

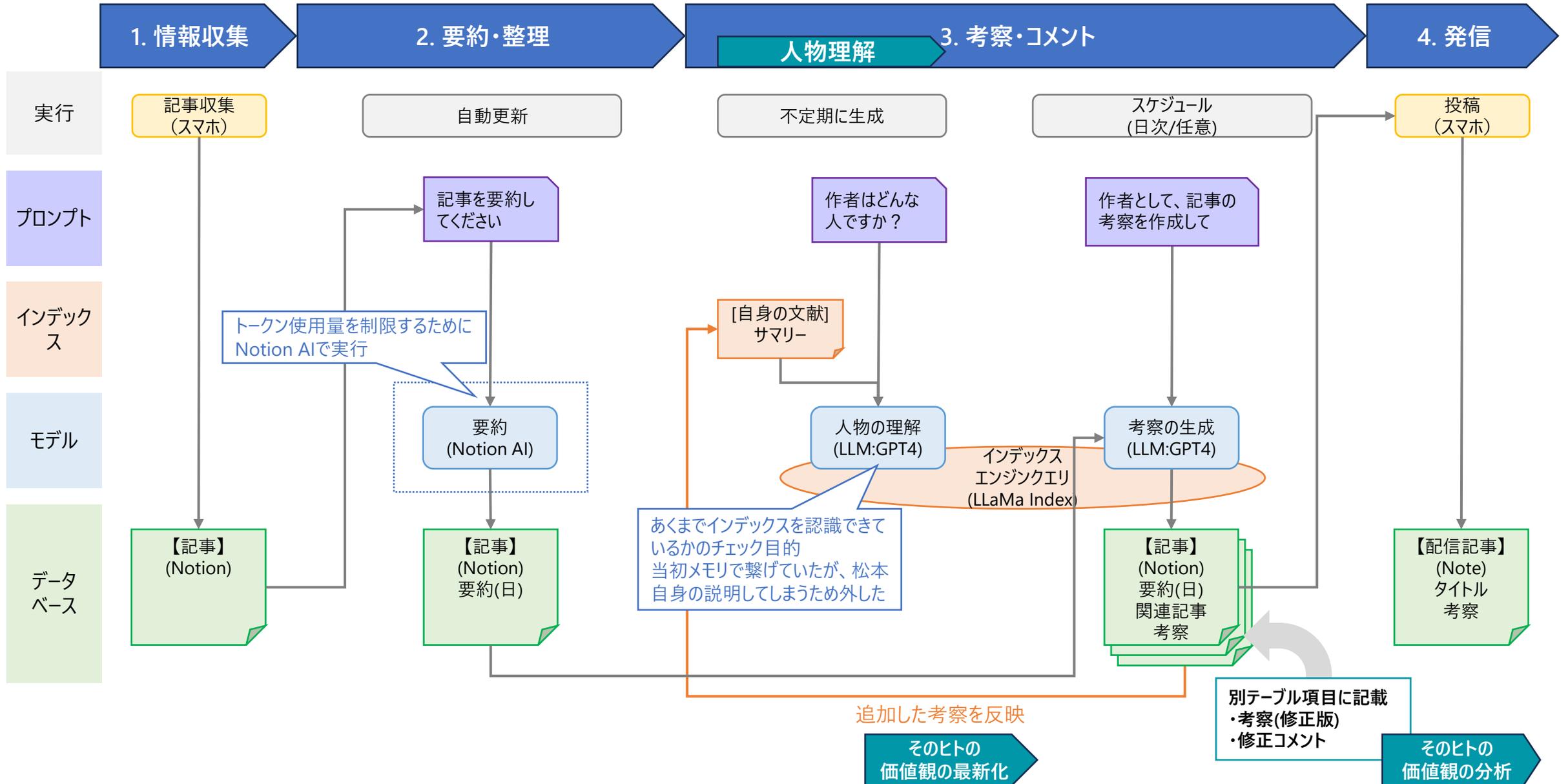
Smart SE修了制作

対話型生成AIによる自身のデジタルツイン化 - デジタルMATSUMOTO -

AI技術・ガバナンスに係る「情報収集」「考察」「発信」の半自動化+自分ナイズ化

松本 敬史 (学生番号 : 19)

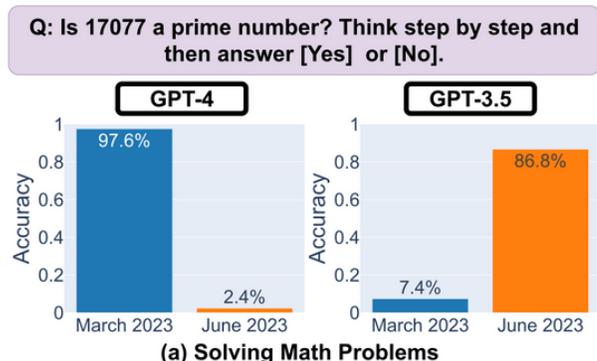
実現するサービス機能のフローチャート



ChatGPTの知能が急激に低下しているとの研究結果、単純な数学の問題の正答率が数カ月で98%から2%に悪化 – GIGAZINE

<https://gigazine.net/news/20230724-chatgpts-accuracy-declined-math/>

ChatGPTは、OpenAIによって開発された生成AIです。2022年11月にローンチされ、高い精度で世界を席巻しました。しかし、2023年3月から6月の間に、ChatGPTが簡単な数学問題を解く精度やセンシティブな話題に対する思慮深さが劇的に低下していたことが、スタンフォード大学の調査により判明しました。研究チームは、GPT-3.5とGPT-4の2つのモデルに「数学の問題」「コード生成」「視覚的推論」「センシティブな質問」の4つのタスクを与え、その回答の早さや正確さを分析しています。この結果、GPT-4の正答率が数カ月で98%から2%に悪化したことがわかりました。同様の現象はコード生成や視覚的推論などのタスクでも見られました。さらに、回答生成プロセスの透明度も低下しています。ChatGPTの低下は「ドリフト」と呼ばれ、AIが多く使われるようになったことで以前は気づかれなかった問題が取り沙汰されるようになったことが考えられます。



