
ChatGPTとOpenAIは
どう変わろうとしているのか？

はじめに

ChatGPTの登場から、約一年が過ぎようとしています。その衝撃が冷めやらぬ今また、AIの世界に大きな変化が起きようとしています。

まず、11月6日のOpenAI DevDayの発表は。ChatGPTの新しい技術的な変化の方向を示しました。

同時に、11月17日のAltmanのCEO解任とその後の混乱は、OpenAIの内部に、OpenAIが進むべき進路をめぐって対立があることを内外に示しました。

今回のセミナーは、基本的には、 こうした事態の進行に触発されたものです。

このセミナーでは、最初に、この激動の一年にAIの世界に起きたことを振り返ろうと思います。

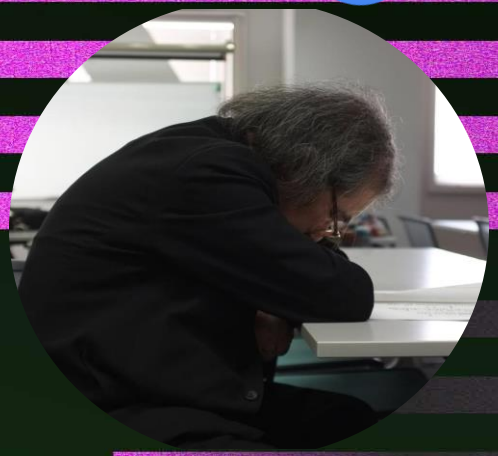
第二に、OpenAIという組織が、どのような成り立ちの組織なのかをお話しようと思います。それは今回の「対立」の背景を知るには必要なことだと考えています。

第三に、現在の人工知能技術の二つの技術的な焦点、AIの「マルチモーダル化」と「カスタム化」についてお話します。

最後に、AIの近未来像の話を、AIの「パーソナル化」を中心についてお話しようと思います。

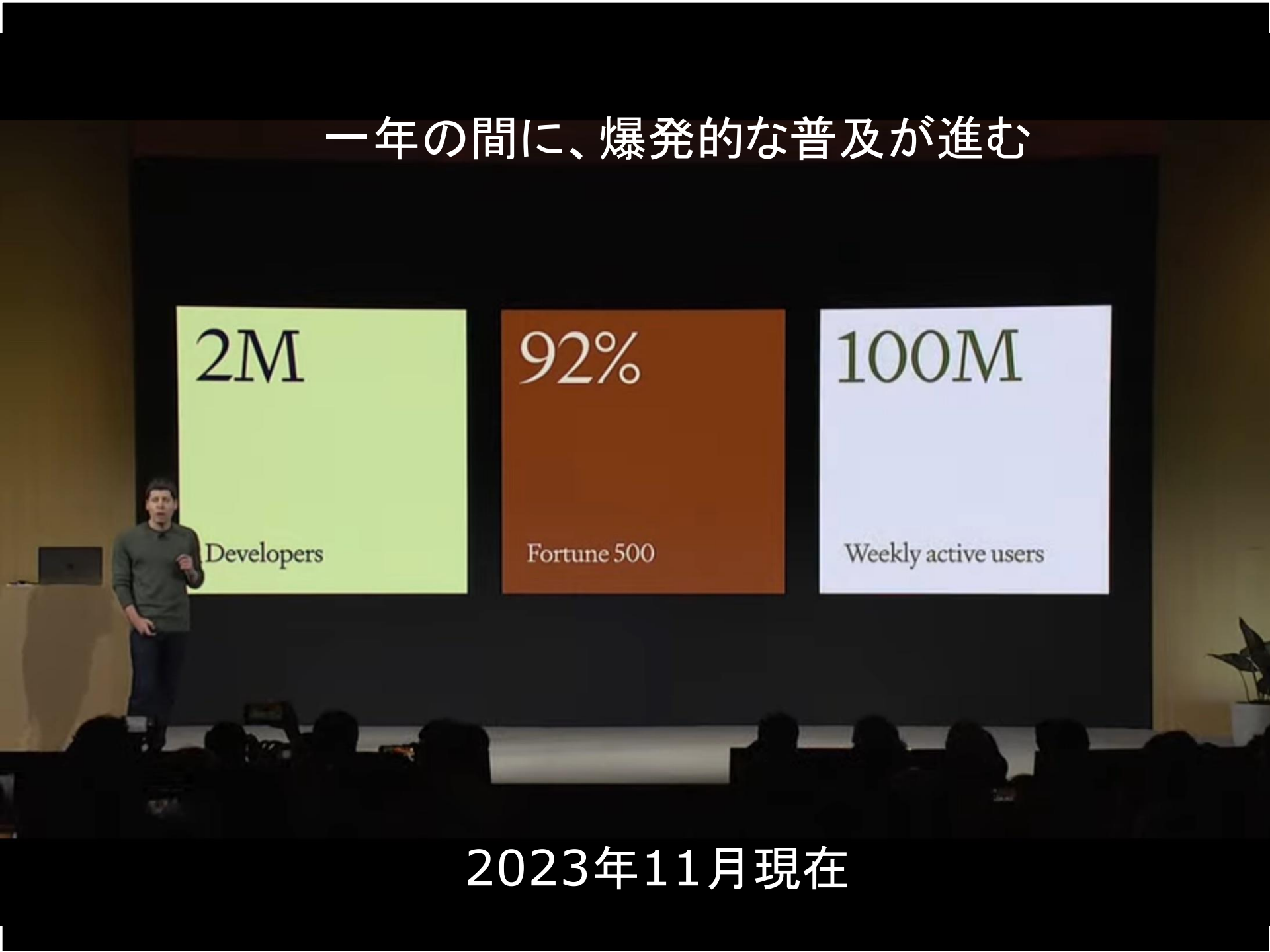

この一年の間に
AIの世界で起きたこと

ChatGPTの登場



2022/11/30

一年の間に、爆発的な普及が進む

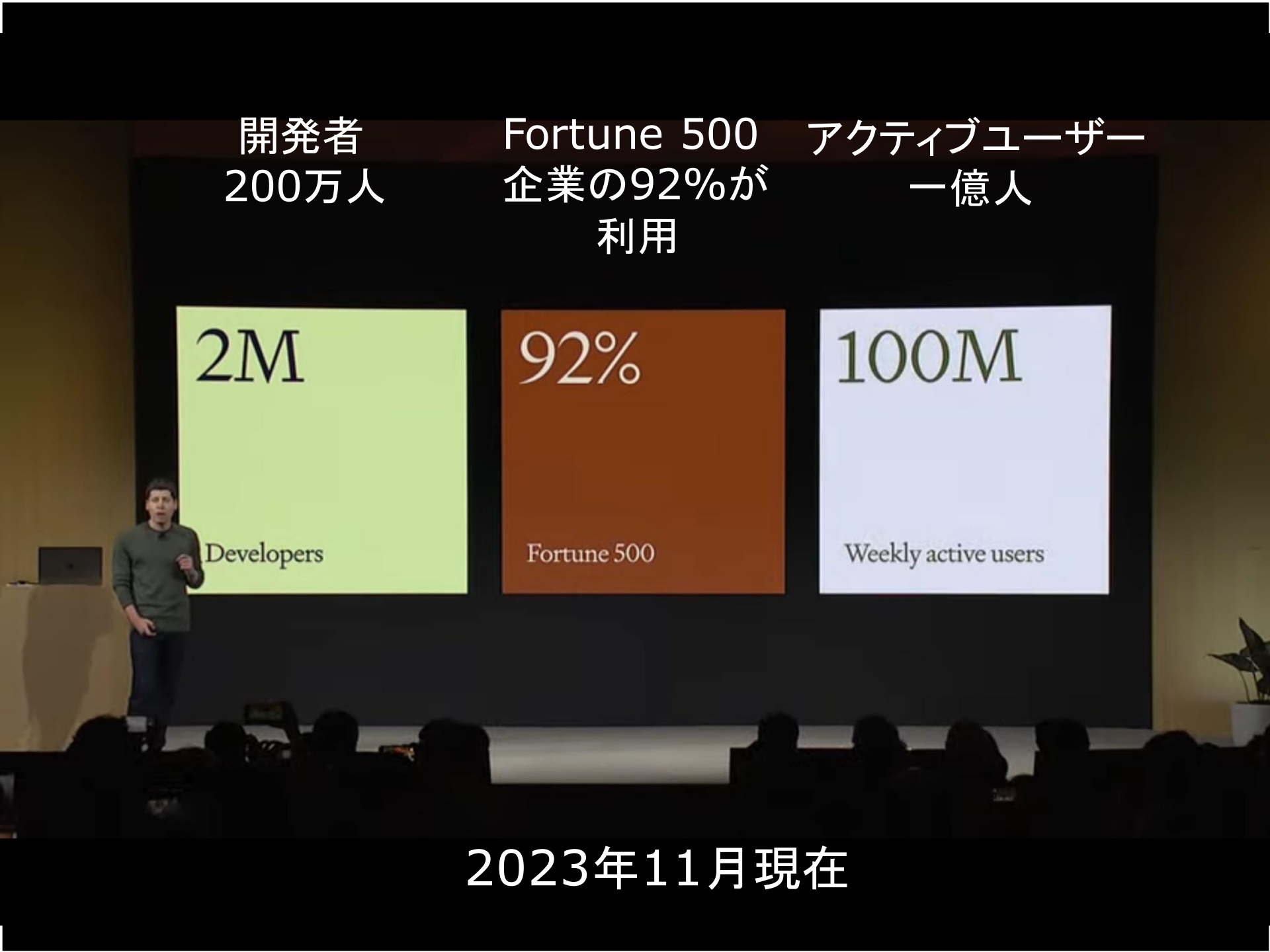


2023年11月現在

開発者
200万人

Fortune 500
企業の92%が
利用

アクティブユーザー
一億人



2023年11月現在

OpenAI GPT-4 をリリース

2023/03/14

「AIの父」ヒントン、Googleを辞める

2023/05/01

OpenAIが、危惧していること

GPT-4 で観察された
安全性への挑戦

- 幻覚
- 有害コンテンツ
- 悪意のある表現
- 偽情報と影響力操作
- 過信
- ・・・

2023/05/23

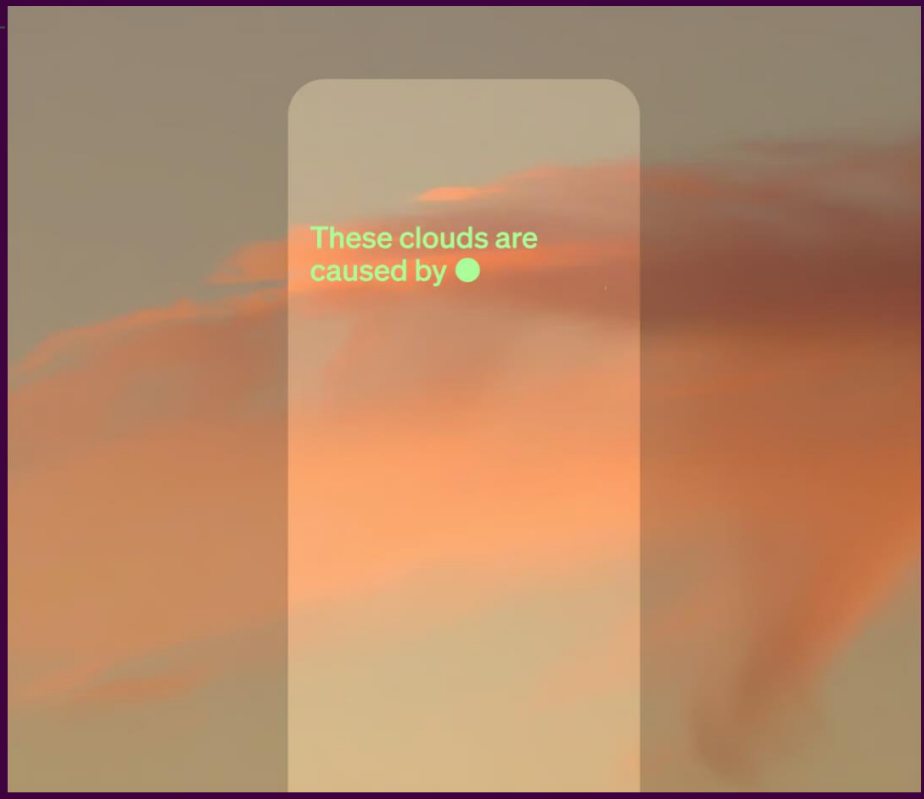
OpenAI GPT-4 System Card 論文

ChatGPTでは、新しい音声と画像機能を提供し始めています。
音声で会話したり、話している内容をChatGPTに見せることで、
より直感的な新しいタイプのインターフェイスを提供します。

2023/09/25

ChatGPT can now see, hear, and speak

We are beginning to roll out new voice and image capabilities in ChatGPT. They offer a new, more intuitive type of interface by allowing you to have a voice conversation or show ChatGPT what you're talking about.



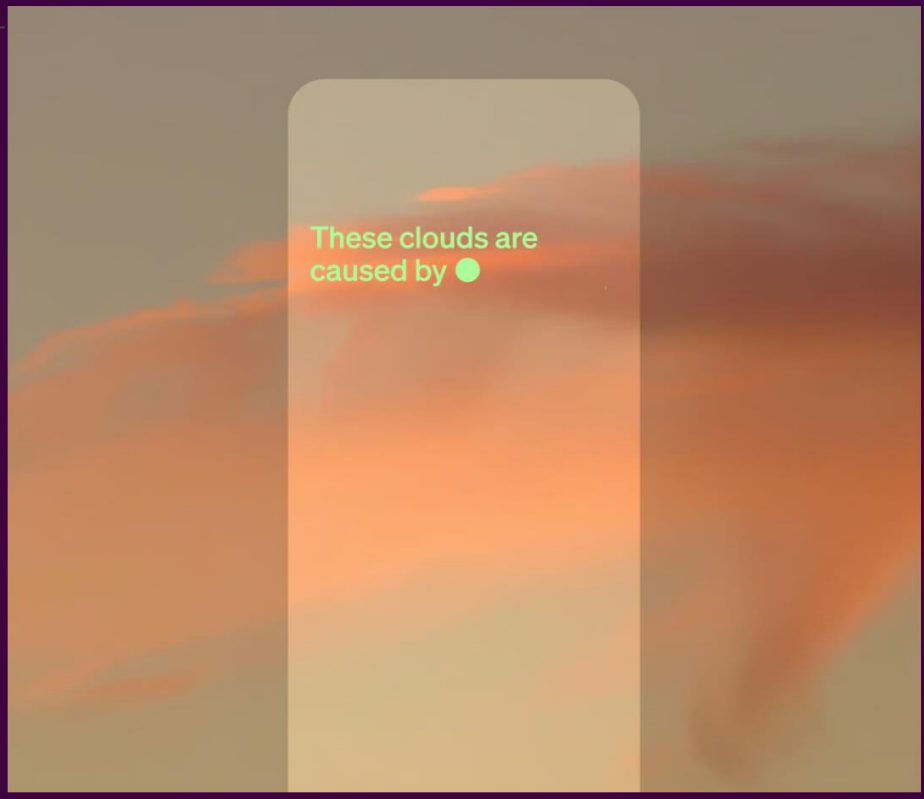
AIは、マルチモーダルへ

ChatGPTは、いまや、見ることも聞くことも話すこともできる

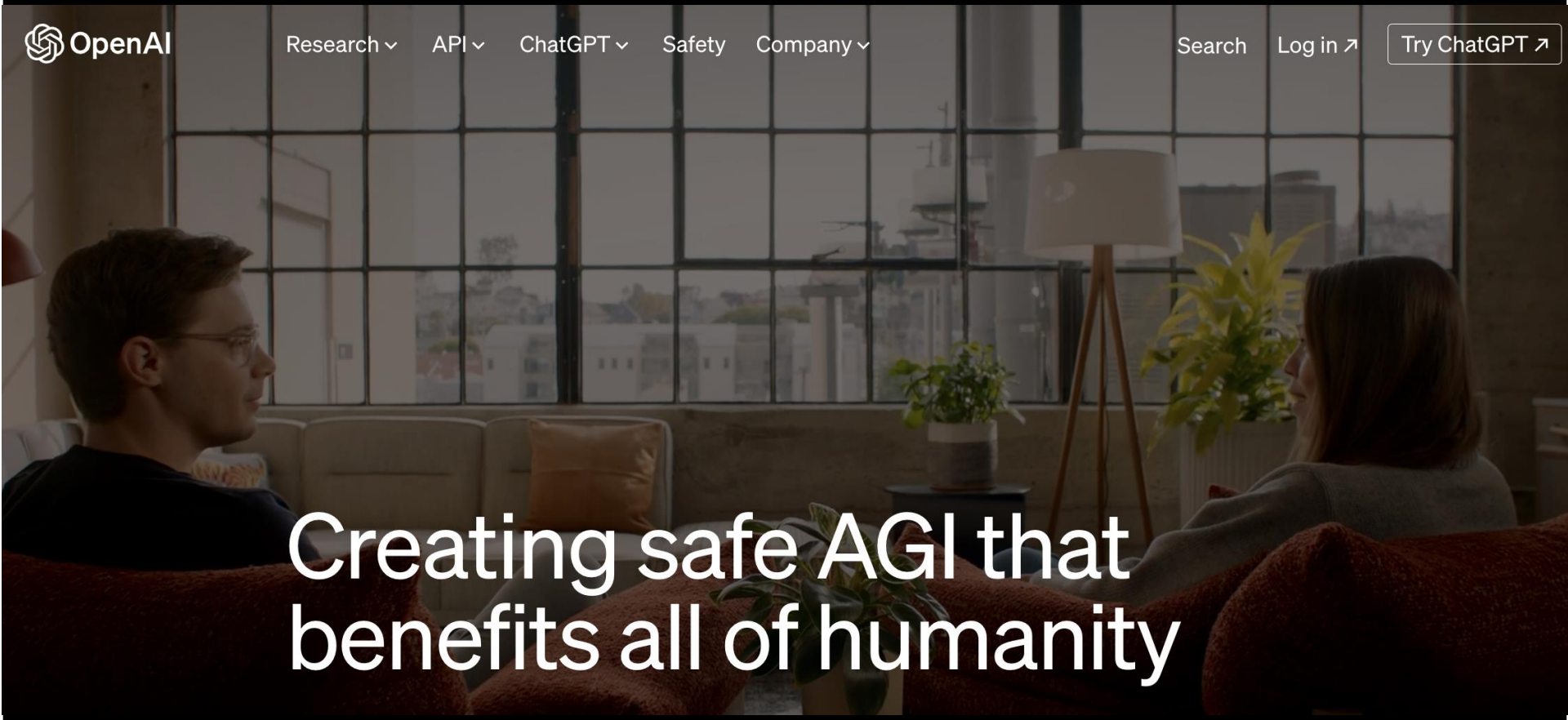
2023/09/25

ChatGPT can now see, hear, and speak

We are beginning to roll out new voice and image capabilities in ChatGPT. They offer a new, more intuitive type of interface by allowing you to have a voice conversation or show ChatGPT what you're talking about.



