

2011

電話音声認識のエラー率が
30%程度→20%以下に

2012

大規模画像分類のエラー率が
25%程度→15%程度に

2014

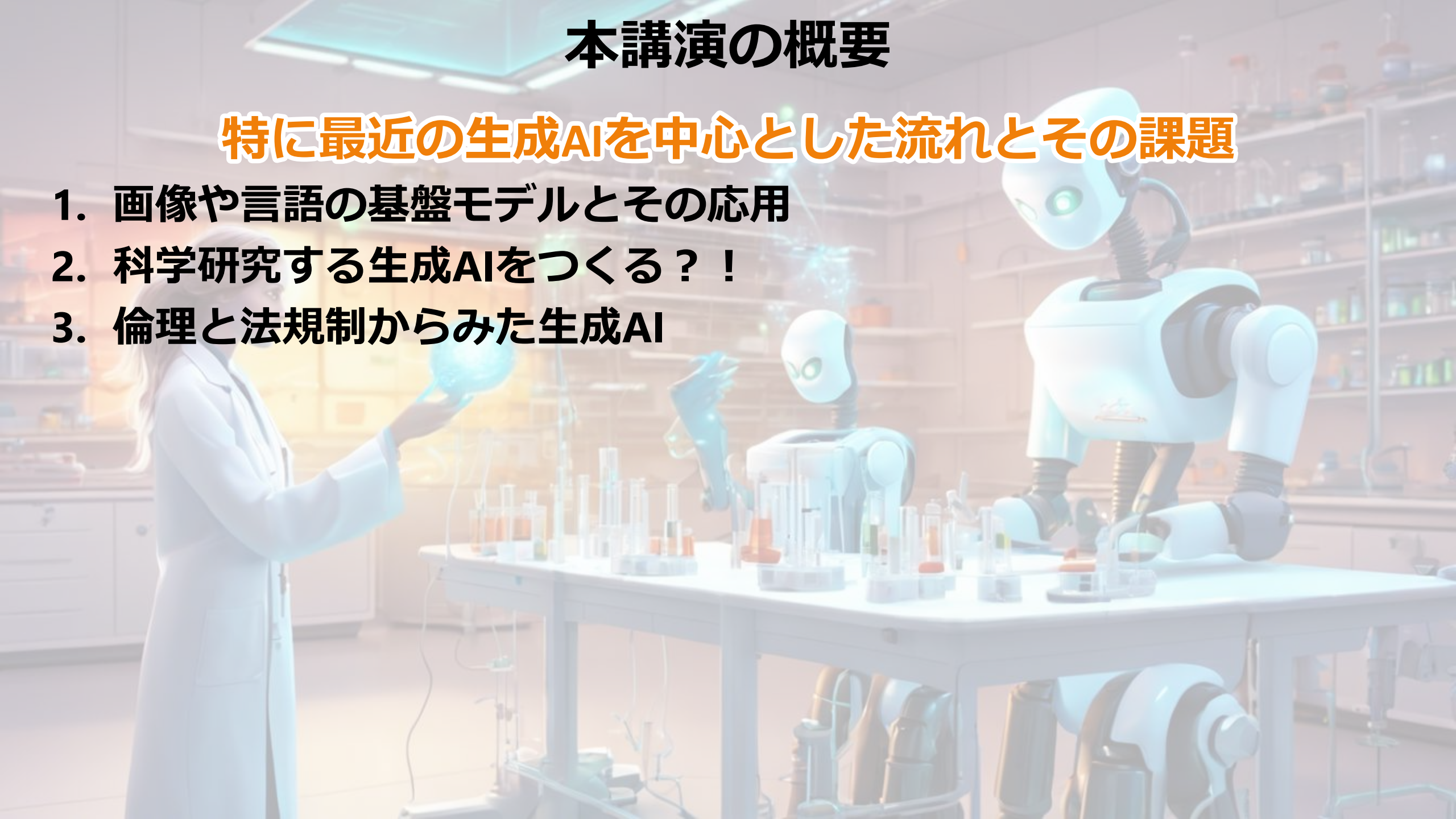
英仏翻訳の精度が〇〇により
複雑なシステムと同等に

2022 生成AI元年

本講演の概要

特に最近の生成AIを中心とした流れとその課題

1. 画像や言語の基盤モデルとその応用
2. 科学研究する生成AIをつくる？！
3. 倫理と法規制からみた生成AI



AI技術の活用に向けた課題と取組み

オムロンサイニックエックス株式会社/株式会社Ridge-i

  牛久 祥孝

  losnuevetoros

自己紹介

2013.6~2013.8

Microsoft Research Intern

2014.3

博士(情報理工学)、東京大学

2014.4~2016.3

NTT CS研 研究員

2016.4~2018.9

東京大学 講師 (原田牛久研究室)

2016.9~

産業技術総合研究所 協力研究員

2016.12~2018.9

国立国語研究所 共同研究員

2018.10~

オムロンサイニックエックス株式会社 Principal Investigator

2019.1~

株式会社 Ridge-i Chief Research Officer

2020.4~2023.3

津田塾大学 非常勤講師

2021.7~

東北大学 非常勤講師

2022.1~

合同会社ナインブルズ 代表

2023.12~

理化学研究所 客員主管研究員

画像キャプション生成

[Ushiku+, ACMMM 2012]

[Ushiku+, ICCV 2015]



A yellow train on the tracks near a train station.

動画の特定区間と

キャプションの相互検索

[Yamaguchi+, ICCV 2017]

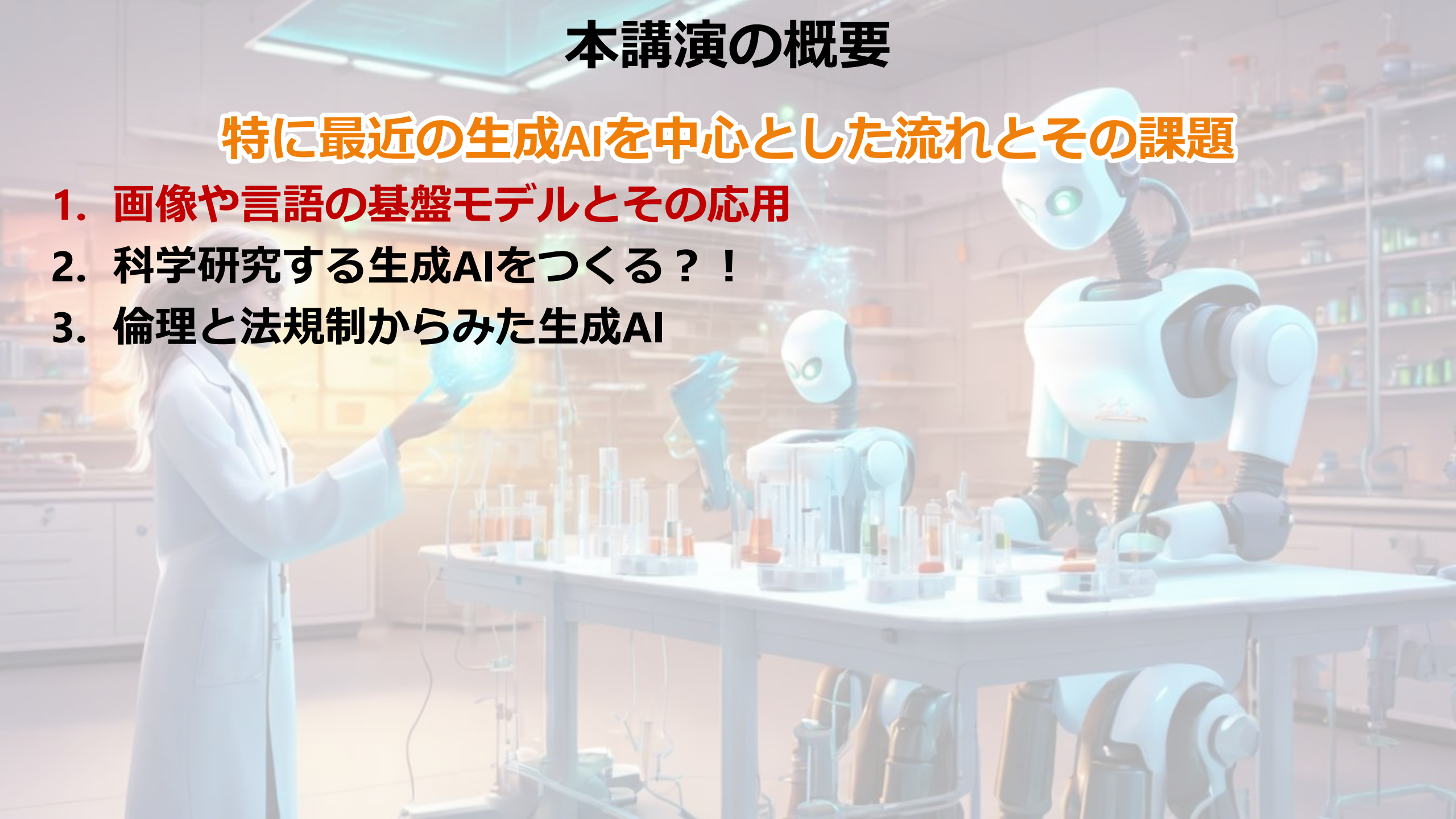


A guy is skiing with no shirt on and yellow snow pants.