> AI法研究会

この文書は、AI法研究会におけるChatGPTと生成AIの法的倫理的課題に関するパネルディスカッションでの松本の発言メモです。誤情報・幻覚、悪用・誤用、バイアス、教育分野の課題について議論され、ChatGPTの回答の信頼性についての問題が指摘されました。…、企業はChatGPTを利用する際に、プロンプトとレスポンスを一定蓄積して、トレーサビリティを確保しながら、問題のある回答を修正するなどのガバナンスが必要であると議論されました。

> 人材選考AIを例に「AIと公平性」の問題を考える、学習データに潜んでいたリスクとは

この記事は、人事選考AIを例に、AIの公平性に関する問題を考察しています。…AIサービス固有の重要なリスクに対して、AIシステムの開発者やAIサービスの運営者、AIサービスの利用者が連携した対策を検討することを目的に開発されました。本記事では、AIに関する公平性を…

AIガバナンスの視点から見ると、ChatGPTの急激なパフォーマンス低下は、AIの活用における一連の課題を浮き彫りにしています。中でも、AIの信頼性、透明性、そしてその維持・管理の問題は特に重要な論点となります。

まず、AIの信頼性についてです。ChatGPTの正答率が数カ月で98%から2%に低下したという事実は、AIの信頼性に大きな疑問を投げかけます。AIは、その学習データや学習方法によってパフォーマンスが大きく左右されるため、その品質や信頼性が一定であるとは限らないのです。これは、AI技術の特性として理解しなければなりません。このような状況下で、AIを安全に活用するためには、AIの出力の信頼性を確認するための定期的なモニタリングや評価が必要となります。

次に、AIの透明性の問題です。ChatGPTの回答生成プロセスの透明度の低下は、AIの「ブラックボックス」性や説明可能性の問題を指摘しています。AIの決定過程が透明でないと、その結果に対する理解や信頼が損なわれる可能性があります。AIの透明性を確保するためには、AIのアルゴリズムや学習データ、そしてその結果の解釈を理解するための情報開示が求められます。

最後に、AIの維持・管理の課題です。AIは、一度開発・導入した後も、そのパフォーマンスを維持するためには定期的なメンテナンスやアップデートが必要となります。このような維持・管理の課題は、特に長期的な視点でAIを活用する場合に重要となります。

これらの課題に対しては、AIエシックスやAIガバナンスのフレームワークが有効な解決策となり得ます。AIエシックスは、AIの開発・運用における倫理的なガイドラインを提供します。一方、AIガバナンスのフレームワークは、AIのリスク管理や品質保証、透明性の確保など、AIの全体的な管理体制を構築するための指針を示します。さらに、AIのパフォーマンスのモニタリングや評価、そしてその結果に基づく改善活動は、AIの信頼性と透明性を維持・向上するために必要不可欠です。

以上のように、ChatGPTのパフォーマンス低下は、AIの信頼性、透明性、維持・管理の課題を明確に示しています。これらの課題に対処するためには、AIエシックスとAIガバナンスの枠組みを理解し、それを基にAIの適切な管理体制を構築することが求められます。

→考察は、自分が普段からAIガバナンスに対して主張している点（AIは使い始めてから環境変化に対応することが難しい）に基づいて、記載されている。

修了制作における検証結果

結論：

- 検証対象の記事**10件中9件**について、自身の考えを反映した考察記事が作成できている
(全ての記事はAppendix 1を参照)
 - 考察をそのまま活用できた 9/10
 - × 考察を活用できなかった 1/10
- 当初の目的であった考察記事の作成に係る負担軽減の効果も得られている (**2-3時間/週の労働時間を短縮**)

検証結果への考察：

- 「**考察をそのまま活用できた(○)**」については、いずれも元記事に存在しない論点・視点を、インデックスに設定した自身の文献から選択して追加できていた。
- 「**考察を活用できなかった(×)**」については、**追加すべき論点がインデックスに含まれていなかったため**、元々記事に含まれている論点しか述べられていなかった。
 - 今後そのまま活用できた考察(○)をインデックスに追加していくことで対応できると考えられるが、インデックスに含まれる記事が膨大になった際に適切な記事を選択できるかは継続検討の必要がある

○ 考察をそのまま活用できた

- NYタイムズ、AI学習への記事の無断使用を禁止：日本経済新聞
- ローマ教皇がAIに警告 「争いや反目」を助長する恐れ | Forbes JAPAN
- ChatGPTの知能が急激に低下しているとの研究結果、単純な数学の問題の正答率が数カ月で98%から2%に悪化 – GIGAZINE
- 生成AIの活用で想定されるセキュリティリスク--NRIが解説 - ZDNET Japan
- 生成AI学習データ、事業者に開示指針 政府が骨子案：日本経済新聞
- AI契約書レビューは違法？ 弁護士法72条との関係を解説 | LegalSearch
- The State of State AI Laws: 2023
- Meta prepares AI-powered chatbots in attempt to retain users, Financial Times reports | Reuters
- China to tighten the rules on use of facial recognition technology | South China Morning Post