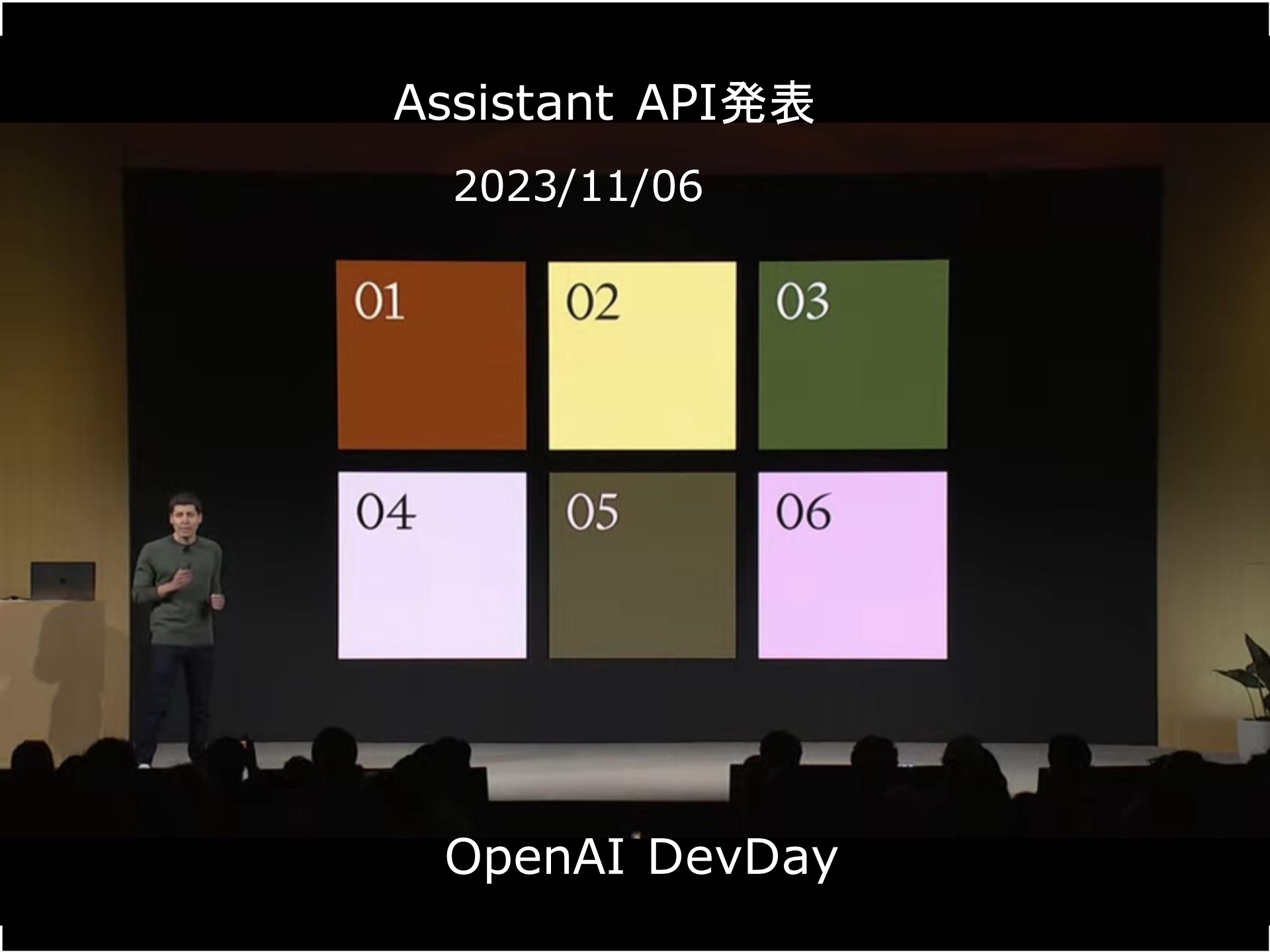
2023/10/23
OpenAIのトップページ



全人類の利益になる安全な汎用の人工知能を創造する

Assistant API発表

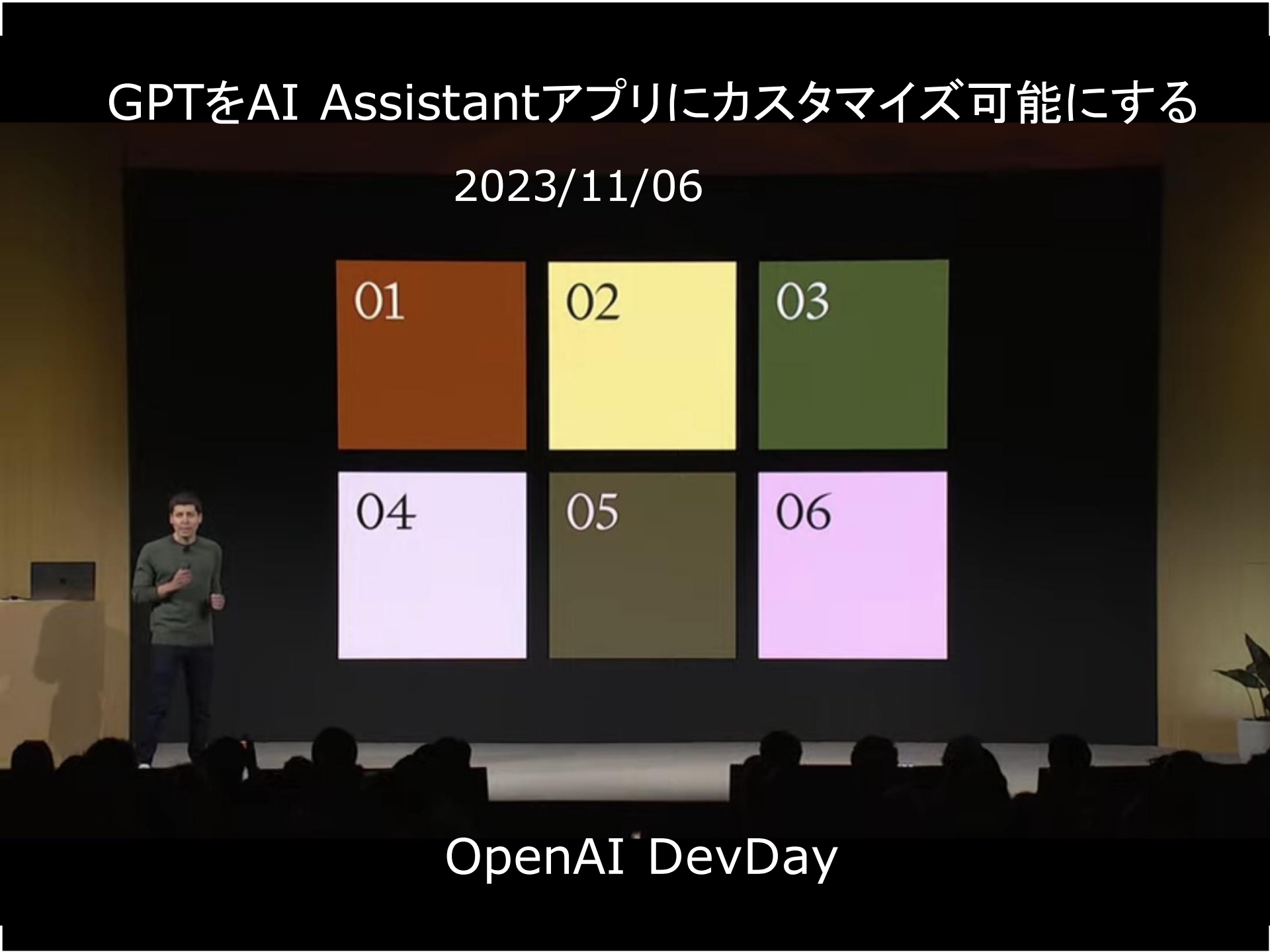
2023/11/06



OpenAI DevDay

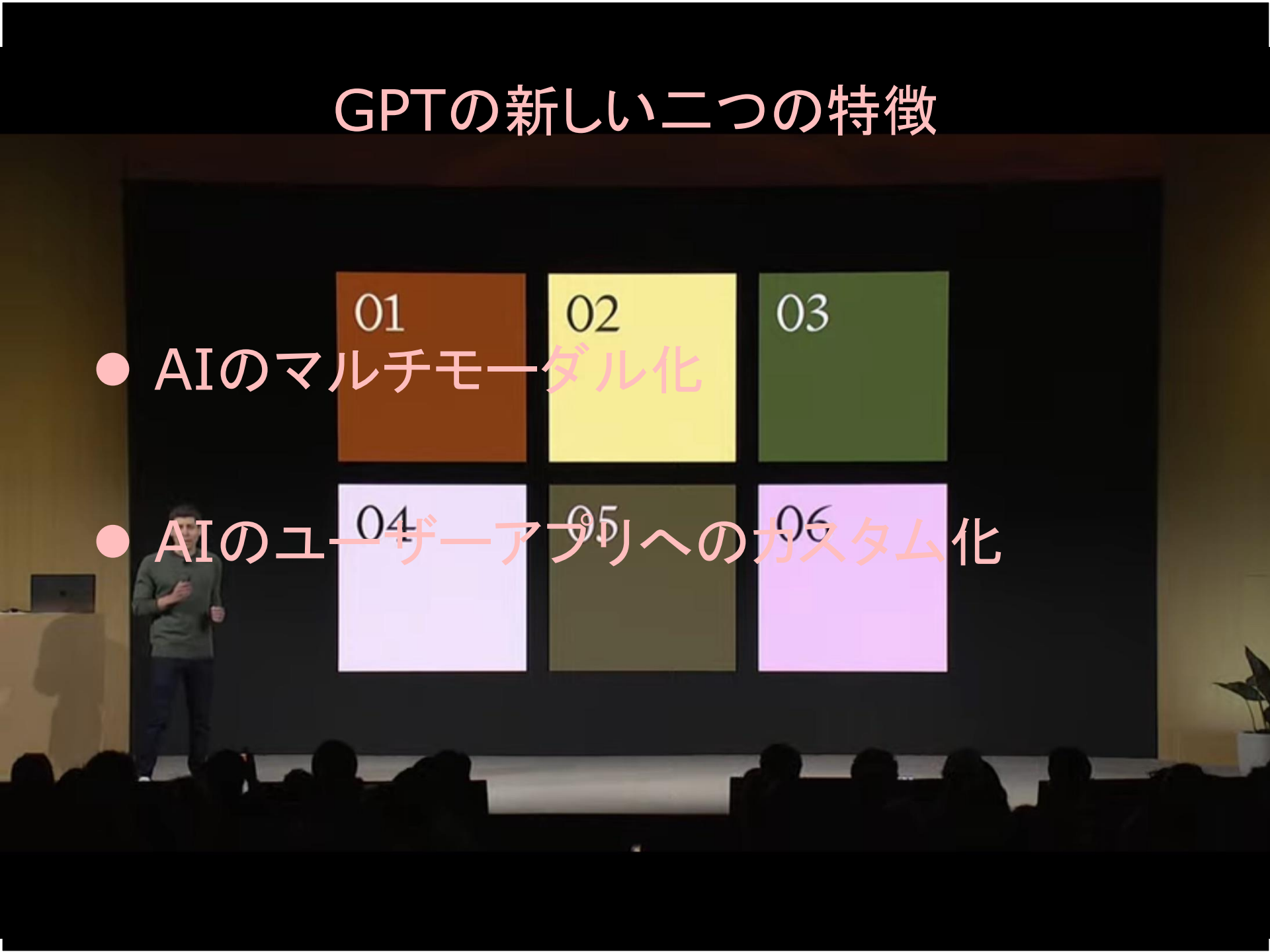
GPTをAI Assistantアプリにカスタマイズ可能にする

2023/11/06



OpenAI DevDay

GPTの新しい二つの特徴



- AIのマルチモーダル化
- AIのユーザーアプリへのカスタム化

GPTの新しい二つの特徴

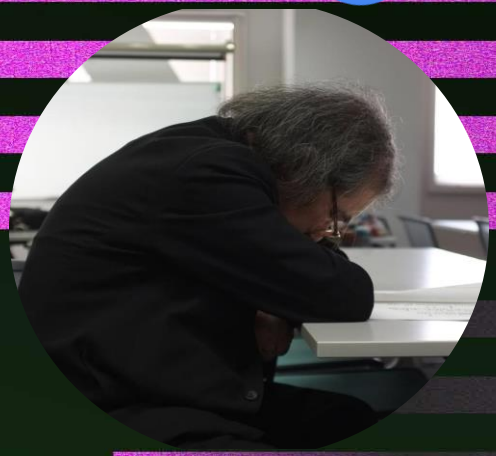
- AIのマルチモーダル化
- AIのユーザーアプリへのカスタム化

OpenAI DevDay


2023/11/17
Altman 解任
2023/11/19
Microsoftへ？

2023/12/06

ChatGPTの登場から
わずか一年である。




OpenAI について
2023年11月の内紛をどう捉えるか

OpenAI について

2023年11月の内紛をどう捉えるか

1. OpenAI創設の経緯
2. OpenAIの組織構造
3. 2023年11月 OpenAIに何が起きたか
4. Helen Toner 論文について

OpenAI創設の経緯

- 2000年の時点でのLarry Page のAIのビジョン
- Elon Musk の悪夢
- 2015年12月 OpenAI 設立

2000年の時点での
Larry Page のAIのビジョン

Google

Larry Page on AI

Larry Page on AI

「Googleの検索エンジンが、AIによって完全なものになったときにのみ、Googleのミッションは、完遂されるだろう。あなたたちは、それが何を意味するのか知っている。それが人工知能なのだ。」

「人工知能は、Googleの最終バージョンになるだろう。Web上のすべてのものを理解するだろう究極の検索エンジンは、あなたが望むものを正確に理解するだろうし、あなたに正しいものを与えるだろう。我々は、今は、そうしたことをするのには、遠いところにいる。ただ、我々は、少しずつ、それに近ずくことはできる。我々が取り組んでいることは、基本的には、そのことなのだ。」

http://goo.gl/OEL1oC

Larry Page on AI

「検索における我々の大きな目標は、人が望むものを、実際に正確に理解し、世界のすべてのものを理解することである。コンピューター科学者として、我々は、それを人工知能と呼ぶ。」

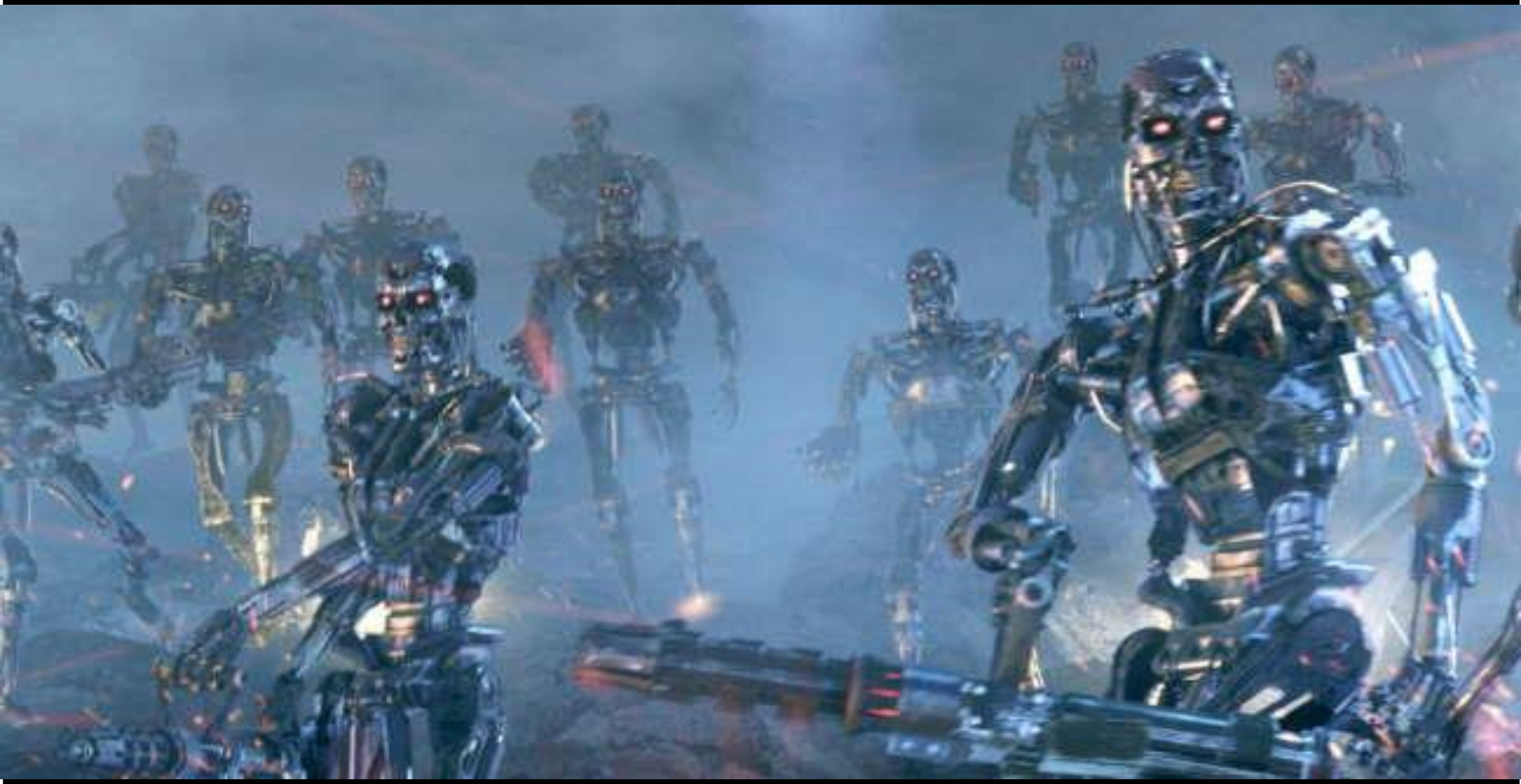
http://goo.gl/OEL1oC

Elon Musk の悪夢

“Elon Musk lives in fear of Google's killer robot army”

2015/05/28

“Elon Musk lives in fear of Google's killer robot army”



http://goo.gl/LdHPZA

Musk is genuinely worried that Page might just lead to the destruction of humanity as we know it.

