

国内における製薬企業での人工知能(AI)の導入状況と課題

—製薬企業へのアンケート結果から見てきたこと—

医薬品評価委員会 データサイエンス部会 タスクフォース1

製薬協医薬品評価委員会データサイエンス部会(以下、DS部会)では、製薬業界向けに人工知能(AI)の基本的な仕組み、特徴および現状等を紹介する報告書を作成しています。その一環として、国内製薬企業におけるAIの導入の状況、課題等を把握すべく、2018年9月6日から10月12日の期間で「AIに関するアンケート」を実施しました。本ニュースレターではその集計結果に若干の考察を加えて紹介します。

はじめに

近年、「人工知能(AI)」の利用を含めて「第4次産業革命」^[1]とさえいわれ、さまざまな分野においてAIが利用され、その効果が大きな話題、ブームとなりました。製薬企業においても、さまざまな部門、業務に新たなブレークスルーを起こさせる技術としてビッグデータ、リアルワールドデータと合わせて期待が高まっていると思います。しかし、そもそも「AIとはなにか?」という明確な定義もされておらず、どういったことが可能になるのか? どういった場合に適さないのか? といった理解よりも漠然としたイメージが先行しているのではないのでしょうか。製薬業界でも創薬分野では数年前から利用されている場合がありますが、それ以外にもたとえば臨床開発、コールセンターの分野にもAIを利用するといった記事が掲載され、経営層等からAIのビジネスへの利用の指示がされるようなことも生じているかもしれません。

しかし、現状のAIは、どんな問題にも答えられ、あるいは人のように考えて判断できるのではなく、ほかのシステムのようにベンダーに発注すれば、仕組みはわからなくても多くの場合にほしいものが導入できるものでもありません。そもそも、あらかじめ仕様書や設計書を作成して、その通りにプログラミング、設定する方法では開発できません。また、背景知識の不足や専門用語等の壁があり、製薬企業の担当者(ユーザー)とAIベンダーの担当者とのコミュニケーションに困難が生じることもあります。こういった課題に対して、社内にある程度のAIに関する知識、スキルをもつ担当者を確保することは1つの解法となると考えられます。このような背景から、製薬企業の日本国内におけるAI利用への取り組み、人材を含む課題について全社的な状況、ならびに回答必須部門として臨床開発部門(生物統計部門を含む)、GVP・GPSP部門、任意として創薬部門(毒性評価を含む)、コールセンター(社外用、社内用を含む)、営業・マーケティング部門、生産・品質保証部門、そのほか についてのアンケートを実施しました。部門については、導入済あるいは導入を検討しているAIについても聞いています。なお、対象とするAIは種類、社外のサービス利用/自社開発は問いません。OCR^[2]、RPA^[2]、汎用的な翻訳ソフトといったごく一般的なものには除くこととしました。

アンケートは設問の順番等そのフォームによって結果が影響を受けるため、アンケート全体を実際の設問を含めて紹介すべきですが、ここでは製薬企業と関連するほかの分野のみなさんに関係が深いと思われる設問に絞ってご紹介します。

より詳細な結果は別途、報告書として製薬協のウェブサイトに掲載いたします^[3]。

[1] 第4次産業革命とは、IoTおよびビッグデータをネットワークでつなげてまとめ、これを解析・利用することで新たな付加価値が生まれ、AIによりコンピューター自らが学習し一定の判断を行うことが可能となるほか、従来のロボット技術もさらに複雑な作業が可能となる等の技術革新を指す。(内閣府 日本経済2016-2017 第2章 第1節 第4次産業革命のインパクト http://www5.cao.go.jp/keizai3/2016/01117nk/n16_2_1.html を要約)

[2] OCR : Optical Character Recognition/Reader (光学的文字認識)、RPA : Robotic Process Automation (ロボットによる業務自動化)

[3] 製薬協 委員会からの情報発信 医薬品評価委員会の成果物
<http://www.jpma.or.jp/medicine/shinyaku/tiken/tiken/expert/results.html>