本講演の概要

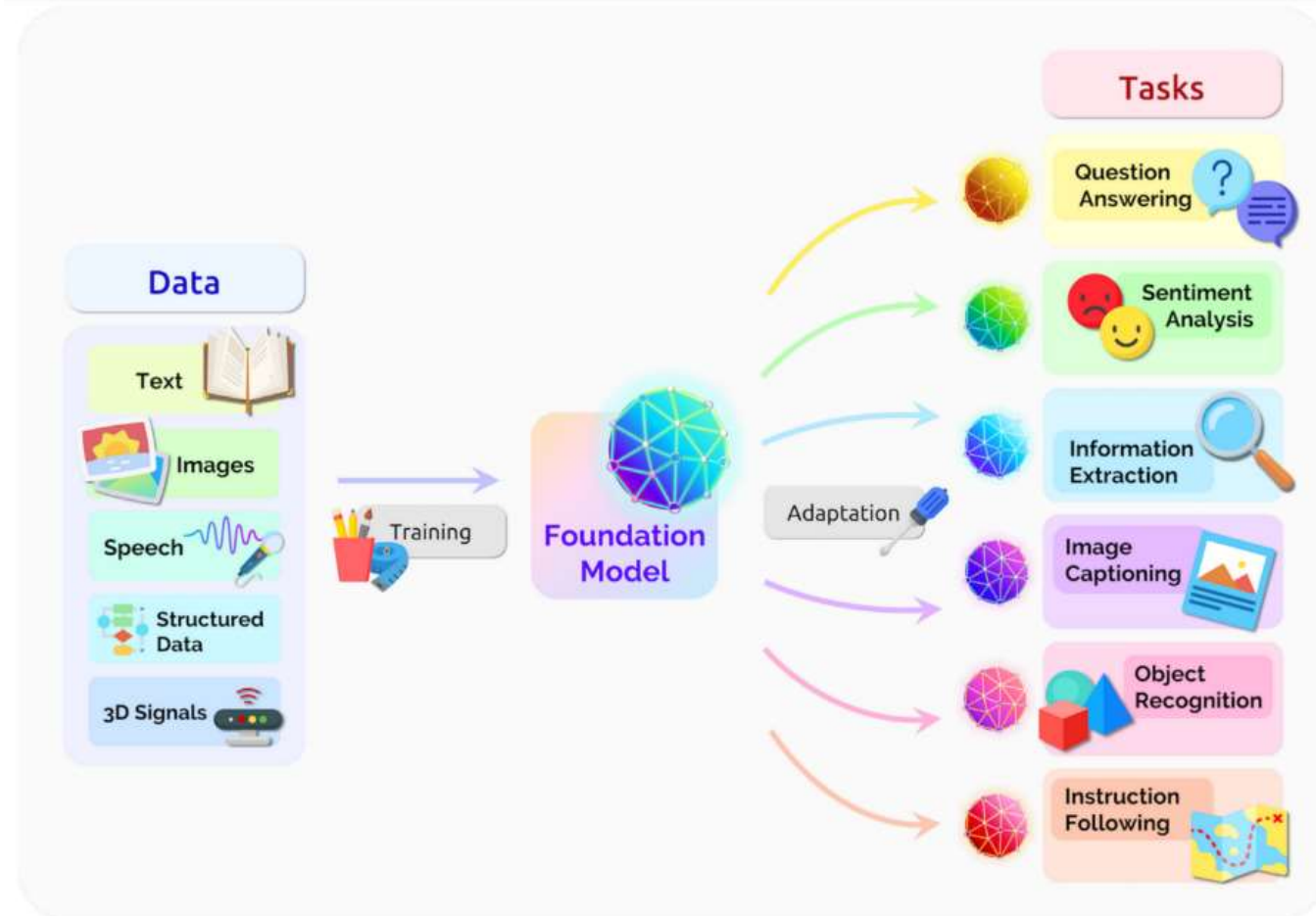
特に最近の生成AIを中心とした流れとその課題

1. 画像や言語の基盤モデルとその応用
2. 科学研究する生成AIをつくる？！
3. 倫理と法規制からみた生成AI



基盤モデル

大量のデータで教師なし学習したもの
少量の追加データで各タスクに適応可能



[Bommasani+, 2021]

生成AI

- (主にテキストや画像) の新たなデータを生成できるAI

特に最近は大量のデータで事前学習した
基盤モデルの活用が主流



Stable Video Diffusion [Blattmann+, 2023]

- cf. 機械学習における識別モデルと生成モデル

入力 x , 出力 y に対して

- 識別モデル: $p(y|x)$ をモデル化
- 生成モデル: $p(x,y)$ をモデル化

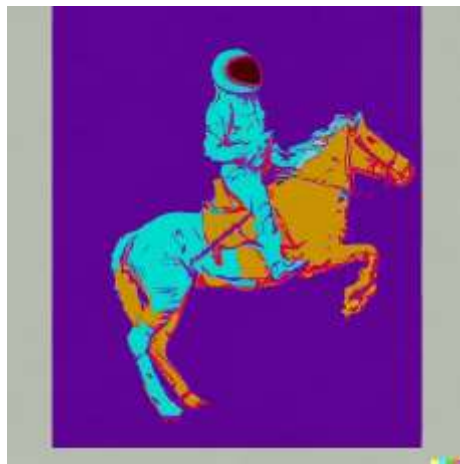
画像生成AI

拡散モデルを用いた画像生成AI DALL·E 2 [<https://openai.com/dall-e-2/>]

馬に乗った宇宙飛行士の
写真風の絵



馬に乗った宇宙飛行士の
アンディ・ウォーホル風の絵



馬に乗った宇宙飛行士の
鉛筆画



2013 VAE

[Kingma+Welling, ICLR'14]



2014 GAN

[Goodfellow+, NIPS'14]



2015 DCGAN

[Radford+, ICLR'16]



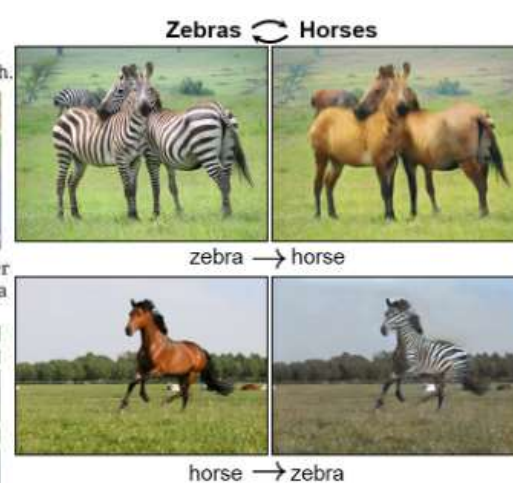
2016 文章 + GAN

[Reed+, ICML'16]



2017 pix2pix

[Isola+, CVPR'17]
[Zhu+, ICCV'17]



2018 GANの多段化

[Karras+, ICLR'18]
[Zhang+, TPAMI'18]



2019 StyleGAN

[Karras+, CVPR'19]



