



AI人材育成のための企業間 データ提供促進検討会

第2回検討会資料

2020年9月17日

本日のアジェンダ

前回検討会の振り返り

第1回検討会以降の検討状況のご報告

ガイドラインの位置づけの再確認

ガイドラインに関する討議

今後の進め方

AI人材育成のための企業間データ提供促進検討会体制

座長

渡部 俊也 東京大学 未来ビジョン研究センター 教授

委員

(五十音順)

内田 誠 iCraft法律事務所 弁護士
福岡 真之介 西村あさひ法律事務所 パートナー 弁護士
松下 外 渥美坂井法律事務所・外国法共同事業 弁護士
渡辺 知晴 渡辺総合知的財産事務所 弁理士

オブザーバー

個人情報保護委員会事務局
文部科学省 高等教育局 専門教育課
数理・データサイエンス教育強化拠点コンソーシアム カリキュラム分科会、教育用データベース分科会
経済産業省 経済産業政策局 知的財産政策室
特許庁 総務部 企画調査課
特許庁 オープンイノベーション推進プロジェクトチーム

事務局

齊藤 友紀 法律事務所 LAB-01 弁護士
経済産業省 商務情報政策局 情報経済課
経済産業省 商務情報政策局 情報技術利用促進課
中小企業庁 経営支援部 技術・経営革新課(イノベーション課)
マッキンゼー・アンド・カンパニー・インコーポレイテッド・ジャパン

第1回検討会の振り返り

合意事項

あ ガイドラインの 対象の明確化

- 本ガイドラインの対象はデータ提供者(例: 企業)と教材作成者(企業、大学)
- データ提供者に自治体等公的機関を含める社会的意義は大きい(現在は提供ルールが不明確なため対応にばらつきが存在)ものの、含めると議論は複雑化...含めるか否かの判断が必要

い 出口(データ・ 教材利用のスキーム)の設計

- データ提供にあたっては教材作成者の属性や信用度が大きく影響(相手が変わると法律も変わる)。ひいてはインセンティブ設計にも関わる...出口の設計が必要
 - 教材作成者の属性(例: 企業、大学)
 - 教材利用者の属性(例: 学生、社会人)・ニーズ
 - 教材利用方法の制限
 - 形態(例: ハッカソン、コンソーシアム、データベース)

う 個別ユース ケースに織り 込む論点

- 外観検査と需要予測で論点を分けるにはユースケースの粒度が不十分。個人情報に関わる場合と関わらない場合に大別し、その先で個別ユースケースに沿って議論してはどうか
- (外観検査は教材にはしづらい...「画像認識」という表現の方が適切ではないか)

第2回検討会に向けた方向性

- 公的機関か否かが特に重要になるのは個人情報の取り扱いであり、「個人情報を含む場合」において、公的機関の場合を触れる

- 教材作成者の追加ヒアリングによる課題・ニーズの深掘り
- 本検討で前提とするデータ提供者と教材利用者(企業・大学)の関係性の整理と、1対1の基本形以外でのインセンティブスキームのベストプラクティスを類型化
- 海外事業者への提供は検討範囲外である旨を明記

- 個別ユースケース毎に論点と解の方向性を述べるのではなく、個人情報の有無で論点と解の方向性を提示

第1回検討会の振り返り詳細(1/2)