アンケート結果

アンケートは医薬品評価委員会DS部会委員を通じて65社に依頼し、34社(うち外資系6社)、108部門(図1)から回答を得られました。全社の状況に関する質問への回答は、可能な限り情報システムの全社的な導入を推進する部門にお願いし、それ以外は各部門の方に回答いただくよう依頼しましたが、全社の状況の12/34は臨床開発部門所属者からの回答でした(図2)。このためか、AIの導入状況は全社部門を問わず導入済と回答したのは10社、部門ごとの導入済は11社、全社の「各部門のAIの導入支援をする全社組織がある」は8社ですが、「あり」としなかったうち7社10部門は「部門では取り組まず、AIの導入支援をする全社組織が取り組んでいる」に「あり」と回答している等、全社の状況が必ずしも把握されていないと思われました。このように、矛盾する場合があります、回答率も高くはありませんが、取り組んでいる状況、件数や具体的な課題を把握する一助となると考えます。

図1 部門区分毎回答数

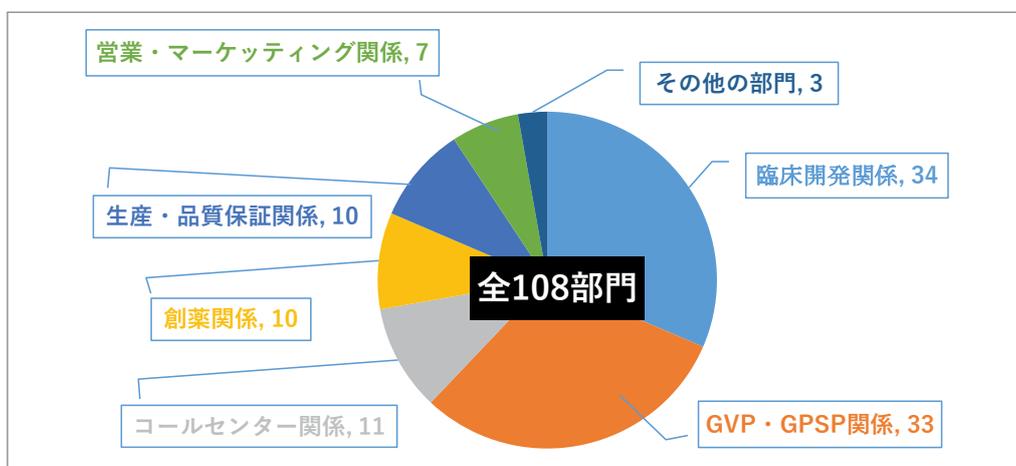
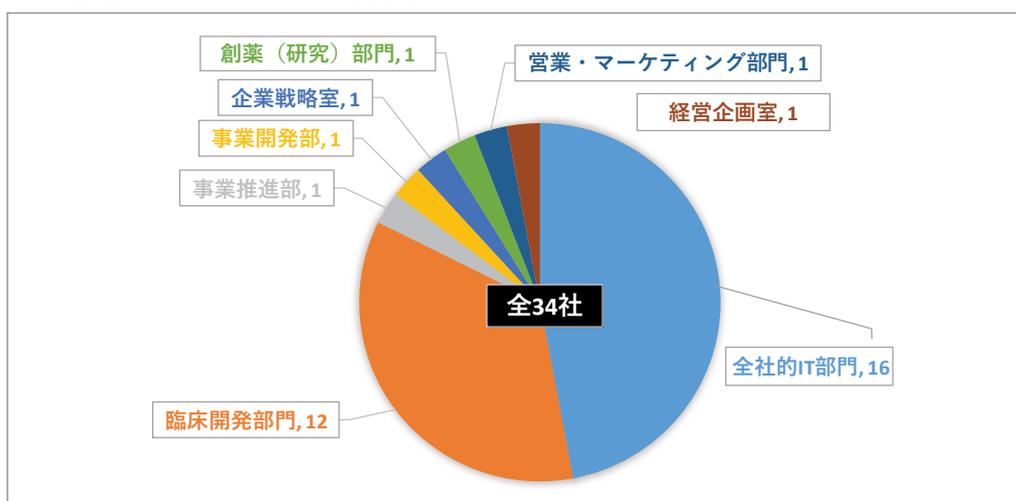


