(動画再生やAIの動作)が問題ないかどうかを確認するために調査を実施。本プロジェクトで開発したネットワーク計測ツールを利用し調査した。	全国の拠点において、AI訓練ツールを用いた訓練の実施が問題なく可能であることを確認。
検証1回目 (警察環境での動作確認)	2024年11月11日～ 2024年11月29日	全国の警察官 計104名	AI訓練ツールが警察の共通基盤上で正常に動作することを確認	全国の警察官に対するAI訓練ツールの動作確認を実施。システム上問題なく動作することを確認。
検証2回目 (訓練効果の検証)	2025年1月20日～ 2025年2月5日	全国の警察官 計155名	全機能の評価 AI訓練ツールの訓練効果の確認	特に初学者において聴取技術が高まることが示されたほか、研修動画等の介入を含む訓練が効果的であることを確認。検証結果から、AI精度とフィードバック活用の課題が明らかとなったため、改善策を講じ再検証が決定。
検証3回目 (ツールの改善を行い再実施)	2025年2月26日～ 2025年3月5日	全国の警察官 計39名	精度向上したAI訓練ツールにより訓練が滞りなく実施できるかを確認	精度を改善したツールが問題なく動作することを確認。

アンケート結果

アンケート結果においても訓練効果や有効性を実感する意見が多く、積極的に活用したいとの声も

アンケートの自由回答欄で得られた意見においては、AI訓練ツールの役立ち度やフィードバックの効果について、特に高く評価しており、今後も使用したいと考える人が多かった。

検証3回目アンケート集計結果

■1(評価低) ■2 ■3 ■4 ■5(評価高)