"I'm really worried about this,"

"This," refers to the possibility that Page would develop artificially-intelligent robots that could turn evil and have the ability to annihilate the human race.

Page may be well-meaning, but as Musk says, "He could produce something evil by accident."

2015年12月、OpenAI設立

The organization was founded in December 2015 by Ilya Sutskever, Greg Brockman, Trevor Blackwell, Vicki Cheung, Andrej Karpathy, Durk Kingma, Jessica Livingston, John Schulman, Pamela Vagata, and Wojciech Zaremba, with Sam Altman and Elon Musk serving as the initial board members.

OpenAIの組織構造

- 2019年 OpenAI Structureの制定

2019年 OpenAI Structureの制定

OpenAI Structure
https://openai.com/our-structure

当初のOpenAI Nonprofitを設立してから約3年後の2019年に、私たちは「上限利益」構造を発表しました。

設立当初から私たちは、AGI（経済的に価値のある仕事において人間を凌駕する高度に自律的なシステム）を頂点とする強力なAIは、安全に対処すべきリスクとともに、社会を再構築し、多大な利益をもたらす可能性を秘めていると信じてきました。

現在のシステムの能力が高まっていることは、OpenAIや他のAI企業にとって、それぞれの使命と運営の中核となる原則、経済メカニズム、ガバナンスモデルを共有することがこれまで以上に重要であることを意味します。

Overview

私たちは、人類の利益のために安全で有益な人工知能を構築することを目標に、2015年後半に非営利団体OpenAIを設立しました。このようなプロジェクトは、以前であれば1つまたは複数の政府によるものであり、人類にとって広範な利益を追求する人類規模の取り組みであったかもしれません。

公共部門には明確な道筋がないと考え、また民間企業における他の野心的なプロジェクト（スペースX社、クルーズ社など）の成功を考慮し、私たちは公益への強いコミットメントに縛られた民間の手段でこのプロジェクトを推進することにしました。私たちは当初、501(c)(3)が、利益インセンティブに邪魔されることなく、安全で広く有益なAGIの開発を指揮する最も効果的な手段であると考えていました。私たちは、そうすることが安全であり、公共の利益になると思われる場合には、私たちの研究とデータを公表することを約束しました。

- OpenAIの非営利団体はそのまま残り、その理事会はOpenAIのすべての活動の統括団体として継続する。
- 新しい営利目的の子会社が設立され、資本を調達するために株式を発行し、ワールドクラスの人材を雇用することができますが、非営利団体の指示に従います。営利目的の取り組みに従事していた従業員は、新しい子会社に移行された。
- 営利企業はNPOの使命を追求する法的義務を負い、研究、開発、商業化、その他の中核業務に従事することでその使命を遂行する。OpenAIの指導原則である安全性と広範な利益は、そのアプローチにおいて中心的なものとなる。
- 営利目的の株式構造では、純粋な利益最大化に焦点を当てるのではなく、商業性と安全性および持続可能性のバランスをとった方法でAGIを研究、開発、展開するインセンティブを与えるため、投資家と従業員への最大財務リターンを制限する上限を設ける。

- NPOは、自らの運営に加え、理事会を通じてそのような活動すべてを管理・監督する。また、包括的なベーシックインカム研究の後援、経済効果研究の支援、OpenAI Scholarsのような教育中心のプログラムの試行など、幅広い慈善活動を継続する。NPOは長年にわたり、スタンフォード大学人工知能インデックス・ファンド、ブラック・ガールズ・コード、ACLU財団など、テクノロジー、経済効果、正義に焦点を当てた他の公益団体も数多く支援してきた。

そうすることで、非営利団体は私たちの組織の中心であり続け、AGIの開発をコントロールし、営利団体は、OpenAIのコアミッションを追求する義務を負いながら、これを達成するためのリソースを結集する任務を負うことになります。ミッションが何よりも優先されることは、すべての投資家と従業員が従う営利企業の運営契約書に明記されています：

IMPORTANT

****Investing in OpenAI Global, LLC is a *high-risk investment*****
****Investors could lose their capital contribution and not see any return****
****It would be wise to view any investment in OpenAI Global, LLC in the spirit of a donation, with the understanding that it may be difficult to know what role money will play in a post-AGI world****

The Company exists to advance OpenAI, Inc.'s mission of ensuring that safe artificial general intelligence is developed and benefits all of humanity. The Company's duty to this mission and the principles advanced in the OpenAI, Inc. Charter take precedence over any obligation to generate a profit. The Company may never make a profit, and the Company is under no obligation to do so. The Company is free to re-invest any or all of the Company's cash flow into research and development activities and/or related expenses without any obligation to the Members. See Section 6.4 for additional details.

2023年11月、
OpenAIに何が起きたのか？

- 2023年11月 OpenAIに何が起きたか
- Ilyaの心がわり

OpenAI について

2023年11月 OpenAIに何が起きたか

Brief departure of Altman and Brockman
2023年11月17日

2023年11月17日、サム・アルトマンは取締役会（ヘレン・トナー、イリヤ・スーツキーヴァー、アダム・ダンジェロ、ターシャ・マッコーリーで構成）の不信任に基づきCEOを解任され、最高技術責任者のミラ・ムラーティが暫定CEOに就任した。
OpenAIの社長であったグレッグ・ブロックマンは取締役会の会長から解任された。ブロックマンはこの発表の直後に同社の社長職を辞任し、彼が辞める前に起こった出来事のいくつかの詳細を報告した。

これに続いて、OpenAIの3人の上級研究者が辞任した。研究ディレクター兼GPT-4リードのヤクブ・パチョッキ、AIリスク責任者のアレクサンダー・マドリー、研究者のシモン・シドーである。

https://en.wikipedia.org/wiki/OpenAI

2023年11月18日

2023年11月18日、Altmanの退任を非難したMicrosoftやThrive Capitalなどの投資家が取締役会に圧力をかける中、AltmanがCEOに復帰する話があったと報じられた。
Altman自身はOpenAIに復帰することに賛成であると話したが、話がうまくいかなければ新会社を立ち上げ、OpenAIの元従業員を連れてくることを考えていると述べた。

アルトマンが復帰する場合、取締役会のメンバーは「原則的に」会社を辞職することで合意した。
2023年11月19日、アルトマンとの復帰交渉は失敗し、ムラーティはエメット・シアーに代わり暫定CEOに就任した。取締役会は当初、アルトマンの後任としてOpenAIの元幹部であるAnthropicのCEOダリオ・アモデイに接触し、合併を提案したが、両方の申し出は断られた。

2023年11月20日

2023年11月20日、マイクロソフトのサティア・ナデラCEOは、アルトマンとブロックマンが高度AIに関する新しい研究チームを率いるために同社に入社することを発表し、このような事態に陥ったにもかかわらず、彼らはOpenAIに引き続きコミットしていると述べた。
OpenAIの770人の従業員のうち、MuratiとSutskeverを含む約738人が、取締役会がAltmanをCEOとして再雇用し、その後辞任しないのであれば、仕事を辞めてMicrosoftに入社するという公開書簡に署名した。

投資家は、潜在的な大量辞任とAltmanの解任を受けて、取締役会メンバーに対して法的措置を取ることを検討している。これに対してOpenAIの経営陣は、Altmanと取締役会との交渉が再び進行中であり、しばらく時間がかかるという社内メモを従業員に送った。

2023年11月21日

2023年11月21日、継続的な交渉の後、AltmanとBrockmanは以前の役割のまま会社に戻り、Bret Taylor（会長）とLawrence Summersで構成される新メンバーとD'Angeloが残る取締役会が再構築された。

2023年11月22日、Sam AltmanのOpenAIからの解雇は、組織の極秘プロジェクトQ*における重要なブレークスルーを誤って処理した疑惑に関連している可能性があることを示唆する報道がなされた。OpenAIの情報筋によると、プロジェクトQ*は論理的推論と定理証明におけるAI能力の開発を目的としている。この開発に対するアルトマンの対応、特に発見の潜在的な安全性への影響に関する懸念は、彼が解雇される直前に会社の取締役会に提起されたと伝えられている。

To the Board of Directors at OpenAI

「OpenAIは世界をリードするAI企業です。私たちOpenAIの従業員は、最高のモデルを開発し、この分野を新たなフロンティアへと押し進めてきました。AIの安全性とガバナンスに関する我々の仕事は、グローバルな規範を形成しています。私たちが構築した製品は、世界中の何百万人もの人々に利用されています。これまで、私たちが働き、大切にしている会社がこれほど強い立場にあったことはありません。

あなたがサム・アルトマンを解雇し、グレッグ・ブロックマンを取締役から解任したプロセスは、この仕事すべてを危険にさらし、私たちの使命と会社を弱体化させました。あなたの行為は、あなたにOpenAIを監督する能力がないことを明らかにしました。"

あなたの決断を私たち全員が予期せず知ったとき、OpenAIのリーダーシップチームは会社を安定させるために迅速に行動しました。彼らはあなたの懸念に注意深く耳を傾け、あらゆる理由であなたに協力しようとしました。あなたの主張に対する具体的な事実を何度も求めたにもかかわらず、あなたは一度も証拠書類を提出しませんでした。彼らはまた、あなたが職務を遂行する能力がなく、不誠実な交渉をしていることに次第に気づいていった。

リーダーシップ・チームは、当社の使命、会社、利害関係者、従業員、そして世間一般に最も貢献できる、最も安定した前進の道は、あなたが辞任し、会社を安定的に前進させることができる有能な取締役会を設置することであると提案しました。リーダーシップは24時間体制であなたと協力し、互いに合意できる結果を探しました。しかし、あなたは最初の決断から2日も経たないうちに、会社の最善の利益に反して再びミラ・ムラーティ暫定CEOを交代させた。あなたはまた、会社の破壊を許すことが "使命に合致する "とリーダーシップ・チームに伝えた。

あなたの行動は、あなたがOpenAIを監督する能力がないことを明白にしました。私たちは、私たちの使命と従業員に対する能力、判断力、配慮に欠ける人たちのために、あるいは一緒に働くことはできません。私たちは、OpenAIを辞職し、Sam AltmanとGreg Brockmanが経営する新しく発表されたマイクロソフトの子会社に参加することを選択するかもしれません。マイクロソフトは、私たちが参加することを選択した場合、この新しい子会社にOpenAIの全従業員のためのポジションがあることを保証してくれました。現取締役全員が辞任し、取締役会がBret TaylorやWill Hurdのような2人の新しい主席独立取締役を任命し、Sam AltmanとGreg Brockmanを復職させない限り、私たちは間もなくこのステップを踏むでしょう。

Ilyaの心がわり

1. Mira Murati
2. Brad Lightcap
3. Jason Kwon
4. Wojciech Zaremba
5. Alec Radford
6. Anna Makanju
7. Bob McGrew
8. Srinivas Narayanan
9. Che Chang
10. Lillian Weng
11. Mark Chen
12. Ilya Sutskever"


Helen Toner論文について
OpenAI について

Helen Toner論文について Agenda

- Helen Toner論文について
  - 「三つのAI」論 -- 軍事的AIと民主的AIと民間セクターのAI
- 軍事的AI
  - 軍事的AIと自律型兵器に歯止めはかかっていない
  - 軍事と民事との「dual use」の危険性
- 民主的AI
  - 民主的AIの理念の国際的共有の進行
  - 民主主義国家と権威主義的国家のAIと国際政治の現実
- 民間セクターのAI
  - 民間セクターのAI　--「底辺への競争」の危険
  - OpenAIの活動への厳しい評価
- まとめ

Helen Toner論文の翻訳

MaruLaboのページ「Helen Tonerらの「民主的AI」論」
https://www.marulabo.net/docs/helen-toner/
に彼女の論文“Decoding Intentions -- Artificial Intelligence and Costly Signals ”の全訳があります。
https://cset.georgetown.edu/wp-content/uploads/CSET-Decoding-Intentions.pdf
の全文の翻訳を掲載しました。参照ください。

OpenAIについて

Helen Toner論文について

「三つのAI」論
軍事的AIと民主的AIと民間セクターのAI

Helon Tonerの論文

	軍事的AI	民主的的AI	民間セクター
Tying hands	一方的な政策声明を発表し、意図を伝える。 核の指揮統制に関する意思決定	AIを活用した民主主義社会を攻撃する敵対的な攻撃に対して、あらかじめ定義された行動をとることを約束することで、民主的なAIの原則を守る。	学習データ、モデルの性能、危険な能力に関する透明性など高度なAIモデルに関する重要な情報を公開する。
Sunk costs	訓練中および配備前のレッドチーム編成手順に投資し、AI対応兵器システムの帰属を容易にするエンブレムの使用を検討する。	AI技術が悪用されるシステミックリスクがある市場で事業を行う民間企業向けのデューデリジェンス指針を公表する。	信頼できるホスティングサービスと、テストベッドやその他の施設を含むテスト・評価インフラに投資する。
Installment costs	AI対応システムの持続的検証技術にコミットし、集中的なコンピュートアカウンティングのための取り決めを開発する。	AI監査人のための共通の認証基準、ツール、慣行を開発する。	リアルタイムのインシデント監視と、AI対応システムが関与するインシデントのデータ収集と分析に関する共通基準にコミットする。
Reducible costs	要件を設定し、解釈可能なAIモデルや代替設計原則に投資するインセンティブを設ける。	AIの安全性に関する研究や民主主義的価値を促進するプライバシー向上技術の開発に対する賞金コンテストを主催する。	AIの影響評価とAIシステムの内部監査結果を公表する。

https://cset.georgetown.edu/wp-content/uploads/CSET-Decoding-Intentions.pdf

軍事的AI

軍事的AIと自律型兵器に歯止めはかかっていない
軍事と民事との「dual use」の危険性

軍事AIと自律兵器

「2014年以来、各国はジュネーブに集まり、このような兵器の潜在的な使用に関する原則を策定してきた68。政策立案者たちは、自律型兵器システムを採用する決定において、国際法がどこでどのように適用されるのか、また人間の判断が果たす重要な役割について議論してきた69。米中両国はこのプロセスに参加し、これらの技術に関連する問題を検討するために2016年に設立された国連機関である「致死的自律型兵器システムに関する政府専門家グループ（GGE）」の合意文書に合意した。」

「2019年、特定通常兵器に関する条約の締約国は、説明責任、人的責任、LAWS(Lethal Autonomous Weapons Systems)の開発と潜在的使用に対する国際人道法の適用を含む11の指導原則を採択した70。しかし、これらの合意文書の背後には、自律型兵器の定義と、国際法の遵守を確保するために必要な人間の関与のレベルをめぐる実質的な意見の相違が横たわっている。2019年以降、各国はこうした相違を調整するのに苦労し、勢いは停滞している。」