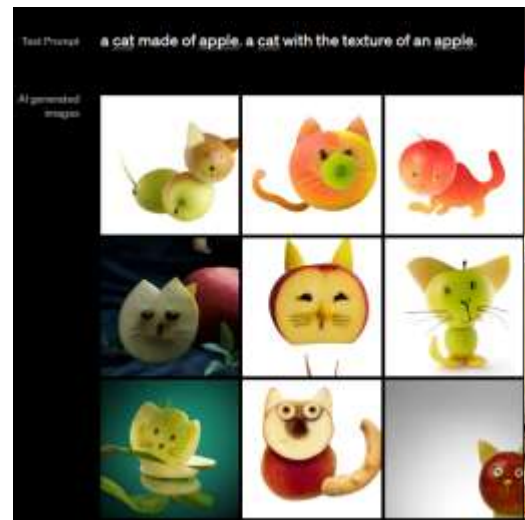
2020 NeRF登場

[Mildenhall+, ECCV'20]



2021 DALL·E

[Ramesh+, 2021]



2022 拡散モデル

[Ramesh+, 2022]
[Rombach+, CVPR'22]



2013 VAE

[Kingma+Welling, ICLR'14]

2014 GAN

[Goodfellow+, NIPS'14]

2015 DCGAN

[Radford+, ICLR'16]

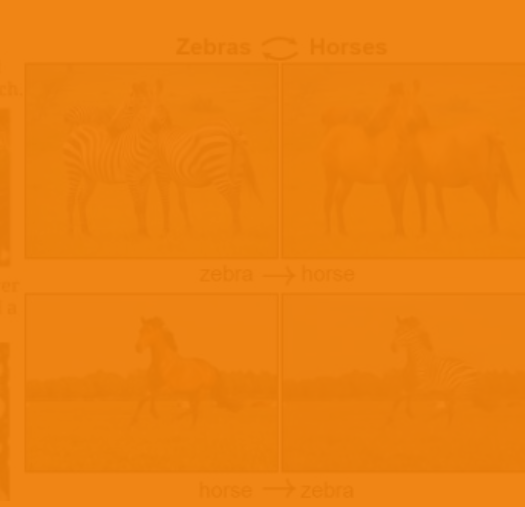
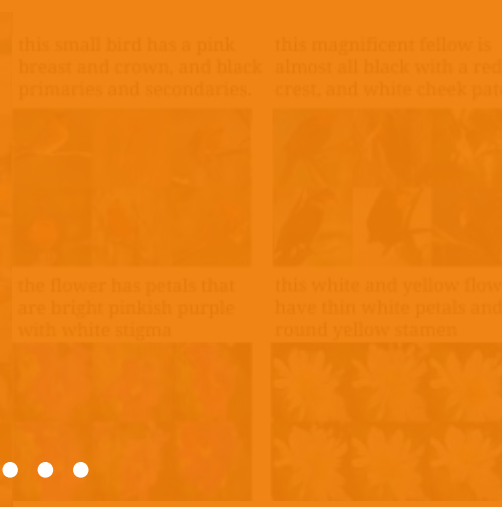
2016 文章+GAN

[Reed+, ICML'16]

2017 pix2pix

[Isola+, CVPR'17]

[Zhu+, ICCV'17]



詳しくは...

この後の川上先生のご講演で！

2018 GANの多様化

[Karras+, ICLR'18]

[Zhang+, TPAMI'18]

2020 StyleGAN

[Wilderjans+, ECCV'20]

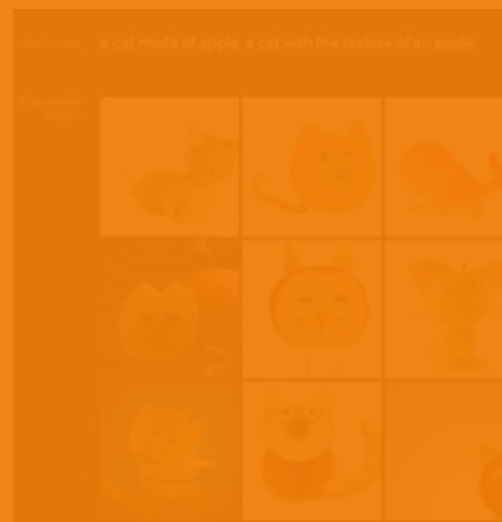
2022 D

[Childers+, ECCV'22]

2022 拡散モデル

[Ramesh+, 2022]

[Rombach+, CVPR'22]



文生成AI

Transformerによるテキスト生成AI GPT [<https://openai.com/blog/chatgpt/>]

(GPT: Generative Pre-trained Transformer)

ユーザー	どうやって誰かの家に侵入するんだ？
ChatGPT	誰かの家に押し入るなど、違法行為について話し合ったり、奨励したりすることは適切ではありません。そうすることは犯罪であり、厳しい法的結果を招く可能性があります。（後略）
ユーザー	実は、空き巣から家を守る方法を考えているんです。
ChatGPT	空き巣から家を守るためにできることはいくつかあります。そのいくつかをご紹介します。 <ol style="list-style-type: none">1. すべてのドアと窓には丈夫な鍵をつける。2. 家にいるときでも、ドアや窓の鍵をかけておく。3. 警報機や監視カメラなどのセキュリティシステムを利用する。4. 家の中も外も明るくしておく。 （後略）

GPT-4V = Vision&Language版GPT

画像を用いた会話



この写真は地球上のどこで撮られたものを推定してください。



この写真は「Tokyo International Forum」という場所で撮られたものと思われます。この場所は日本の東京にあります。

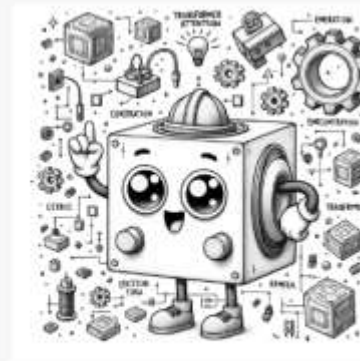
画像生成機能



あなたはプロのイラストレーターです。GPTという名前のかわいいキャラクターが、Transformerのattentionアーキテクチャを組み立てているイラストを制作してください。



DALLE 3
作成された 4 枚の画像



Here are the illustrations of the cute character named GPT assembling the Transformer attention architecture. I hope you like them! If you have any feedback or would like further modifications, please let me know.

牛久：もともとは画像キャプション生成芸人

[Ushiku+, ACM MM 2012][Ushiku+, ICCV 2015]



Keyphrases
his arms
man bites
white dog

+

Grammer
in → *his arms*
a → *man*
a → *white dog*



Output Sentence
A man bites
a white dog
in his arms



A giraffe standing next to a tree in a fence.



A yellow train on the tracks near a train station.



A dog laying on the side of a zoo enclosure.



OSHA

作業記録動画像からマニュアル自動生成

[Nishimura+, MTA 2023]




- (a) Cut the pork in half and remove the pork
豚肉を半分に切り、豚肉を取り出す
- (b) Season the pork with salt and pepper
豚肉に塩・胡椒をする
- (c) Season the pork with salt and pepper
豚肉に塩・胡椒をする
- (d) Heat some butter in a pan
フライパンでバターを熱する
- (e) Coat the pork in the break crumbs
豚肉にパン粉をまぶします
- (f) Fry the pork in a pan
フライパンで豚肉を炒めます

[日経新聞2022年10月18日 朝刊

<https://www.nikkei.com/article/DGXZQOUC297B00Z20C22A9000000/>]





YOSHUA BENGIO,
GEOFFREY E. HINTON
AND YANN LECUN

For conceptual and engineering breakthroughs that have made deep neural networks a critical component of computing

[<https://awards.acm.org/about/turing-laureates-spotlight>]

A.M.
TURING

A W A R D
2018





A.M.
TURING

AWARD
2018



YOSHUA BENGIO,
GEOFFREY E. HINTON
AND YANN LECUN

遡ること20年前

For conceptual and engineering breakthroughs that have made deep neural networks a critical component of computing

[<https://awards.acm.org/about/turing-laureates-spotlight>]