図2 全社的事項についての設問回答者の所属

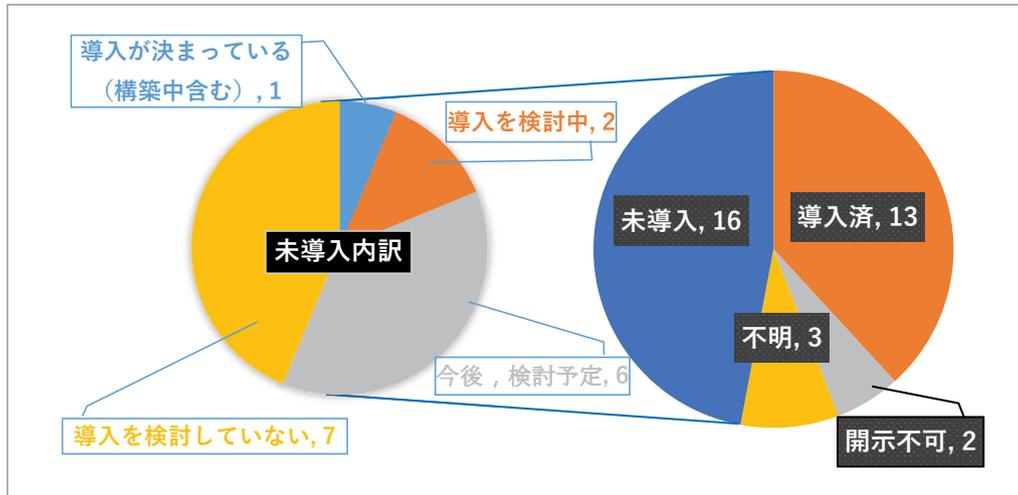


AIの導入状況について

全社、特定の部門を問わないAIの導入状況の設問の結果は図3の通りで、導入済は38%ですが導入を検討中まで含めれば47%(16社)に達します。反対に、明確に導入を検討していないと回答したのは7社に過ぎませんでした。その7社中3社は人材面の課題について「社内にAIについて企画立案できる人材がない」と答えていました。同様に部門ごとのアンケートでは、全108部門中「導入を検討していない」と回答した43部門のうち、課題として「社内にAIについて企画立案できる人材がない」が12部門、「社内育成したいが適当な人材がない」が5部門、「通常の情報システムと異なり、目的が達成できるか

わからない」が11部門、「必要なデータが社内外にあるが、充足しているかわからない」「競合他社のデータが必要(社内データでは不足、偏りがあるなど)」がおのおの4部門等、21部門は人材、データまたは情報に課題があると回答しており、まったく検討していないのではなく、具体的に検討するリソースや情報の不足がうかがえました。

図3 全社、部門を問わずAIの導入状況(会社数)



導入されているまたは検討されているAIについて

各部門では、導入済または予定しているAIについても調査しました。導入状況が導入済、導入が決まっている、導入を検討中または今後検討予定と答えたのは計56部門あり、分野ごとの件数および導入方法は、臨床開発関係が21部門(自社開発3部門、ベンダーに依頼5部門、既存サービスの利用4部門[4])、GVP・GPSP関係が14部門(4部門、5部門、1部門)、創薬関係5部門(4部門、1部門、1部門)、コールセンター関係5部門(0部門、2部門、1部門)で、他部門を含む全部門では34部門が社内で開発またはベンダーに依頼しての開発であり、独自のものが多くありました。

表1は導入済～今後検討予定56部門のAIの分野の集計結果です。自然言語処理が50%、最適化や推論が43%、自動制御が27%でした。自動制御と答えた部門には偏りがあり、GVP・GPSP関係の5部門(36%)、生産・品質保証関係の3部門(60%)は部門内では最多でしたが、創薬関係(3部門、50%)、生産・品質保証関係(3部門、60%)はいずれも3番目でした。

表1 各部門の導入済～検討中のAIの分野(部門数)

導入済または予定するAIについて—AIの分野は何ですか	
①画像解析(分類, 診断)	9
②音声解析	5
③自然言語処理	28
④自動制御(生産や培養設備などの自動運転や, AIを用いたデータ処理を用いるRPAなど)	15
⑤最適化や推論	24
⑥異常検知	10
⑦汎用的な機械学習の環境	12
⑧その他	4
⑨開示不可	3