ご意見

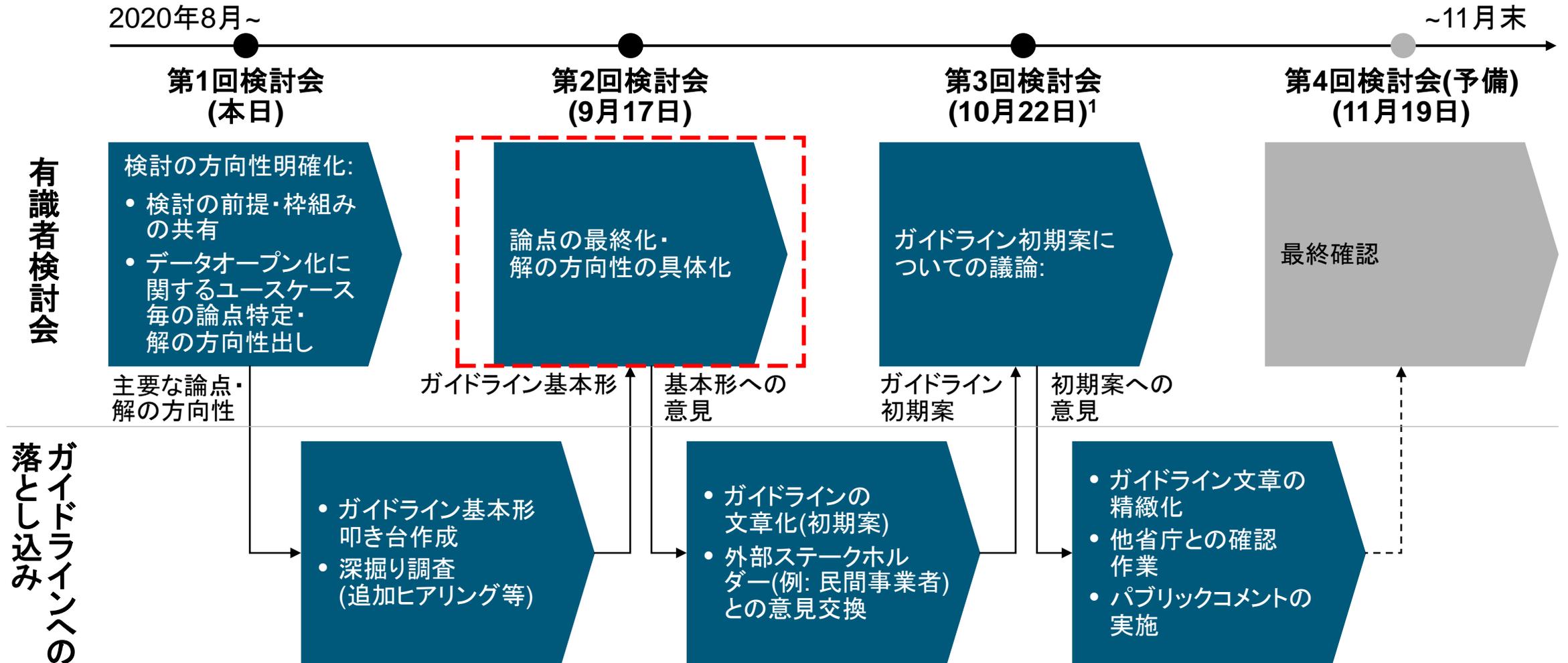
あ ガイドラインの対象の明確化	ガイドラインの対象	<ul style="list-style-type: none">教材作成者が企業なのか、大学なのか。教材を作った後のユーザーが学生なのか、その先の一般企業に属するプログラマーなのか。これによってだいぶ立て付けは変わると思う個人情報絡む場合は受け取り側が安全管理してくれるという信用がないと出せない。受け取り側がさほど大きな組織でない場合は、情報の管理体制が重要ではないか
	データ提供者の範囲	<ul style="list-style-type: none">データ提供者に公的機関は含まれるか。スマートシティ等を考えると、公的機関の持っているデータも必要ではないか地方自治体等は、情報提供に対する方針がばらばら。ルールが決まっていないと、基本的には何もしないという方向になりがち
い 出口(データ・教材利用のスキーム)の設計		<ul style="list-style-type: none">コンソーシアムのようなものを主催して、企業はそこで自分のデータを使って参加者にモデルの構築等、自由に議論していただく。ただし、そこで出た成果物に関しては、その企業が自由に使えるようにする。こうすると、データ提供者側もただで新しい学習済みモデルの開発をしてもらえることになり、データ提供をする意味が出てくるハッカソンで優秀な人が作ったモデルの著作権をもらうのか、使用权をライセンスできるようにするのは、議論すべき

第1回検討会の振り返り詳細 (2/2)