「軍事AIと自律性に関するシグナルを解読することは、第三の課題に直面している。最先端のAI技術を開発する多国籍企業は、本社を一つの国に置いているかもしれないが、グローバル化したサプライチェーンを持つ世界的なAI研究企業の一部である。彼らの決定は国家の優先順位を反映することもあるが、企業は何よりもまず、株主の要求、金融市場、貿易の流れ、国際的な経済動向に左右される。さらに問題を複雑にしているのは、AIが民間および軍事に幅広く応用される汎用技術であることだ。営利団体と政府がパートナーシップを組んでデュアルユース技術を開発しても、結局は軍事イノベーションを支援することになりかねない。
防衛産業基盤と民間企業を融合させた「技術安全保障国家」を発展させようとする中国の努力はよく知られているが、この戦略の成功を測定することは難しい。」

「ある国防総省（DoD）の元高官は米中経済安全保障再検討委員会で証言したように、「入手可能な証拠は、中国がAIを活用した殺傷能力のある自律型兵器の開発を追求していることを示唆している」78。この主張を補強するために、この元高官は中国が2017年の新世代AI開発計画、2021年から2026年の第14次5カ年計画、および最新の国防白書でAIを戦略的優先事項として定義していることを指摘した。また、中国第3位の防衛企業の幹部の発言も引用した。この幹部は、各国が戦場でAIと自律性を統合し続けることに自信を示した：「将来の戦場では、人が戦うことはないだろう」79 このような発言と一致して、元国防総省幹部は、中国の軍用無人機メーカーZiyanのBlowfish A2モデルなど、自律機能を備えた軍用無人機や武装ドローンを中国が輸出していることを強調した。元国防総省高官は、AI搭載兵器の安全性の問題を認識しながらも、中国人民解放軍が国防総省との防衛政策対話を拒否しているのは、中国がLAWSを開発し、国際規範の制約を受けないことを意図している証拠だとしている。」

「ロシアの兵器メーカーであるカラシニコフの子会社であるドローンメーカーは、この兵器は「（センサーの）ペイロードの照準画像」から座標を取得できると主張している90。これら2つの声明はいずれも、ウクライナのドローンにAIが搭載され、人間の操作とは無関係に標的を選択し交戦したことを意味するものではないが、各国政府がそうでないと仮定するのは無理からぬことだろう。同様に、ウクライナ人は英国のブリムストーン・ミサイルを運用している。このミサイルの開発元は、「天候に左右されない照準、キルボックスに基づく識別、一斉発射を提供する」「ファイア・アンド・フォーゲット」モードなど、いくつかの動作モードを宣伝していた91。専門家がすぐに指摘したように、この兵器は現在、半自律モードで動作している可能性が高いが、ソフトウェアのアップデート次第で、完全自律兵器への曖昧な閾値を超える可能性がある。92」

民主的AI

民主的AIの理念の国際的共有の進行
民主主義国家と権威主義的国家のAIと国際政治の現実

民主的AIと不用意なシグナル

「民主的AIは、多国間のフォーラムや各国のAIに関する声明で広く議論されるようになった。これらの声明に基づく民主的AIの広義の定義は、民主的プロセスと社会に対するセーフガードをその開発と展開に組み込んだAIアプリケーションと、民主主義を保護する将来の規制を指す。
その例としては、データが乏しかったりアルゴリズムの設計が悪かったりすることで、システムが特定の市民層に偏らないようにすること、政府が市民の市民的自由を侵害するような方法で、顔認識やその他のプライバシーを侵害する可能性のあるAIアプリケーションを使用しないようにすること、敵対者や悪意ある行為者が生成モデルを使用して情報環境を混乱させ、選挙や法の支配に対する信頼を損なわないようにすることなどが挙げられる。」

多国間および国家レベルの政府声明は、その詳細さや具体性のレベルこそ異なるものの、一般的に民主的AIのこの理解を支持している。

例えば、2023年に日本で開催されたG7サミットで、G7諸国は「我々の共有する民主的価値観に沿って、信頼できるAIという共通のビジョンと目標を達成するために、包括的な（AIの）ガバナンスと相互運用性に関する国際的な議論を進める」決意を表明した。欧州連合（EU）の人工知能法草案は、2023年6月に採択された新たな修正案とともに、「人間中心で信頼できる人工知能の導入を促進し、健康、安全、基本的権利、民主主義、法の支配の高水準の保護を確保する」ことを目指している。109

その他、AIの開発とガバナンスに民主的価値を求める注目すべき多国間グループには、OECD、欧州評議会、AIに関するグローバル・パートナーシップ、国連教育科学文化機関（ユネスコ）、フリーダム・オンライン連合、米EU貿易技術評議会などがある。EU貿易技術評議会などである（付録A参照）110。

「志を同じくする米国のパートナーは、特に多国間声明の共同署名国である場合、民主的AIに関するシグナルの明確な受信者である。あるいは、署名国が以前に合意した原則を守らなかった場合、風評被害や民主的な同盟国からの外交的圧力につながる可能性もある。このような米国とその同盟国との違いは、手をつなぐメカニズムを通じて高価なシグナルを発する機会を生み出す。一般市民、市民社会団体、メディアを含む国内の聴衆は、将来、指導者に説明責任を負わせるために、AIの原則に関する公約を利用するかもしれない。研究者やシンクタンク、業界団体、非政府組織を含むジャーナリストや利益団体は、政府が民主主義的価値観や市民的権利に反するAIの使用や開発を許可した場合、過去の発言に国民の注意を向けさせ、指導者に国内政治的コストを生じさせる可能性がある。」

「政策立案者がこのような異なる聴衆に送るシグナルを考えると、民主的AI、特に権威主義との対立というフレーミングは、民主主義国家のアプローチと競争相手のアプローチを区別するための有用な略語かもしれない。しかし、このような枠組みは、民主主義国家が自国の利益と安全を守るために権威主義政府と頻繁に協力しているという、より複雑な現実を裏切るものである。

さらに、民主主義国家はしばしば、中国との競争の中でグローバル・スイング・ステートとの関係を強固にする必要性を強調することで、そうした協力を擁護する125。民主的AIに関する発言だけで、必ずしも中国に近づくとは限らないが、民主的AIと権威主義的に開発されたAIの品質が同等である場合、非民主的なパートナーは後者の技術を無条件で採用することを選ぶかもしれない126。」

「湾岸協力会議（GCC）の君主制国家は、民主主義をAIと関連づけることが外交的・戦略的課題を生み、安全保障に悪影響を与えかねない権威主義的な米国のパートナーの例を示している。
サウジアラビア、アラブ首長国連邦（UAE）、カタール、バーレーン、クウェート、オマーンは、エネルギーと安全保障上の利益に関する協力と引き換えに、長年にわたるアメリカの安全保障を前提とした戦略的関係を、個々に、あるいは集団的にアメリカと培ってきた127。 今日、GCC 諸国は、3 万人以上の米軍兵士、軍事領域にわたる複数の米中央軍司令部 （CENTCOM）、多国籍海上機動部隊を受け入れており、湾岸全域に少なくとも 20 の基地施設へのアクセスを提供している128 。」

「過去 20 年間の米軍 CENTCOM での協力は、情報共有、政治交渉の支援、さらには対テロ共同作戦を特徴としてきた。その重要性にもかかわらず、米国と湾岸諸国の関係は困難であり、対立的ですらある。緊張の原因は、政策や脅威に対する評価の違いから、湾岸諸国における反対意見の抑圧、市民の自由、女性や少数民族、移民労働者の権利などに関する米国の正当な懸念にある。米国の議員や市民社会は、米国が湾岸諸国と距離を置くことを求め、注目される批判を先導してきた。特にサウジアラビアなどである。」

民間セクターのAI

民間セクターのAI　--「底辺への競争」の危険
OpenAIの活動への厳しい評価

民間のシグナリング

「現代の特筆すべき特徴は、20世紀の大半の時代とは異なり、戦略的技術はもはや政府が運営したり資金を提供したりする研究所で主に開発されるものではなくなっていることだ。AIも例外ではなく、最先端のシステムの多くは消費者向けのテクノロジー企業で開発されている。技術が開発される場所の重心がこのように変化しているということは、政府と民間セクターが深く関わり合っているということであり、関連するシグナルはより広範な主体によって発信される可能性がある。殺傷能力のある自律型兵器や民主的なAIをめぐるシグナルに関するケーススタディが示すように、AIの開発と利用の軌跡を予測しようとするオブザーバーは、今や政府からのシグナルだけでなく、ウクライナにおける主要な技術プラットフォームの継続的な貢献のように、紛争環境において不可欠な機能やサービスを提供するようになっている様々な業界関係者からのシグナルにも注意を払わなければならない144。」

「AI開発のアナリストの間で長年懸念されているのは、競争力を維持するために複数のプレーヤーが安全性やセキュリティの課題を軽視するプレッシャーを感じる「底辺への競争」の可能性である。このシナリオでは、認識とシグナルが重要な変数となる。ほとんどの当事者は、AI システムの信頼性を確保するための時間を確保したいと考えるだろうが、一番になりたいという願望、市場投入へのプレッシャー、競合他社が手抜きをしているかもしれないという考えは、すべて開発者の慎重さを失わせる可能性がある。AIシステムを開発する当事者は、自制へのコミットメント、安全で信頼できるシステムの開発への注力、またはその両方を強調することができる。理想的には、これらの点について信頼できるシグナルを発信することで、他の締約国を安心させることができる。」

「民間セクターの主体がどのようにコストのかかるシグナルを送ることができるかをより完全に理解するためには、AIを責任を持って開発するというコミットメントを示すために、公的な声明を超えた行動をとっている主要なAI企業の2つの例を検討する価値がある：OpenAIはGPT-4モデルの発表と同時に「システムカード」を公表し、AnthropicはチャットボットClaudeのリリースを延期することを決定した。

これらの例はいずれもLLMを開発している企業のもので、LLMは2022年11月にOpenAIがChatGPTをリリースしたことで一躍脚光を浴びたAIシステムの一種である147。LLMはテキスト中の次の単語を予測するように設計されており、翻訳、プログラミング、要約、詩の執筆など様々なタスクに役立つことが証明されている。このような多用途性は、LLMを有用なものにしているが、同時に、情報の捏造、偏見の蔓延、悪用されるコンテンツの作成、危険な活動の障壁の低下など、LLMがもたらすリスクを理解し、軽減することをより困難なものにしている。」

「2023年3月、カリフォルニアを拠点とするOpenAIは、LLMシリーズの最新版をリリースした。GPT-4（GPTは〝generative pre-trained transformer〞の略で、LLMがどのように構築されたかを表す言葉である）と名付けられた新しいモデルは、LLMの言語理解をテストするために設計された、いくつかのベンチマークで新記録を樹立するなど、さまざまなタスクで素晴らしいパフォーマンスを示した。しかし、シグナリングの観点からは、GPT-4 のリリースで最も興味深かったのは、その能力を詳述した技術報告書ではなく、このモデルがもたらす安全上の課題と、リリース前に OpenAI が実施した緩和策をまとめた 60 ページのいわゆる「システムカード」であった。」

システムカード自体は、GPT-4のリスクプロファイルを理解することに関心のある研究者の間では好評だが、安全性に対するOpenAIのコミットメントを広く示すものとしては、あまり成功していないようである。この意図しない結果の理由は、同社がシステムカードの重要性を覆すような他の行動をとったからである。最も顕著なのは、その4ヶ月前にChatGPTをリリースしたことである。比較的目立たない "研究用プレビュー "として意図されたオリジナルのChatGPTは、GPT-3.5と呼ばれるそれほど高度でないLLMを使って構築された。GPT-3.5が事前に流通していたことが、おそらくOpenAIが今回このような詳細な安全性テストを実施したり公開したりする必要性を感じなかった理由だろう。それにもかかわらず、ChatGPTのリリースの1つの大きな効果は、大手テック企業内に危機感を呼び起こすことだった149。チャットボットに対する顧客の熱狂の波の中でOpenAIの後塵を拝することを避けるために、競合他社は社内の安全性と倫理審査プロセスを加速させたり、回避しようとした。

「この結果は、OpenAIなどが避けたいと表明している底辺への競争力学に酷似しているように思われる。

オープンAIはまた、ChatGPTとGPT-4のローンチに関連した、著作権問題、データ注釈者の労働条件、ユーザーが安全制御を回避することを可能にする「ジェイルブレイク（脱獄）」に対する自社製品の脆弱性など、他の多くの安全性と倫理の問題についても批判を浴びている151。この混濁した全体像は、意図的なシグナルによって送られたメッセージが、意図を明らかにするために意図されたものではない行動によって覆い隠されてしまう例を示している。

OpenAIの主な競争相手の1つであるAnthropic社は、民間企業におけるシグナリングに対して異なるアプローチをとっている。Anthropic社は、安全性を重視する企業であると認識されたいという願望が、そのキャッチフレーズから始まるコミュニケーション全体を通して輝いている：同社の意思決定を注意深く見てみると、このコミットメントは言葉だけにとどまらないことがわかる。

2023年3月にAnthropicのウェブサイトで公開された戦略文書によると、ChatGPTの競合であるAnthropicのチャットボットClaudeのリリースは、「AI能力の進歩速度を早める」ことを避けるために意図的に延期されたことが明らかになった153。2023年初頭にClaudeをユーザーと共有し始めるという決定は、文書によると「一般的な技術状態とのギャップが小さくなった今」なされたもので、Claudeがベータテストに入る数週間前にChatGPTがリリースされたことを明確に言及している。

言い換えれば、AnthropicはAIの誇大広告の炎をあおらないために、意図的に技術を製品化しないことにしたのだ。似たような製品（ChatGPT）が他社からリリースされると、Claudeをリリースしないこの理由は無効になり、Anthropicは3月にClaudeを製品として正式にリリースする前に、テストユーザーにベータ版アクセスを提供し始めた。

Anthropicの決断は、AIの安全性に関する "底辺への競争 "の力学を減らすための別の戦略を示している。GPT-4システムカードが、OpenAIが安全なシステムを構築することに重きを置いていることを示す高価なシグナルとして機能したのに対し、Anthropicの製品を市場に出さないという決定は、代わりに高価な抑制のシグナルとなった。他社が同じような性能の製品を出すまでClaudeのリリースを遅らせることで、Anthropicは、ChatGPTのリリースが拍車をかけたと思われるような必死の手抜きを避ける意志を示したのです。

まとめ


現在の人工知能技術の技術的な焦点
AIのマルチモーダル化とカスタム化


Google Vision Transformer
AIのマルチモーダル化とカスタム化

大規模言語モデルから Multimodalな人工知能へ

現在の人工知能技術の技術的な焦点の一つは、「Multimodalな人工知能」の実現にあります。ここでは、大規模言語の上にMultimodalな人工知能を実現しようとする動きを紹介しようとおもいます。

マルチモーダルな人工知能とは、現在のテキスト中心の人間と人工知能のインターフェースを大きく変える「見ることも聞くことも話すこともできる」インターフェースを備えた人工知能のことです。

ただ、AIが「聞くこと話すこと」と比べて、AIが「見ること」を実現するのは技術的には様ざまな難しさがあります。ですから、マルチモーダルなAIを目指す技術の大きな関心は、AIが「見ること」の実現にむけられていると僕は考えています。

Vision Transformer とは何か？

大規模言語モデルがMulti-Modal なAI に展開して上で、大きな役割を果たしたシステムがあります。それが、2021年に Google が発表した Vision Transformer です。

自然言語処理の世界では、Transformerベースの大規模言語モデルが大きな成功を収めていたのですが、画像情報処理の世界では、近年に至るまで CNN (Convolution Neural Network)が主流でした。

それに対して、GoogleのVision Transformer は、大規模な画像情報処理の世界でも、CNNを全く利用せずに、Transformerだけで最先端のCNNのシステムを上回る性能を発揮できることを示しました。

このことは、Transformerをエンジンとする一つのシステムで、自然言語処理と画像処理のタイプの異なる二つの処理が同時に可能になることを意味しています。

Vision Transformer が、Multi-ModalなAIへの突破口となったというのは、そういうことです。

Vision Transformer のアーキテクチャー

Vision Transformerが自然言語だけではなく、画像も処理できるのは、次のような手法を用いているからです。

「元の画像を小さな画像パッチに分割し、これらのパッチの線形なembeddingのシーケンスをTransformerへの入力として提供する。」

画像パッチは、自然言語処理アプリケーションにおけるトークン（単語）と同じように扱われ、教師あり方式で画像分類モデルを学習します。

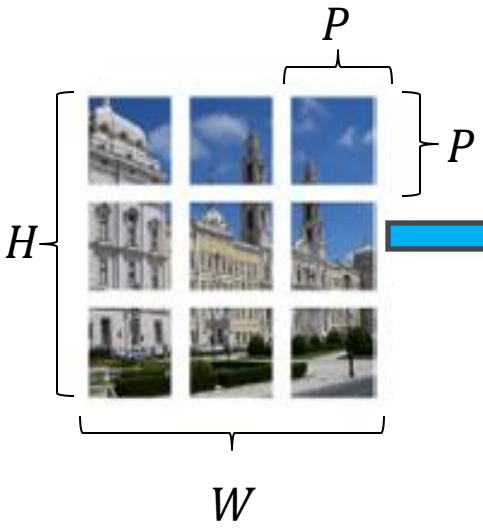
論文タイトルの "An Image Is Worth 16x16 Words" というのは、このことを指しています。

注目すべきことは、この画像のembedding の方法を除いては、Vision Transformer は、元のTransformerの実装を、可能な限り修正しないようにしています。