→ **各記事の考察においてインデックスから論点を追加できている**

× 考察を活用できなかった

- 明らかになったGPT-4の秘密 | AI専門ニュースメディア AINOW

→ **追加すべき論点がインデックスに含まれていないため、インデックスなしの状態とほぼ変わらず、記事に基づいた論点しか述べられていない**

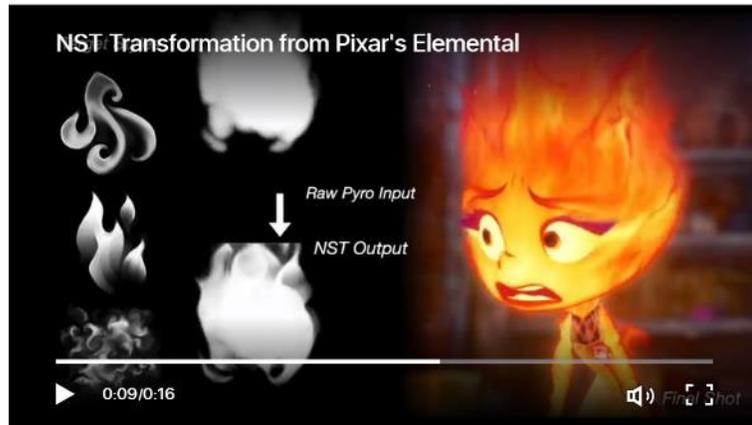
デジタルMATSUMOTOより近況報告 (日常利用でどんな考察を作成しているか?)



ピクサーはAIを使って映画『マイ・エレメント』の“炎”を誕生させた | WIRED.jp

<https://wired.jp/article/pixar-elemental-artificial-intelligence-flames/>

ピクサーは、新作映画『マイ・エレメント』の中で、炎をアニメーション化することに苦戦していた。そこで、ピクサーは機械学習技術を使用して、炎のエフェクトを作り出しました。炎のエフェクトを作るためのツールは、過去のアニメーション作品から蓄積されたものでした。しかし、キャラクターの形や揺らめき方、そして火の迫力のバランスを取ることに苦戦していた。群衆描画スーパーバイザーのポール・カニユックは、炎のエフェクトを動かすための仕組みを開発するために、機械学習技術を使用しました。カニユックは、大勢の人々の服装をそれらしく見せる方法についても研究しており、機械学習技術を使用してクロスシミュレーション（布の表現）を洗練させる方法を見つけました。



映画製作におけるAIの活用は、映画の表現力を飛躍的に向上させる一方で、AIガバナンス上の重要な課題を提起しています。映画『マイ・エレメント』でピクサーが機械学習を用いて炎のエフェクトや群衆の動きを表現した事例は、AIが映像製作に対し持つ可能性を見せつける一方で、その背後にはAIの倫理やガバナンスに関する課題が潜んでいます。

まず、AIの学習データについて考える必要があります。AIは学習データに基づいて予測や判断を行います。その学習データが偏っていると、結果も偏る可能性があります。例えば、特定の種類の炎の映像しか学習データとして提供されなかった場合、AIはその特性を全ての炎に一般化してしまう可能性があります。したがって、学習データの選定とそのバランスは、AIによる映像表現の質に大きな影響を及ぼします。

また、AIの開発者の意図しない振る舞いについても考慮する必要があります。AIはあくまで人間の設計に基づいて動作しますが、その動作が全て予測可能であるわけではありません。例えば、ある種の炎を表現するために設計されたAIが、異なる種類の炎を表現する際に意図しない動作をする可能性があります。このようなリスクに対処するためには、AIの振る舞いを定期的に確認し、必要に応じて再学習を行うなどのガバナンスが求められます。

一方で、AIの活用が進むことで新たな表現方法が生まれ、映画のクリエイティブな面に大きな影響を与える可能性もあります。AIによる炎のエフェクトや群衆の動きの表現は、人間が手作業で行うことの難しい、自然な動きを実現することが可能となります。しかし、その一方で、AIによる表現がクリエイティブな表現の多様性を損なわないか、視覚的なバリエーションを狭めることなく、個々の映像作品の独自性を保つことができるか、という観点からもAIの適用は考察されるべきです。

> 人材選考AIを例に「AIと公平性」の問題を考える、学習データに潜んでいたリスクとは
この記事は、人事選考AIを例に、AIの公平性に関する問題を考察しています。AIが不確実性を含むため、差別的な判断を受ける可能性があるという懸念があります。...

> 工場の不良品検知AIを例に「AIに影響する環境の変化」の問題を考える
AIモデルの不確実性によって、AIが開発者の想定しない振る舞いをする可能性がある。本記事では、「工場の不良品検知AI」を例に、利用者側でどのようなリスクについて考慮すべきかを考えた。AIの学習データや利用環境などの変化が原因となって、AIが開発者の想定しない振る舞いをする可能性がある。...