ご意見

う 個別ユースケースに織り込む論点	ユースケースの特定 <ul style="list-style-type: none">外観検査は教材として使うのは難しいのでは。需要予測やデータマーケティングは答え合わせしやすいが、外観検査は個別のノウハウが重要なため、あまり教材には向かない外観検査というよりも画像認識では。外観検査というと、カメラの設置等メカニカルな要素が強く、そこまでは教材として使えないと思うので
	データ開示の原則 <ul style="list-style-type: none">ノウハウはどう守るかという問題設定が、果たして適切なものか。守りたいなら開示しないという方法しかないのではデータ提供をする場合、教材に利用制限をかけても流出するリスクは残るため、データ提供者から見て、ノウハウや営業秘密が含まれない生データだけを開示してもらうという立て付けが適当ではないか
	不正競争防止法(ノウハウ・営業秘密の保護) <ul style="list-style-type: none">ノウハウが入っているのであれば事業者側がそれをきちんと加工する義務を契約上、負わせるようにした方がいいのでは(データ提供企業側の)何となく嫌だという心理が厳しくなってるから、限定提供データで出せるものをどのように判断するかという点も重要
	ビジネスモデル特許による保護 <ul style="list-style-type: none">特許でデータを守るのは相当難しい。データ構造は特許の対象になるが、純粋なデータの羅列では特許は取れないビジネスモデル特許でビジネスを守り、その特許で守られているビジネスモデル特許でしか使えないデータであれば、事実上そのデータも特許で守られていることになる
	個人情報の保護 <ul style="list-style-type: none">データマーケティングの中でも、POSデータ等個人情報が含まれないパターンもあるマスキングしただけでは匿名加工情報に必ずなる訳ではないので注意が必要。匿名加工情報にならない場合は当然、個人情報として扱わなければならない最大3段階にわたる第三者提供の同意を誰がどのように同意を取るのかを検討しないと個人情報保護法の問題はクリアできない。1つの方法はどこかのタイミングで委託スキームを使い、匿名加工情報の状態で提供していく方法。また、現行法であれば、オプトアウトの方式で出す方法もある

AI人材育成のための企業間データ提供促進検討会の全体タイムライン



1. 第3回以降のスケジュールについては別途ご相談させていただきたい(後述)

第2回検討会で達成したい事項



「討議の前提」の再確認

第1回検討会で不明確だった点を明確化し、討議の土台を固める

- 教材作成者側ヒアリングを実施し、課題・ニーズを整理
- データ提供者としての公的機関の扱い方
- インセンティブスキームを類型化



ガイドラインの目的・位置づけの再確認

本ガイドラインの目的や対象に関して、再度確認

- データ提供の目的の広がり
- 上記の結果として、「ガイドラインの名称」への反映



ガイドラインに織り込む論点の最終化と解の方向性出し

第1回検討会を基に作成したガイドラインたたき台について討議

- 特に第4章に織り込むべき論点の全体像を確認
- 論点毎の初期的な解の方向性の修正・明確化(特に「インセンティブスキームの類型」、「契約の雛形」)

本日のアジェンダ

前回検討会の振り返り

第1回検討会以降の検討状況のご報告

ガイドラインの位置づけの再確認

ガイドラインに関する討議

今後の進め方

再掲)第1回検討会の振り返り

次頁以降詳細

「ガイドラインに関する討議」にてご説明

合意事項

ガイドラインの対象の明確化(特にデータ提供者の範囲)

- 本ガイドラインの対象はデータ提供者(例: 企業)と教材作成者(企業、大学)
- データ提供者に自治体等公的機関を含める社会的意義は大きい(現在は提供ルールが不明確なため対応にばらつきが存在)ものの、含めると議論は複雑化...含めるか否かの判断が必要

出口(データ・教材利用のスキーム)の設計

- データ提供にあたっては教材作成者の属性や信用度が大きく影響(相手が変わると法律も変わる)。ひいてはインセンティブ設計にも関わる...出口の設計が必要
 - 教材作成者の属性(例: 企業、大学)
 - 教材利用者の属性(例: 学生、社会人)・ニーズ
 - 教材利用方法の制限
 - 形態(例: ハッカソン、コンソーシアム、データベース)

個別ユースケースに織り込む論点

- 外観検査と需要予測で論点を分けるにはユースケースの粒度が不十分。個人情報に関わる場合と関わらない場合に大別し、その先で個別ユースケースに沿って議論してはどうか
- (外観検査は教材にはしづらい...「画像認識」という表現の方が適切ではないか)

第2回検討会に向けた方向性

- 公的機関か否かが特に重要になるのは個人情報の取り扱いであり、「個人情報を含む場合」において、公的機関の場合に触れる

A 教材作成者の追加ヒアリングによる課題・ニーズの深掘り

B 本検討で前提とするデータ提供者と教材利用者(企業・大学)の関係性の整理と、1対1の基本形以外でのインセンティブスキームのベストプラクティスを類型化

- 海外事業者への提供は検討範囲外である旨を明記

- 個別ユースケース毎に論点と解の方向性を述べるのではなく、個人情報の有無で論点と解の方向性を提示(詳細は別添)

