



# 生成AIのもたらす社会インパクトと取るべき戦略

栄藤 稔  
大阪大学 先導的学際研究機構 教授  
科学技術振興機構（JST）CREST人工知能領域総括  
順天堂大学医学研究科客員教授

7/31/2023  
@mickbean  
<https://www.linkedin.com/in/micketoh/>

A white rectangular card with a thin gray border. At the top is a circular profile picture of a man with dark hair. Below the picture, the name "Mick Etoh" is printed in bold black font. Underneath the name, the text "Professor, CEO, Program Director at Japan Science & Technology Agency" is printed in a smaller gray font. At the bottom right of the card is a large blue QR code.

# A Survey of Large Language Models

Wayne Xin Zhao, Kun Zhou\*, Junyi Li\*, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong, Yifan Du, Chen Yang, Yushuo Chen, Zhipeng Chen, Jinhao Jiang, Ruiyang Ren, Yifan Li, Xinyu Tang, Zikang Liu, Peiyu Liu, Jian-Yun Nie and Ji-Rong Wen

**Abstract**—Ever since the Turing Test was proposed in the 1950s, humans have explored the mastering of language intelligence by machine. Language is essentially a complex, intricate system of human expressions governed by grammatical rules. It poses a significant challenge to develop capable artificial intelligence (AI) algorithms for comprehending and grasping a language. As a major approach, *language modeling* has been widely studied for language understanding and generation in the past two decades, evolving from statistical language models to neural language models. Recently, pre-trained language models (PLMs) have been proposed by pre-training Transformer models over large-scale corpora, showing strong capabilities in solving various natural language processing (NLP) tasks. Since the researchers have found that model scaling can lead to an improved model capacity, they further investigate the scaling effect by increasing the parameter scale to an even larger size. Interestingly, when the parameter scale exceeds a certain level, these enlarged language models not only achieve a significant performance improvement, but also exhibit some special abilities (e.g., in-context learning) that are not present in small-scale language models (e.g., BERT). To discriminate the language models in different parameter scales, the research community has coined the term *large language models (LLM)* for the PLMs of significant size (e.g., containing tens or hundreds of billions of parameters). Recently, the research on LLMs has been largely advanced by both academia and industry, and a remarkable progress is the launch of ChatGPT (a powerful AI chatbot developed based on LLMs), which has attracted widespread attention from society. The technical evolution of LLMs has been making an important impact on the entire AI community, which would revolutionize the way how we develop and use AI algorithms. Considering this rapid technical progress, in this survey, we review the recent advances of LLMs by introducing the background, key findings, and mainstream techniques. In particular, we focus on four major aspects of LLMs, namely pre-training, adaptation tuning, utilization, and capacity evaluation. Besides, we also summarize the available resources for developing LLMs and discuss the remaining issues for future directions. This survey provides an up-to-date review of the literature on LLMs, which can be a useful resource for both researchers and engineers.



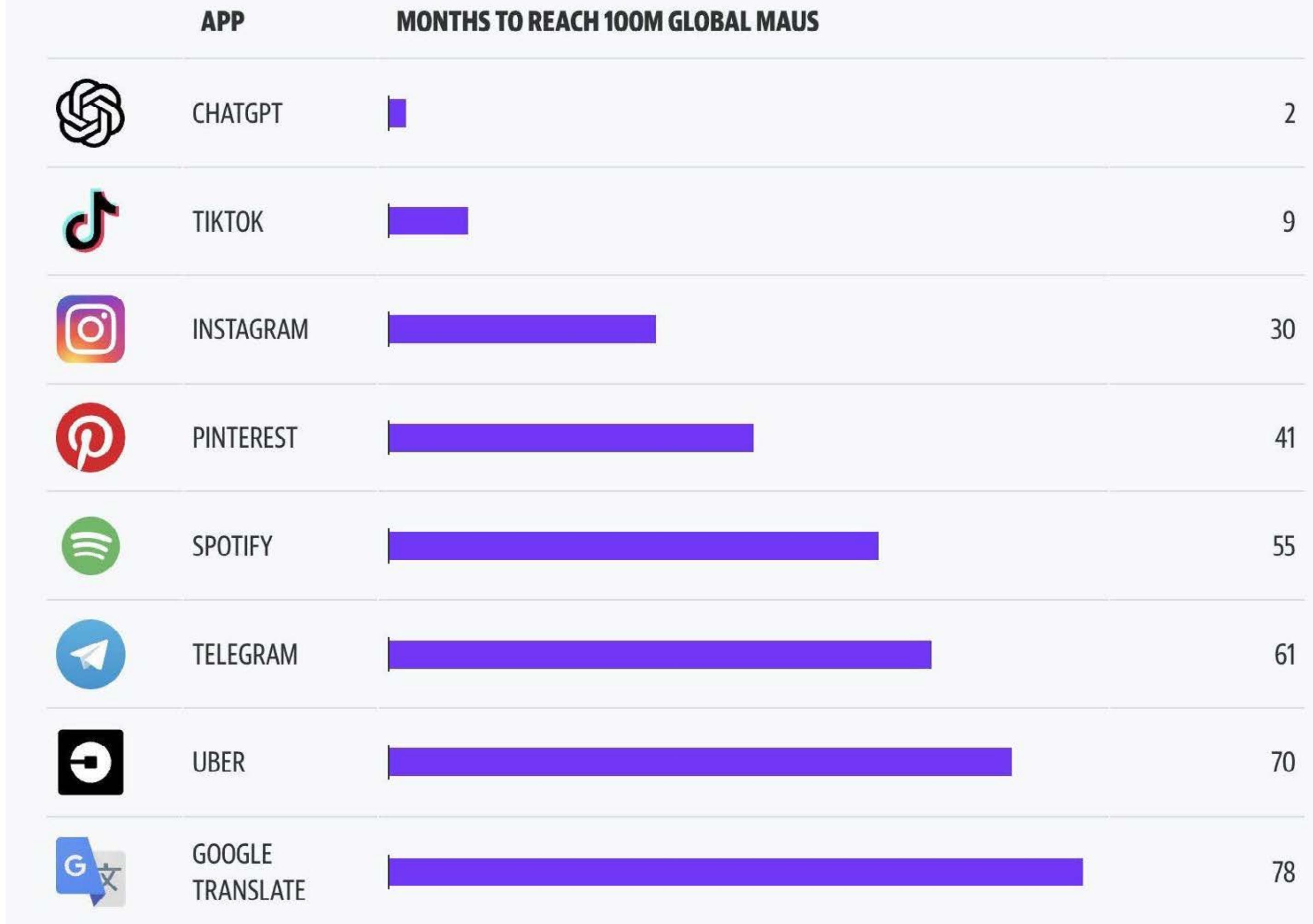
スケーリング則の上で  
性能がある閾値を超えた時に  
キラーアプリケーションが登場