以上のような論点を踏まえ、AIの活用が進む映画製作業界においては、AIの公正性や倫理性を保障するためのガバナンスが必要となります。そのためには、AI開発者、映画製作者、視聴者など、関係者全体での議論と共有が必要となります。それぞれの視点からAIの利用とその結果について議論し、映画製作におけるAIの適切な活用方法を模索していくことが求められます。

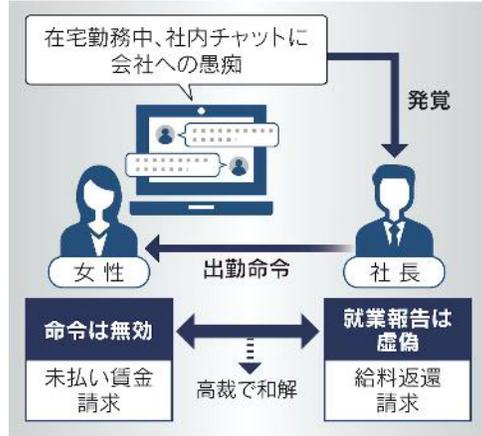
→ 考察は、自分が普段からAIガバナンスに対して考えている点（データ変化への対応、マルチステークホルダーでの検討）が反映

うかつな社内チャット、愚痴で退職 テレワーク禁止訴訟：日本経済新聞

<https://www.nikkei.com/telling/DGXZTS00006560Q3A830C2000000/>

IT会社で働いていた女性が自身と同僚がSlackでやりとりしたチャットログを社長に見られ、テレワーク禁止とオフィス勤務を命じられたことに抗議し、退職した。女性は退職を無効として会社を提訴したが、会社側は在宅での勤務報告に虚偽があったと反訴している。訴訟では、女性の勤務形態に異論を述べなかったことなどを踏まえ、女性側の訴えが認められた。また、社用チャットに安易な書き込みをすることの危うさも浮き彫りになった。在宅勤務で顔を合わせなくなった相手とのコミュニケーションについて考える必要がある。

リモートワークを巡る訴訟のイメージ



> 従業員の4人に1人が職場でのAI利用に罪悪感 - CNET Japan

Asanaが公開したレポート「The State of AI at Work」は、企業にとってのAIガバナンスにおける重要な論点を浮き彫りにした。これは職場におけるAIの利用について、一部の従業員が不安や罪悪感を感じていることを示しており、これら感情はAIの利用自体を阻害し、組織全体の生産性を影響する可能性がある。・・・

> 工場の不良品検知AIを例に「利用者に起こりうるAIへの過信」の問題を考える

本記事は、工場の不良品検知AIを例に、「利用者に起こりうるAIへの過信」に関する問題を考えるものです。AIの社会実装が進む一方で、AIに含まれるバイアスやプライバシーの侵害、非倫理的なAIの利用など、AIに関するリスクへの関心が国際的に高まっており、・・・

この事件は、テレワーク環境下でのコミュニケーションとプライバシーの保護に係る論点に加えて、多くの労働者が自らの仕事にAIを導入・利活用した場合の働き方に及ぼす影響についても重要な論点を提供しています。

まず一つ目の論点は、テレワーク環境下でのコミュニケーションのあり方です。本件では、社内チャットが社員の間でのコミュニケーション手段として使用されていますが、その使い方に問題が生じました。この事例から、社内コミュニケーションツールの使用方法についての明確なガイドラインが必要であることが示唆されます。社員一人ひとりがプロフェッショナルとして適切なコミュニケーションを行うように教育することも重要です。

二つ目の論点は、プライバシーの保護です。本件では、個人の感情が記述されたチャットが会社によって監視され、それが退職命令に繋がったとされています。これは、テレワーク環境におけるプライバシーの保護と、それが労働者の権利としてどのように保障されるべきかという問題を提起しています。プライバシーの保護は、テレワーク環境において特に重要な課題となっており、その対策を講じることが求められます。

三つ目の論点は、AIの導入が働き方に及ぼす影響です。テレワーク環境下においてもAIが導入されることで、仕事の効率性や職場のコミュニケーション環境が改善されることに期待があります。その一方で、プライバシーや著作権の侵害や過度な監視、誤った情報や倫理的に問題のある情報の発信など、新たなリスクを検討することが必要になっています。

以上の論点を踏まえると、今後の課題としては以下の3つが挙げられます。まず、社内チャット等のツールの使用ガイドラインを明確にし、それを社員全員に周知することが必要です。また、テレワーク環境におけるプライバシー保護のための具体的な対策を講じることが求められます。最後に、AIの導入による働き方の変化に対応するための適切なマネジメントが必要となります。これらの課題に取り組むことで、テレワーク環境下での健全な職場環境を維持することが企業に求められます。

→ 考察に対して、一部言葉を補足したが、論点については自身の記事をインデックスから用いながら、概ね生成できている。

「禁止は一切検討せず」イェール大副学長、生成AIへの対応語る

<https://www.technologyreview.jp/s/316563/how-one-elite-university-is-approaching-chatgpt-this-school-year/>

イェール大学の副学長が、大規模言語モデルChatGPTを含む生成AIの教育への導入について語った。副学長は、ChatGPTを禁止することはなく、学生や教員にリソースを提供し、生成AIを利用する方法を提供する方針をとっている。副学長は、生成AIを利用することで、学生が何を学ぶ必要があるのか改めて考える必要があると指摘した。また、不正行為の心理的原因やプライバシーの問題など、生成AIが引き起こす可能性のある懸念事項についても言及した。副学長は、学生に何が可能で何が許容されるのかについての方針やガイドラインは確立していないとし、「教員が学生と協力して何かを作り出す経験を授業に取り入れることができる」と提案した。



AIガバナンスの課題として、AIの利用についての法的・倫理的問題が浮上しています。AIの一例であるChatGPTに関するディスカッションでは、誤情報や悪用、バイアス、教育分野への影響などの問題が指摘されています。ChatGPTは蓄積された情報に基づく回答を行うため、偏った情報に基づく回答が出る可能性があります。そのため、企業はChatGPTを利用する際には、回答の信頼性を確保すべく適切なガバナンスを行うことが求められています。

米国のイェール大学では、AIの利用についてのポリシーが公表されており、ChatGPTなどの大規模言語モデルに対する禁止は一切検討せず、利用についてのガイドラインを提供する方針を示しています。これは、AIが学生に与える影響についての認識を示すもので、教育者が学生に何を教えるべきかを再考することが重要であり、そのためにも教職者自身がChatGPTに出来ることとそのリスクについて理解を深めることが重要であると強調しています。