A 教材作成側ヒアリングの目的とヒアリング先

ヒアリングの目的

現在どのように教材を作成・利用しているか。
その中でどのような課題・ニーズがあるか

- どのような教材を作成・使用しているか (対象、目的、公開範囲)、
- 教材作成にどのようなデータ(例: 提供元、種類、加工度合い)を使っているか、どのような条件(提供側へのベネフィットを含む)でデータを提供してもらっているか
- データを提供してもらうにあたり、どのような課題があるか

教材作成側ヒアリング先一覧

大 学

- 数理・データサイエンス領域 教授4名

民間事業者

- 株式会社グリッド
- 株式会社SIGNATE
- Data Robot Japan株式会社

計3社

A 教材作成者への追加ヒアリングによる「教材」へのニーズの確認(大学)

教材の目的・対象は多岐に渡り、利用されるデータの種類も多様

“” 大学で作成する教材は文理問わずあらゆる学生を対象とするものから、データサイエンス選考の学生のものまで様々。テーマも、教授の専門領域によって自然系から産業系まで幅広い

“” データ付き教材を用意するために、民間企業からデータを購入したり、企業の古くなった型落ちデータを活用するなどもしている

意欲ある学生へのアプローチや、より専門性の高い教育では実データの重要性が増す

“” 大学1、2年生向けのリテラシーレベルの教育では、フェイクデータやネット上の既存公開データでも代替は可能。但しフェイクデータを作る場合にもベースとなる実データは必要

“” 実データの必要性は学問分野によっても異なるものの、かなり上級レベルの学生になると、実データから出た結果の考察等の重要性は増す

“” ビジネスへの応用や実務レベルのスキル育成を目的に据えたと、必要となるデータはほぼオープンには存在しない...企業からの実データ提供が増えれば教材の幅は広がる

大学で作成される教材の目的や対象、基となるデータは多岐に渡るものの、企業が提供した実データに対するニーズは存在

A 教材作成者への追加ヒアリングによるニーズと機会

スキームを含めたインセンティブの明確化の必要性

“ ” データ提供企業へ一般的に示せるインセンティブはCSRや採用等、非常に限定的。複数企業との連携等を通じた完全非公開のデータ利用スキームが充実するとよいのではないか

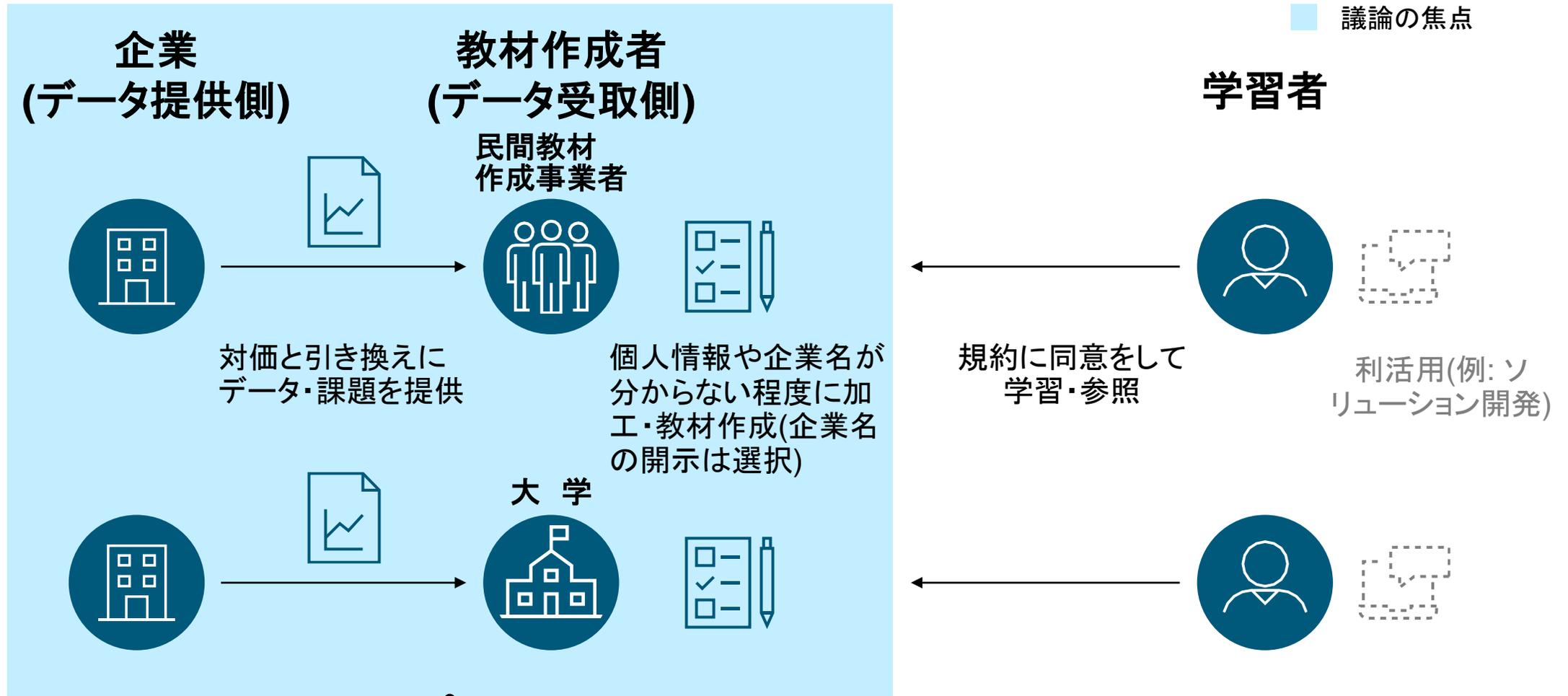
“ ” 共同研究等かなり踏み込んだ関係になれば、提供されたデータを使用した分析の成果としての還元も可能になってくる