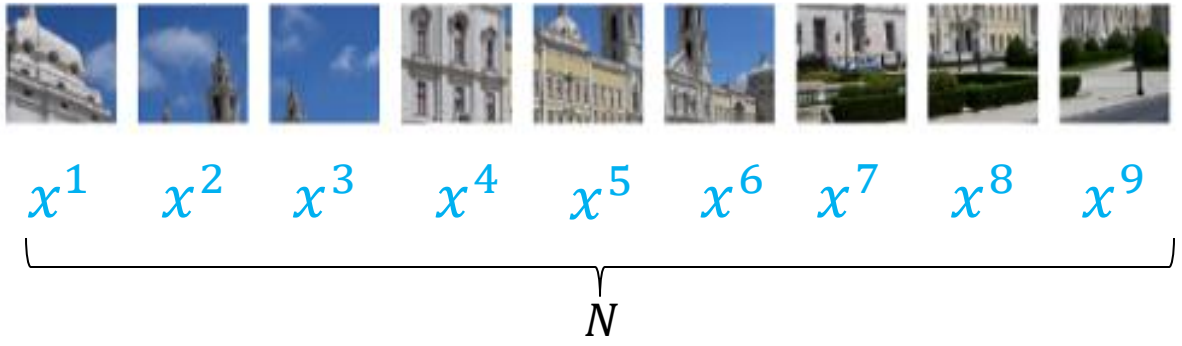
ですから、もしも、自然言語処理での標準的なTransformerの実装を知っていれば、この画像のembedding の方法さえ理解すれば、ほとんど、Vision Transformer の振る舞いを理解できることになります。

画像 x をパッチx^iのシークエンスへ

画像 x



パッチx^iのシークエンス



$N = HWC \,/\, P^2\mathrm{C}$

$x \in \mathbb{R}^{H\times W\times C}$

Cは(R, G, B)等の
カラーチャンネル

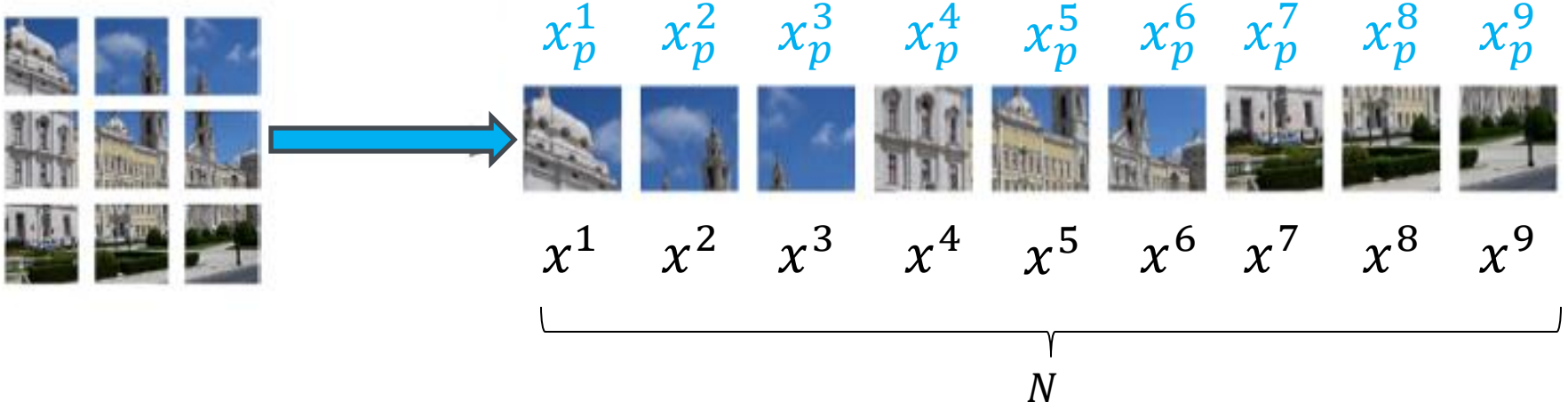
$x_p \in \mathbb{R}^{N\times(P^2\cdot C)}$

P^2C次元の行ベクトルのN個の並びへ

$x_p \in \mathbb{R}^{N \times (P^2 \cdot C)}$

$x_p^i \in \mathbb{R}^{1 \times (P^2 \cdot C)}$

パッチ x^iのP^2C 個のピクセルからなる
行ベクトルx_p^iのN個の並び



画像 x

パッチx^iのN個のシークエンス

D次元のembeddingを構成する

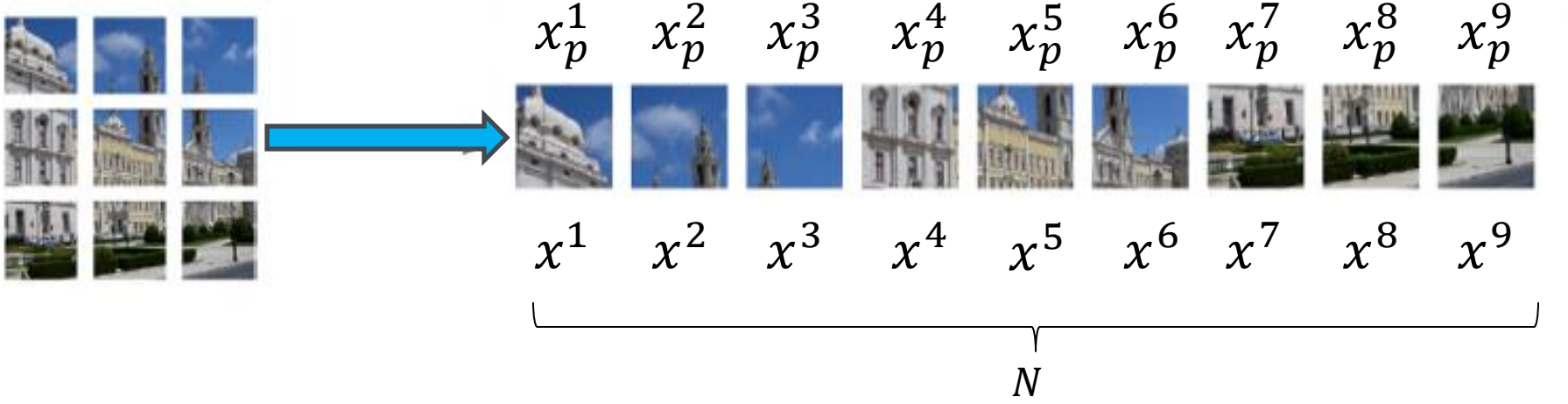
Linear Projection of Flattened Patches

$x_p^i \in \mathbb{R}^{1\times(P^2\cdot C)}$

$E \in \mathbb{R}^{(P^2C)\times D}$

$x_p^i E \in \mathbb{R}^{1\times D}$

D次元の行ベクトルのN個の並び



画像 x

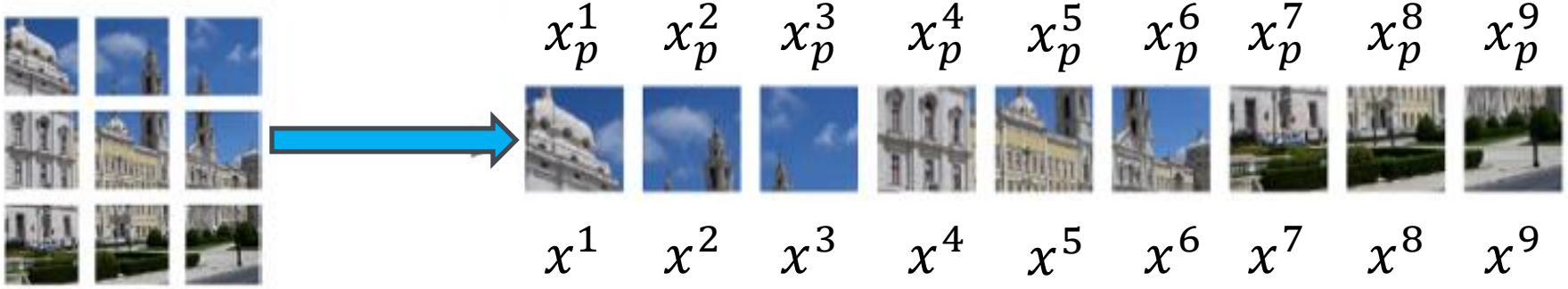
パッチx^iのN個のシークエンス

学習可能な [class]トークンの追加

$$[\, x_{class}; x_p^1 E; x_p^2 E; x_p^3 E; x_p^4 E; x_p^5 E; x_p^6 E; x_p^7 E; x_p^8 E; x_p^9 E \,]$$

$x_p^1 E$ $x_p^2 E$ $x_p^3 E$ $x_p^4 E$ $x_p^5 E$ $x_p^6 E$ $x_p^7 E$ $x_p^8 E$ $x_p^9 E$

Linear Projection of Flattened Patches



画像 x

パッチx^iのN個のシークエンス

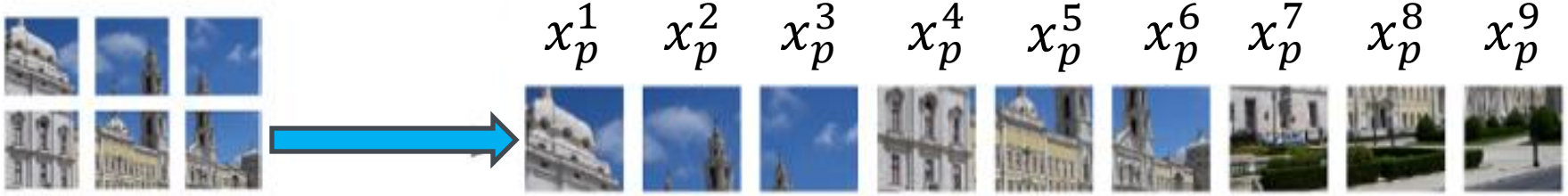
位置埋め込みの追加

$$[\, x_{class};\ x_p^1E; x_p^2E; x_p^3E;\ x_p^4E; x_p^5E; x_p^6E; x_p^7E; x_p^8E;\ x_p^9E \,] + x_{pos}$$

$$[\, x_{class}; x_p^1E; x_p^2E; x_p^3E;\ x_p^4E; x_p^5E; x_p^6E; x_p^7E; x_p^8E;\ x_p^9E \,]$$

x_p^1E x_p^2E x_p^3E x_p^4E x_p^5E x_p^6E x_p^7E x_p^8E x_p^9E

Linear Projection of Flattened Patches



x_p^1 x_p^2 x_p^3 x_p^4 x_p^5 x_p^6 x_p^7 x_p^8 x_p^9

x^1 x^2 x^3 x^4 x^5 x^6 x^7 x^8 x^9

N

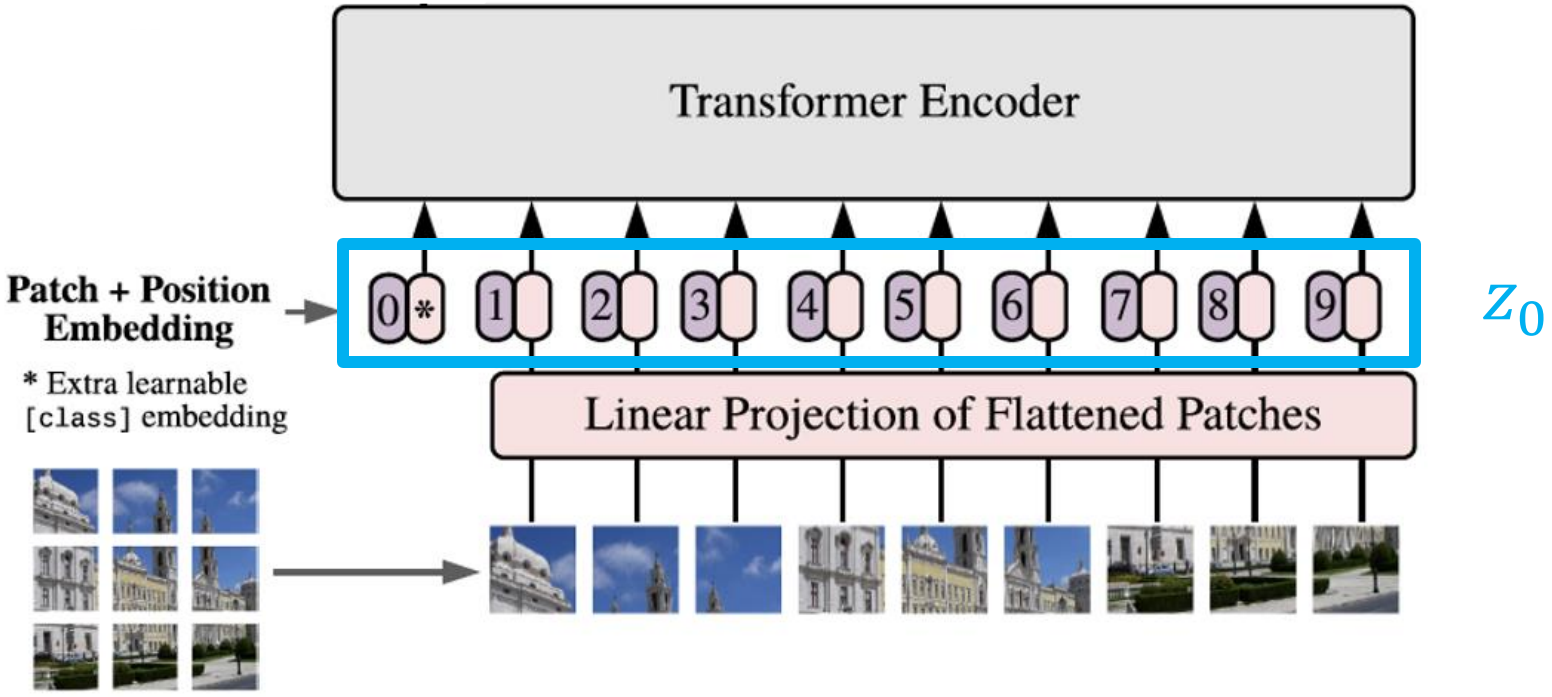
画像 x

パッチx^iのN個のシークエンス

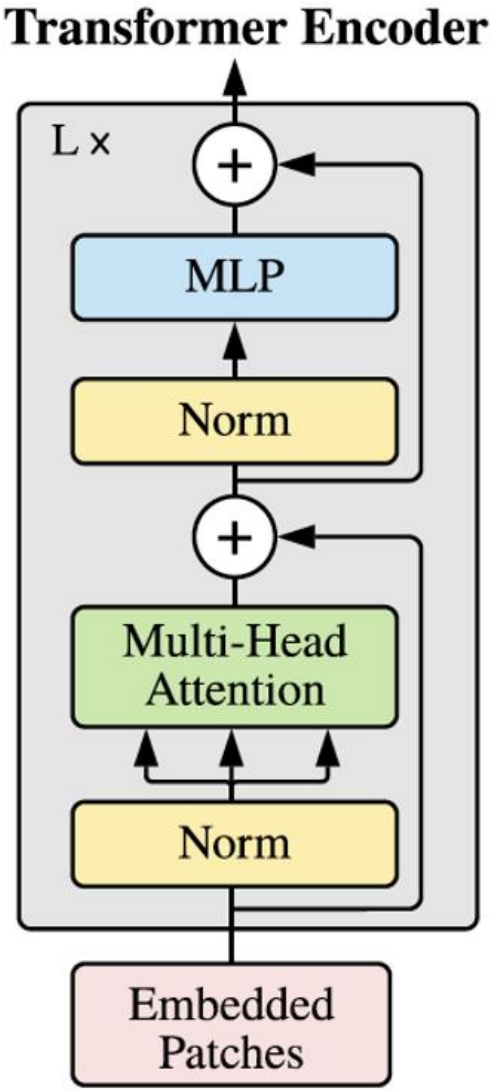
エンコーダへの入力 z_0

$$z_0 = [\, x_{class}; x_p^1 E; x_p^2 E; x_p^3 E; x_p^4 E; x_p^5 E; x_p^6 E; x_p^7 E; x_p^8 E; x_p^9 E \,] + x_{pos}$$