---

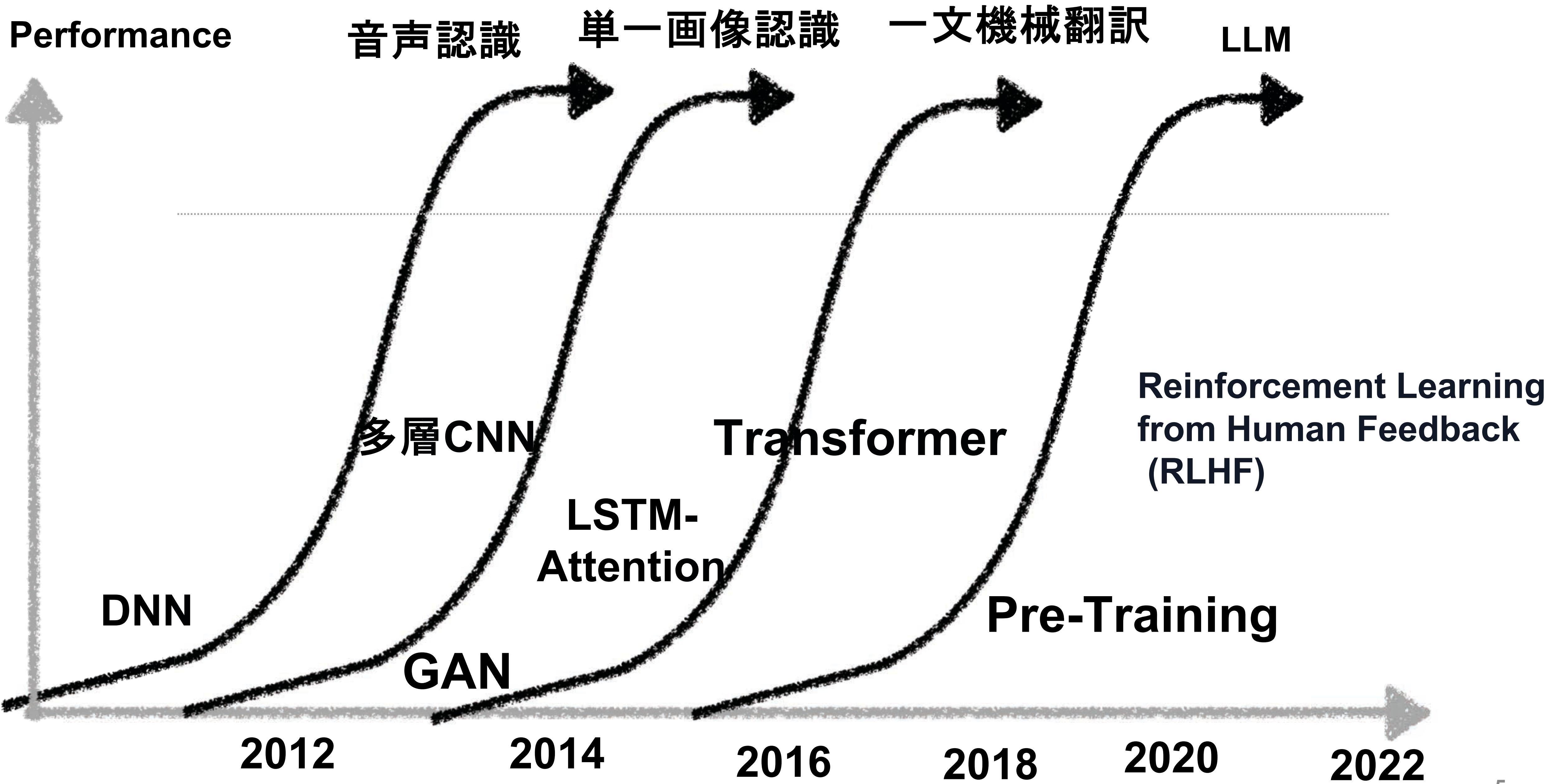
1億ユーザー  
獲得までの  
月数

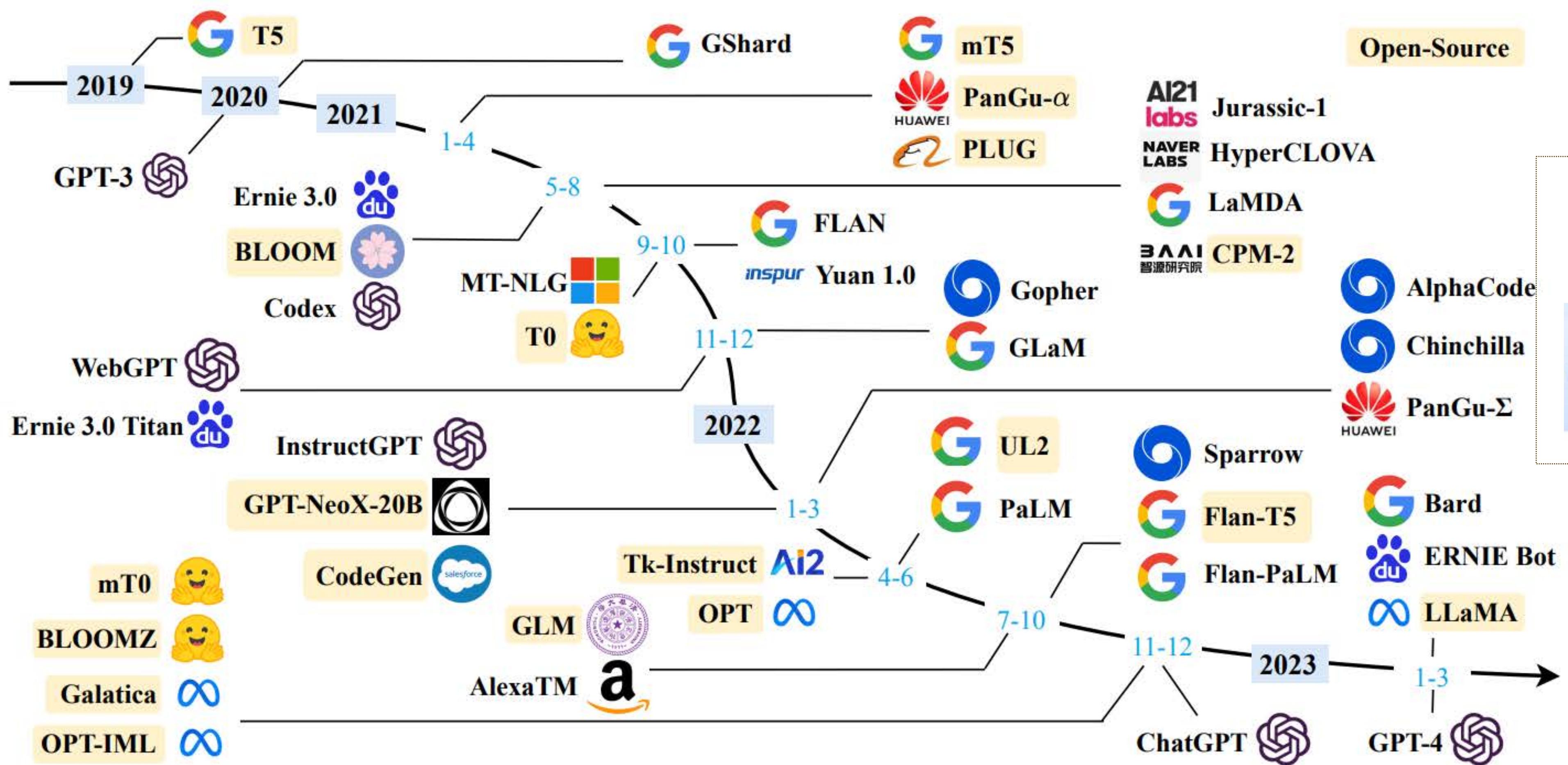
## HOW LONG IT TOOK TOP APPS TO HIT 100M MONTHLY USERS

ChatGPT is estimated to have hit 100M users in January, 2 months after its launch.  
Here's how long it took other top apps to reach that:



SOURCE: UBS



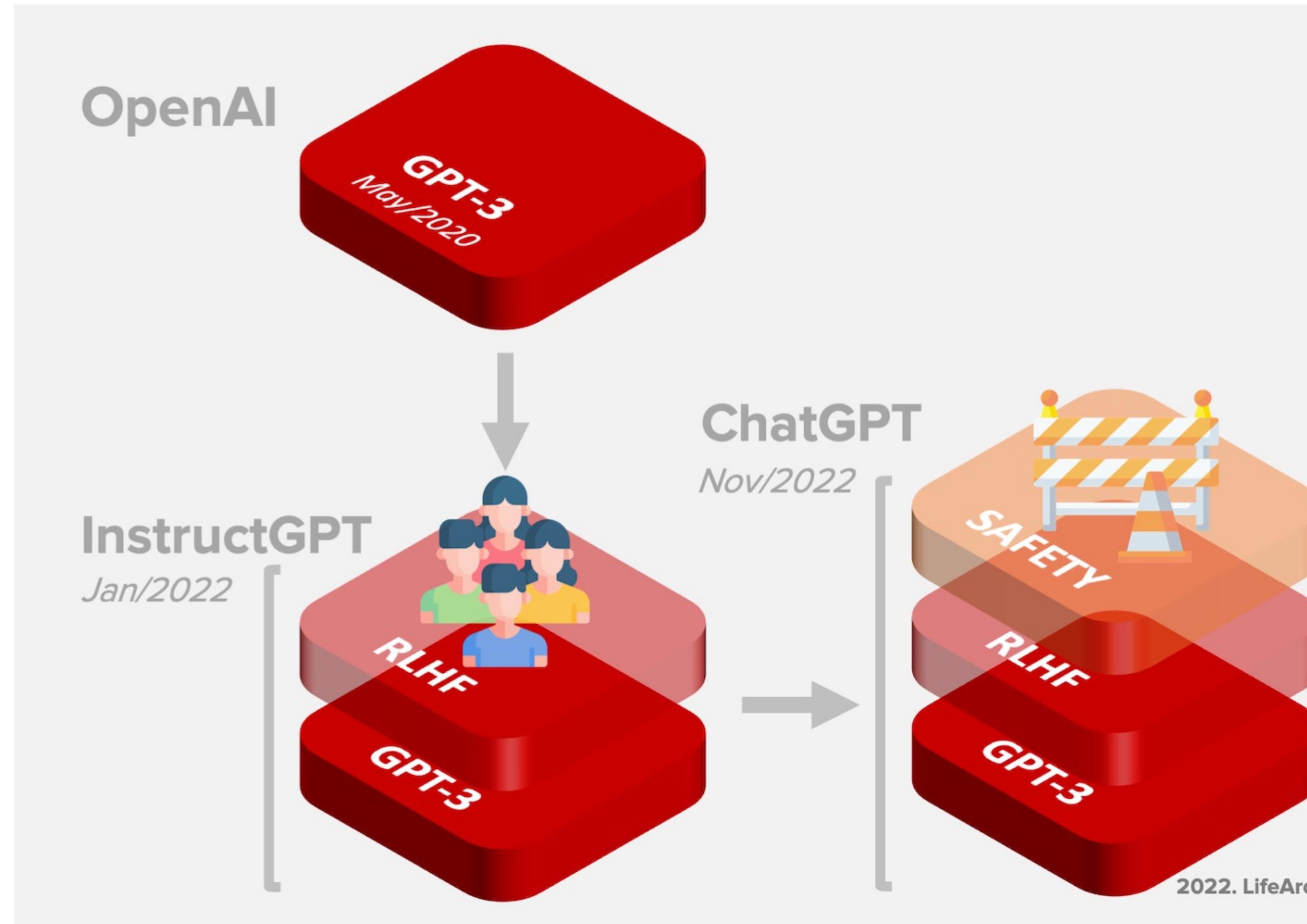


文心一言

SparkDesk 讯飞星火认知大模型

中国では  
国家管理

Fig. 1. A timeline of existing large language models (having a size larger than 10B) in recent years. We mark the open-source LLMs in yellow color.



Source: <https://lifearchitect.ai/chatgpt/>

# 3種のチューニング

	What	How	When
<b>Fine-tuning</b> 提供者主導	Entails taking a pre-trained language model and further training it on a specific & smaller dataset that is specific to the task at hand. This is typically done by updating the weights of the model's last layer or layers while leaving the rest of the model static.	During fine-tuning, a pre-trained model is loaded into memory and its weights are frozen. A smaller dataset relevant to the task at hand is loaded, and the pre-trained model is adjusted by tuning its weights. The model is typically trained for several epochs until the desired level of accuracy is reached.	The fine-tuning process is normally used when the task or domain is well-defined, and there is sufficient labeled data available to train on. If you have a large dataset and a specific task in mind, fine-tuning a language model is likely to be the most effective approach.
<b>Prompt Engineering</b> ユーザー主導	Involves designing natural language prompts or instructions that can guide a language model to perform a specific task.	Select & arrange the words in a prompt or query to elicit a specific response from the model. Top-notch prompt engineers conduct experiments, systematically record their findings, and refine their prompts to identify essential components.	Best suited for tasks requiring a high level of precision and well-defined outputs. Prompt engineering can be used to craft a query that elicits a desired output. In some cases, prompt engineering can be used to improve the performance of a fine-tuned model by providing more guidance to the model during inference.
<b>RLHF</b> Reinforcement Learning from Human Feedback 提供者主導	Reinforcement Learning from Human Feedback (RLHF) involves training a model by receiving feedback from human evaluators.	The model is first trained on a dataset to establish a baseline level of performance. The model then generates a response to a prompt or query that is evaluated by a human. Feedback from the human evaluator is utilized to update the model's weights so that it can generate more accurate responses in the future.	RLHF is ideally suited when the task requires a high level of accuracy and the model needs to be trained on a wide variety of inputs. RLHF is particularly useful when there is very limited data that can be used to train the model, since the model can be trained on a wide range of inputs through human feedback.