1対多のスキームを通じた工数の削減

“ ” 個別に企業とデータ提供のための守秘義務契約を締結するのは非常に工数がかかる。ある程度同じ枠組みの中で、まとめてデータが手に入ると嬉しい。企業側としても、個別のやり取りに比べ社内への説明が楽になるのではないかと

データ提供者と教材作成者1対1の関係性では、提供できるインセンティブが限定的...基本形以外のインセンティブスキームが必要

B 再掲) 検討の前提: 企業の実課題・実データに基づく教材作成



民間教材作成事業者や大学がデータを収集し教材を作成
(=1対1の関係性が基本前提。但し、データ提供側または受取側が複数の1対Nは派生形として、インセンティブスキームの類型の中で検討)

本日のアジェンダ

前回検討会の振り返り

第1回検討会以降の検討状況のご報告

ガイドラインの位置づけの再確認

ガイドラインに関する討議

今後の進め方

ガイドラインの目的と位置づけの再確認

ガイドラインの目的に関する考え方

AI人材育成にとどまらず、中長期的には幅広い目的で企業等によるデータ提供とその利活用が活発に行われている状況を創出することに資するガイドラインとしたい

(ガイドラインの目的を「人材育成」や「そのための教材」と制限しすぎると、企業への訴求が阻害される恐れ)

上記を達成するために、実務上指標となる項目を分かりやすくまとめるガイドラインを作成すべく、
題材として「教材作成のためのデータ提供」を扱う



一方で、人材育成・教材等、特定の目的を据えた方が他ガイドライン・検討との差別化ができ、インパクトの創出は可能かもしれない

ガイドラインの位置づけ

左記目的を達成するために実務上指標となる項目を分かりやすくまとめる。そのために、既存のガイドラインや手引き等を必要に応じて引用する等、既存の検討内容を活用する形で、実務上指標となる項目をとりまとめる

ガイドラインの中身も踏まえ、目的をどこに据えるかについてご意見をいただきたい

本日のアジェンダ

前回検討会の振り返り

第1回検討会以降の検討状況のご報告

ガイドラインの位置づけの再確認

ガイドラインに関する討議

今後の進め方

「企業の実課題・データ提供ガイドライン(仮)」の章立て (1/2)

ガイドライン章立て (案)

