AIをどう監査に統合していくか？
～具体的なLLMの活用事例と将来への課題～

株式会社X-Regulation

代表取締役CEO　小田切 洋介
代表取締役CBO　東海林 和広

X-Regulation
クロスレギュレーション

www.x-regulation.com
info@x-regulation.com

目次

1. イントロダクション
2. 実際のLLM活用の具体例
3. LLM活用時の留意点
4. Q&A

１．イントロダクション

1. イントロダクション（1/4）会社紹介

X-Regulation
クロスレギュレーション

2023年5月創業のTech×Professionalのスタートアップで、LLM統合型ガバナンスツールを開発、サービス提供しています。

Mission
すべてのガバナンス・コンプライアンスをAIトランスフォームする。

Vision
誠実な人・企業がより評価され、成長できる世界。

Purpose
世界を今よりもフェアにする（ステークホルダー資本主義を実現する）

1. イントロダクション（2/4）講師紹介

代表取締役CEO　小田切 洋介

- 大手監査法人を経て、X-Regulation社を創業
- 公認情報システム監査人（CISA）
- 東京大学修了（学術修士）

取締役CBO　東海林 和広

- 大手監査法人および中央省庁を経て、X-Regulation社を創業
- 公認情報システム監査人（CISA）、Certified Data Privacy Solutions Engineer （CDPSE）
- 上智大学修了（経済学修士）

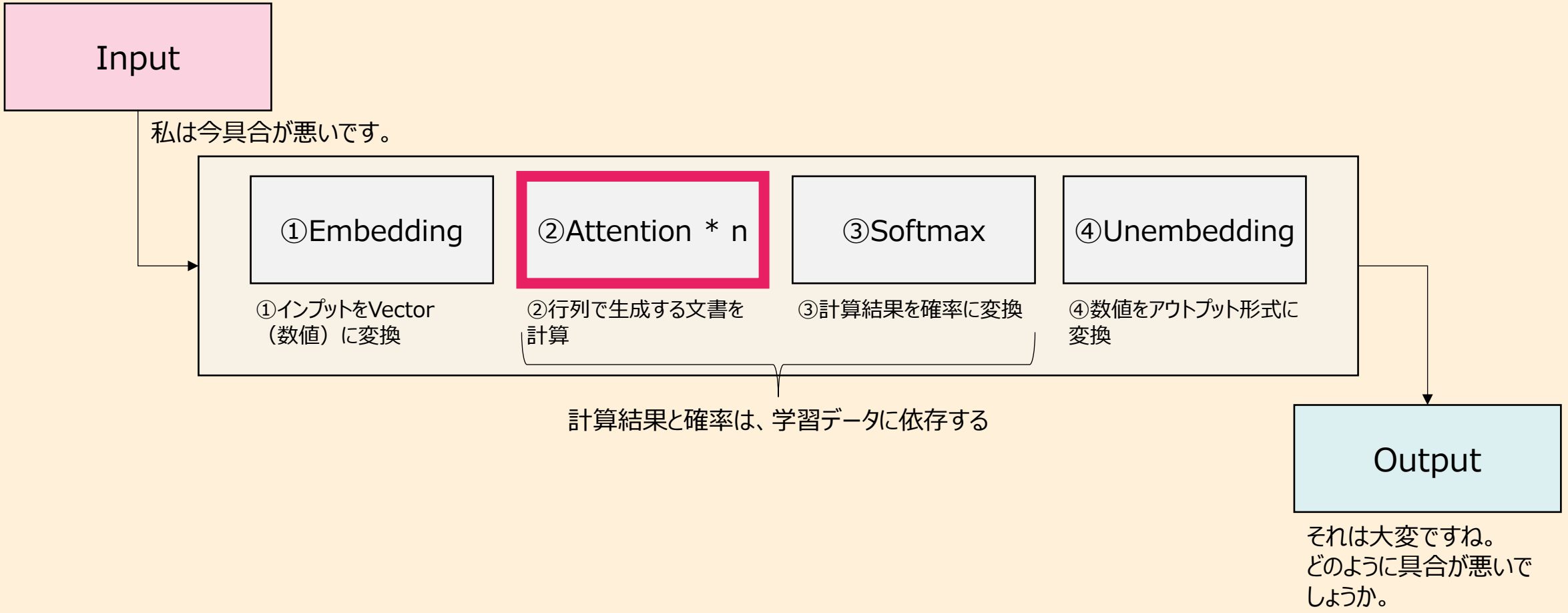
1. イントロダクション（3/4）前回の振り返り

- 監査の未来を見据えると、LLMの監査への利用は不可逆的に進行する。
- LLMは、大量のデータセットを基に学習した確率遷移（アテンション＋ソフトマックス）により、人間の思考・判断を模倣する。
- LLMと監査は非常に相性が良く、いくつか考慮点はあるものの、統合のメリットは大きい。