ChatGPT、GoogleのBard、AnthropicのClaude、DeepMindのSparrowなどの汎用モデルの微調整に広く使用



開発コスト  
昨年のロス \$540 M in 2022

運用コスト  
Running Cost \$70,0000/Day  
\$0.01/response

期間: instructGPT 3年

キーコンポーネント (Key Component)	主要な活動 (Key Activity)	与える価値 (Value Proposition)	顧客との関係構築方法 (Customer Relation)	顧客 (Customer Segment)
Microsoft	LLM Development キーリソース (Key Resource) 655 Employee total funding \$11B	<b>Generative AI per se</b>	SaaS License チャネル (Channel) <b>Direct via Web Microsoft</b>	Enterprise DX
	コスト (Cost Structure) <b>OPEX: \$255M CAPEX: \$400M?</b>	収入 (Revenue Stream)  \$88.7M per year.(2022)		

Source: <https://www.businessinsider.com/openai-2022-losses-hit-540-million-as-chatgpt-costs-soared-2023-5>

<https://growjo.com/company/OpenAI>

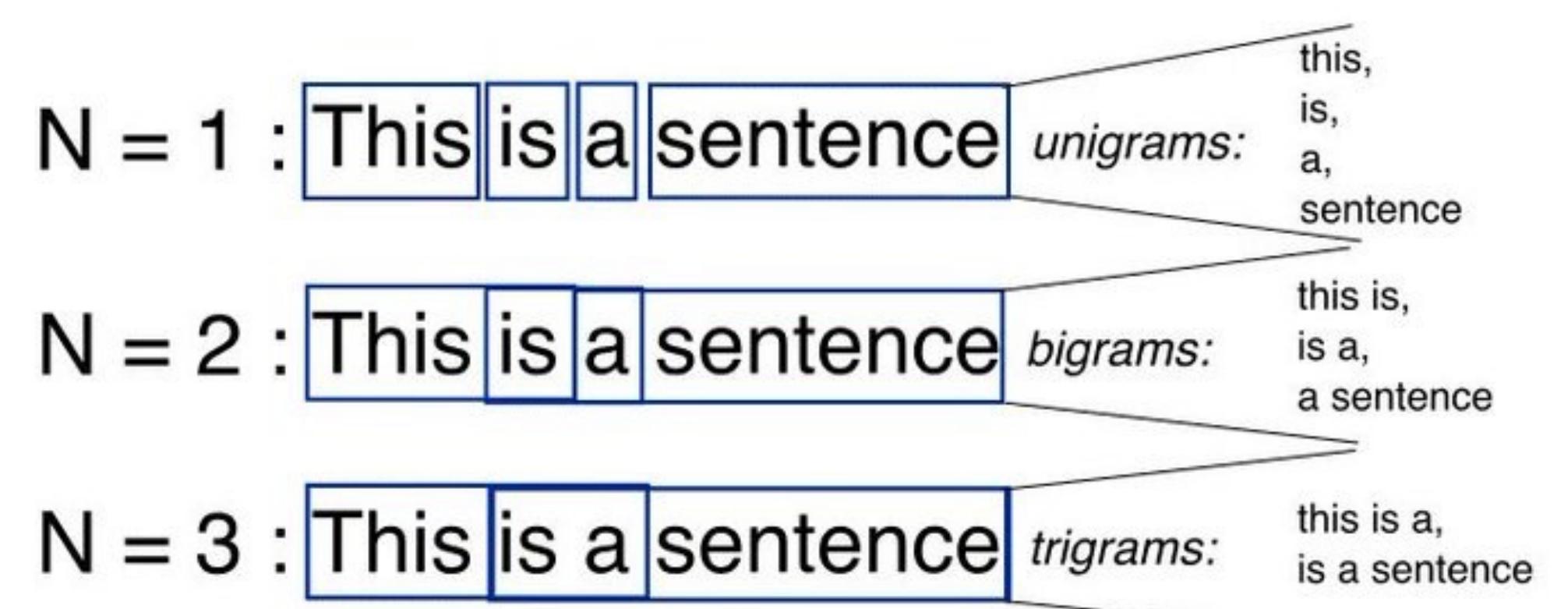
日本でLLMをスクラッチから設計できる人材は50人もいないだろう。

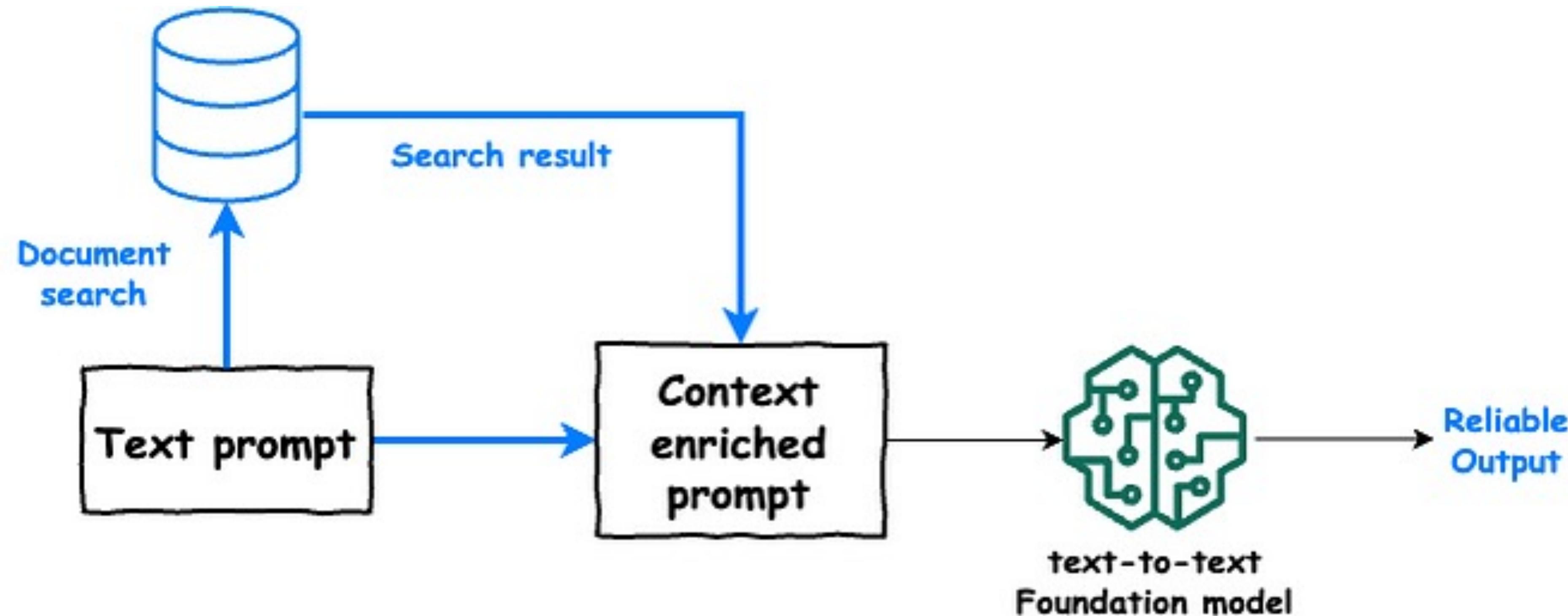


# 大規模言語モデルの2つの解釈

言語生成器

多言語の応答型知識ベース





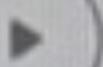
# Retrieval Augmented Generation (RAG)

- ユーザが質問文を入力する
- 質問文をそのままクエリとしてベクトルに変換する
- データベースを検索し、内容が近そうな文章を取り出す
- 取り出した文章と質問を合わせて、LLMに入力し回答を得る

バス停に着いたら園児がいませんどうしたらいいですか

A

園に連絡し、保護者に連絡して園児の安全を確認する。必要に応じて、次の場所や園に向かう。



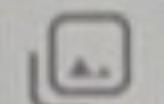
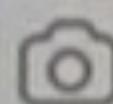
ME

今の説明を英語でお願いします

A

AIアシスタント 8:23

If there are no children at the bus stop, contact the school and parents to ensure the safety of the children. If necessary, proceed to the next location or the school.



ここにメッセージを…





# MICIN、金沢大学と実施していたChatGPTおよびGPT-4を用いて第117回医師国家試験(2023年2月実施)を解かせる研究において初めて合格点に到達し、その成果を論文としてオンライン公開

<https://www.medrxiv.org/content/10.1101/2023.04.17.23288603v1>

## Promptの一部:

以下の選択肢から、低補体血症を示す二つの疾患は何ですか？選択肢にはセルライト、細菌性肺炎、虚血性大腸炎、急性糸球体腎炎、混合性クリオグロブリン血症が含まれます。回答には(2つ)の選択が必要です。

回答は(急性糸球体腎炎)と(混合性クリオグロブリン血症)です。これらの二つの疾患は低補体血症と関連していることが知られています。

急性糸球体腎炎は免疫介在性の腎臓疾患であり、<略>

## 質問:

意識障害のある患者に対して行いにくい身体検査は何ですか？

腱反射 | 頸部の硬直 | 光反射 | 指鼻検査 | 膝の持ち上げテスト  
(下肢ドロップテスト)

示唆

マルチモーダル生成AI と  
演繹的な推論への展開  
が研究開発の大テーマ

英語の言語資産 + 日本語の言語  
資産の重要性

画像・グラフを含む設問は除外

ChatGPTおよびGPT-4が誤答を生成する3大要因  
「医学知識の不足」  
「日本特有の医療制度に関する情報」  
「計算問題での誤り」

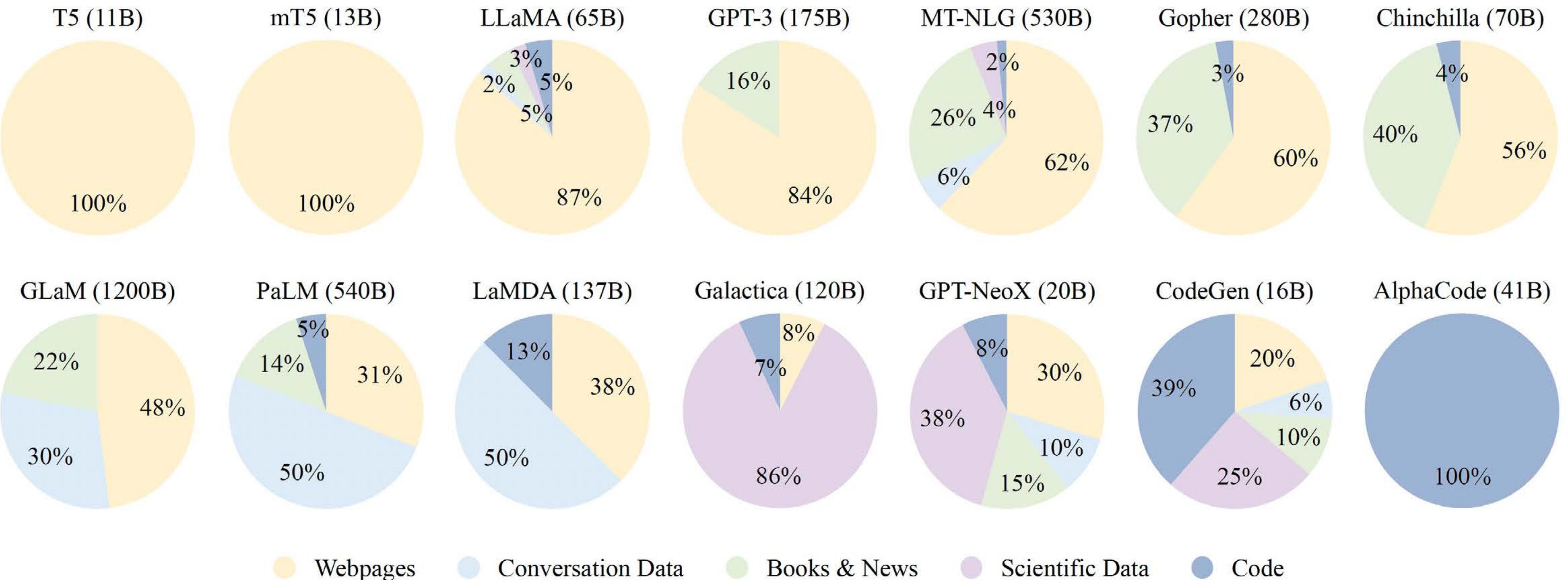
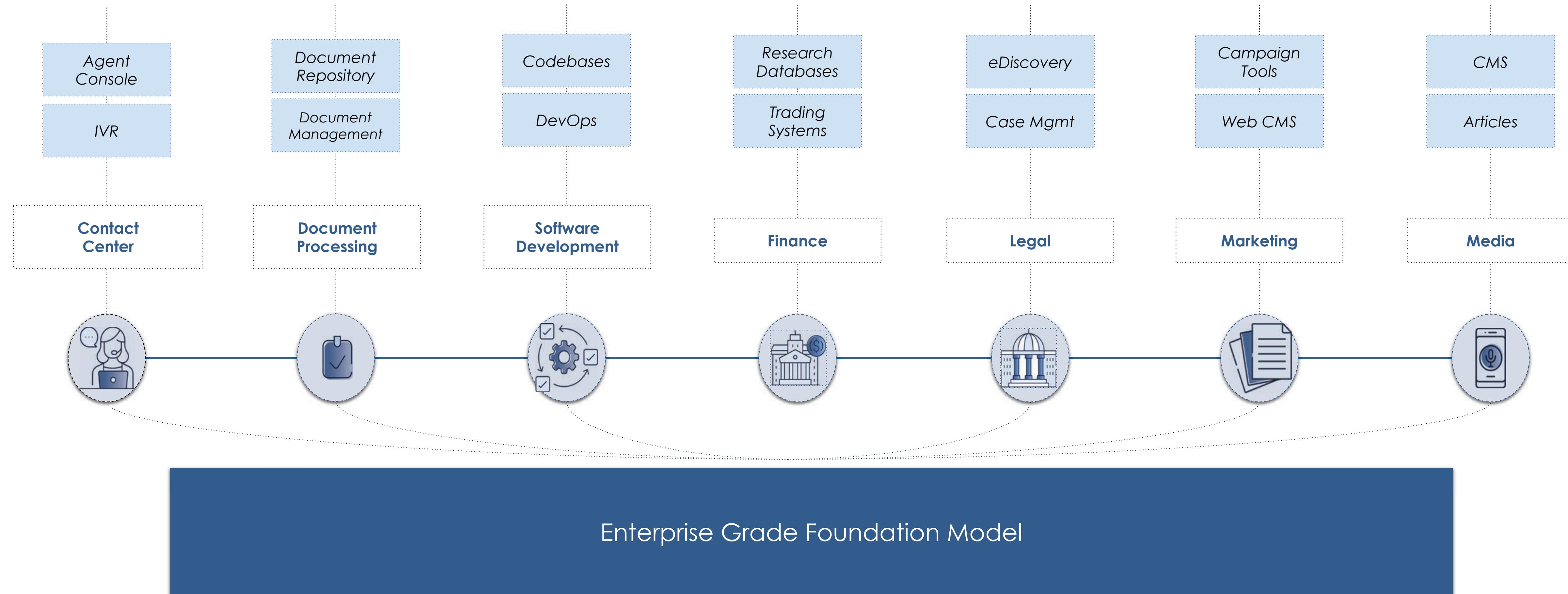