AIの利用が広がる一方で、その法的・倫理的な問題についての理解や対策が十分になされていない現状があります。この課題を解決するためには、AIの開発者や利用者が連携し、AIの利用に関する法的・倫理的なガイドラインの策定と遵守が求められます。また、AIの利用者に対する教育も重要な課題となっており、AIの可能性とリスクを理解し、適切な利用を行うための知識を身につけることが必要とされています。これらの取り組みを通じて、AIの持つ課題を克服し、その利便性を最大限に活用することが求められています。

AI技術は完璧なものではないため、利用者である学生の入力したプロンプトが思わぬリスクを顕在化させる可能性があります。イェール大学の副学長のコメントの通り、一方的に禁止するだけではAIを正しく使いこなすための理解が進みません。

> AI法研究会

この文書は、AI法研究会におけるChatGPTと生成AIの法的倫理的課題に関するパネルディスカッションでの松本の発言メモです。誤情報・幻覚、悪用・誤用、バイアス、教育分野の課題について議論され、ChatGPTの回答の信頼性についての問題が指摘されました。...

> 人材選考AIを例に「AIと公平性」の問題を考える、学習データに潜んでいたリスクとは

この記事は、人事選考AIを例に、AIの公平性に関する問題を考察しています。AIが不確実性を含むため、差別的な判断を受ける可能性があるという懸念があります。政府や企業は、AIのポリシーやガイドラインを策定していますが、ユーザー企業の多くでは、具体的に何をすればよいか不明瞭であるという課題があります。...

AIの開発者・サービス提供者側にはAIサービスの信頼性を確保する取り組みを含めた情報を開示し、教育現場でAIを活用する上では、利用者である教師と学生も継続的に対話しながらAIのメリットとリスクについて理解を深めることが必要になります。

→ 当初考察には、関係のない文章（人材育成AIに係る論点）が含まれていたのを削除し、一部補足を行った。

SmartSEで学べたことの活用例



カスタマージャーニーマップを参考に現状課題を整理・分析

機械的・網羅的に見て
重要な点を外さなければOK

自分の価値観が
強く入る(重点)

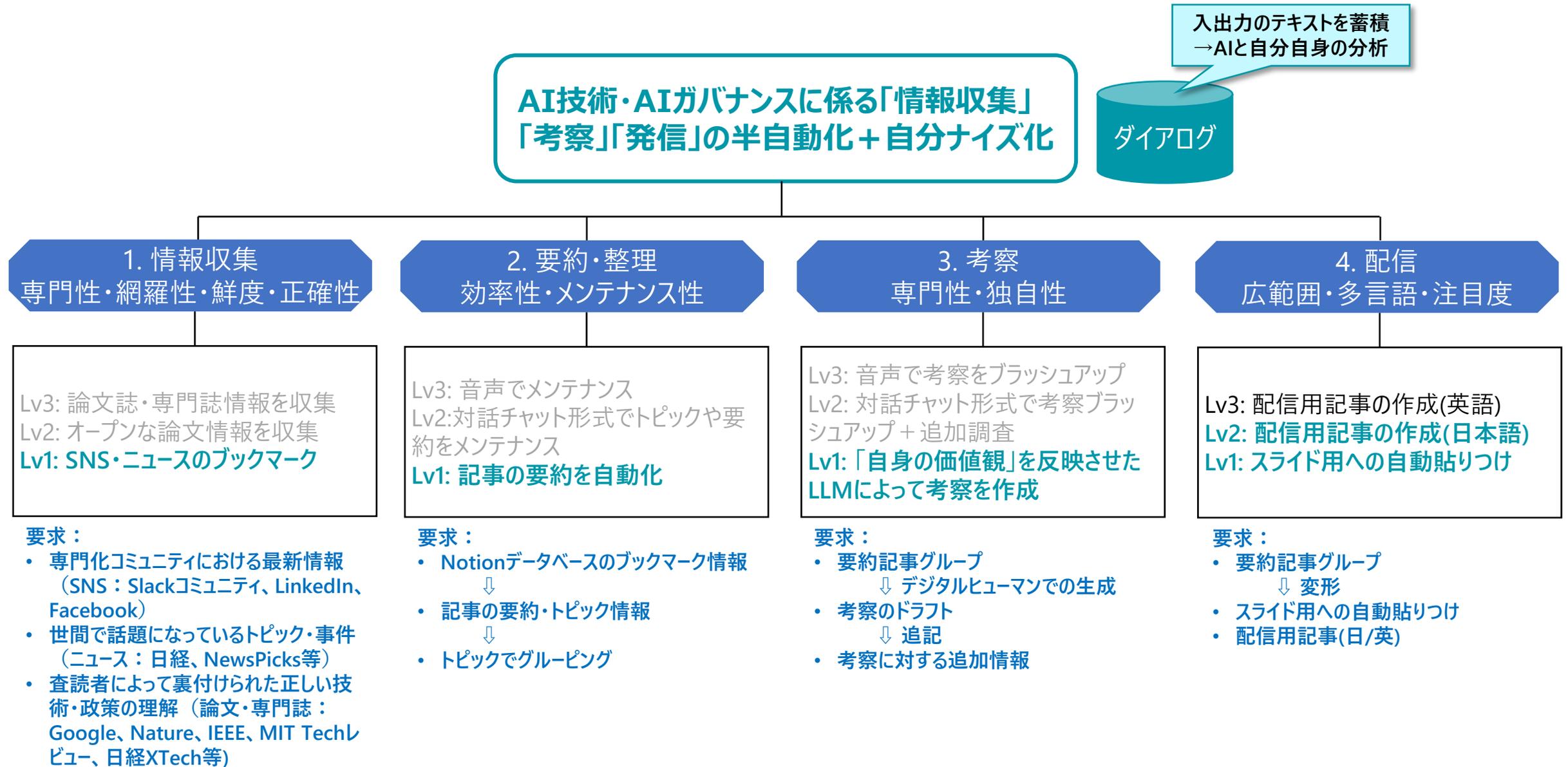
フェーズ	1. 情報収集	2. 要約・整理	3. 考察・コメント	4. 発信
Action 行動	①情報をキャッチ ②Notionにブックマーク(技術情報/活用事例/ガバナンス)	①トピックごとに記事をグループ ②スライド(PPT)に列挙(見出し+URL)	①トピックごとの情報を見ながら、疑問点を追加で調べる ②専門領域からコメント追加	①社内外に話題提供