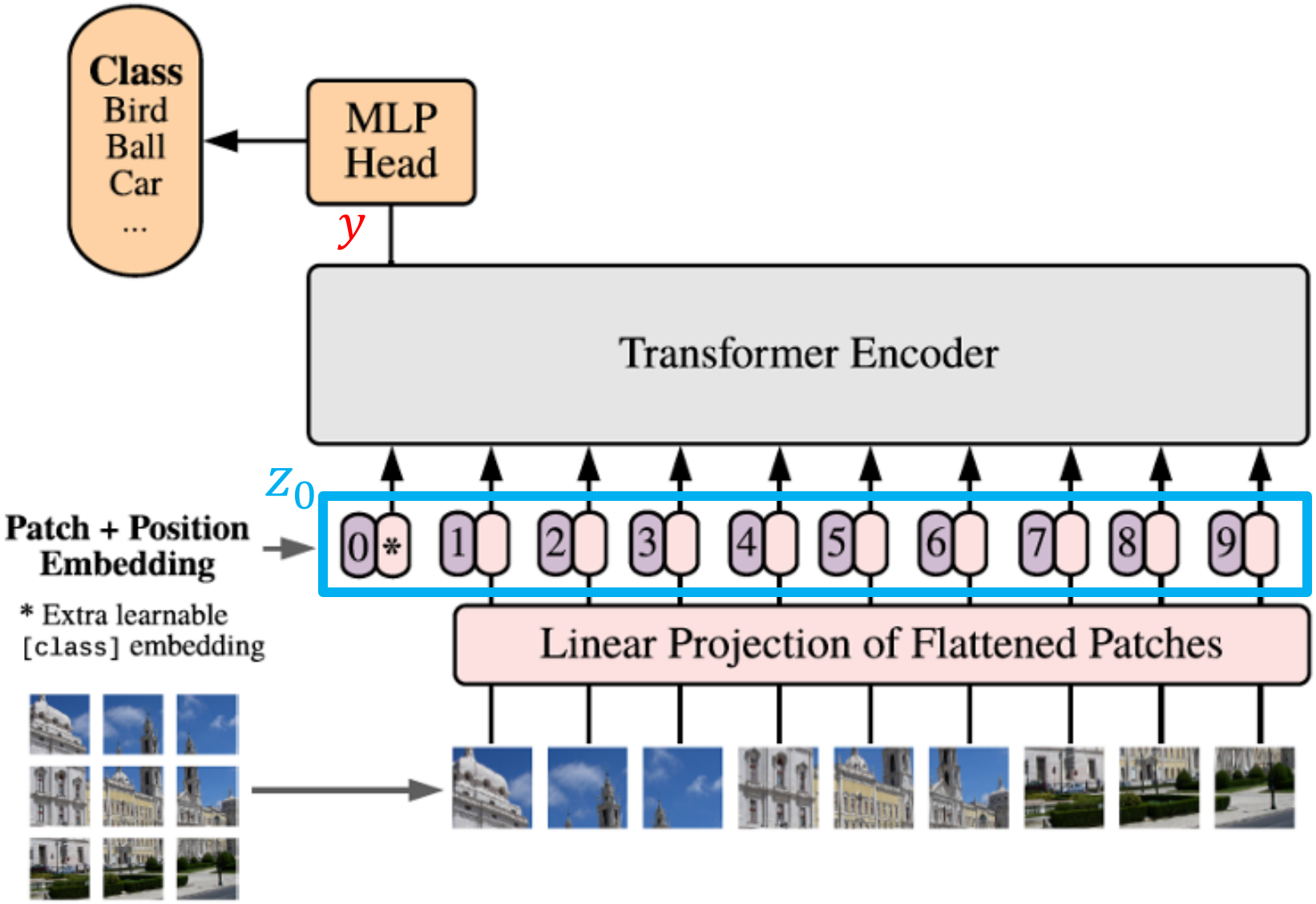
一般にエンコーダへの入力　$z_0 = [\, x_{class}; x_p^1 E; x_p^2 E; \cdots; x_p^N E \,] + x_{pos}$



Transfomer Encoder



Transformerエンコーダは、Multi-head Self Attention（MSA)とMLP(Multi Layer Perceptron)ブロック（の交互の層で構成される。各ブロックの前にはLayer-norm（LN）が適用され、各ブロックの後には残差接続が適用される。MLPにはGELU非線形性を持つ2つの層が含まれる。


Class
Bird
Ball
Car
...
MLP Head
y
Transformer Encoder
z_0
Patch + Position Embedding
* Extra learnable [class] embedding
0 * 1 2 3 4 5 6 7 8 9
Linear Projection of Flattened Patches

Vision Transformerが登場した時代

留意して欲しいのは、2021年のこのモデルが、Transformerを搭載した大規模言語モデルの規模拡大による快進撃の中で生まれたことです。

「Transformersの計算効率とスケーラビリティのおかげで、100Bを超えるパラメータを持つ前例のないサイズのモデルを訓練することが可能になった（Brown et al.）モデルとデータセットが増大する中、性能が飽和する兆候はまだない。」

「NLPにおけるTransformerのスケーリングの成功に触発され、我々は標準的なTransformerを、可能な限り少ない修正で、画像に直接適用する実験を行う。」

Vision Transformerが発見したこと
Inductive Bias Free！

「ImageNetのような中規模のデータセットを強力な正規化なしで学習した場合、これらのモデルの精度は、同程度のサイズのResNetsを数%下回る。

この一見がっかりするような結果は予想通りかもしれない：Transformerは、変換の等価性や局所性といったCNNに固有の帰納的バイアスのいくつかを欠いているため、十分な量のデータで訓練してもうまく汎化できない。

しかし、より大規模なデータセット（1,400万～3,000万画像）でモデルを学習させると、様相は一変する。我々は、大規模訓練が帰納的バイアスに勝ることを発見した。」


OpenAI CLIP
AIのマルチモーダル化とカスタム化

OpenAIのCLIPのアプローチ

CLIPは、GoogleのVision Transfomer のすこし後に、OpenAIによって公開された「テキストとイメージを結合する」を目標とするプロジェクトです。

それは、「見ることも聞くことも話すこともできる」ChatGPTとして最近公開されたGPT-4Vや、テキストから自由に画像を生成することのできるDall E-3の基礎技術です。

OpenAIのCLIPの一つの特徴は、現在のコンピュータによる画像処理技術の現状に満足できないことを率直に語ることから始めていることです。

「ディープラーニングはコンピュータ・ビジョンに革命をもたらしたが、現在のアプローチにはいくつかの大きな問題がある。」

最大のものは、データセットの問題だとOpenAIは言います。

先に見た Vision Transformer は、”Inductive Bias Free”なシンプルなアーキテクチャーでも、データセットの規模を拡大すると、画像認識の性能を上げられることを強調し、「大規模訓練が帰納的バイアスに勝ることを発見した。」と豪語していたのですが、OpenAIのCLIPのアプローチは、すこし違ったものです。

「典型的なビジョン・データセットは、作成に労力とコストがかかる一方で、狭い範囲の視覚概念しか教えない。標準的なビジョン・モデルは、1つのタスクと1つのタスクにしか向いておらず、新しいタスクに適応させるためには多大な労力を必要とする。」

「また、ベンチマークでは優れた性能を発揮するモデルも、ストレス・テストでは失望するほど低い性能しか発揮できず、コンピュータ・ビジョンへのディープラーニング・アプローチ全体に疑問を投げかけている。」

なかなか辛辣です。

「我々はこのような問題を解決することを目的としたニューラルネットワークを発表する。」

それがCLIPだといいます。

「それは、インターネット上に豊富に存在する多種多様なnatural language supervisionを用いて、多種多様な画像で学習される。これは重要な変更点である。」

GoogleとOpenAIで、少しマルチモーダルAIの実装の方向性について、違いがあることは、留意してもらえたらと思います。

CLIP論文　はじめに
画像処理処理の現状

しかし、コンピュータビジョンのような分野では、ImageNetのようなクラウドラベル付きデータセットでモデルを事前学習するのが標準的なやり方である。

ウェブテキストから直接学習するスケーラブルな事前学習法は、コンピュータビジョンにおいても同様のブレークスルーをもたらすのだろうか？

さまざまな、有望な先行研究がある。

はじめに
本研究の課題

本研究では、大規模なnatural language supervisionで訓練された画像分類器の振る舞いを研究する。

インターネット上で公開されている大量のこの形式のデータを利用し、4億の（画像とテキストの）ペアからなる新しいデータセットを作成し、ゼロから学習したCLIP（Contrastive Language-Image Pre-training）が、natural language supervisionから学習する効率的な手法であることを実証する。

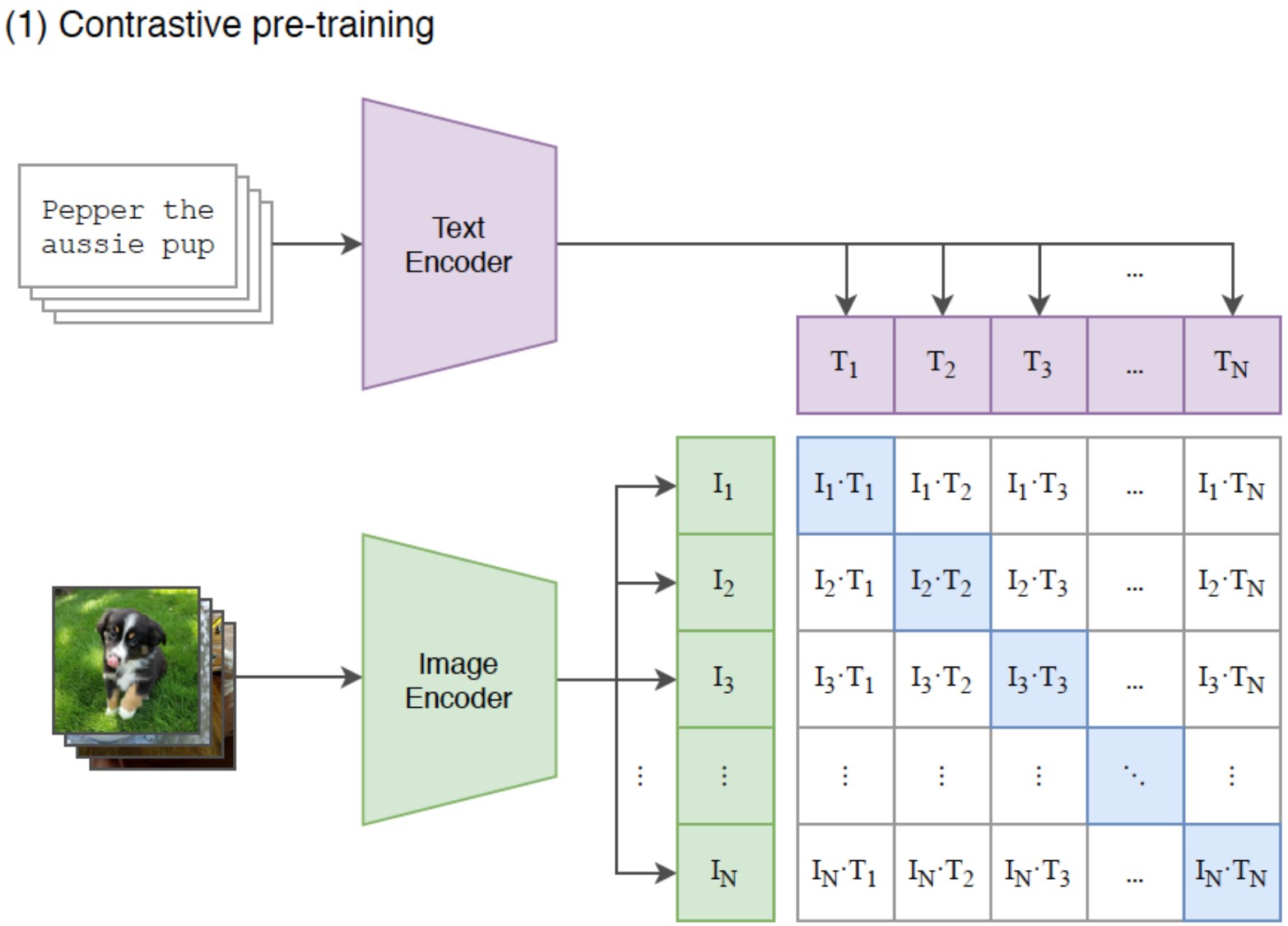
我々は、ほぼ2桁の計算量に及ぶ一連の8つのモデルを訓練することによってCLIPのスケーラビリティを研究し、転送性能が計算量の滑らかに予測可能な関数であることを観察する。

我々のアプローチの概要

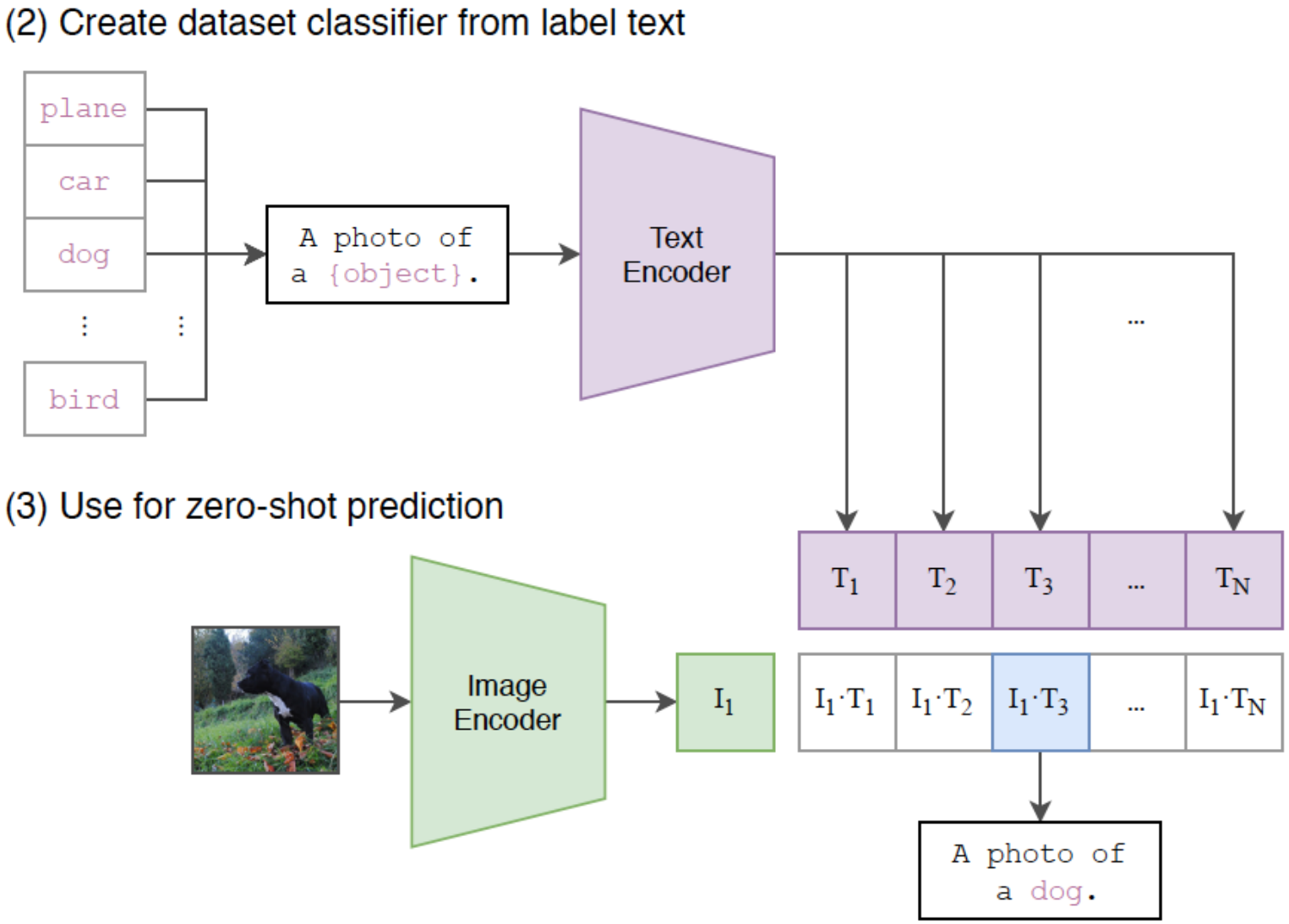
CLIPは画像エンコーダとテキストエンコーダを共同で学習し、（画像とテキストの）バッチ学習例の正しいペアリングを予測する。

テスト時に、学習されたテキストエンコーダは、ターゲットデータセットのクラスの名前や説明を埋め込むことで、ゼロショットの線形分類器を合成する。

(1) Contrastive pre-training



(2) Create dataset classifier from label text
(3) Use for zero-shot prediction



アプローチ
Natural Language Supervision

我々のアプローチの核心は、自然言語に含まれる監督情報から知覚を学習するという考え方である。

これは全く新しいアイデアではないが、この領域での研究を説明するために使用される用語は様々で、一見矛盾しているようにさえ見え、述べられた動機も多様である。

Zhangら(2020)、Gomezら(2017)、Joulinら(2016)、Desai & Johnson(2020)はいずれも、画像と対になったテキストから視覚表現を学習する手法を紹介しているが、それぞれのアプローチを教師なし、自己教師あり、弱教師あり、教師ありと表現している。

私たちは、この一連の研究に共通しているのは、使用されている特定の手法の詳細ではなく、自然言語を学習の信号として評価していることであることを強調する。

これらのアプローチはすべて、Natural Language Supervisionから学習している。初期の研究では、トピックモデルとn-gram表現を使用する場合、自然言語の複雑さと格闘していたが、深い文脈表現学習の改善は、我々は現在、この豊富な監督ソースを効果的に活用するためのツールを持っていることを示唆している（McCannら、2017）。

自然言語からの学習には、他の学習方法と比較していくつかの潜在的な強みがある。画像分類のための標準的なクラウドソースラベリングと比較して、自然言語監視のスケールははるかに簡単です。なぜなら、正統的な1対Nの多数決「ゴールドラベル」のような古典的な「機械学習互換フォーマット」である注釈を必要としないからです。

その代わりに、自然言語で動作するメソッドは、インターネット上の膨大な量のテキストに含まれる監視から受動的に学習することができる。

また、自然言語からの学習は、「単に」表現を学習するだけでなく、その表現を言語と結びつけることで、柔軟なゼロショット転送を可能にするという点で、ほとんどの教師なし学習や自己教師あり学習アプローチよりも重要な利点がある。


OpenAI
AI Assistant アプリ
AIのマルチモーダル化とカスタム化

AI利用のインターフェースを
劇的に変えるAI Assistant アプリ

このセッションでは、OpenAi DevDayで発表された、GPTの能力をユーザーが開発したアプリの上で自由に生かすことを可能にするAssistant APIの概要を見ていきます。

また次回以降のセッションでは、OpenAIが同時に開発を進めていたAIのマルチモーダル化の成果を、今や、AI Assistant アプリの形で、ユーザーが利用できることを紹介したいと思います。