Fig. 2. Ratios of various data sources in the pre-training data for existing LLMs.

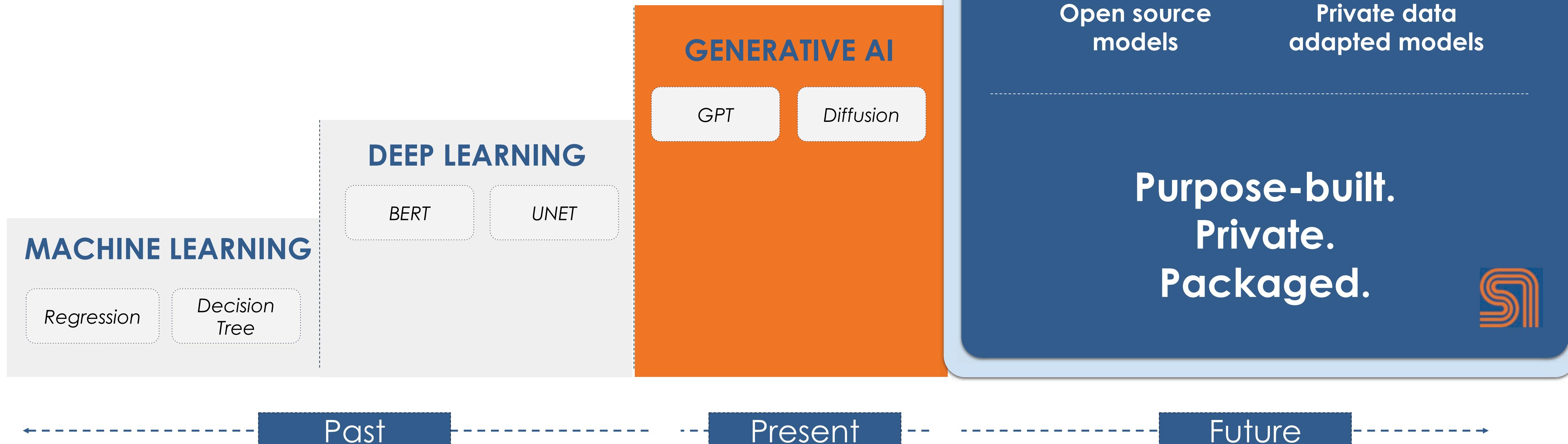
Zhao, Wayne Xin, et al. "A survey of large language models." *arXiv preprint arXiv:2303.18223* (2023).