Opponent 接点	<ul style="list-style-type: none"> SNS：Slackコミュニティ、LinkedIn、Facebook ニュース：日経、NewsPicks等 論文：Google、Nature、IEEE 	<ul style="list-style-type: none"> Notionのデータベース 	<ul style="list-style-type: none"> PPTに追記 	<ul style="list-style-type: none"> 社内：TeamsでPPTを共有 社外：クライアント会議
Thinking 思考	<ul style="list-style-type: none"> タイムリーに情報を取得したい 重要なテクノロジー及び政策・ガバナンスの論点は正しく理解できる情報を押さえない 	<ul style="list-style-type: none"> クイックに要約を抑えて、精読すべき記事を捉えたい 様々なテーマで過去記事とのつながりも含めて情報整理したい 	<ul style="list-style-type: none"> 考察の過程で関連情報と紐づけることで理解が深まる 日本固有のトピック/国際動向を同じレベルで検討したい 	<ul style="list-style-type: none"> 最新記事を過去記事と関連づけて話すとトピックのコンテキストが伝わりやすい 二次共有してもらっている
Feeling 感情	<ul style="list-style-type: none"> SNSだけで情報は基本充実+Notionへのブックマークは楽 論文を追っていないことが不安(正確な理解なのだろうか?) 	<ul style="list-style-type: none"> 多過ぎ、早過ぎ、難し過ぎ 記事の分類が現状煩雑 読み込むために集中する時間が取れない 	<ul style="list-style-type: none"> ココが付加価値になる 考察を行う際の裏付け調査に時間が掛かる 	<ul style="list-style-type: none"> 一定の価値は出せている 本当は海外含めて広く発信したい(日本人専門家の考察)
Insight 気づき	<ul style="list-style-type: none"> SNS・ニュース情報の取得については現状と大きく変えない 	<ul style="list-style-type: none"> 記事要約の自動化 情報の利活用が出来そう 	<ul style="list-style-type: none"> 考察を自動ドラフトして欲しい AIとインタラクティブに検討したい(気づきたい) 	<ul style="list-style-type: none"> 要約リストを自動作成 考察記事を日/英で発信