1. イントロダクション（3/4）前回の振り返り

アテンションの役割を理解することが、LLMを理解する上で最も重要です。

LLMのInputからOutputまでの大まかなイメージ



1. イントロダクション（3/4）前回の振り返り

LLMを監査に応用する際の大きな障壁として、
アカウンタビリティとバイアスが挙げられます。

1．アカウンタビリティ

- 一般的な監査・評価業務では、その性質上、経営者や監査責任者等が、実施された監査手続結果に対して責任を持つことが求められます。

- 当該責任を果たすため、 深度は監査テーマ等によって差はありつつも、**LLMが行った監査手続結果にミスが内在していないか、誰かが確認を行う必要**があるでしょう。

2．バイアス

- LLMのアウトプットには、バイアスが含まれる可能性があります。大別すると、**①活用するLLMのモデルそのものが孕むバイアス、②インプット情報によるバイアス**、の2つが挙げられるでしょう。

- 一般的に、①はLLMの供給元において非開示なことが多いです。②は例えば、インプットした情報に応じて、本当は出来ていないことを出来ていると言ってしまう（厳格に見るとNGにも関わらずOKと出力してしまう）等の現象が起こり得るということです。

1. イントロダクション（4/4）セッションの目的

当セッションの目的は、**「結局、LLMを監査に取り込むとどうなる!?」について、現在の実装を基に体感する**ことです。

＜当セッションで体験するLLMの現在地＞

- LLMはすでに、我々が想像している以上に、人間を強力に補佐することができる
- しかし、監査の世界では、結局、まだまだ人間の力は必要である（「全て自動化」という発想は危険）
- なお、当セッションでは以下の項目は取り扱わない
  - LLM活用に関する理論的枠組みの形成やリスクの網羅的検討
  - プロンプト講座（プロンプトの細かい説明や実装方法等）