[4] 導入方法は8つの選択肢について複数選択可としており、多かつた3つを記載しています。

表2は設問「導入済または予定するAIの機能・種類」の回答です。この設問は表3の選択肢または自由記載によるもので、47件中15件はリストからの選択でした。回答が臨床開発、GVP・GPSP分野が中心であったこともあり、臨床試験の効率化(逸脱監視、ドキュメントの自動作成)、データ処理(入力、コーディング)、ドキュメントの翻訳、テキスト文書のマイニングに関するものが多く見受けられました。これらはすでに海外または他分野で類似のサービスがあり、追加学習(用いる分野のデータで追加の学習等を行う)が可能と思われるものが少なくなく、比較的容易な分野からAIの導入が進められていることがうかがわれました。また、矛盾があるため集計から除外していますが、「今後、検討予定」の回答に治験の逸脱・有害事象検知、治験の各種書類作成の自動化、治験データから主要帳票作成の自動化、申請資料の品質管理、各施設の医療情報システムからElectronic Data Captureシステム(EDC)への出力に関するもの、データマイニング手法による副作用情報検出、消費者の声の分析がありました。

表2 導入済または導入を予定しているAIの機能

<p>臨床開発関係</p> <ul style="list-style-type: none"> ●●から学習して主要エンドポイントの標準臨床経過モデルを作成する。 ●●のデータから治験リクルート対象者を選択する。 ●●画像診断の自動化 翻訳 データを元に文章を作成する 治験実施施設ごとの各種書類を自動作成する。 議事録作成 プロトコル等、治験関連書類の自動作成 自動翻訳 プロトコル骨子(開発戦略)作成支援 予測モデルを生成する 関連因子を探索する 定型業務を自動化する AI機能を使った自動翻訳システム メディカルコーディング(MedDRA, WHO-DD) リスクアセスメント(RBMの一端) 申請資料(eCTD)の翻訳 ●●から学習して主要エンドポイントの標準臨床経過モデルを作成する。 治験データ(ADaM)から主要帳票作成の自動化 <p>創薬関係</p> <ul style="list-style-type: none"> ●●画像診断の自動化 ●●のデータから治験リクルート対象者を選択する。 被験者の日誌, ePROのマイニング 画像による細胞の品質管理 活性やADMET予測による化合物構造・プロファイルの最適化支援 文献情報解析による、ターゲット探索、ターゲットバリデーション支援 各種開発・申請関連資料の翻訳 	<p>GVP・GPSP関係</p> <ul style="list-style-type: none"> 自社・他社の薬剤の使われ方を分析 ●●から学習して主要エンドポイントの標準臨床経過モデルを作成する。 被験者個々のデータのモニタリングを行い、逸脱、有害事象の発生などを予測し、通知する。 ●●画像診断の自動化 安全性情報のオートコーディング 安全性情報の入力支援 ●●画像診断の自動化 ●●から学習して主要エンドポイントの標準臨床経過モデルを作成する。 MR日報から未報告の有害事象情報の抽出 有害事象症例処理の一部に利用 <p>コールセンター関係</p> <ul style="list-style-type: none"> 音声認識システム(Amivoice等) 学習AI(WATSON AI-Q)によるFAQの提示 チャットボット データマイニング手法による副作用情報検出 Voice of Customerの抽出 <p>生産・品質保証関係</p> <ul style="list-style-type: none"> 生産量の最適化 設備稼働データのモニタリングを行い、逸脱、有害事象の発生などを予測し、通知する。 既にあるデータベースを参考にして事象に関連する対応策を予測提案する。 <p>営業・マーケティング関係</p> <ul style="list-style-type: none"> Medical Representativeの日々の行動計画の最適化・効率化 <p>その他の部門</p> <ul style="list-style-type: none"> 営業部門からの問い合わせ対応 講演会スライドのレビュー
---	--

表3 部門ごとアンケート4.2.の選択肢

<ul style="list-style-type: none"> ●●から学習して主要エンドポイントの標準臨床経過モデルを作成する。 ●●のデータから治験リクルート対象者を選択する。 治験実施施設ごとの各種書類を自動作成する。 各施設の医療情報システムからEDC(CDASH)に出力するデータ標準化/クリーニングプロトコルの自動作成 CDASHからSDTM, SDTMからADaM作成プログラムの自動化 治験データ(ADaM)から主要帳票作成の自動化 被験者個々のデータのモニタリングを行い、逸脱、有害事象の発生などを予測し、通知する。 ●●画像診断の自動化 被験者の日誌, ePROのマイニング 被験者の活動量計の解析 申請資料(eCTD)の品質管理(表記の統一, データチェックなど) 申請資料(eCTD)の翻訳
--