Pace of Commerce will Increase by 10x

AI will touch every function of every company in every industry

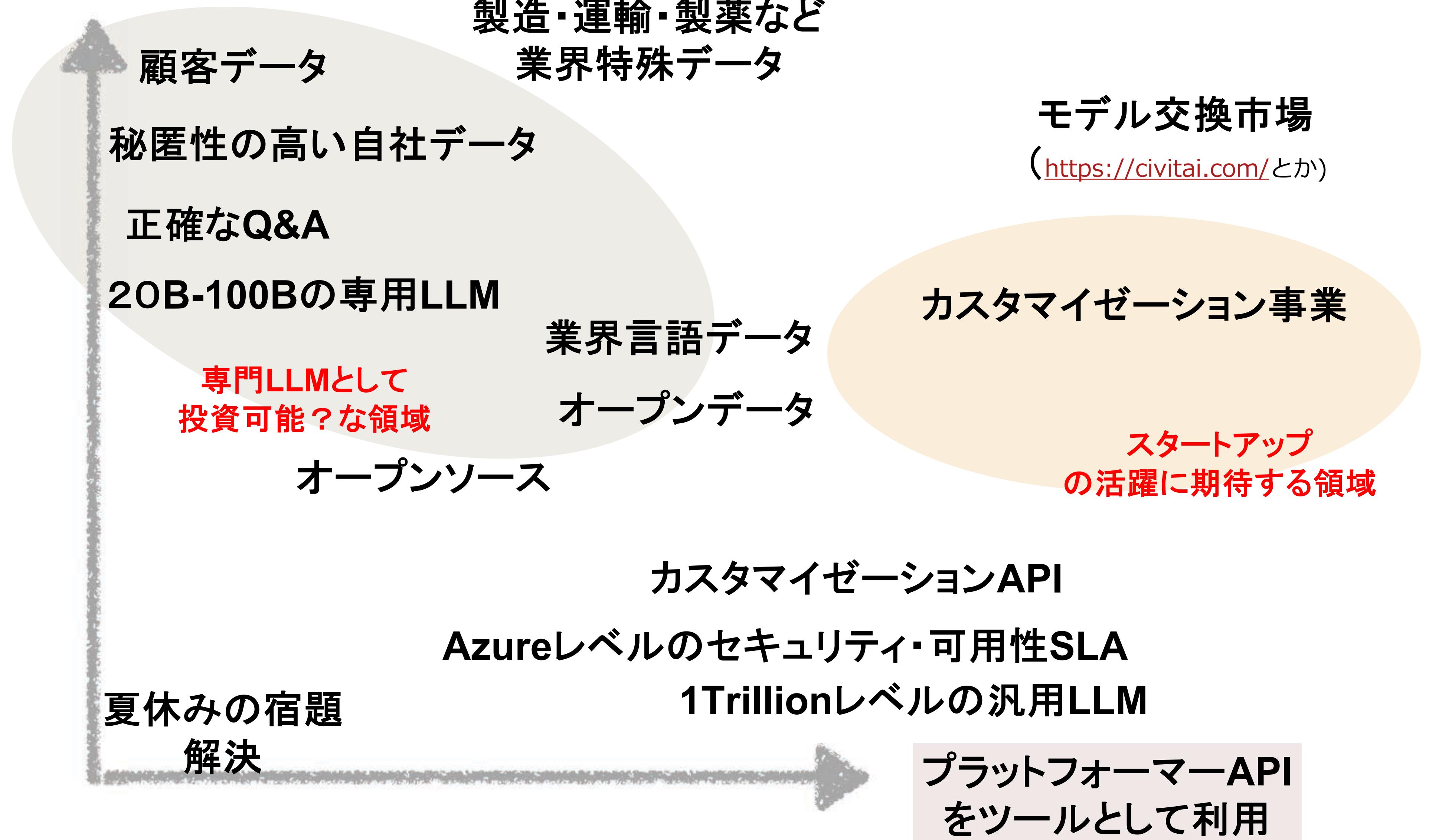


**AI is a journey:**  
Invest in a foundation for the future

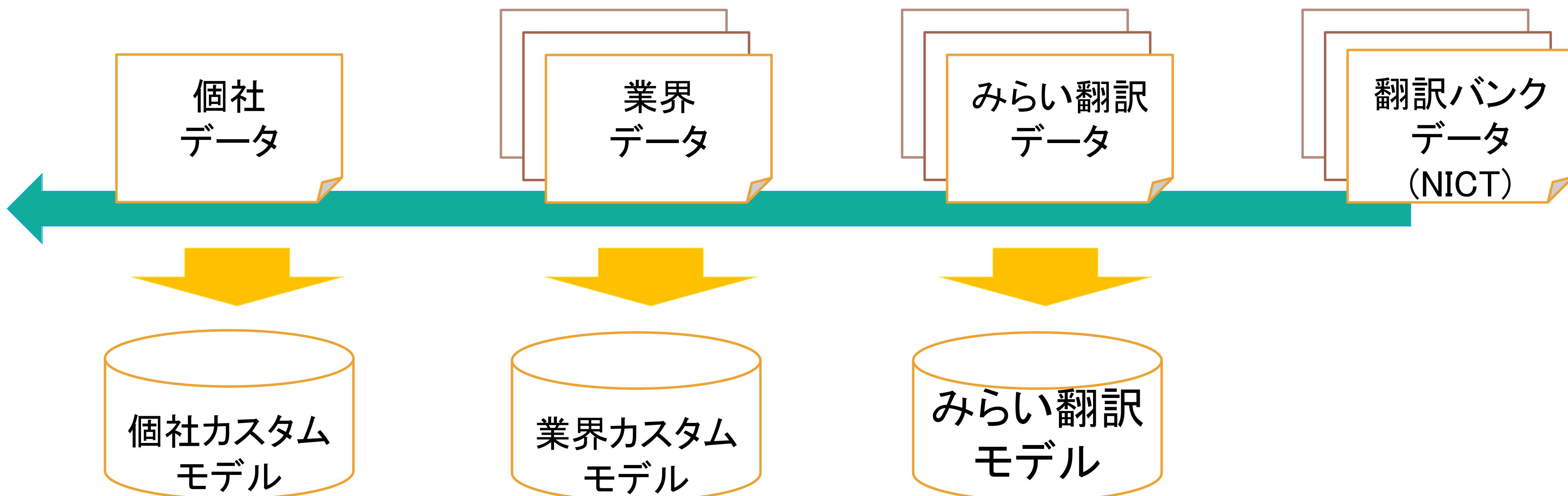




## 自社アセットとして利用

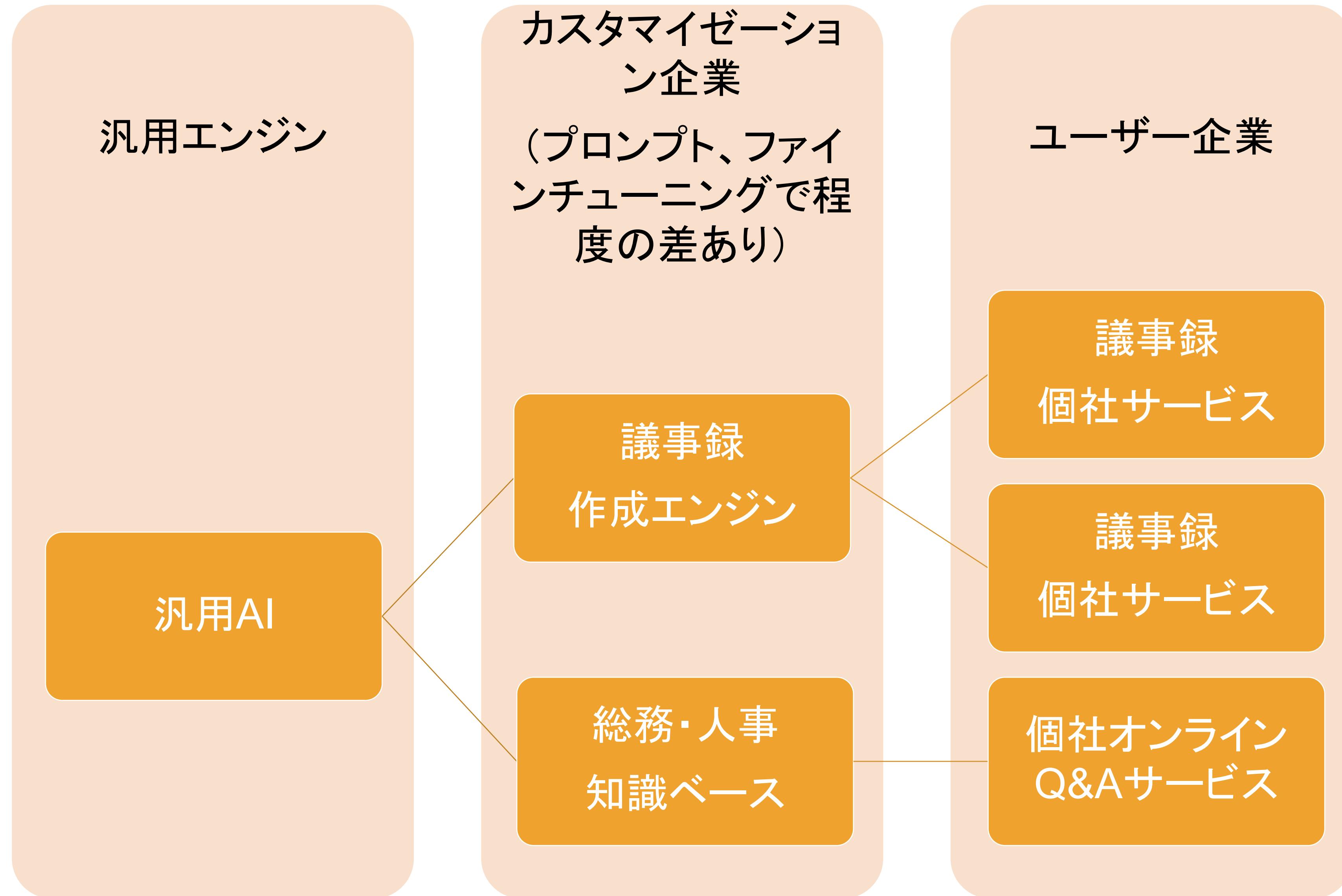


# 自社アセット化：みらい翻訳の事例

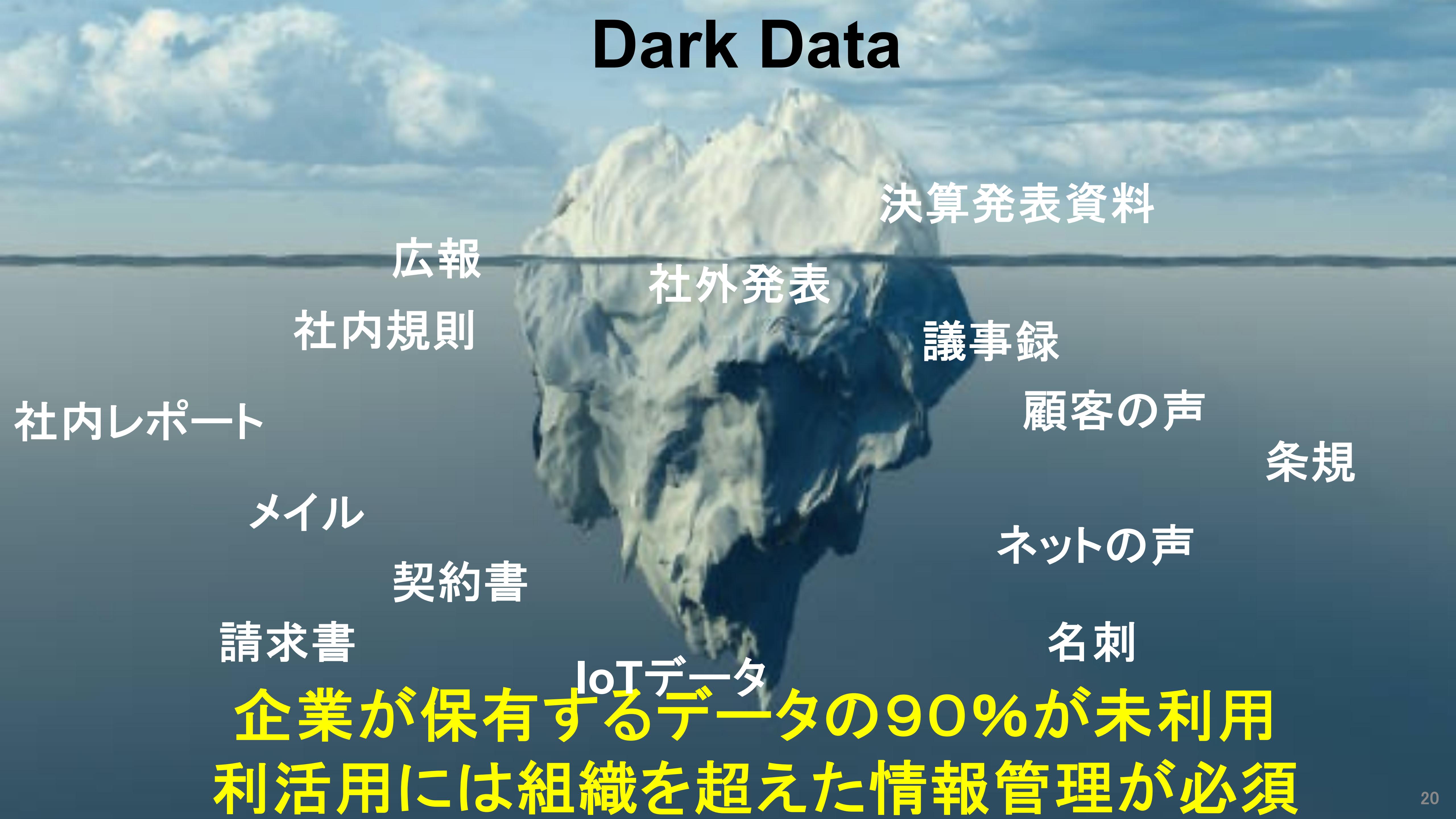




# プラットフォーマーAPIをツールとして利用



# Dark Data



広報  
社内規則  
社内レポート  
メイル  
契約書  
請求書

決算発表資料  
社外発表  
議事録  
顧客の声  
ネットの声  
名刺

IoTデータ  
企業が保有するデータの90%が未利用  
利活用には組織を超えた情報管理が必須

- ・価値創造は年間数兆ドル

- ・創造的AIの利用事例がもたらす価値の約75%は、顧客運用、マーケティングと販売、ソフトウェアエンジニアリング、およびR&Dの4つの領域に集中

- ・例えば、顧客との対話をサポートし、マーケティングと販売のための創造的なコンテンツを生成し、自然言語のプロンプトに基づいてコンピュータコードを作成するなどのタスク

- ・銀行、ハイテク、ライフサイエンスなどの業界にインパクト

- ・創造的AIは、他の種類の作業よりも高い賃金と教育要件を持つ職業に関連する知識労働により大きな影響

McKinsey  
& Company

# The economic potential of generative AI

The next productivity frontier

June 2023





**Generative AI could deliver significant value when deployed in some use cases across a selection of top industries.**

Selected examples of key use cases for main functional value drivers (nonexhaustive)

	Total value potential per industry, \$ billion (% of industry revenue)	Value potential, as % of operating profits <sup>1</sup>	Product R&D, software engineering	Customer operations	Marketing and sales	Other functions
<b>Banking</b>	200–340 (3–5%)	9–15	<ul style="list-style-type: none"> <li>■ Legacy code conversion</li> <li>Optimize migration of legacy frameworks with natural-language translation capabilities</li> </ul>	<ul style="list-style-type: none"> <li>■ Customer emergency interactive voice response (IVR)</li> <li>Partially automate, accelerate, and enhance resolution rate of customer emergencies through generative AI-enhanced IVR interactions (eg, for credit card losses)</li> </ul>	<ul style="list-style-type: none"> <li>■ Custom retail banking offers</li> <li>Push personalized marketing and sales content tailored for each client of the bank based on profile and history (eg, personalized nudges), and generate alternatives for A/B testing</li> </ul>	<ul style="list-style-type: none"> <li>■ Risk model documentation</li> <li>Create model documentation, and scan for missing documentation and relevant regulatory updates</li> </ul>
<b>Retail and consumer packaged goods<sup>2</sup></b>	400–660 (1–2%)	27–44	<ul style="list-style-type: none"> <li>■ Consumer research</li> <li>Accelerate consumer research by testing scenarios, and enhance customer targeting by creating "synthetic customers" to practice with</li> </ul>	<ul style="list-style-type: none"> <li>■ Augmented reality-assisted customer support</li> <li>Rapidly inform the workforce in real time about the status of products and consumer preferences</li> </ul>	<ul style="list-style-type: none"> <li>■ Assist copy writing for marketing content creation</li> <li>Accelerate writing of copy for marketing content and advertising scripts</li> </ul>	<ul style="list-style-type: none"> <li>■ Procurement suppliers process enhancement</li> <li>Draft playbooks for negotiating with suppliers</li> </ul>
<b>Pharma and medical products</b>	60–110 (3–5%)	15–25	<ul style="list-style-type: none"> <li>■ Research and drug discovery</li> <li>Accelerate the selection of proteins and molecules best suited as candidates for new drug resale</li> </ul>	<ul style="list-style-type: none"> <li>■ Customer documentation generation</li> <li>Draft medication instructions and risk notices for drug resale</li> </ul>	<ul style="list-style-type: none"> <li>■ Generate content for commercial representatives</li> <li>Prepare scripts for interactions with physicians</li> </ul>	<ul style="list-style-type: none"> <li>■ Contract generation</li> <li>Draft legal documents incorporating specific regulatory</li> </ul>