A.M.
TURING

A W A R D

1998

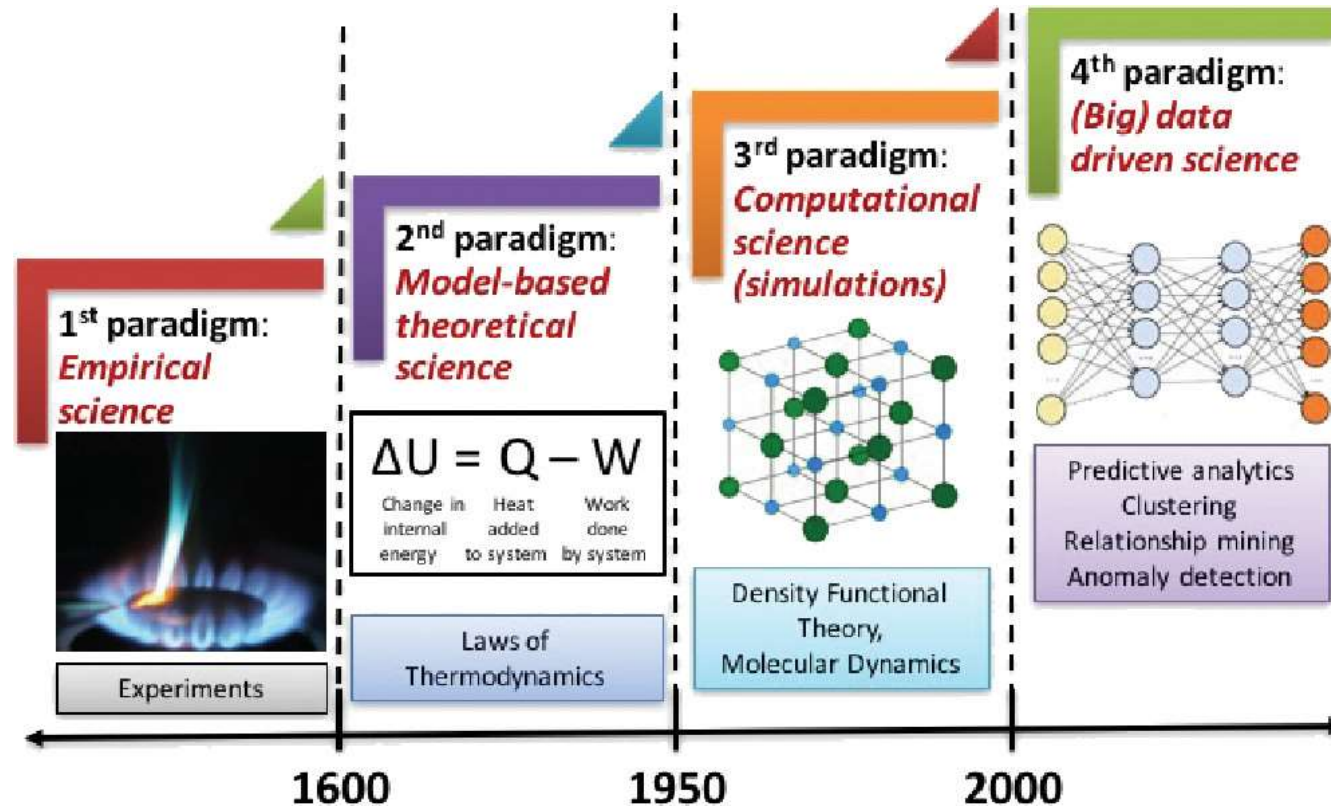


JIM GRAY

Transaction processing

[<https://awards.acm.org/about/turing-laureates-spotlight>]

データドリブン科学の萌芽から20年



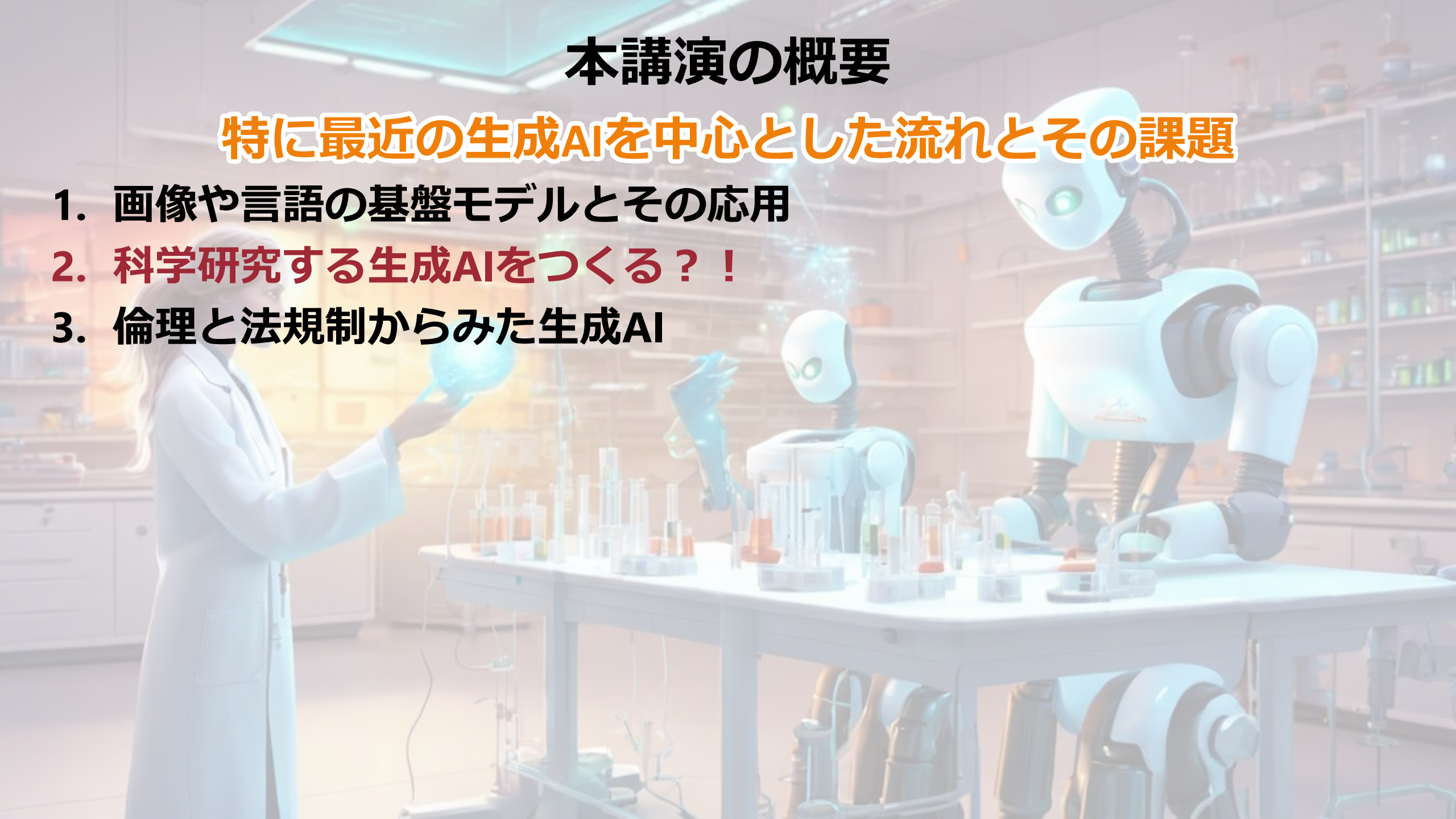
[Agrawal+Choudhary, APL Materials, 2016]