導入・開発体制について

導入済～今後検討予定の56部門の導入構築体制は表4の通りで、最も多い「ベンダーに依頼して開発」は36%、次いで「社内で開発、運用」は23%でしたが、カスタマイズを含む既存サービスの利用も23%でした。導入済の17部門に限定すると「ベンダーに依頼して開発」、「社内で開発、運用」がそれぞれ41%(7部門)、カスタマイズを含む既存サービスの利用は9部門(53%)ですが、導入が決まっている4部門はすべて「ベンダーに依頼して開発」または「社内で開発、運用」であることから、既存サービスを利用することで早期に導入したことがうかがわれます。

表4 導入済～今後検討予定の部門における導入方法(複数回答可)

導入方法 (AIを用いた仕組みを構築するなど) について	
①社内で開発, 運用 (社外のコンサルタント利用含む)	13
②ベンダーに依頼して開発	20
③既存サービスをカスタマイズして導入	6
④既存サービスの利用 (ASPなど)	7
⑤学習済みモデルを利用 (カスタマイズ, 転移学習を含む)	5
⑥その他	3
⑦わからない (把握していない, 担当でない)	4
⑧開示不可	0
未回答	13

課題について

人材、データ、AIに関する情報に分けて設問を設定しました。課題は導入状況によって異なっており、導入状況が導入済～今後検討予定の56部門の主な課題(20%以上)を表5にまとめました。導入済はデータに関する課題が多く、ほかにはAIでできないか、目的を達成できるかわからないといった課題が多いようです。AIの開発、導入に関して、社会全般で人材不足が指摘されています。3種類の課題のうち、人材面での課題の設問を導入状況ごとに集計した結果が表6です。この表から、導入の進んでいる部門は人材が確保されている場合もありそうですが、全体としては人材が不足しているようです。なお、「既に人材を確保できており、課題はない」と答えた6部門のうち創業関係は2部門で、それ以外はGVP・GPSP関係、生産・品質保証関係等が1部門ずつでした。

表5 導入済～今後検討予定の部門の主な課題 [5]

	導入済	導入が決まっている (構築中含む) + 導入を検討中	今後, 検討予定	総計
部門数	17	21	18	56
社内育成したいが適当な人材がない	2	6	7	15
社内にAIについて企画立案できる人材がない	6	7	9	22
必要なデータが社外にあるが, 利用できない (法制度などの制約)	8	4	1	13
必要なデータが社内外にあるが, 散在している, 品質が低いなど, 利用が困難	7	8	4	19
必要なデータが社内外にあるが, 充足しているかわからない	4	9	12	25
競合他社のデータが必要 (社内データでは不足, 偏りがあるなど)	8	3	5	16
AIで何ができるのかよくわからない	6	8	9	23
どのベンダーに相談してよいかわからない	3	8	5	16
通常の情報システムと異なり, 目的が達成できるかわからない	6	14	12	32

[5] 表中の集計値は「部門数」、バーグラフは導入状況の部門数に占める各課題の割合。表6も同じ

表6 導入状況別の人材に関する課題

	導入済	導入が決まっている (構築中含む) + 導入を検討中	今後, 検討予定	総計
部門数	17	21	18	56
①既に人材を確保できており, 課題はない	4	2	0	6
②全て社外に委託するので, 課題はない	0	4	0	4
③大学での教育が不十分 (学部, 学科が少ない, カリキュラムが不十分) で新卒採用が難しい.	1	0	0	1
④AIの知識を有した人材は世の中にあるものの, 製薬企業で使うスキルは特殊なので, 新規雇用は難しい	3	4	2	9
⑤進歩が速すぎて育成が間に合わない	3	2	2	7
⑥社内育成したいが適当な人材がない	2	6	7	15
⑦教育機関と連携して育成 (社員の派遣, 留学) したが, 教育機関の環境が不十分	1	0	1	2
⑧社内にAIについて企画立案できる人材がない	6	7	9	22
⑨導入を考えていない	1	0	0	1
⑩わからない (把握していない, 担当でない)	2	5	5	12

クラウドセキュリティおよびAIの信頼性について

AIはクラウド環境を用いることが多いため、クラウドのセキュリティを含め3つの設問を設定しました。全社に関する設問の結果は図4の通りで、いずれも「わからない」が10社以上あり、AIの結果の信頼性について不安はないと答えたのは3社に過ぎませんでした。製薬企業は医薬品という生命にかかわる製品と情報を扱い、また、薬事法令の規制を受けており、信頼性に対する考え方はほかの分野より厳しいことが影響しているのかもしれない。AIの利用において信頼性だけでなく、著作権等も通常のシステム等とは異なる部分があります。経済産業省から公開されているAI開発・利用に関する権利関係・責任関係等の考え方を含めた「AI・データの利用に関する契約ガイドライン」^[6]は信頼性の考え方も含めて参考になるでしょう。