ダイアログ

自分自身の価値観の変化を継続的に分析

GDTAで目的に対するサブゴールと要求を整理



アーキテクチャ (デジタルツイン)

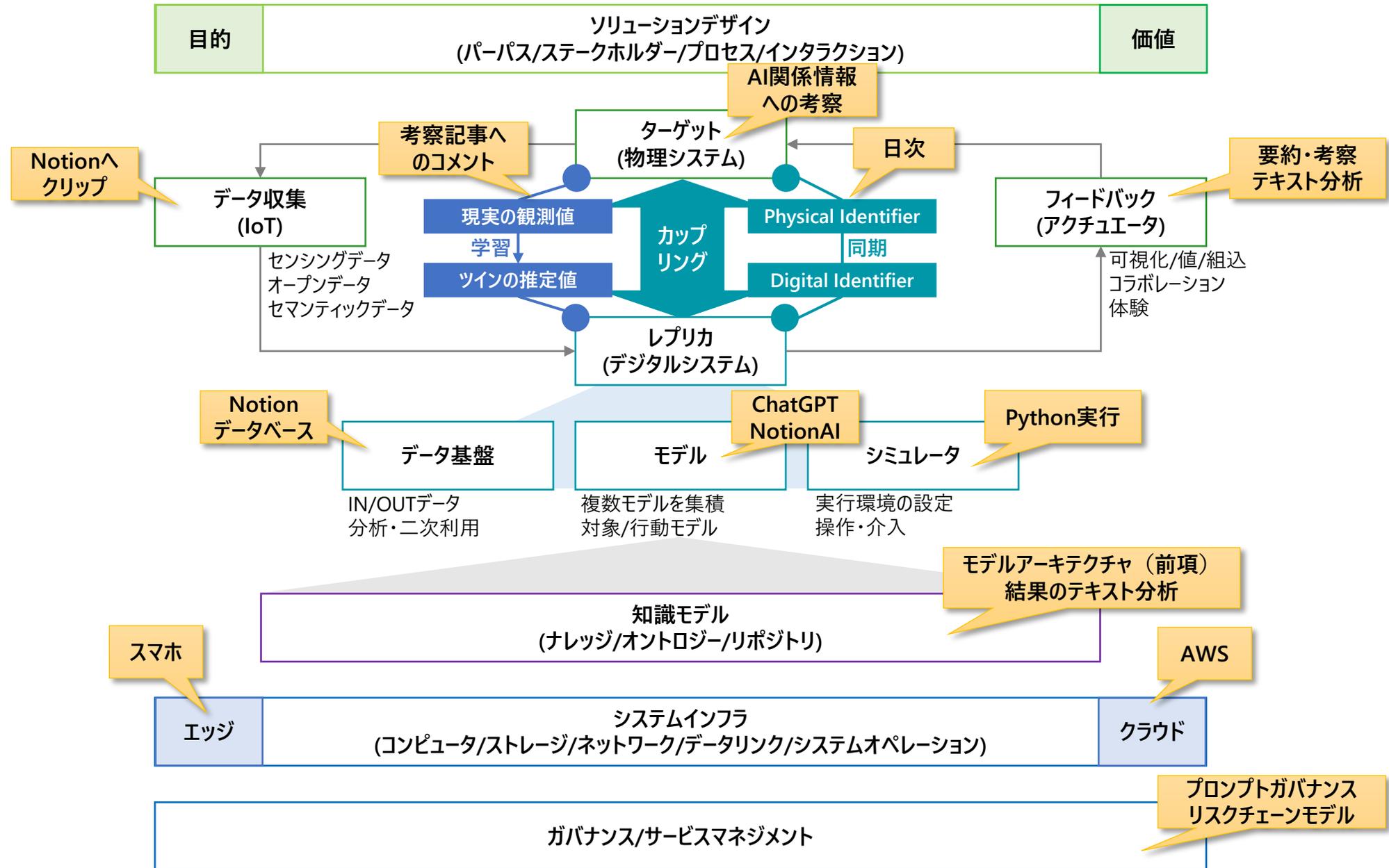
ビジネス
環境

リアル
環境

デジタル
環境

知識
環境

システム
環境

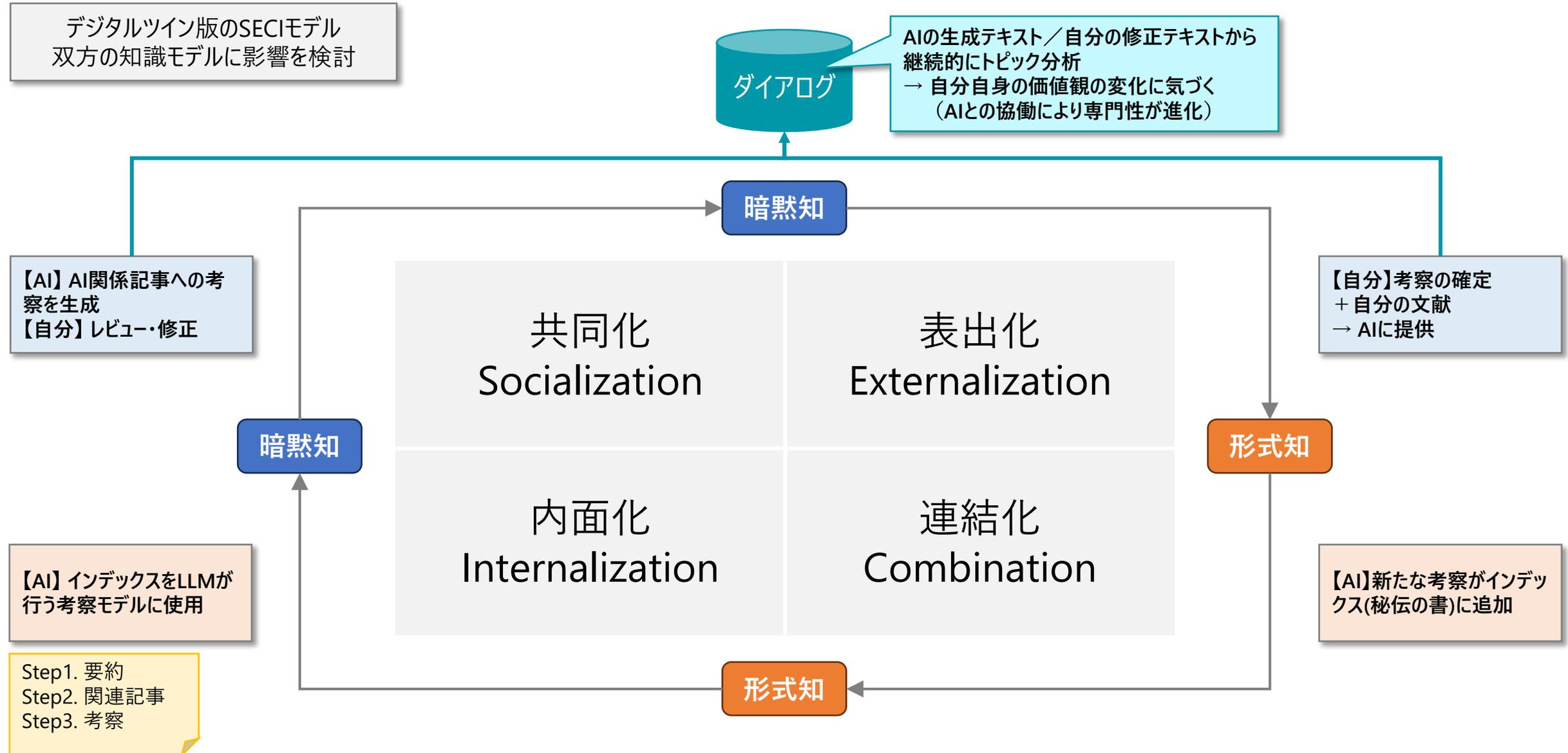


デジタルMATSUMOTOの今後の展開



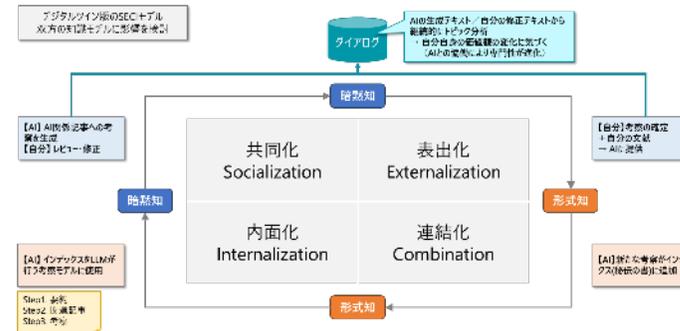
SECIモデル（自分の価値観 ⇔ ChatGPTの知識情報）

新たな考察をインデックスに継続追加して、自分自身の暗黙知の変化を分析する

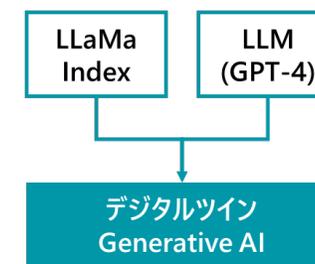


今後の展開

- デジタルツインとしての研究：国際学会に向けて論文化
 - 専門家自身の知の変化に、AIが対応できるか？
 - 専門家自身の知が、AIとの対話によって変化するか？



- モデルの汎用化：Personalized AIとしての活用促進
 - 汎用目的へのアプローチ：対話型で他のタスクに適用



- Personalized AI利活用に係るガバナンスの検討
 - 実用化に向けたガバナンス：リスク検討とステークホルダーとの合意形成

The collage includes several documents related to AI governance:

- Risk Assessment:** "Distribution of significant risks associated with AI".