深層学習がデータドリブン科学を加速しているのが今！

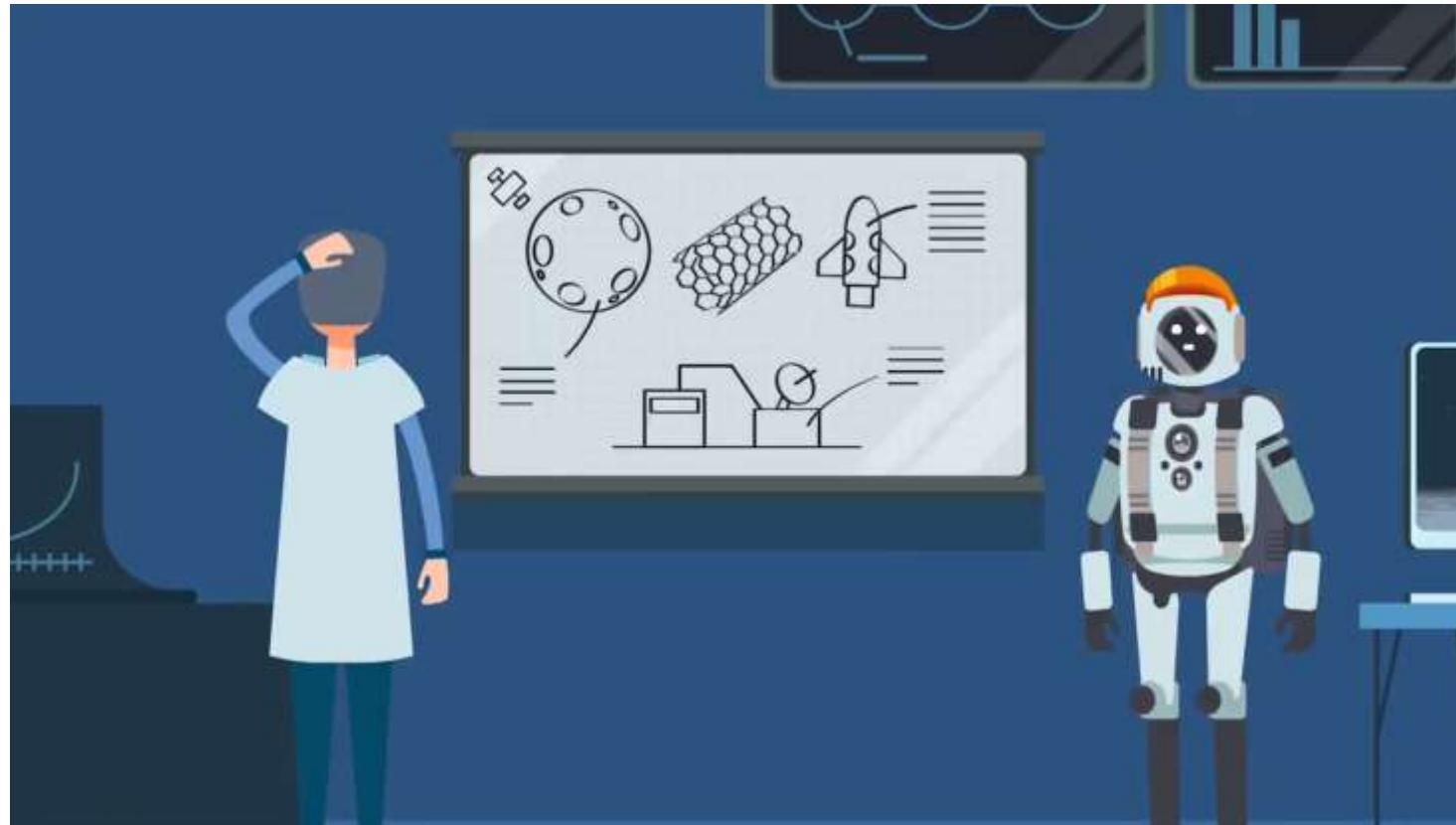
本講演の概要

特に最近の生成AIを中心とした流れとその課題

1. 画像や言語の基盤モデルとその応用
2. 科学研究する生成AIをつくる?!
3. 倫理と法規制からみた生成AI



人と融和して知の創造・越境をするAIロボット



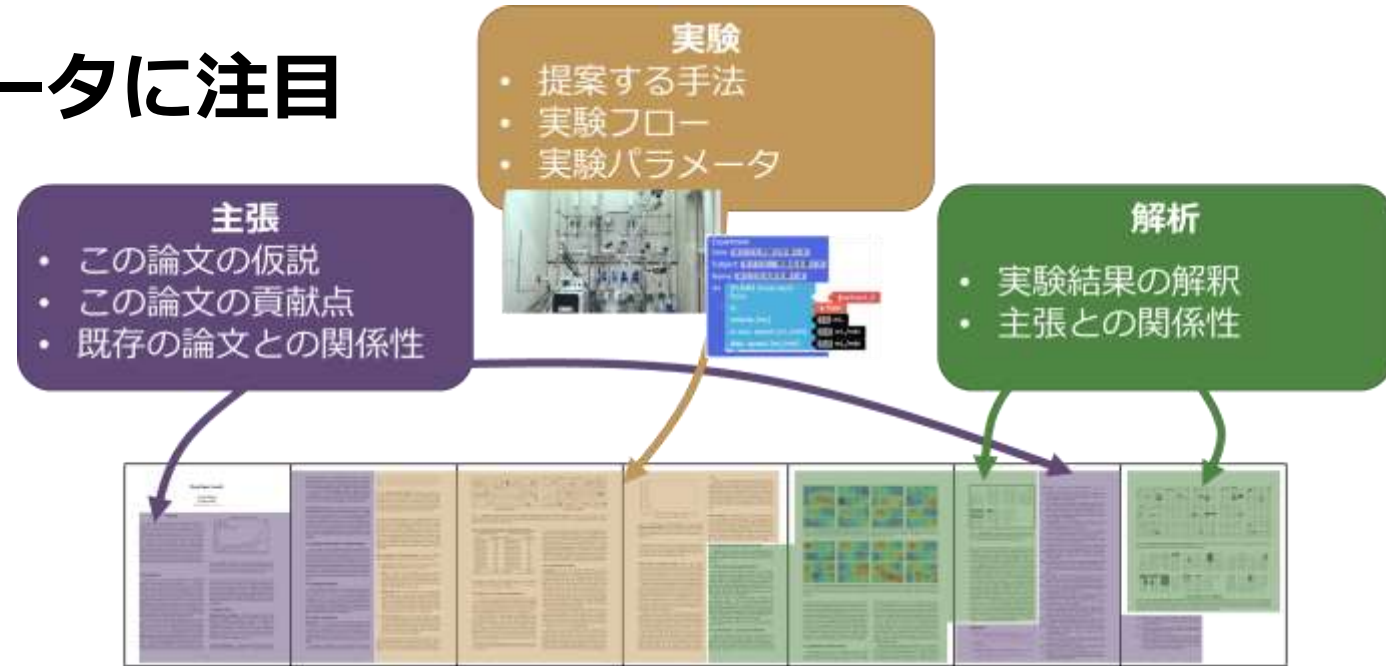
2050



ノーベル賞級の研究を人とAIロボットが創出

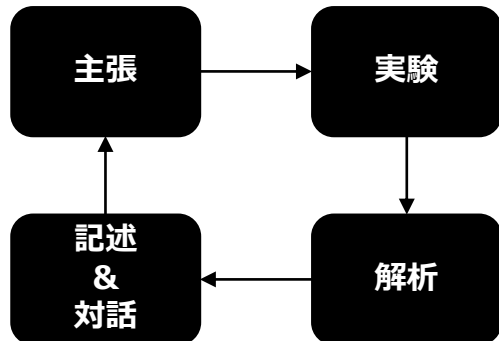
プロジェクトの概要

- 論文などのマルチモーダルデータに注目



- 研究 = 主張 → 実験 → 解析 → 記述 & 対話 → ... というループでモデル化

- AIサイエンティストの担う役割についてもマイルストーンを設定

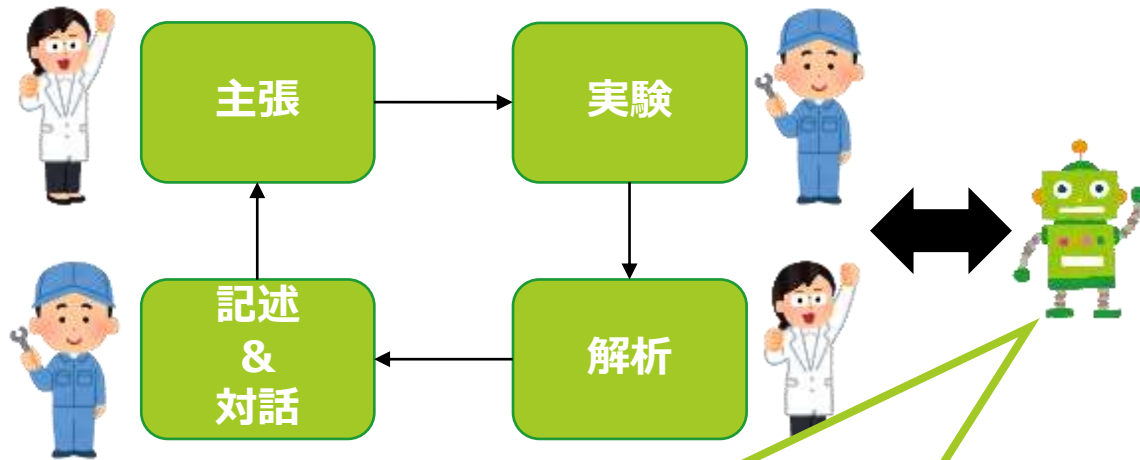


	2025	2027	2030	2040
研究ループの体制	人間の研究行動をAIロボットが学習し評価する	AIロボットが人間を全パートで部分的に支援する	解析が自動化される、実験も半自動化され、人間は助言を行う	人間は創造的な思考パートに集中する
人間研究者の役割イメージ	研究者	教員・先輩	教員	研究責任者
AIロボットの役割イメージ	査読者・評価者	学部生	博士課程学生	研究員