概要

第1章 本ガイドラインに関する 基本事項	<ol style="list-style-type: none">1. 本ガイドライン策定の背景・課題認識2. 本ガイドラインの趣旨・目的3. 本ガイドラインの対象・想定される活用方法4. 本ガイドラインの構成	本ガイドラインは「データを提供する企業と、データを受け取る教材作成者を想定し、教材作成のための企業の実課題・データ提供に向けて実務上の指標となる項目をまとめる」ことを記述
第2章 企業の実データを取り巻く 現状と課題	<ol style="list-style-type: none">1. 企業の実データに基づく教材に対するニーズの現状2. 実データに基づく教材作成のためのデータ提供に関する課題	実データに基づく教材に対するニーズの高まりの一方で、データ提供に関する課題(例:「ノウハウ」の概念のあいまいさ)が存在することを、データ提供企業・教材作成者それぞれへのヒアリング結果を基に記述
第3章 データ提供に関する法的・ ビジネス基礎知識	<ol style="list-style-type: none">1. 本ガイドラインで想定する前提2. データ提供に伴うリスク・ベネフィット概要3. データ流出や不正利用を防止する各種手段	本ガイドラインで前提とするデータ提供者と教材作成者について説明した上で、データ提供に伴い一般的に想定されるリスクとベネフィットとデータ流出や不正利用の各種防止手段について記述
第4章 データ提供に関する手引き		
第5章 総括		

「企業の実課題・データ提供ガイドライン(仮)」の章立て (2/2)

第4章で述べる論点の全体像を紹介

■ 本日特に議論したい点

ガイドライン章立て (案)

第1章

本ガイドラインに関する
基本事項

第2章

企業の実データを取り巻く
現状と課題

第3章

データ提供に関する法的・
ビジネス基礎知識

第4章

データ提供に関する手引き

第5章

総括

概要

1. データ提供にあたっての基本的な考え方と想定され得る法的・ビジネス論点全体像	技術上・経営上重要なデータは提供しないという原則を説明した上で、教材の目的に応じて必要なデータ加工の方向性と、教材の利用範囲の制限方法のベストプラクティスについて記述
2. データ提供にあたって想定され得る論点に対する解の方向性	保護手段として①契約による保護、②不正競争防止法による保護、③知的財産法による間接的保護について記述
(1) 総論(個人情報含む・含まない場合両方) (ア) 教材のあり方 (イ) データを保護するための法的手段 (ウ) インセンティブスキームの類型	データ提供側のコストをベネフィットが上回るようにするインセンティブスキームを紹介(次頁詳細)
(2) 個人情報を含む場合に追加で考慮すべき点 (ア) 個人情報保護法が適用される場合: 個人情報を含むデータを教材に使うための法的スキーム(教材の利用制限の話も含む) (イ) 個人情報保護法が適用されない場合	データが個人情報を含む場合の考慮点として、データの取り扱い、個人情報の対象、第三者提供の要件、「匿名化」の意味の誤解等について説明
(3) 実務で活用可能なモデル契約書 (ア) 教材としてのデータ提供に関するモデル契約書 (イ) 教材作成者が学習者に教材を提供する際のモデル契約書	データ提供企業と教材作成者間、教材作成者と学習者間の契約の雛形とその解説を掲載

討議の進め方

第1~3章 : 事前にいただいた意見を基に更新したガイドラインについて確認
第4章 : 本検討会の中で主に議論

1対1の基本形以外でのインセンティブスキームのベストプラクティス(議論用)

→ データ・教材 → 契約 → 成果物・サービス

概要	データ提供者へのインセンティブ	企業 (データ提供側)	教材作成者 (データ受取側)	学習者
ハッカソン型 企業が提供したデータを基に学習者が競い合う形で開発を実施	<ul style="list-style-type: none"> データから創出された成果物の還元 先端人材との接点の増加 ハッカソン参加という企業としてのPR 		ハッカソン 	
クローズドデータプラットフォーム型 複数企業が参加し、データを提供(企業間のデータや知見の共有も可能)	<ul style="list-style-type: none"> 先端人材との接点の増加 他社データや利活用に関する知見の獲得、データの相互利用による新たなビジネス・サービスの創出 		プラットフォーム 	
データ創出・活用エコシステム型 教材作成者が、事業の中で(自然に)発生するデータを取得し、企業側へのサービスへ利用	<ul style="list-style-type: none"> データから創出された成果物の還元 他社データや利活用に関する知見の獲得、データの相互利用による新たなビジネス・サービスの創出 		エコシステム 	

1. Non-Disclosure Agreement: 秘密保持契約

再掲) ガイドラインの目的と位置づけの再確認

ガイドラインの目的に関する考え方

AI人材育成にとどまらず、中長期的には幅広い目的で企業等によるデータ提供とその利活用が活発に行われている状況を創出することに資するガイドラインとしたい

(ガイドラインの目的を「人材育成」や「そのための教材」と制限しすぎると、企業への訴求が阻害される恐れ)