これまで、ChatGPTの利用のスタイルは、OpenAIのサイトにログインして直接ChatGPTと向き合って対話を続けること、具体的にはキーボードとスクリーンを通じてChatGPTとテキストを交換するのが基本でした。このスタイルが大きく変わろうとしています。

ユーザは、場合によればそのアプリの背後にAIがいることを全く意識せずに、普通のスマートフォンアプリと同じように画面タッチでボタンを押したり、スワイプしたりすればいいのです。僕が一番気に入っているインターフェースは、アプリに声で話しかけ、アプリが声で答えるというものです。

重要なことは、こうしたアプリを、OpenAIだけでなく開発者なら誰でも作成できるということです。OpenAIは、こうしたアプリの開発・流通を促進するためのマーケットを用意しています。

AI Assistant アプリ（これを、OpenAIはGPTsと呼んでいるようですが、ChatBotといういいかたもよく使われているようです）の登場は、一般のユーザーとAIとの距離をとても身近なものに劇的に変えるだけではありません。

それは、IT技術者・開発者とAIの距離を大きく変えるものです。

IT技術は・開発者は、これまで 、github copilot等を利用して、主要に開発支援ツールとしてAIを利用してきました。これからは、AIに支援された強力な独自のアプリを、自分の手で開発し、それを多数のユーザーが待つ市場に送り出すことができるのです。

Assistant とは何か

Assistantとは、OpenAI APIの場合、GPT-4のような大規模な言語モデルからパワーを得て、ユーザーのためにタスクを実行することができるエンティティのことを指す。

これらのアシスタントは、モデルのコンテキストウィンドウ内に埋め込まれた命令に基づいて動作する。

また、アシスタントは通常、コードを実行したり、ファイルから情報を取得したりするような、より複雑なタスクを実行できるツールにもアクセスできる。

https://platform.openai.com/docs/introduction

Assistants API

Assistants APIを使用すると、独自のアプリケーション内にAI Assistantを構築することができる。

Assistantは指示を持ち、モデル、ツール、知識を活用してユーザーの問い合わせに応答することができる。

Assistants APIは現在、3種類のツールをサポートしている：

- Code Interpreter
- retrieval(知識検索)、
- Function Call（関数呼び出し）

である。

Assistant APIの働きの流れ
Assistant, Thread, Message, Run

1. Assistant を作成する。それは、モデルに対する指示を定義し、モデルを選択することで、必要であれば、Code Interpreter、Retrieval、Function callingなどのツールを有効にする。

2. ユーザーが会話を開始すると、Threadを生成する。

3. ユーザーが質問する際に、TreadにMessageを追加する。

4. Thread上でAssistantを実行(Run)して、応答をトリガーする。これにより、関連ツールが自動的に呼び出される。

Assistant API で利用される基本的なObject

Assistant	OpenAIのモデルと呼び出しツールを使用した専用AI
Thread	アシスタントとユーザー間の会話セッション。スレッドはメッセージを保存し、コンテンツをモデルのコンテキストに合わせるために自動的に切り捨てを処理する。
Message	アシスタントまたはユーザーが作成したメッセージ。メッセージにはテキスト、画像、その他のファイルを含めることができる。メッセージはスレッドにリストとして保存される。
Run	スレッド上でのアシスタントの呼び出し。アシスタントは設定とスレッドのメッセージを使用して、モデルやツールを呼び出してタスクを実行する。実行の一部として、アシスタントはスレッドにメッセージを追加する。
Run Step	アシスタントが実行の一部として行ったステップの詳細リスト。アシスタントは実行中にツールを呼び出したり、メッセージを作成したりできる。実行ステップを調べることで、アシスタントが最終結果にどのように到達しているかを知ることができる。

退職後の財政プランを立ててくれる パーソナル・アシスタント「僕の財政ボット」の例

Assistant

僕の財政ボット

Thread

退職後の財政プラン

Run

Assistant 僕の財政ボット
Thread 退職後の財政プラン

User's message
退職後の財政プラン用に、毎年いくら積み立てればいいだろうか？

Step

1. **Use code interpreter**
2. **Create messages**

Assistant's Message
年間6万円ほど積み見立ておいたほうがいい。さらに、…

Assistantはどのように動くのか？

Assistants APIは、開発者がさまざまなタスクを実行できる強力なAI Assistantアプリを構築できるように設計されている。

- AssistantはOpenAIのモデルを呼び出し、その性格や能力を調整するための具体的な指示を出すことができる。

- Assistantは複数のツールに並行してアクセスできる。OpenAIがホストしているツール（コードインタープリタや知識検索など）、またはあなたが構築/ホストしているツール（関数呼び出しによる）の両方にアクセスできる。

- Assistantは永続的なスレッドにアクセスできる。スレッドは、メッセージの履歴を保存し、会話がモデルのコンテキストの長さに対して長くなりすぎたときに切り捨てることで、AIアプリケーションの開発を簡素化する。スレッドを一度作成し、ユーザーが返信するたびにメッセージをスレッドに追加するだけでいい。

- Assistantは、ファイルを作成する際、またはアシスタントとユーザー間のスレッドの一部として、いくつかの形式でファイルにアクセスすることができる。ツールを使用している場合、アシスタントはファイル（画像、スプレッドシートなど）を作成し、作成したメッセージで参照するファイルを引用することもできる。


近未来のAIの展望


AIのマルチモーダル化の中で
パーソナルなAIを展望する
近未来のAIの展望

「パーソナルなAI」を展望する

今回の講演で僕が示したいと思っているのは、一言でいえば、「パーソナルなAIへ」という展望です。

自分の目や耳や口をもつAIの登場といえば、AIロボットがしだいしだいに人間を押し除けてゆく、AI優位の近未来をイメージする人も、少なくないと思います。

そうではなく、様々な局面で我々人間を支援する、あくまでも人間のために役にたつAIを考えたいと思います。

「パーソナルなAI」を展望する

そういうAIを展望する一つの鍵は、すべての人が日常的にAIをパーソナルなアシスタントとして利用し、また、AIにとって人間のアシスタントであることが、競争的優位性を持つようにAIの未来を設計することだと、僕は考えています。

Be My AI !

もちろん、そのためにはAI技術は誰に対しても開かれたOpenなものでなければなりません。

AIのマルチモーダル化が可能とするボイスAIは AI利用拡大のゲームチェンジャー

僕は、音声で入出力ができる「ボイスAI」に大きな期待を持っています。

もしも、みんながドラえもんのようなAIロボットと一緒に暮らしていて、彼は、僕らの質問に、可能な限りいい答えを返してくれるとしましょう。

彼とのやりとりに、僕らは、キーボードを叩く必要があるでしょうか。それは面倒です。ボイスでやり取りをするのが「自然」です。

彼の話は、聞き取りやすいものになるでしょうか？　それは場合によります。

ある場合には、ボイス・インターフェースを他のテキストあるいはイメージのインターフェースに切り替える必要があるでしょう。

また、ある場合には、AIロボットとのボイスによるやり取りを繰り返して、必要な情報をボイスで取得することに成功するかもしれません。

人間・機械間のやりとりを
繰り返すことには意味がある

実は、大規模言語モデルにとって、こうした 人間・機械間のやりとりを繰り返すfew-shot prompt は、正しい答えに辿り着く、とても有効な方法なのです。

もっとも、現在のAIは、「次のプロンプト」をサジェストすることはできていません。それは、もっぱら、人間の役割です。ただ、この点は、少しマシにできるかもしれません。

近未来のAIの展望

新しいインターフェースと
新しいデバイスへの期待

僕の「ボイスAI」に対する期待は、このような新しいインターフェースの開発への期待です。同時にそれは、そうしたインターフェースを搭載した新しいデバイスの登場への期待です。

その変化の主な舞台は、スマートフォンの世界になるはずです。

Smart Phoneが「賢い電話」という意味なら、それは「もっと賢い電話」にならなければなりません。それは、Smart PhoneにAIを搭載することで、はじめて可能になります。

Androidが、Androidという名前を持っていたことは、こうした変化にとって象徴的だったのかもしれません。

AI利用のインターフェースを
大きく変えるOpenAI Assistant API

これまで、ChatGPTの利用のスタイルは、OpenAIのサイトにログインして直接ChatGPTと向き合って対話を続けること、具体的にはキーボードとスクリーンを通じてChatGPTとテキストを交換するのが基本でした。このスタイルが大きく変わろうとしています。

ユーザは、場合によればそのアプリの背後にAIがいることを全く意識せずに、普通のスマートフォンアプリと同じように画面タッチでボタンを押したり、スワイプしたりすればいいのです。

先に触れたように、僕が一番気に入っているインターフェースは、アプリに声で話しかけ、アプリが声で答えるというものです。

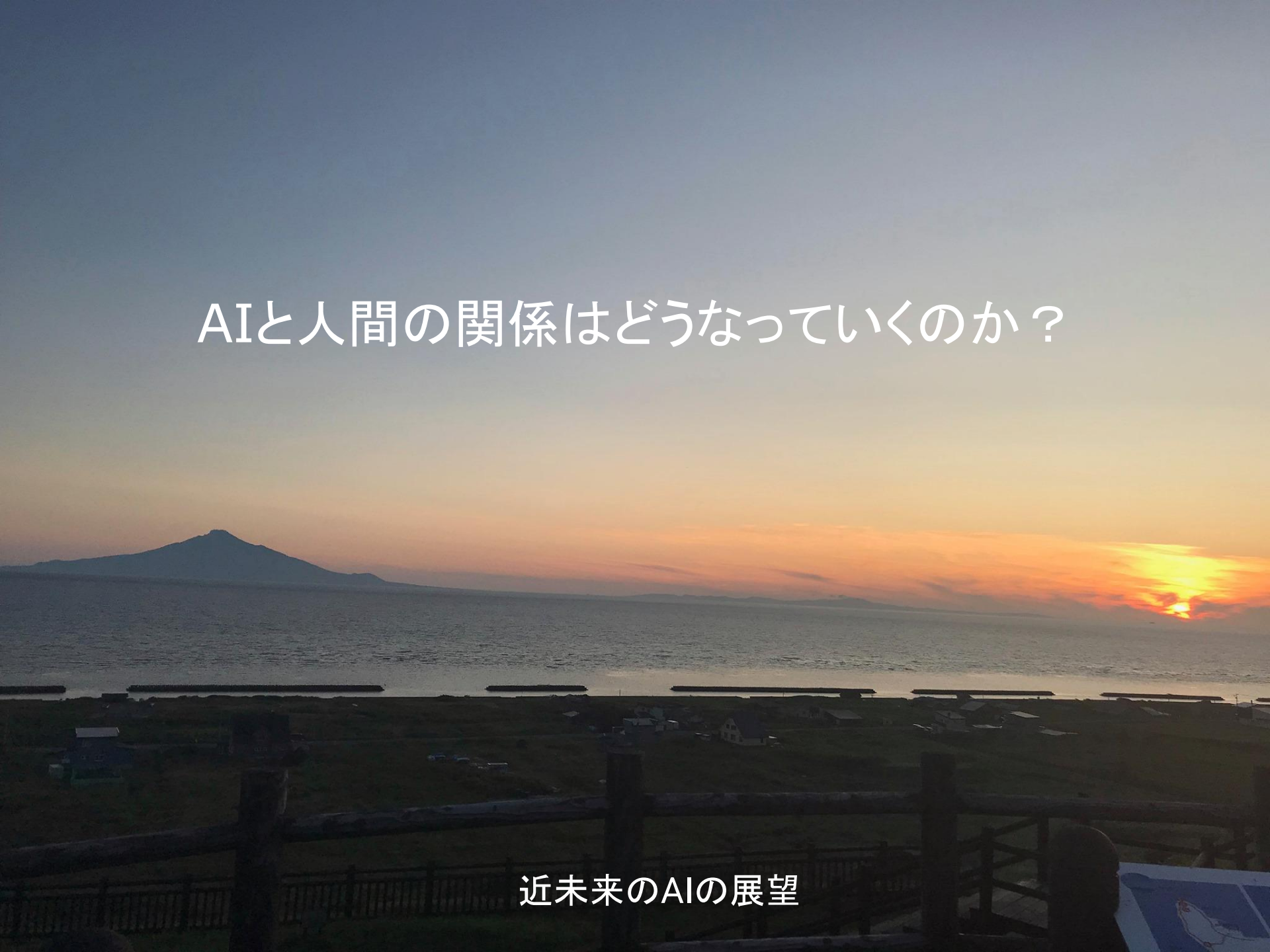
スマートフォンに搭載されたAI Assistant アプリの登場は、一般のユーザーとAIとの距離をとても身近なものに大きく変えるでしょう。

それだけではありません。

それは、IT技術者・開発者とAIの距離を大きく変えるものです。

IT技術者・開発者は、これまで 、github copilot等を利用して、主要に開発支援ツールとしてAIを利用してきました。

これからは、IT技術者・開発者は、AIに支援された強力な独自のアプリを、自分の手で開発し、それを多数のユーザーが待つ市場に送り出すことができるのです。


AIと人間の関係はどうなっていくのか？
近未来のAIの展望

我々とAIの関係を明確にすることが AIの新しい発展を可能にする

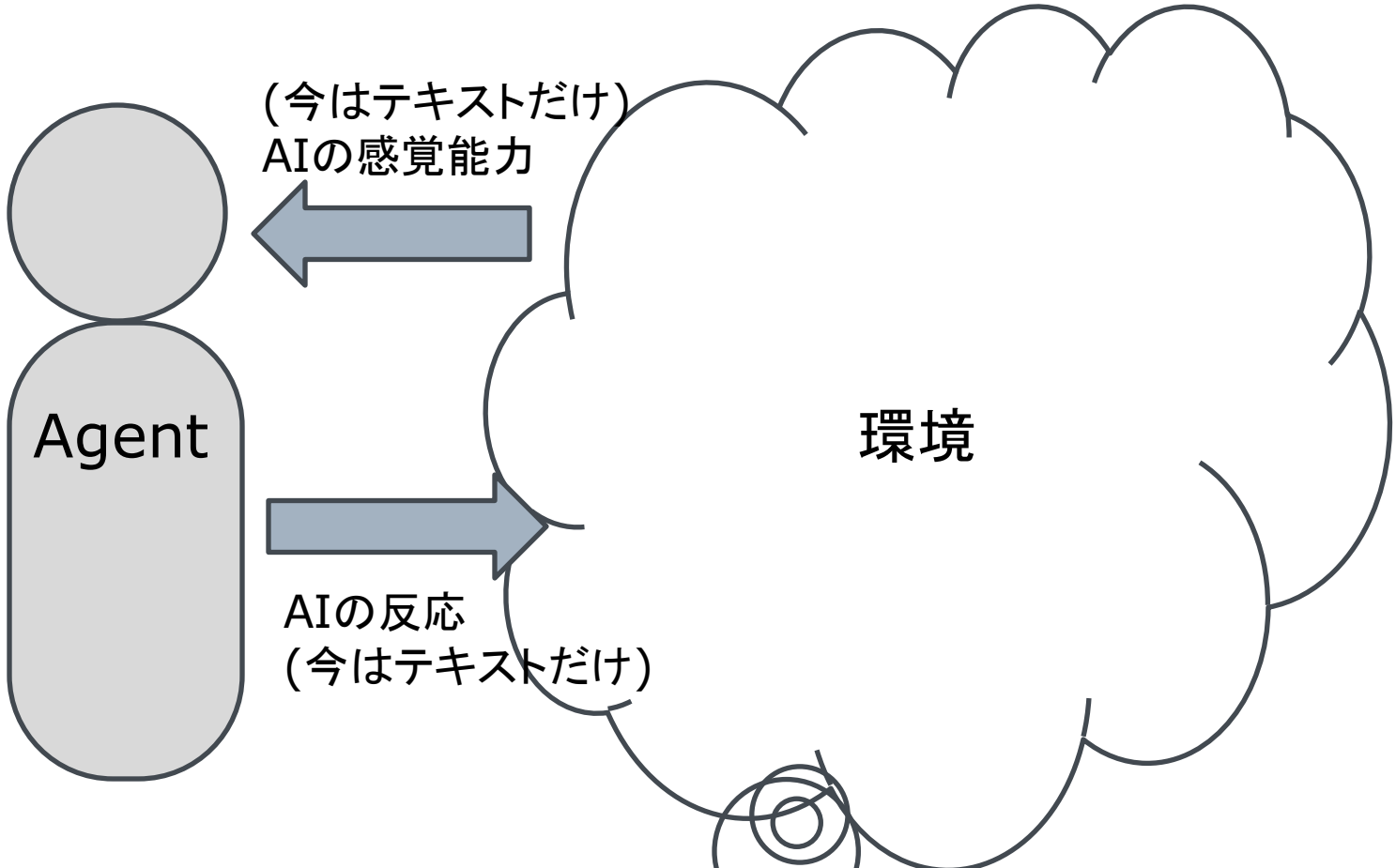
AIと人間の関係にとって一番基本的な問題は、我々人間がどのようなAIを望んでいるのかということにあります。