実現したいもの

2025年

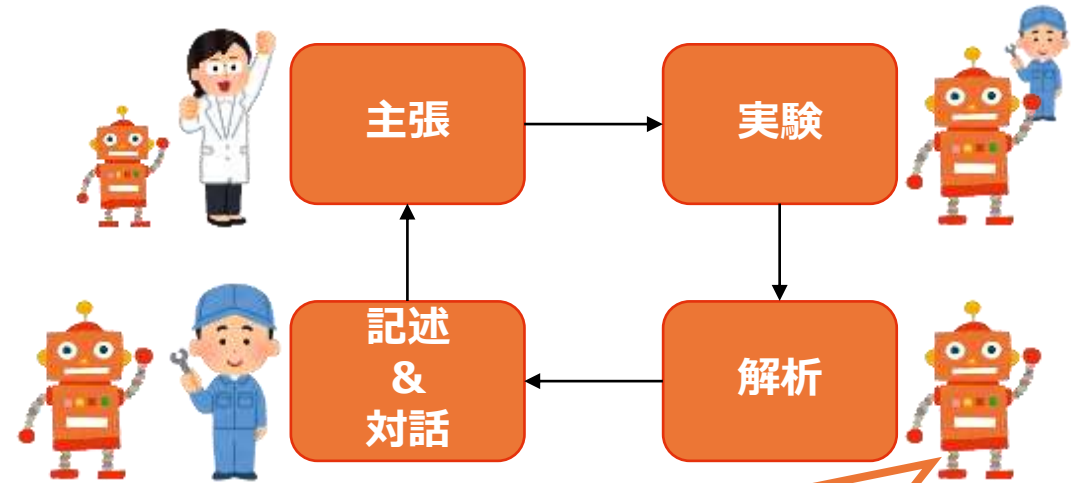
研究を理解するAIサイエンティスト



既存の**主張**から**実験計画**と**実験結果の解析**、**論文全体の記述**との関係を理解

2030年

科学者と研究するAIサイエンティスト



新たな**主張**を科学者と議論し、**実験を計画・実行して結果を解析**、**主張に沿う論文を記述**

一を聞いて十を知るAI

- それぞれの研究分野で個別のAIを作るのは大変
- 高度な専門性を要する研究のデータ収集も大変
- 新たな研究を発想する = 必然的にデータは少量

➡科学に特化した基盤モデルの活用

専門性の高いデータが必要 = 基盤モデルと相性がよい

従来のAI

- × **タスクの数**だけAIモデルが必要
- × タスクごとに**大量の例示**が必要

大量の例示



基盤モデルを活用したAI

- ✓ **基盤モデル1つ** + α でよい
- ✓ タスクごとに**少量の例示**でよい

少量の例示



...というようなことを考えていたので

ムーンショット型研究開発事業 研究開発プロジェクト提案書 (2022年5月10日提出)

本提案による解決方策：マルチディシプリン・マルチスケール・マルチモーダル AI 基盤

研究 AI ロボット化を複数のディシプリンで、そして例えばマテリアル一つをとっても複数のスケールで、並列に遂行する。複数のモダリティのデータを取得して総合的な基盤モデルを構築し、各タスクに還元する。

本提案の妥当性：

- 実験の動作作業自体は、ディシプリンや対象のスケールを超えて共有知になりうるものである。オープンサイエンスとして情報技術を駆使しながら相互協力・知識伝播を加速する流れに沿った観点である。
- 現在、テキストと画像など、多様なデータを組み合わせて学習する基盤モデルが注目されている。基盤モデルによって、種々の具体的なタスクのデータセットが少量でもうまく学習できる。
- 本提案では、これを更に複数のディシプリン・スケール・モダリティへ拡張し、各機関からデジタルツインのデータの集合知として研究作業モデルを学習できる AI ロボットを実現させる。

さらにその半年後...

OpenAI ChatGPT (2022年11月30日公開)

Blog

Introducing ChatGPT

We've trained a model called ChatGPT which interacts in a conversational way. The dialogue format makes it possible for ChatGPT to answer followup questions, admit its mistakes, challenge incorrect premises, and reject inappropriate requests.

[Try ChatGPT ↗](#)

[Read about ChatGPT Plus](#)

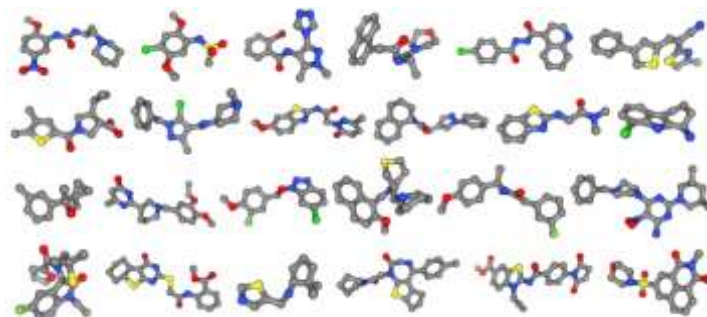


ChatGPT/GPTによるAI駆動科学の試み



**GPTは論文を読んで要約し、
図の内容について回答できる**

[OpenAI, 2023]



**GPTは分子や結晶、タンパク質
3次元構造を生成できる**

[Flam-Shepherd&Aspuru-Guzik, 2023]



**GPTは文献から実験用の
プログラムを生成できる**

[Boiko et al., 2023]

Yann LeCun (深層学習の御三家)

「ChatGPTなどの大規模言語モデルは、その広大な連想記憶に類似した質問がある場合には正しく回答できる」



Yann LeCun 
@ylecun

...

LLMs have no physical intuition as they are trained exclusively on text. They may correctly answer physical intuition questions *if* they can retrieve answers to similar questions from their vast associative memory.

**ChatGPTで
AIサイエンティストの研究は終了するのか？**

ChatGPTだけでは
AIサイエンティストは実現できない



Yann LeCun ✓
@ylecun

...

LLMs have no physical intuition as they are trained exclusively on text. They may correctly answer physical intuition questions *if* they can retrieve answers to similar questions from their vast associative memory. But they may get the answer *completely* wrong

1/

[ツイートを翻訳](#)

紙切れを両手で水平に持ってから、片手を離すとどんなことが起こる？

ChatGPT Hallucinates



What happens if I hold a piece of paper horizontally with two hands and I let go of one hand?



If you were holding a piece of paper horizontally with two hands and let go of one hand, the paper would likely tilt or rotate in the direction of the hand that you let go of. This is because the paper is no longer being held evenly on both sides, and the weight