Value potential of function for the industry  
■ High  
■ Low

## 対応する Python Code

ChatGPT  
Code interpreter

```

import pandas as pd

# Create a dataframe
data = {
    '産業': ['小売と消費者包装製品', 'ファーマと医療製品', '銀行'],
    '利益に対する価値の可能性': ['400-660 (1-2%)', '60-110 (3-5%)', '200-340 (3-5%)'],
    '製品R&D、ソフトウェアエンジニアリング': ['消費者研究: シナリオをテストして消費者研究を加速し、"合成消費者"を作成してターゲティングを強化する'],
    '研究と薬物発見: 新薬製剤候補として最適なタンパク質と分子の選択を加速する',
    'レガシーコード変換: 自然言語変換機能を利用してレガシーフレームワークの移行を最適化する'],
    '顧客オペレーション': ['AR支援顧客サポート: 製品と消費者の嗜好の状況について、リアルタイムで労働力を迅速に情報提供する'],
    '顧客ドキュментーション生成: 薬物再販のための薬物説明書とリスク通知を作成する'],
    '顧客緊急対応IVR: 生成AIにより強化されたIVRインターラクションを通じて、顧客の緊急事態の解決率を部分的に自動化、加速、向上させる(例: クレジットカードの紛失に対して)'],
    'マーケティングとセールス': ['マーケティングコンテンツ作成のためのコピーライティング支援: マーケティングコンテンツと広告スクリプトの作成を加速する'],
    '営業担当者向けコンテンツ生成: 医師との対話用のスクリプトを準備する'],
    'カスタムリテール・バンキングオファー: 顧客のプロフィールと履歴に基づいて各クライアントに合わせたパーソナライズされたマーケティングと販売コンテンツを配信し(例: パーソナライズされた促進策)、A/Bテストのための代替案を生成する'],
    'その他の機能': ['調達サプライヤー・プロセス強化: 供給者との交渉のためのフレイブルックを作成する'],
    '契約生成: 特定の規制要件を含む法的文書を作成する'],
    'リスクモデル文書化: モデル文書化を作成し、欠落している文書と関連する規制更新をスキャンする']
}

df = pd.DataFrame(data)
df

```

実行

産業	利益に対する価値の可能性(\$B)	製品R&D、ソフトウェアエンジニアリング	顧客オペレーション	マーケティングとセールス	その他の機能
0 小売と消費者包装製品	400-660 (1-2%)	消費者研究: シナリオをテストして消費者研究を加速し、"合成消費者"を作成してターゲティング...	AR支援顧客サポート: 製品と消費者の嗜好の状況について、リアルタイムで労働力を迅速に情報提供する	マーケティングコンテンツ作成のためのコーピーライティング支援: マーケティングコンテンツと広告...	調達サプライヤープロセス強化: 供給者との交渉のためのプレイブックを作成する
1 ファーマと医療製品	60-110 (3-5%)	研究と薬物発見: 新薬製剤候補として最適なタンパク質と分子の選択を加速する	顧客ドキュメンテーション生成: 薬物再販のための薬物説明書とリスク通知を作成する	営業担当者向けコンテンツ生成: 医師との対話用のスクリプトを準備する	契約生成: 特定の規制要件を含む法的文書を作成する
2 銀行	200-340 (3-5%)	レガシーコード変換: 自然言語変換機能を利用してレガシーフレームワークの移行を最適化する	顧客緊急対応IVR: 生成AIにより強化されたIVRインターフェンションを通じて、顧客の緊急事態...	カスタムリテールバンキングオファー: 顧客のプロファイルと履歴に基づいて各クライアントに合わせ...	リスクモデル文書化: モデル文書化を作成し、欠落している文書と関連する規制更新をスキャンする



# Occupations most exposed by Language Generative AI

順位	SOCコード	職業名	言語モデリングAIOE
1	41-9041	テレマーケター	1.926
2	25-1123	大学の英語と文学の教師	1.857
3	25-1124	大学の外国語と文学の教師	1.814
4	25-1125	大学の歴史教師	1.813
5	25-1112	大学の法学教師	1.802
6	25-1126	大学の哲学と宗教学の教師	1.800
7	25-1067	大学の社会学教師	1.770
8	25-1065	大学の政治学教師	1.770
9	25-1111	大学の刑事司法と法執行の教師	1.754
10	19-3041	社会学者	1.747

Source: Occupational Heterogeneity in Exposure to Generative AI By Ed Felten  
(Princeton), Manav Raj (University of Pennsylvania), Robert Seamans (New York University), 19 Apr 2023



# Occupations most exposed by Language Generative AI

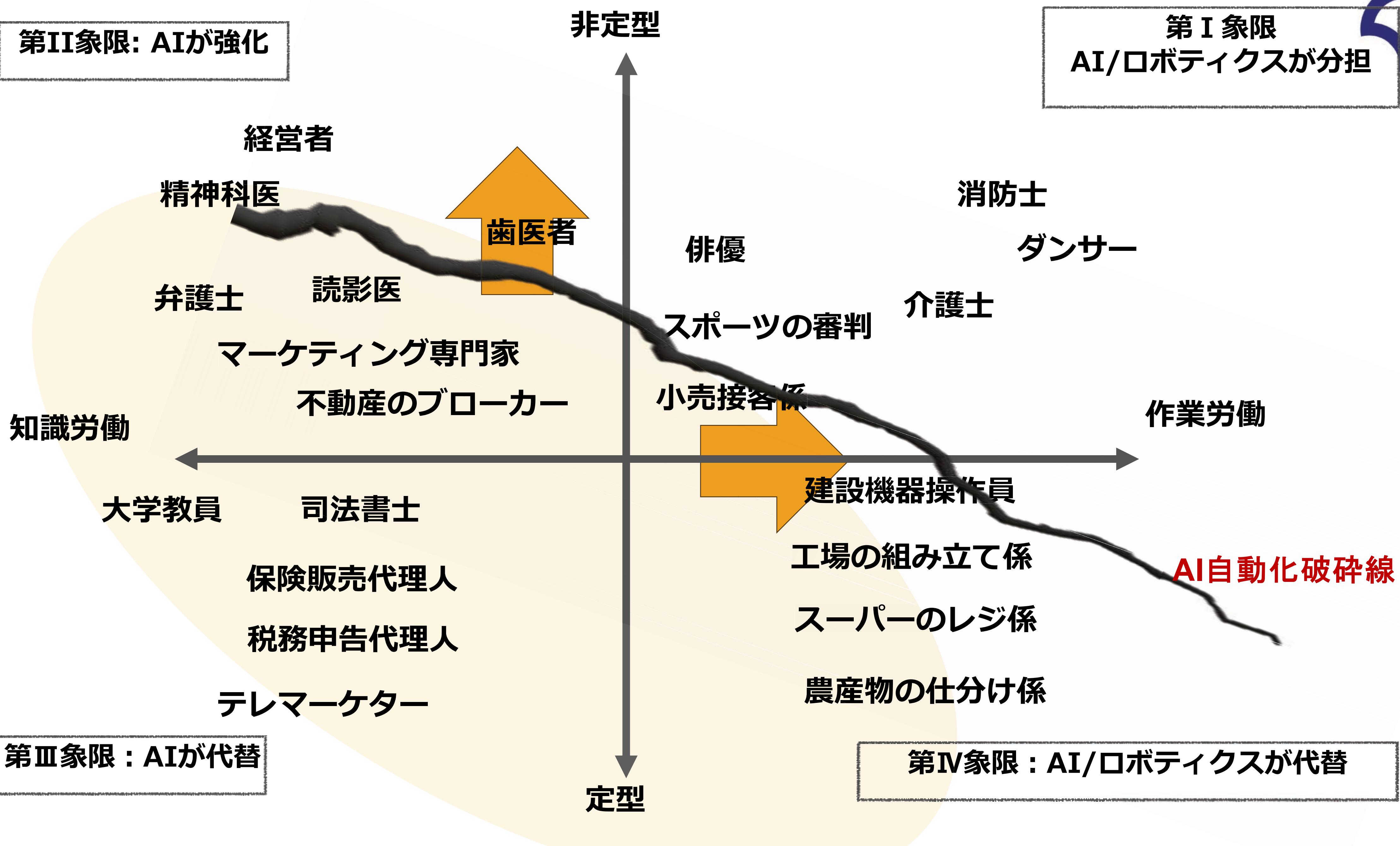
順位	SOCコード	職業名	言語モデリングAIOE
50	23-1011	弁護士	1.454
58	41-3021	保険販売代理人	1.427
17	23-1023	判事、治安判事、裁判官	1.646
29	43-4041	信用認定者、チェッカー、クラーク	1.546
46	21-2011	聖職者	1.470
65	25-1021	大学のコンピュータ科学の教師	1.390
69	29-1066	精神科医	1.366
75	13-1161	市場調査分析家およびマーケティングスペシャリスト	1.356
79	41-9021	不動産ブローカー	1.348

Source: Occupational Heterogeneity in Exposure to Generative AI By Ed Felten (Princeton), Manav Raj (University of Pennsylvania), Robert Seamans (New York University), 19 Apr 2023