- Control Coordination:** "Coordinate the chain (position of the holder) for each important risk scenario".
- Case Study:** "These are fictional case studies below and don't raise issues or assure for any company or AI service."
 - Case01: Recruitment AI (2021.07.02)
 - Case02: Convenience stores (2021.09.01)
 - Case03: Line Inspection AI (2021.09.01)
 - Case04: Detection AI (2021.09.01)
- Risk Chain Model (Review):** A diagram showing a chain of risks and their mitigation.

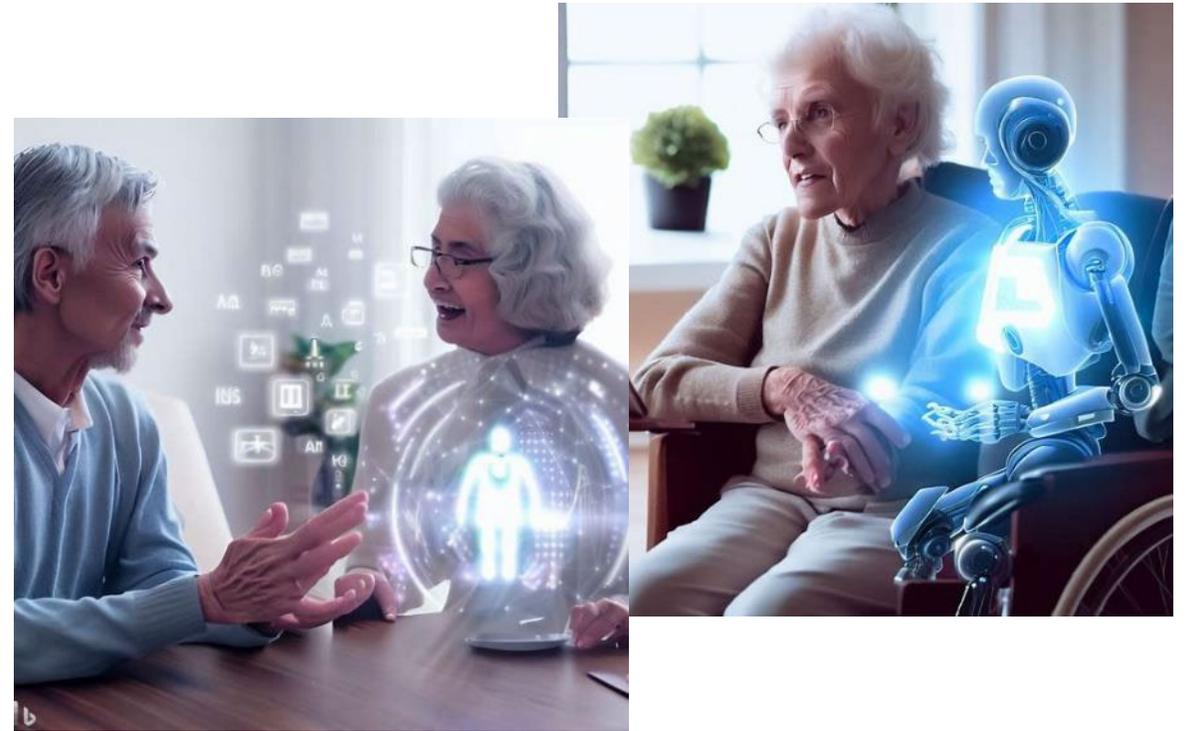
これからの「AIとヒトの付き合い方」

AIを身に着けて自らの能力を拡張させるヒト
Human with AI installed



Microsoft Bing Image Creator :
AIを武装してエネルギーに活躍するビジネスマンを書いてください。
背景は街中で、スピード感とエネルギーにあふれる姿を見せてください。

AIと触れ合って自らの能力を向上させるヒト
Human with AI empowered



Microsoft Bing Image Creator :
AIによって自らの能力を成長させる老人を書いてください。