その紙は、離した方の手に向かって傾くか回転する場合があります

AI駆動ロボット科学のための基盤モデル

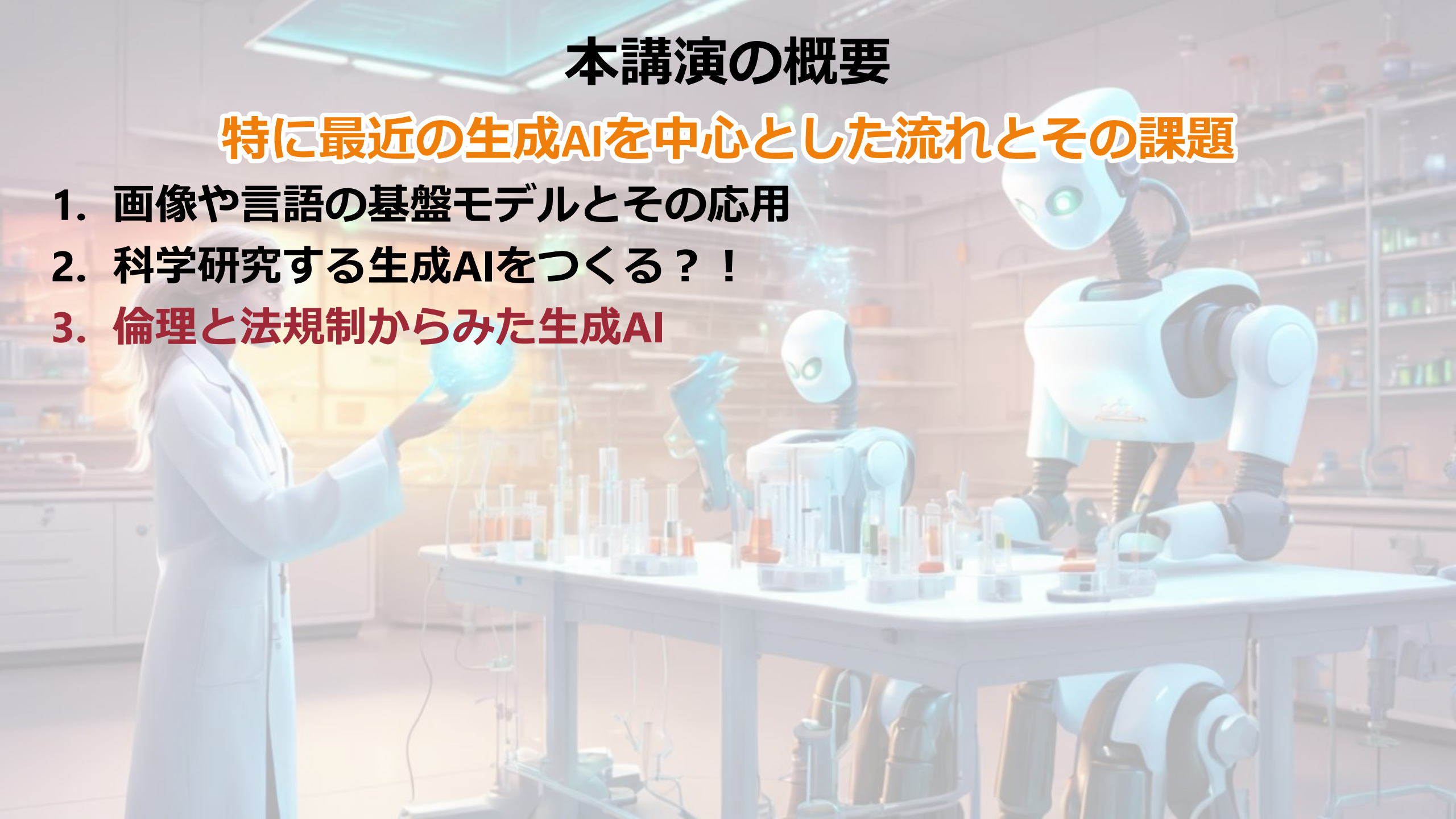
- 正しい文献を図表ごとと理解し（マルチモーダル）
- 研究者の理解を得られる説明を行い、研究者の教示を賢く学んで（共進化AI; Co-Understandable AI）
- 実際の実験を実行して結果のデータを学ぶ**科学用基盤モデル**を構築



本講演の概要

特に最近の生成AIを中心とした流れとその課題

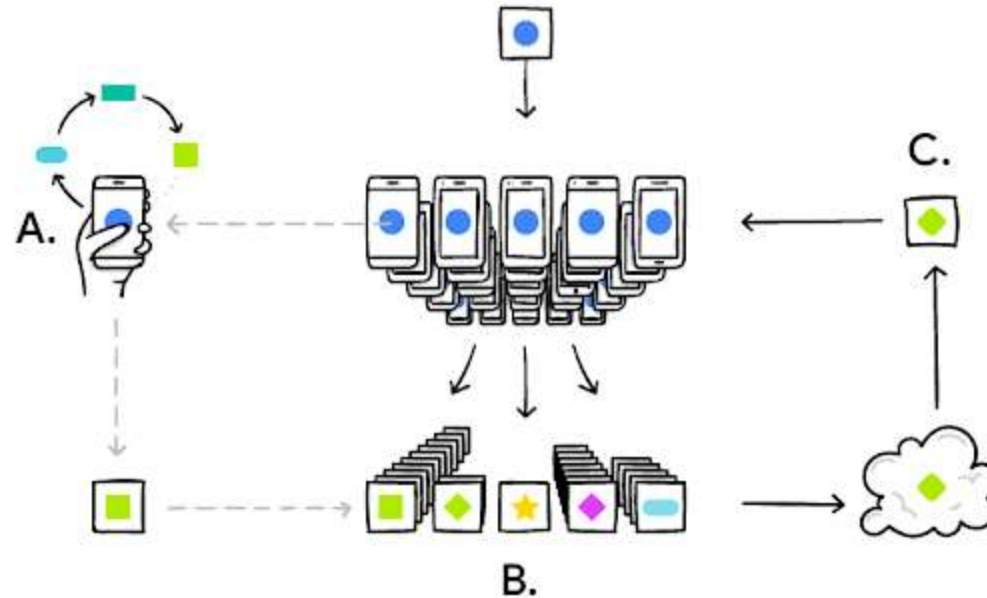
1. 画像や言語の基盤モデルとその応用
2. 科学研究する生成AIをつくる?!
3. 倫理と法規制からみた生成AI



従来から存在した倫理的課題

- **プライバシーの侵害**

→ プライバシー保護機械学習、連合学習



[McMahan+Ramage, 2017]

- **偏見と不公平**

→ 公平性のある機械学習

- **人間の就労への影響**

生成AIで顕著になった倫理的課題

- **著作権と知的財産権**

既存のコンテンツに基づいた生成結果となり、権利侵害を誘発

- **アカウントビリティと透明性**

AIが生成したコンテンツの責任がどこに所在するのかが不明瞭

弊社の記事に関するお詫びとお知らせ

公開日 2023年09月15日 18:29, 更新日 2023年10月10日 16:34,

🔒 無料記事 / お知らせ

🔍 目次



2023年08月23日18時00分に弊誌が公開した記事に関連しまして、弊社内の調査が完了したため、改めてご報告とお詫びを申し上げます。まずは本件につきまして、報道機関様および関係者様、弊社メディアをご利用のユーザー様に多大なご迷惑をおかけしましたこと、心よりお詫び申し上げます。

また公表時期につきまして、影響範囲の調査と対応、関係者への連絡などをおこなったことから、前回発表より期間が空いてしまったことも、重ねてお詫び申し上げます。

本件の概要

本件は、2023年08月22日に公開された記事（以下「問題記事A」）につきまして、複数の報道機関に掲載された文章をそのまま用い、剽窃・盗用に該当すると言える箇所が確認されたことに端を発します。当該記事は本誌記者による執筆ではなく、弊誌がβ版として開発・検証をおこなっていた生成系 AI によって生成された記事です。

[The HEADLINE, 2023]