上記を達成するために、実務上指標となる項目を分かりやすくまとめるガイドラインを作成すべく、
題材として「教材作成のためのデータ提供」を扱う



一方で、人材育成・教材等、特定の目的を据えた方が他ガイドライン・検討との差別化ができ、インパクトの創出は可能かもしれない

ガイドラインの位置づけ

左記目的を達成するために実務上指標となる項目を分かりやすくまとめる。そのために、既存のガイドラインや手引き等を必要に応じて引用する等、既存の検討内容を活用する形で、実務上指標となる項目をとりまとめる

ガイドラインの中身も踏まえ、目的をどこに据えるか。
あわせて、ガイドラインの名称についてもご意見を伺いたい

本日のアジェンダ

前回検討会の振り返り

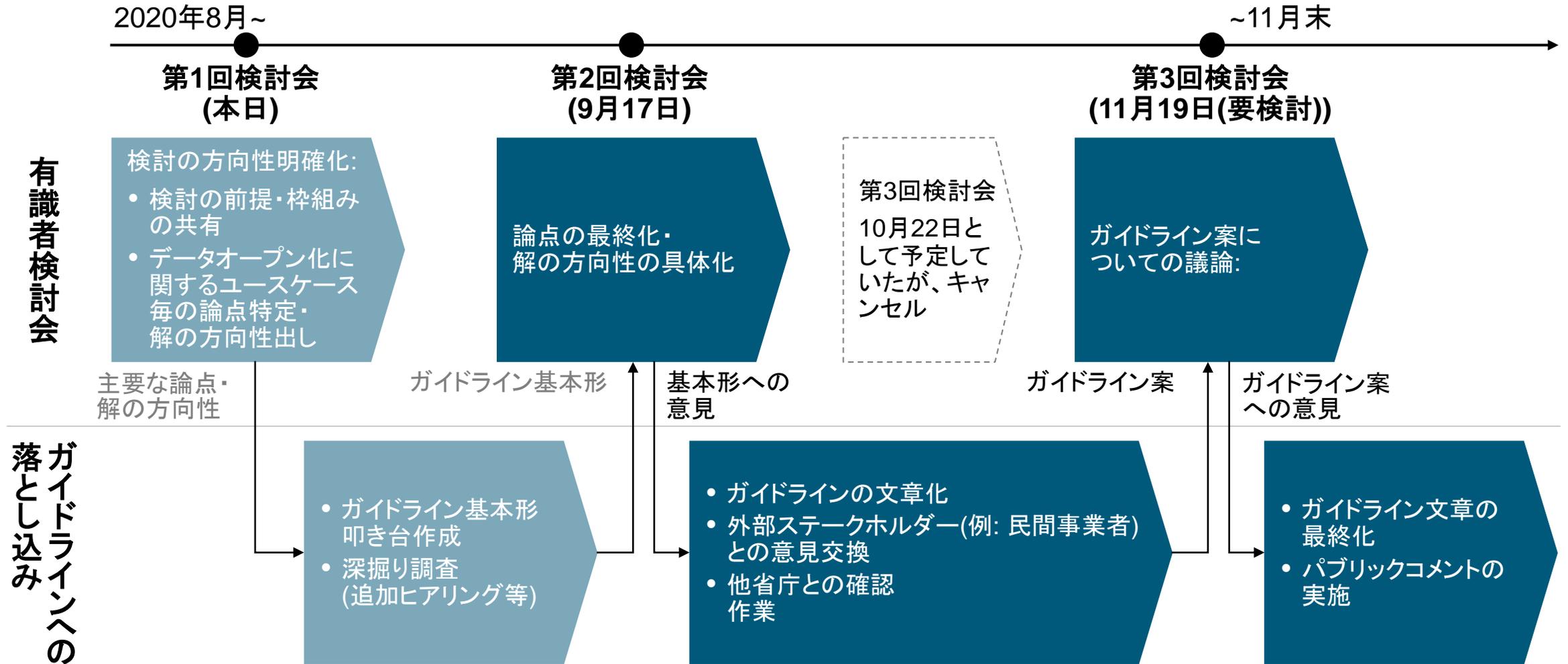
第1回検討会以降の検討状況のご報告

ガイドラインの位置づけの再確認

ガイドラインに関する討議

今後の進め方

AI人材育成のための企業間データ提供促進検討会の全体タイムライン改訂版



11月19日を第3回としてほぼ文章化が完了したガイドラインの確認を行うという進め方に対しご意見を伺いたい(その後パブリックコメントを実施)