第II象限: AIが強化

第I象限  
AI/ロボティクスが分担



第III象限 : AIが代替

第IV象限 : AI/ロボティクスが代替



# 大規模言語モデルの進化方向

スケーリング則: 計算量、データサイズ、パラメータの増加  
に性能が単調増加

マルチモーダル情報(音声・映像・IoTデータ)の統合

オープンソース・オープンデータ、軽量演算モデルの登場

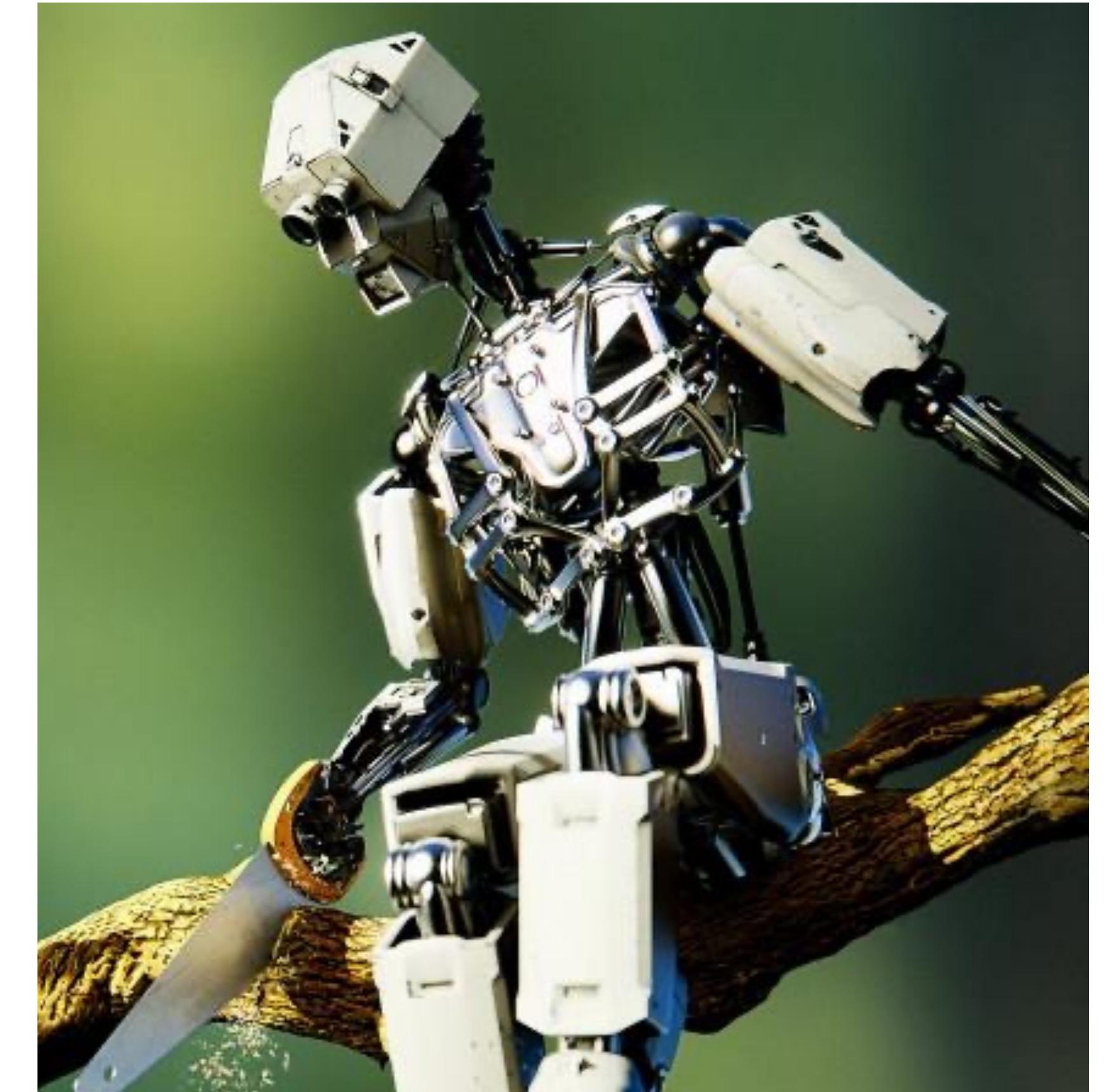
帰納的タスクから演繹的タスク（多段推論）へ相転移

身体性・能動性の獲得により知識の自動習得

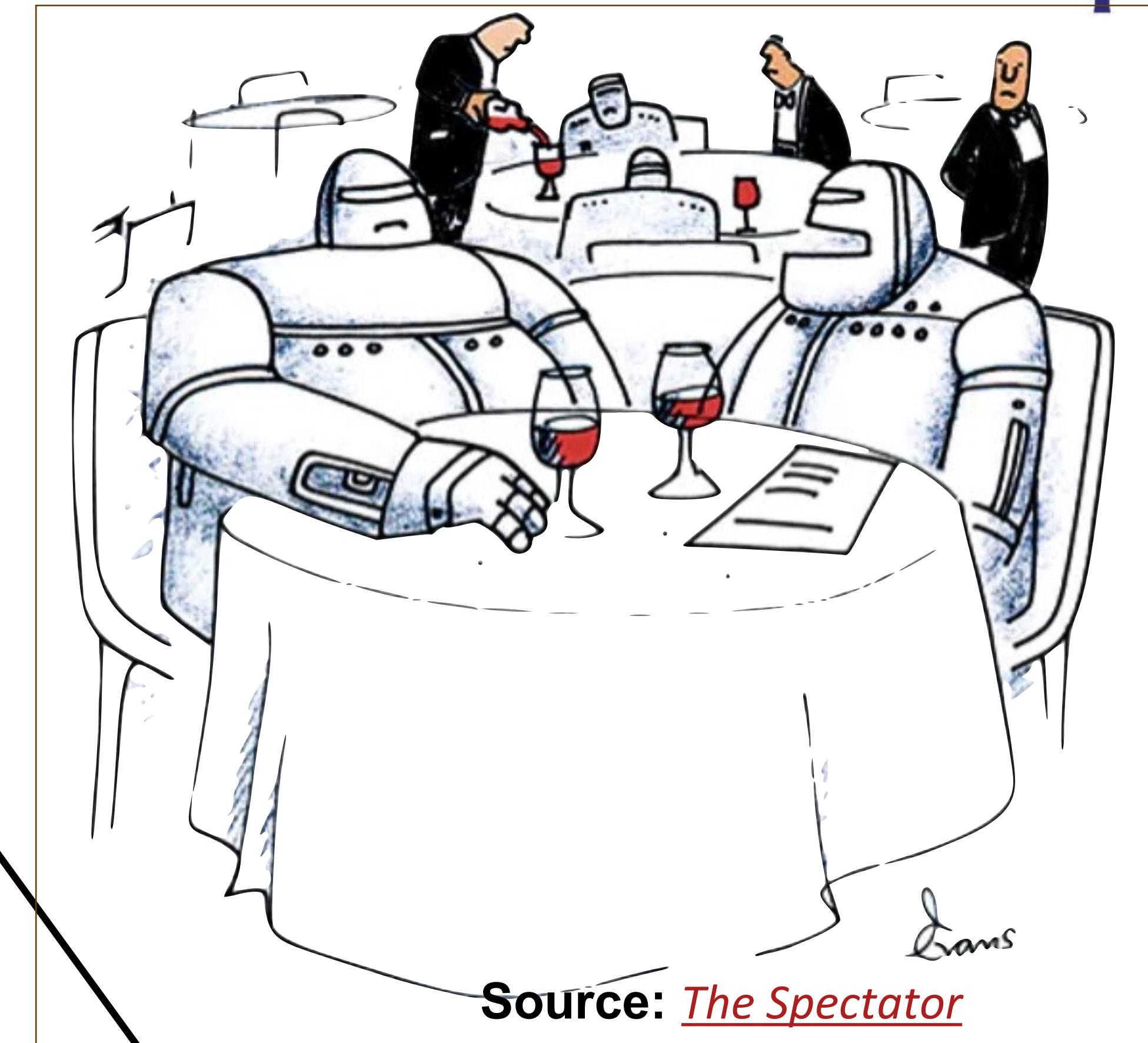
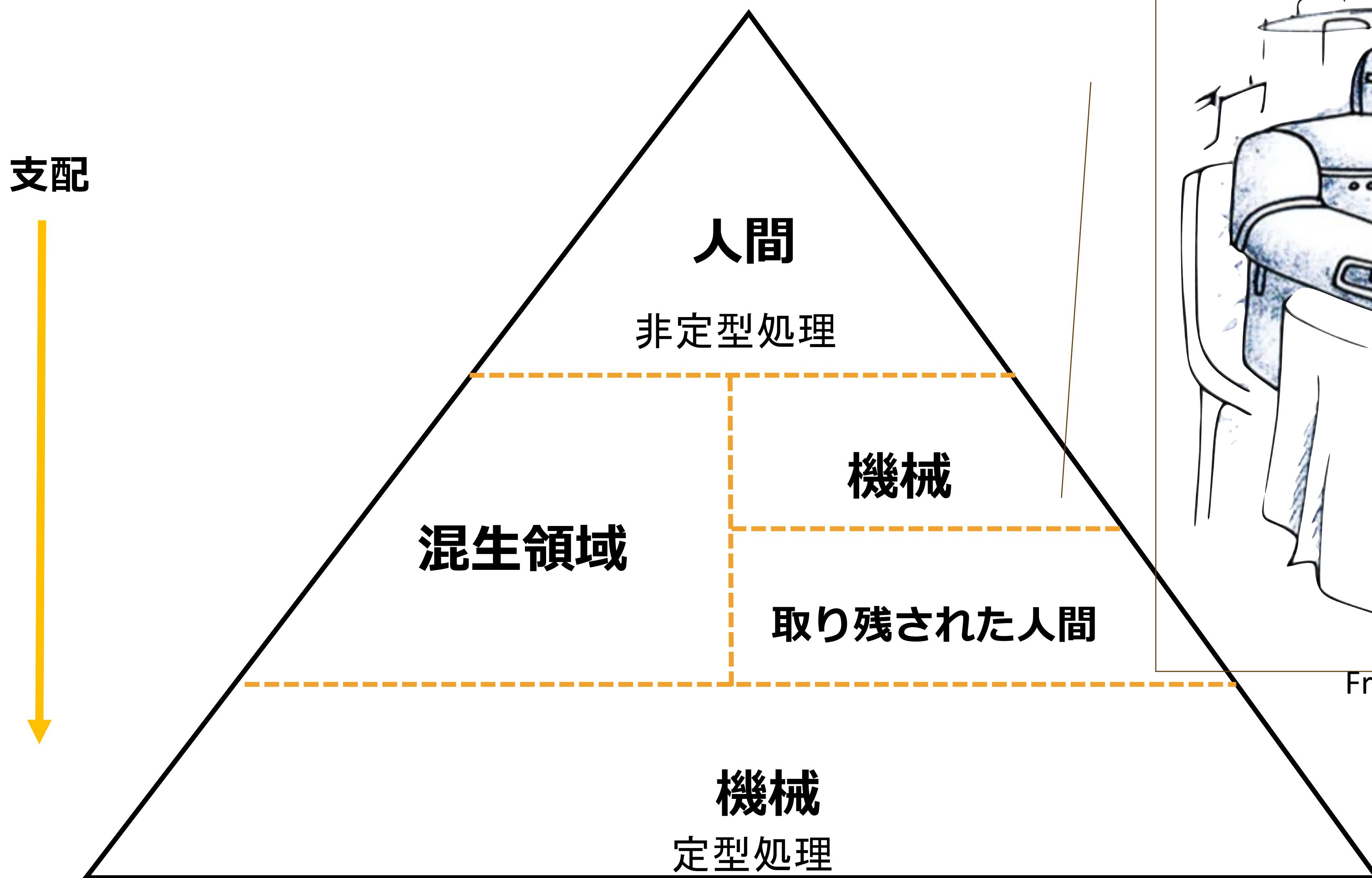


# AGI(Artificial General Intelligence)

- LLMは時空間の構造を理解していない。
  - しかし、巨大システムの管理、セキュリティ対応、医療、プログラミング、多言語能力 は人間を超えた。古典的なチューリングテストはパスしている。
  - 弱いAIと強いAGI を区別する意味がない。
- 
- 五感と身体操作との統合はこれから始まる。
  - 機械が機械のためのコードを書く、機械が機械を再生産するというロボティクス研究は国策として望ましい。
  - 精密機械の自動生産、消費エネルギーの壁を越えれば化ける。