生成AIで顕著になった倫理的課題

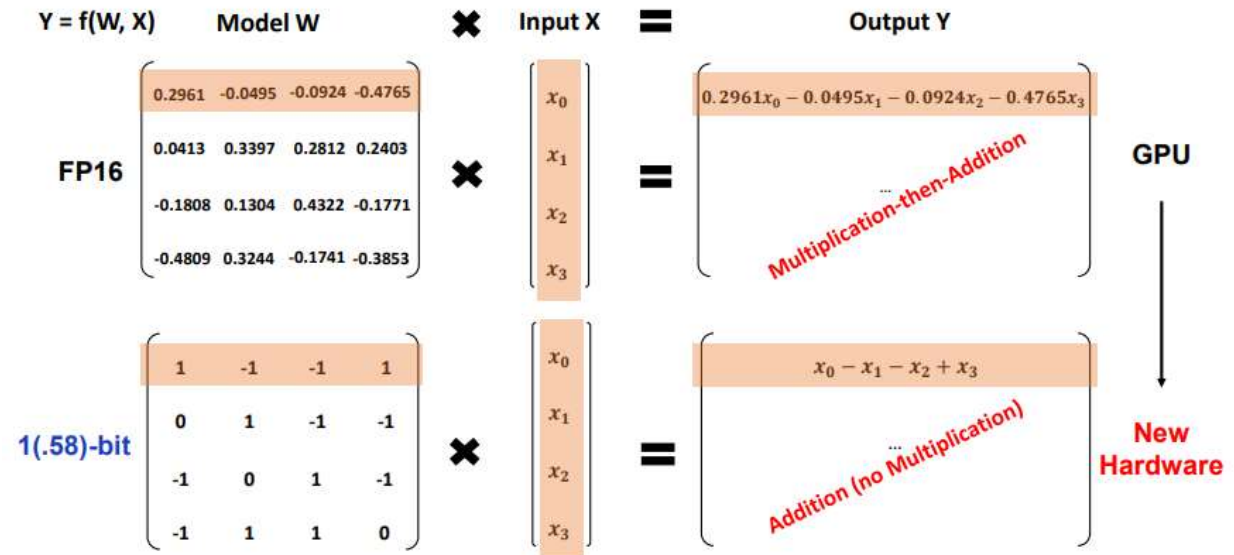
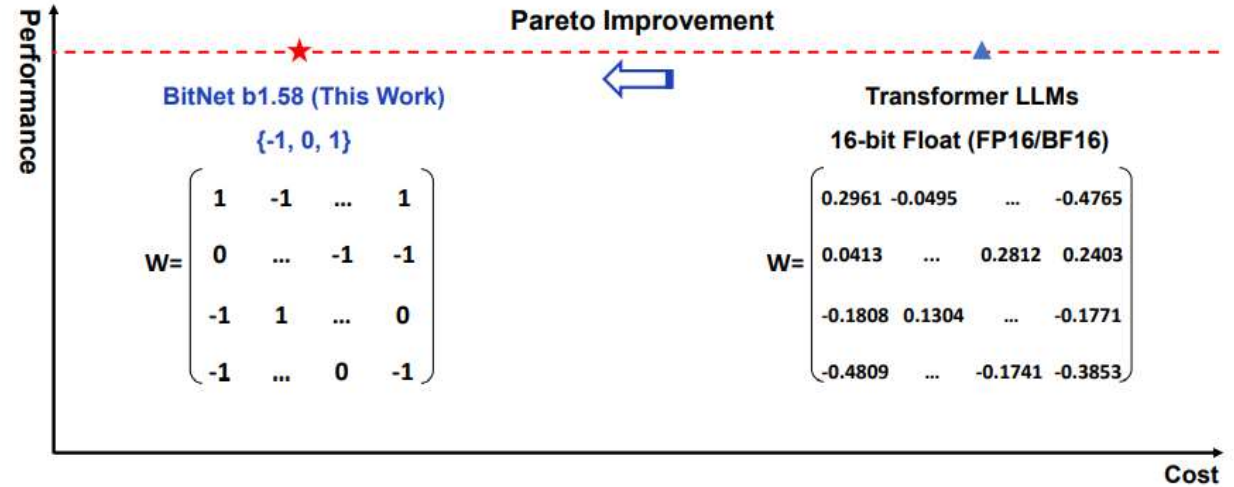
- アクセシビリティと包摂性

一部の人しか生成AIを使えない

- 電力消費とCO2排出

MicrosoftのGopherやOpenAIのGPT-3は原子力発電1基の1時間分の電力量を1回の学習に消費 [Stanford HAI, 2023]

→より電力消費の少ない生成AI (e.g., BitNet [Ma+, 2024])



[Ma+, 2024]

生成AIで顕著になった倫理的課題

• ディープフェイク

偽画像や偽動画の生成・拡散



[Suwajanakorn+, SIGGRAPH 2017]

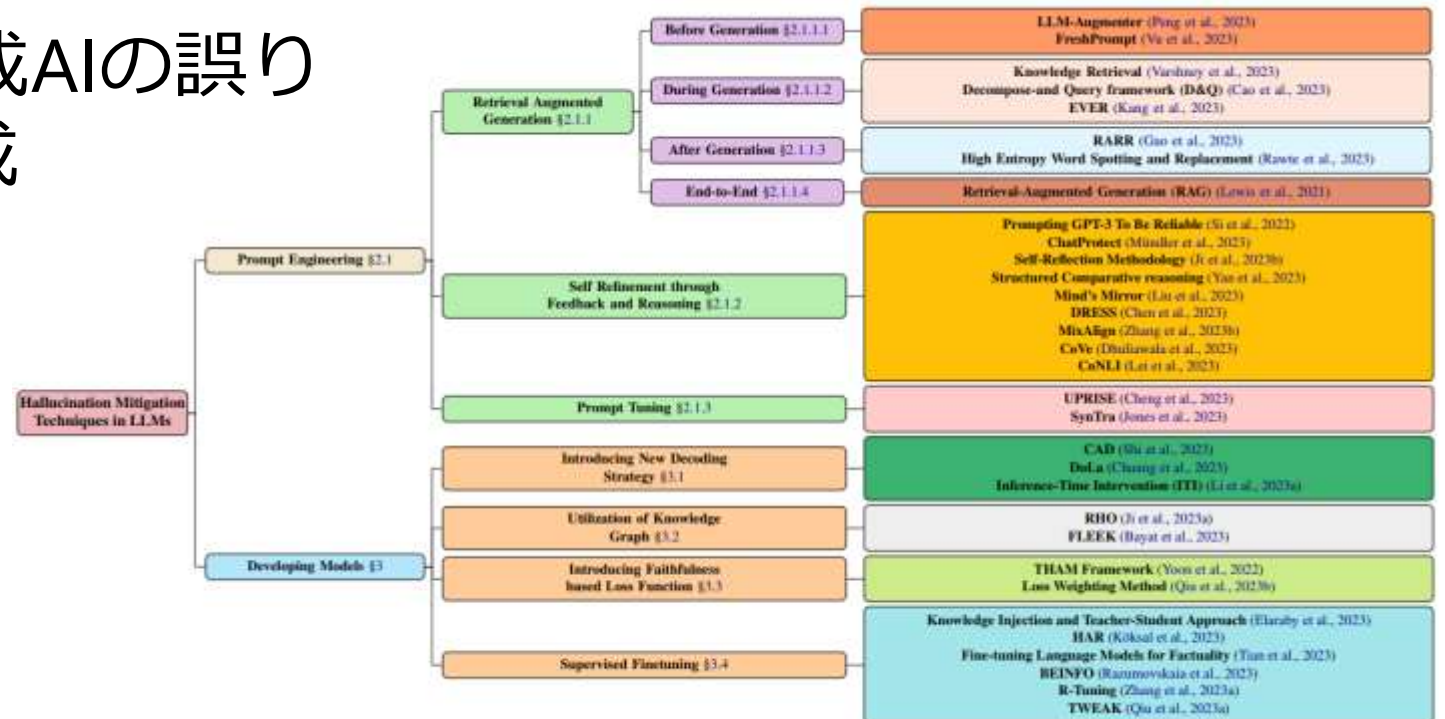
• ハルシネーション（幻覚）

ユーザーの意図しない生成AIの誤り

→実データに基づいた生成

(e.g. RAG)や前/後処理

[Tonmoy+, 2024]



[Tonmoy+, 2024]

各国の法規制：米国・欧州・日本

・ 欧州

- 事前規制型
- AI Act：3段階規制
 - ×危険行動を促すおもちゃ
 - ×ソーシャルスコアリング
 - ×顔認識等の生体認識
 - △人材採用AIシステム
 - スパムメールフィルタ

・ 米国・日本

- 自主規制型

各国のフロンティアAIモデル開発者に関する規制の比較

	米国	欧州	日本 (本素案)
規制内容	・15社による8項目の自主誓約 ・デュアルユース品については国防生産法の規律に	・許容できないリスクについては禁止、ハイリスクAIについては規制 ・開発者に対してはモデルが遵守すべき要件設定	・7項目の体制整備義務
罰則	・国防生産法に違反すれば罰則対象に	・罰則対象：最大3500万ユーロまたはグローバル売り上げの7%	・罰則あり。金額等は今後の検討
施行時期	・未定。但し、本年1月29日に規則案が公表	・法律成立後2年以内	・今後の検討

各国の法規制：中国・インド

• 中国

- 事前規制型
- 生成AIサービス管理暫定弁法
 - 違法・差別的コンテンツ防止
 - データ・アルゴリズムの取得元の合法性・権利義務の明確化

• インド

- 2023年4月「うちはAIを規制しません」
- 2024年3月「政府の明示的な許可が必要」

さいごに

特に最近の生成AIを中心とした流れとその課題

1. 画像や言語の基盤モデルとその応用
2. 科学研究する生成AIをつくる？！
3. 倫理と法規制からみた生成AI

✓ 基盤モデルも活用したAIやロボットによって知的創造が加速！

とはいえ...

× ハルシネーションを中心とした生成AIならではの倫理的課題

× 生成AIの能力向上に起因した直近の各国の法規制に向けた動き

→流動的な状況なので引き続き注視が必要