２．実際のLLM活用の具体例

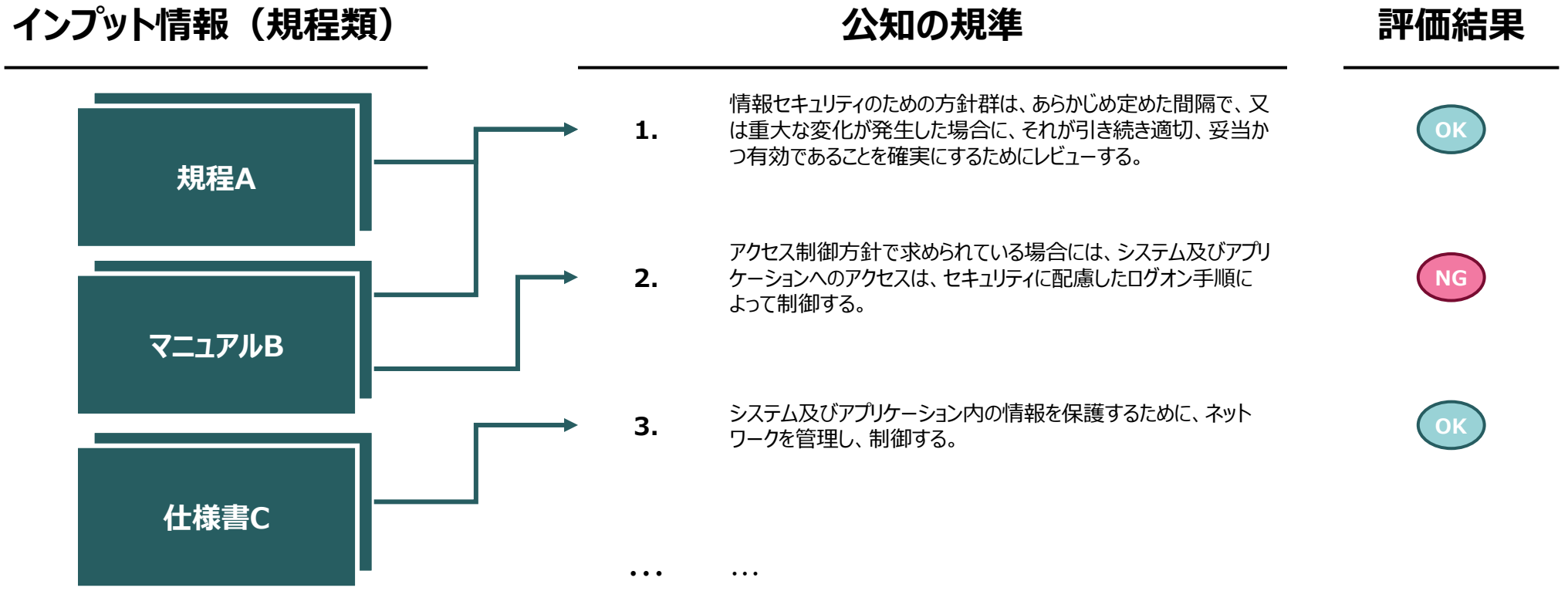
2．実際のLLM活用の具体例　ーはじめにー

まずは、簡単な事例で検証してみましょう

デモ

2．実際のLLM活用の具体例　ー公的規準への適合性評価ー

適合性評価業務とは、公知の基準に対する内部統制のデザインの適切性を
社内規則を基準にマッピングすることで評価する業務です。



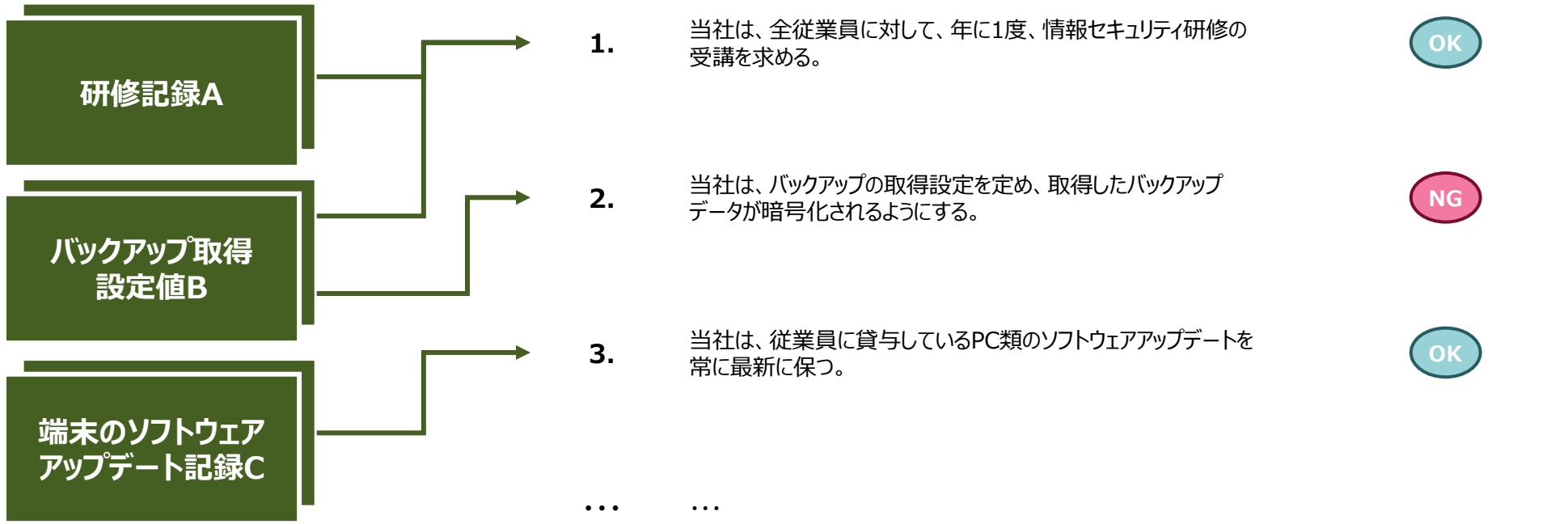
2．実際のLLM活用の具体例　ー公的規準への適合性評価ー

監査・点検等にLLMを統合したプロダクト上で、適合性評価を実施してみます。

デモ

2．実際のLLM活用の具体例　ー内規の点検業務ー

ここでの点検業務とは、評価規準に基づき、質問・閲覧手続を組み合わせ、
内部統制が実務に適用されていることの心証を得る業務を指します。



2．実際のLLM活用の具体例　ー内規の点検業務ー

監査・点検等にLLMを統合したプロダクト上で、点検業務を実施してみます。

デモ

2．実際のLLM活用の具体例　ーLLMと証跡の相性ー

LLMで取り扱うことができる証跡種別や、各証跡との相性について、これまでの経験ベースで列挙してみました。

種別	証跡（例示）	親和性	備考
テキストデータ	ドキュメント情報 （例：Word等の社内規程、・・・）	◎	• 社内の規程や運用手順書、Wikiのような情報については、かなりの精度で正確に読み取り、アウトプットすることが可能
	表形式の情報 （例：csvやExcel、、、）	○	• システムログや組織図など、表形式の中の情報の相関性が比較的見やすいものについては得意 • カラム同士の関係性が見えづらい（人間でも上手く説明できないようなもの）表形式の情報は、場合によっては正確性に欠けることがある
イメージデータ	画像データ （例：比較的テキストが多いもの）	◎	• ワークフローチケットを画像形式で落としたデータ等において、例えば承認者の氏名や承認日付のような、比較的言語化が容易なものについてはかなりの精度でアウトプットが可能
	画像データ （例：印鑑の読み取り）	△	• 例えば申請書の承認欄に印鑑が押されており、その印鑑が人間の目から見ると明らかに後から不正に押されたように見える、等というプロならではの視点のような領域をLLMが汲み取るのは比較的困難（LLMというよりは特化型AIの領域）