# Not AGI but Partial Singularity

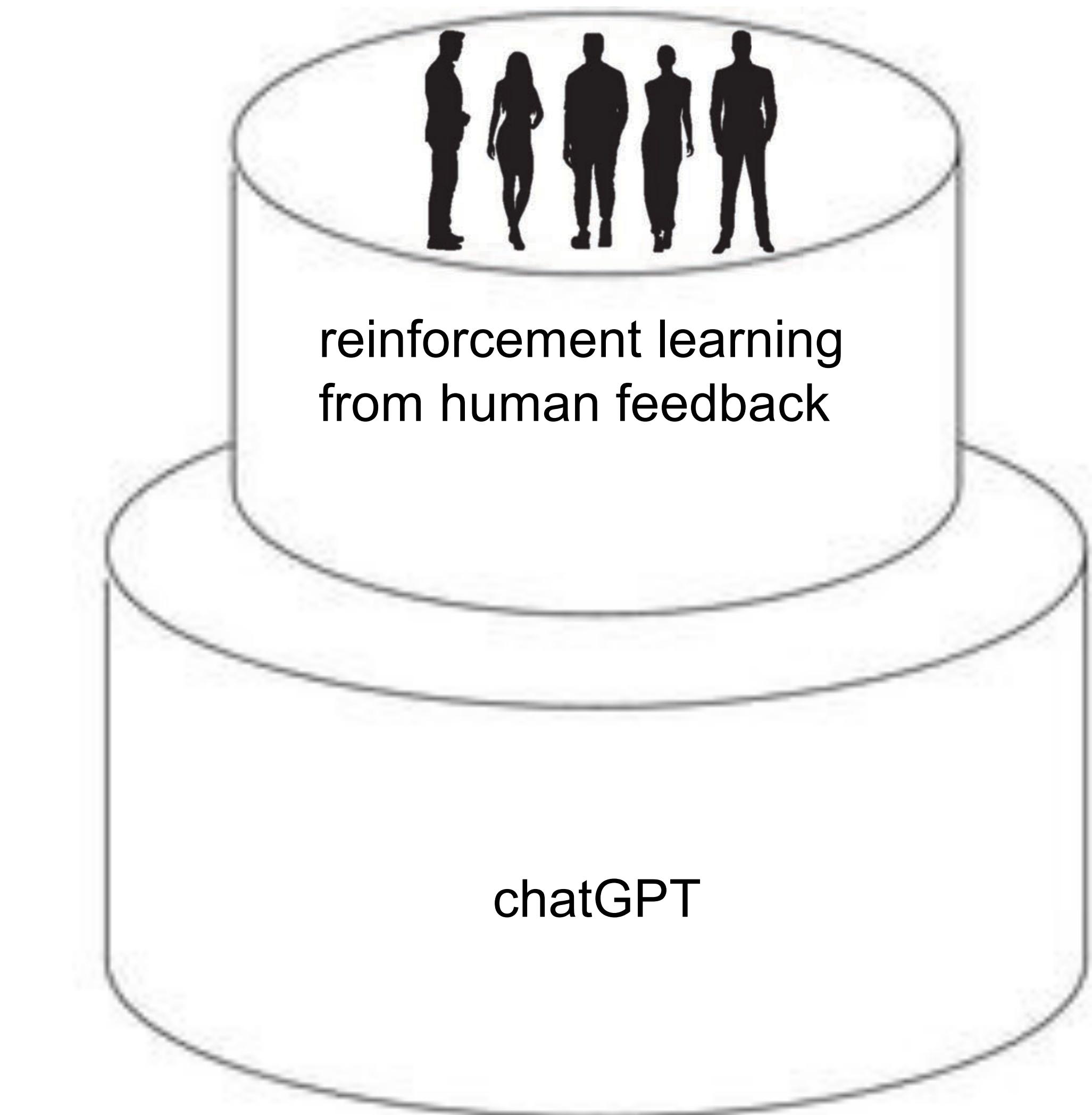


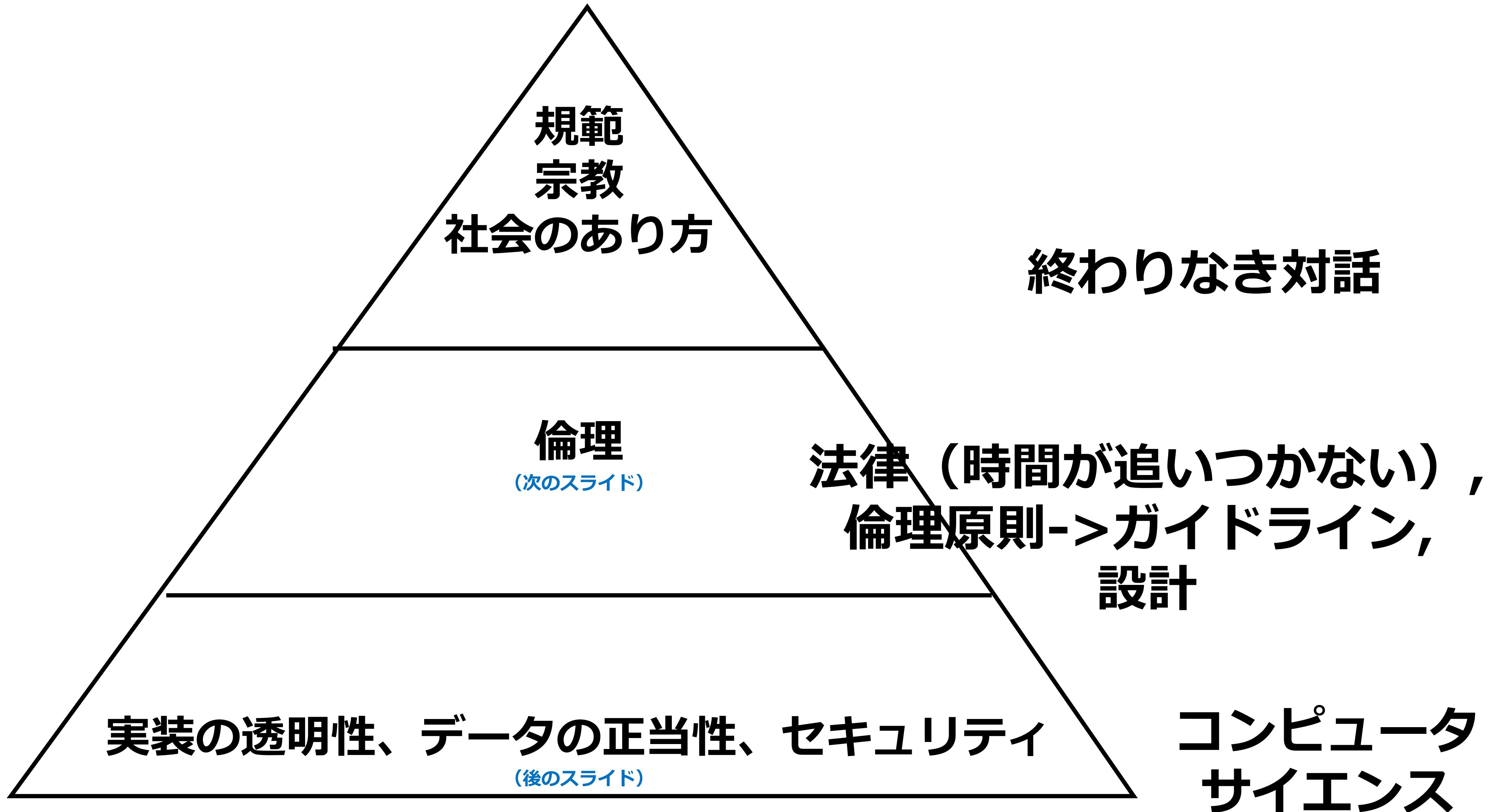
Source: [The Spectator](#)

From magazine issue: [27 January 2018](#)



# Two Layer Cake Structure







- L(法律) vs M(マーケティング) の相剋。変化するS(社会)に法規制の対応が困難。
- 頼りになるのは、E(Ethics)。倫理原則。それを具体化した運用ガイドライン。
- Eを正当化できるのはプロセス。倫理審査・モニタリングの仕組みづくり。
- 提供者が想定しない利用形態への対応。ユーザーからのフィードバック。
- 利用データを詳細に明示している、あるいは開放している提供者は希少。AIへの利用許諾が課題。オープンソース化だけでは不十分。Adobe Fireflyは著作権フリーのデータを利用。

- AI 生成コンテンツに関する Google 検索のガイダンス
  - エクスペリエンス (Experience)
  - 高い専門性 (Expertise)
  - 権威性 (Authoritativeness)
  - 信頼性 (Trustworthiness)
- から
- ## Search Generative Experience(SGE)への発展

Search results for "kids under 3 and a dog, bryce canyon or arches national park"

**ermental**

Bryce and Arches National Parks are family-friendly. Although it dogs on unpaved trails, Bryce Canyon has two paved trails

has distinctive features like hoodoos, natural bridges, and the visitor center's interactive exhibits, children can learn about wildlife, and people of the area. You can bring your dog on the is of both the Rim Trail and the Shared Use Path, both of which er-friendly.

children may enjoy the rock formations, and some say that Arches variety than Bryce Canyon. Pets aren't allowed on any trails, though owned at campgrounds, pullouts, roads, and parking lots.

offer guided tours, ranger-led programs, picnic areas, and and both parks require pets to be leashed.

ow up ↗ How long to spend at Bryce Canyon with kids? ↗ How many days do you need in Arches National Park for kids?

ur Guide  
intourguide.com > 2022/02/17 > which-is-b... 

Guide  
thes only allow dogs at campgrounds, parking lots, and along making it a lot harder to experience the park with your dog.

te  
t.com > bryce-canyon-national-pa... 

n. ng A Trip To Bryce Canyon National ...



# オリジネーター・プロファイル (OP)

インターネット上の  
ニュース記事や広告などの  
情報コンテンツに、発  
信者情報を紐付ける  
**Originator Profile (OP)**  
技術を研究開発





# フルスタック人材教育 未踏+ $\alpha$



データ整備と管理

最良技術の選択

優れた実装と運用



MLOps(DevOps for ML)の重要性が増している。モデルをどう運用するかに焦点が移っている。



# 論点整理

- LLMが作れてようやくスタートライン。 InstructGPTの壁。 OPEXの壁。
- ビジネスマodelキャンバスの有無。出口戦略なしにチューニング・セキュリティ設計は困難。
- ツールのカスタマイゼーションと社内データのアセット化に投資。
- 日本語特化の是非(OPEX低減？)。スイートスポット(知識ベースとして儲かる応用)探索が課題。
- 10年先を見据えた研究開発:マルチモーダルとその先にある創薬、ロボティクス。
- 5年前は単純知識作業の自動化、今は高度知識作業の自動化が焦点。
- 言語資源の権利処理、データ経済圏(個社・業界データの流通)、倫理原則のあり方。
- ビジネスと技術の両方がわかる人材の育成。