図4 全社的事項の設問中のセキュリティ、信頼性に関する回答集計(会社数)

クラウドのAI環境は情報セキュリティの面で不安はありますか	会社数
不安がある	7
不安があるが容認できる	4
不安はない (どちらかと言えない)	12
わからない	10
未回答	1

AIの結果の信頼性に不安はありますか	会社数
不安がある	6
不安があるが容認できる	13
不安はない (どちらかと言えない)	4
わからない	10
未回答	1

AIの結果が誤っていた場合の責任は主にどちらにあると考えますか?	会社数
利用者	9
運用者	6
開発・構築者	4
学習データ提供者	0
わからない	15
未回答	0

^[6] 経済産業省 ニュースリリース「AI・データの利用に関する契約ガイドライン」を策定しました (2018年6月15日)
<http://www.meti.go.jp/press/2018/06/20180615001/20180615001.html>

AIにかかわるみなさんへ

AIにかかわるデータサイエンティストやベンダーにとって、製薬業界はよくわからない、特殊な業界のように思われるかもしれません。創薬分野はAIの対象と認識されて技術開発、導入は進みつつありますが、それ以外の分野ではほとんど進んでいません。その背景に、人を対象とした試験や市販後の安全性情報等の臨床データは複雑である、データ数(人数)は数百から数千とそれほど多くない、結果の信頼性等の要求が高い、という特性が原因かもしれません。しかし、一般の診療用に開発されたものに追加学習や転移学習を行う等、応用可能な場合は少なくないでしょう。加えて、膨大なドキュメントを複数言語で扱いますので、文書校正、文書作成の自動化や翻訳等のニーズも少なくありません。また、AI利用の成果は、良い薬を早く医療現場に届けることにつながると期待されます [7]。

終わりに

AIはなんでもできるのではなく、また通常のコンピューターシステムと異なり、入力に対する出力を仕様・設計により決めることができないだけでなく、なぜそのような出力が得られたのかの説明も容易でない場合がありますし、入力する情報の変化に合わせて継続的な更新が必要といった、信頼性が求められる分野での利用には厄介な特性があります。しかし、入力に対する出力のブレや揺らぎは人間でも生じますが、AIには疲れな、飽きない、面倒がらないといった人間にはないメリットもあります。特性を理解しうまく用いれば、既存業務の効率化や改善にとどまらず、これまでできなかったこと、業務プロセスとして省略していたことも実現できる場合があるでしょう。

また、AIを構成する手法(データの前処理、学習方法等)は特定の目的ごとにあるのではなく、手法と学習させるデータによって変わります。ベンダーはAIの技術はあってもデータがなければAIサービスを開発することはできませんので、データの入手が容易な分野や大量のデータの保有者と共同で開発しています。このため、特殊なデータや用途が多くなる製薬業の分野では独自にAIの開発が必要となる場合が多くなります。しかし、学習で作られたモデル(データを処理するアルゴリズム)も異なるデータで追加の学習をさせる等により、目的にあったものに改良(転移学習と呼ばれます)することもできる場合がありますので、一から開発するのではなく、異分野を含めて広く情報を収集して流用可能なものを見つけることで、開発期間を短縮できるだけでなく、少ない学習データで実装したり、開発の成功率も高めることができる場合があります。

そのためには、AIの導入者側はAIがなにをできるのか、そのために必要な学習データはなにか等がある程度理解している必要があります。一般財団法人日本ディープラーニング協会では、このような事業応用する人材向けの資格検定(ジェネラリスト検定 [8])が行われているほどです。DS部会タスクフォース6では、AIに関する基本的な理解と、ビジネスで導入する際の手がかりに役立てていただけるような報告書の作成を進めています。このアンケート結果と合わせて目的中心のAI導入の一助となれば幸いです。

[7] 厚生労働省 保健医療分野におけるAI活用推進懇談会 報告書(平成29年6月27日)
<https://www.mhlw.go.jp/stf/shingi2/0000169233.html>

[8] <https://www.jdla.org/business/certificate/#education>