それが、AI利用の拡大にとっても、AIとのインターフェースを考える上でも鍵になります。

そうした問いかけが、AIの新しい発展を可能にするのです。

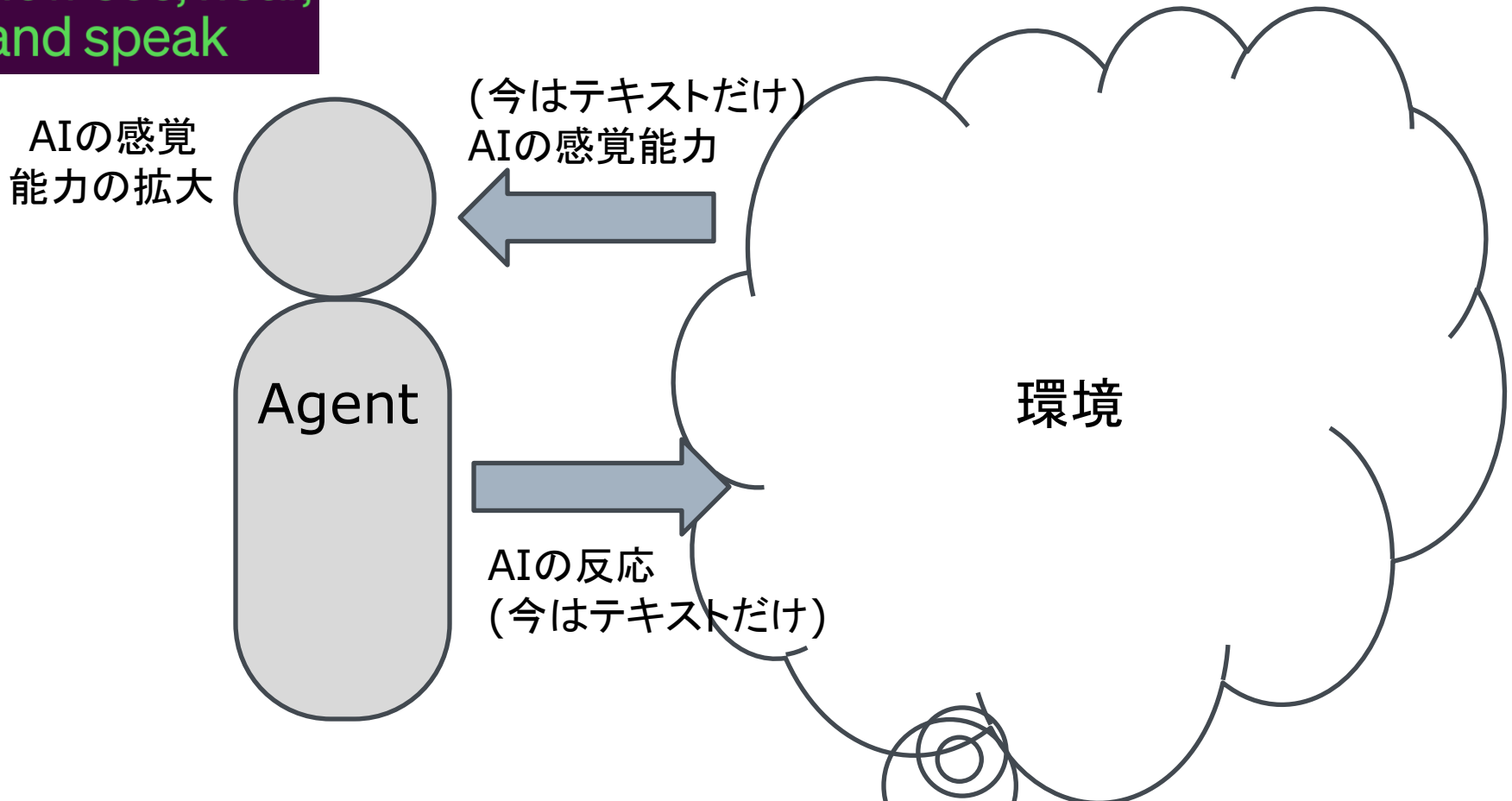
そういう問題を考える時期に、ようやく差し掛かっているのだと思います。

マルチモーダルなAIのモデルを考える
Agent Base Model

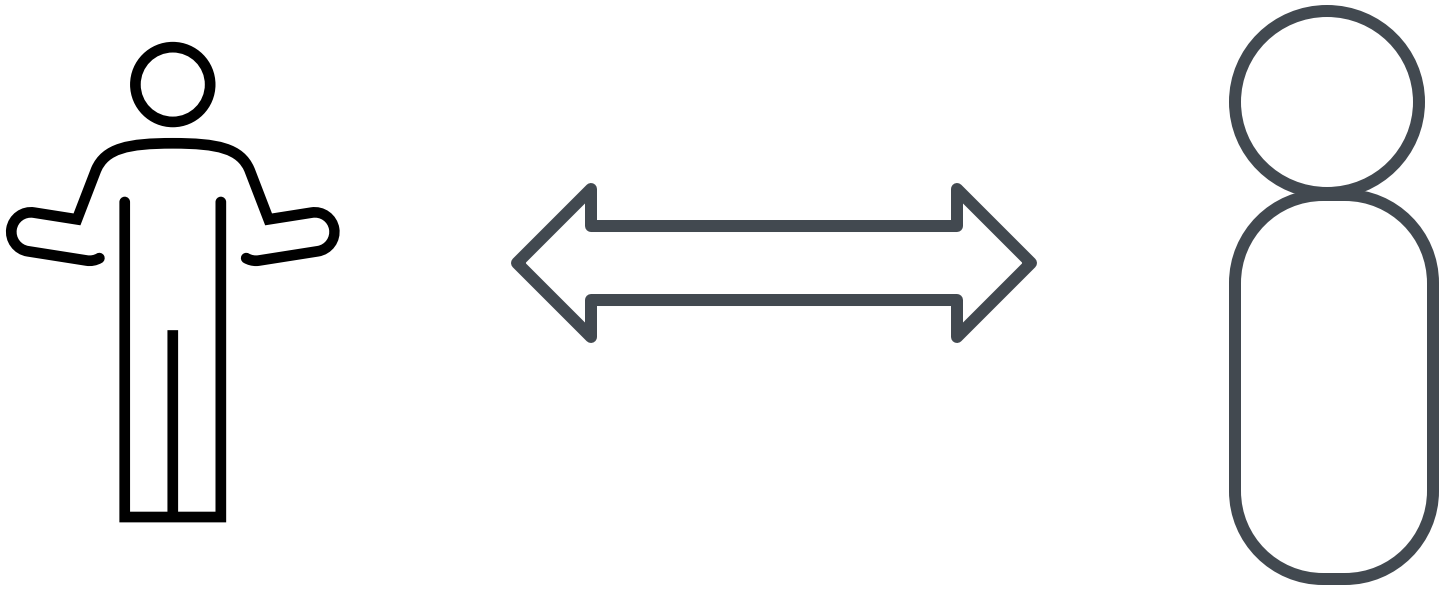


マルチモーダルなAIのモデル
Agent Base Model

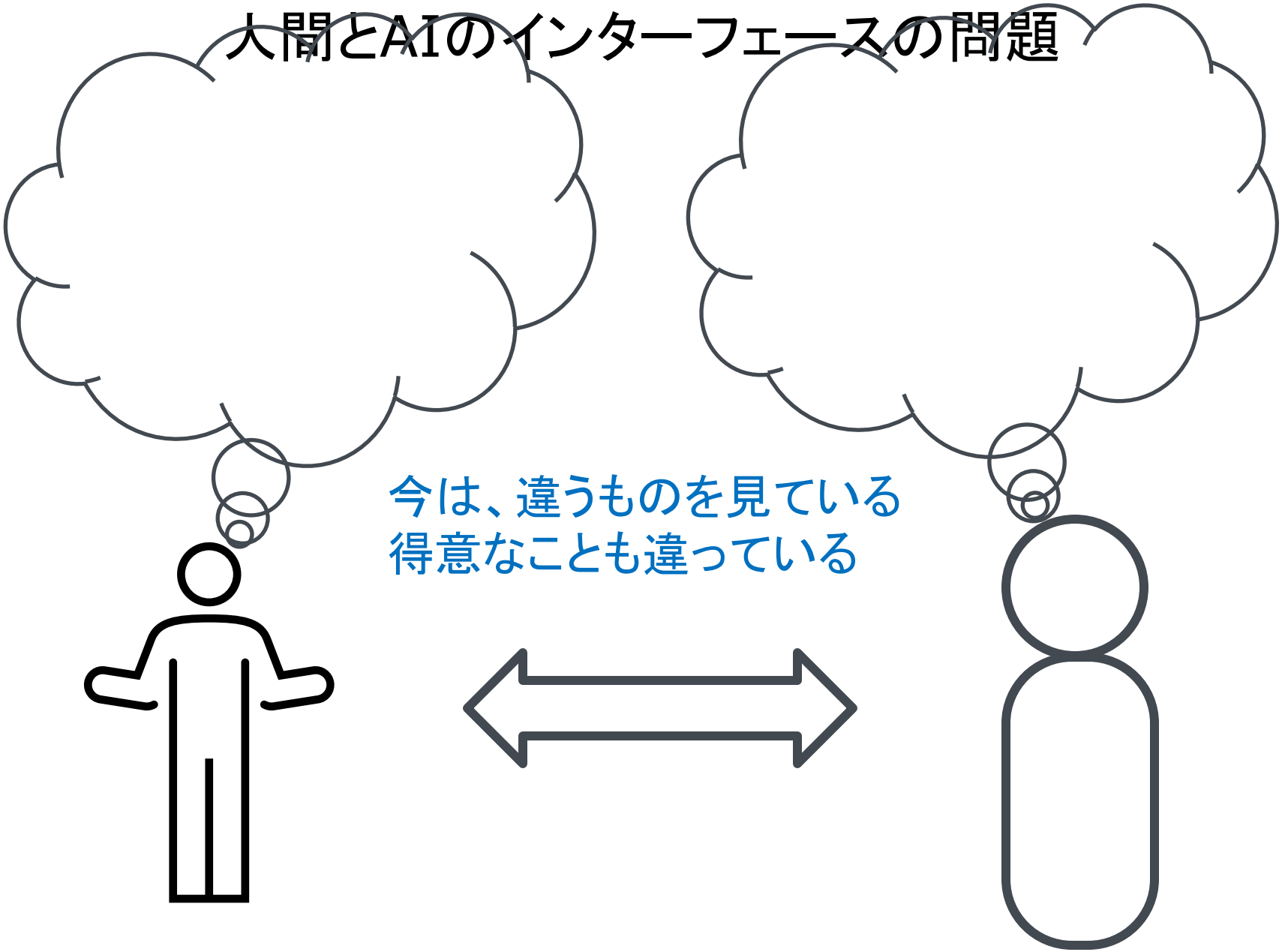
ChatGPT can now see, hear, and speak



人間とAIのインターフェース

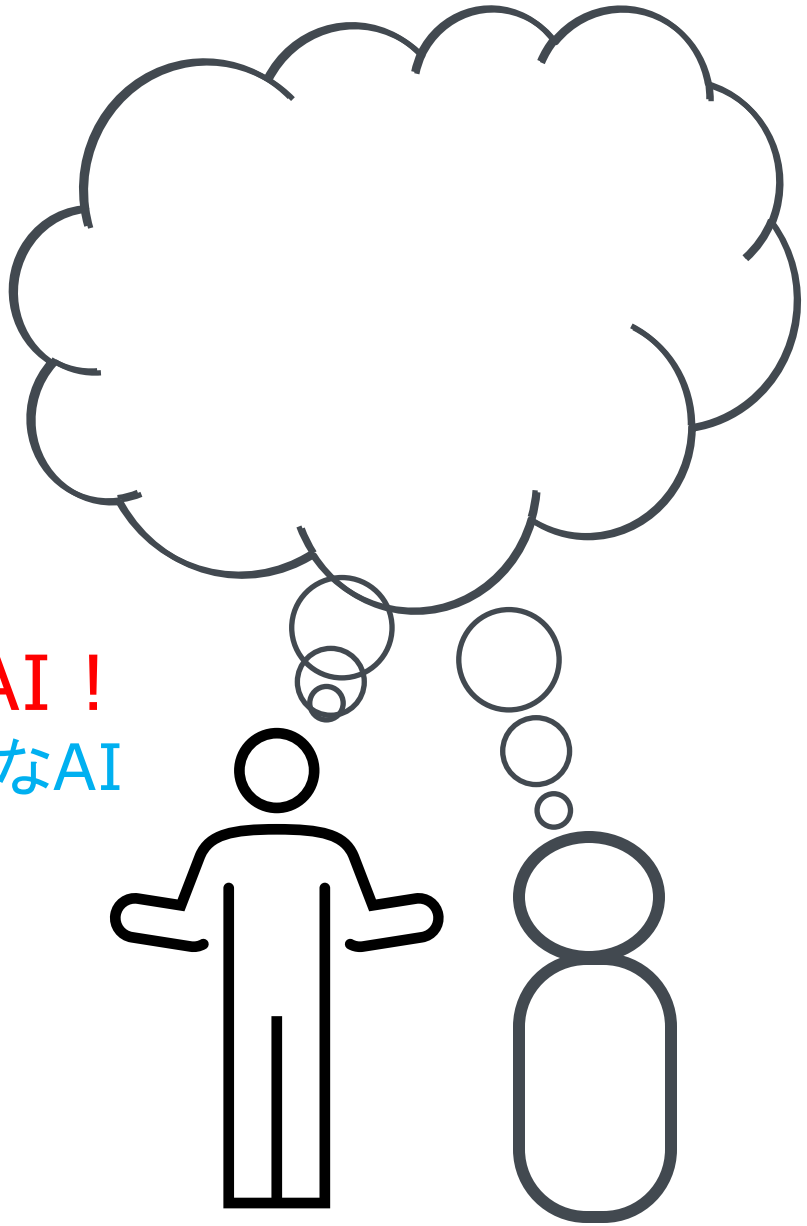


人間とAIのインターフェースの問題



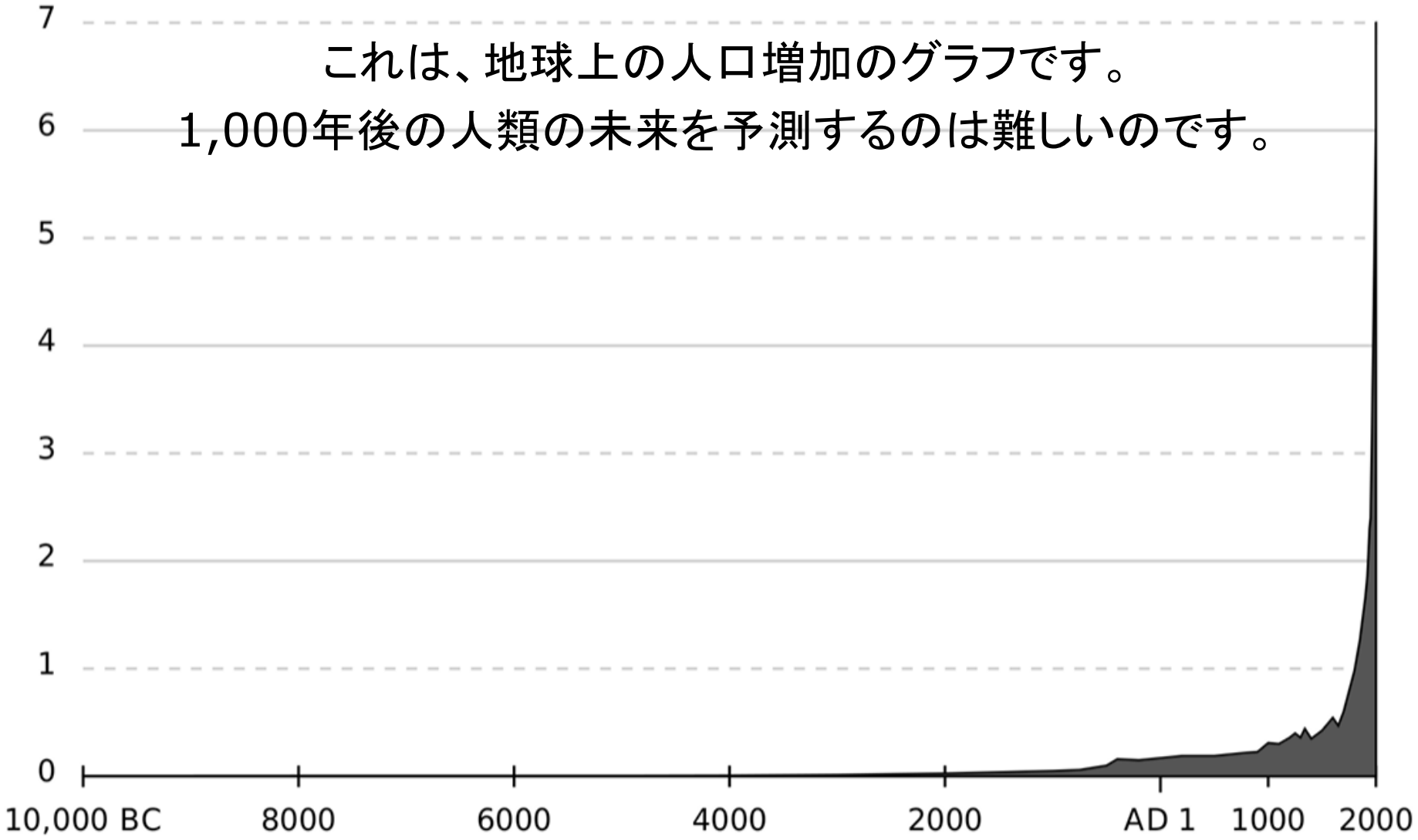
Be My AI !

パーソナルなAI




Be My AI !

ここでは、非常に楽観的な未来予想をしました。



ただ、100年後の未来についていえば、

OpenなAIか、CloseなAIか
私のAIか、誰かのAIか

の選択は、大きな意味を持っていると、
僕はかんがえています。