2．実際のLLM活用の具体例　ー所感ー

デモからも、充分に実用に耐えるレベルである一方で、
直ちに評価業務”全体”を代替するものではないことがわかります。

観点	評価	所感
LLMのアウトプット精度	★★★☆☆	• 各規準に対する分析・作業については、想像以上の精度を発揮する • その領域×監査のプロフェッショナルと比較すると、まだ足りない部分が存在する（グレーゾーンの判断等）
LLMの処理スピード	★★★★★	• 約1,000項目を数分程度でドラフトするため、人間の誰よりも早い
トータルの工数削減量	★★★★☆	• 圧倒的なスピードで分析・作業自体が完了するが、人間が作業結果をレビューする手間は引き続き残り続ける

2．実際のLLM活用の具体例　ーレビュー活動の必要性ー

人間・AI問わず、評価者には「偽陰性のリスク」が潜在するため、
AI活用の世界においても、引き続きレビュー活動を行う必要性は一定程度残り続けると考えられます。

パターン		実際の状態	
		例外事項あり	例外事項なし
評価者（人間・AI）の分類	例外事項あり	**（真）陽性** *「例外事項あり」と正しく分類	**（偽）陽性** *実際の状態は「例外事項なし」だが評価者は「例外事項あり」と分類
	例外事項なし	**（偽）陰性** *実際の状態は「例外事項あり」だが評価者は「例外事項なし」と分類	**（真）陰性** *「例外事項なし」と正しく分類

2．実際のLLM活用の具体例　ー事例紹介ー

実際にLLMを監査・点検活動に組み込んでいる事例をご紹介します。

＜成功事例＞

- 金融業者において、金融庁の要求等によって監督指針や各種ガイドラインへの適合を頻度高く要求される
- 準拠すべき規準1つ当たり、概ね2～3ヶ月程度の評価期間を要していたが、約1か月程で評価業務を完了できるサイクルを回せるようになった
- 今後、さらなる成功のための論点として、以下のような検討の余地がある
    - 将来的な、監査証跡の自動収集（例：LLMエージェントが自律的に証跡を収集する、等）

＜よくある失敗事例＞

- 自身が持つデータの全てをLLMに与えると、例えばリスクの予兆管理において、質の良い結果を出してくれるのでは!?という誤解
- Garbage In Garbage Outの原則がLLMにも適用される
- LLMに対して期待するアウトプットを明確に定義し、そのために必要なインプット情報が整備されていなければ、LLMも当然良い結果を返さない

３．LLM活用時の留意点

3．LLM活用時の留意点

LLM活用時に想定される一般的な懸念は、主要なクラウド事業者（例：AWS, Azure, GCP）によってクリアされていることがほとんどです。

パターン	想定される懸念	対策例
LLMのモデルそのもの	• データの再学習 • データの保管場所 • その他、セキュリティ問題 etc…	• SOC2やISMAPといった、一般的に認められている公知の報告書・認証制度への準拠状況の確認

LLMの利用者	• ハルシネーション問題（以下、例示）
	  • 人間による無意識的な誘導
	  • アウトプット（期待値）の明確な定義の欠如による回答のブレ
	  • etc…
		• フィードバックループを整備する
		  • 人間によるレビュー活動の挿入
		  • AIに期待する業務要件の明確化
		  • LLM Opsが確立されたサービスを活用する
		  • etc…

3．LLM活用時の留意点

LLM活用においては、LLMで何がどこまで出来るのか、何をさせるべきなのかを考えた上で利用することが大切です。

- **人間が保有すべき専門知識**
  - LLMが専門的判断をある程度のレベルで行えるとはいえ、倫理的な決定は人間が行うべき
  - 特に例外処理やグレーゾーンに対する判断は人間の役割
- **LLMの補完的役割**
  - LLMは分析や作業を支援するが、最終的な責任は人間にあり、「完全な自動化」に夢を見られるわけではない
- **LLMを活用していく上での今後の論点**
  - LLMの判断を承認できる人間が必要
  - こうした人間のスキルは、分析・作業を通じて養われるものだが、その業務自体をAIが代替するようになる中で、どう人間を教育していくべきか

⇒今後、AI時代において、責任を取れる最終決定を行う人間をどう育てていくのかが、焦点となります

4. Q&A