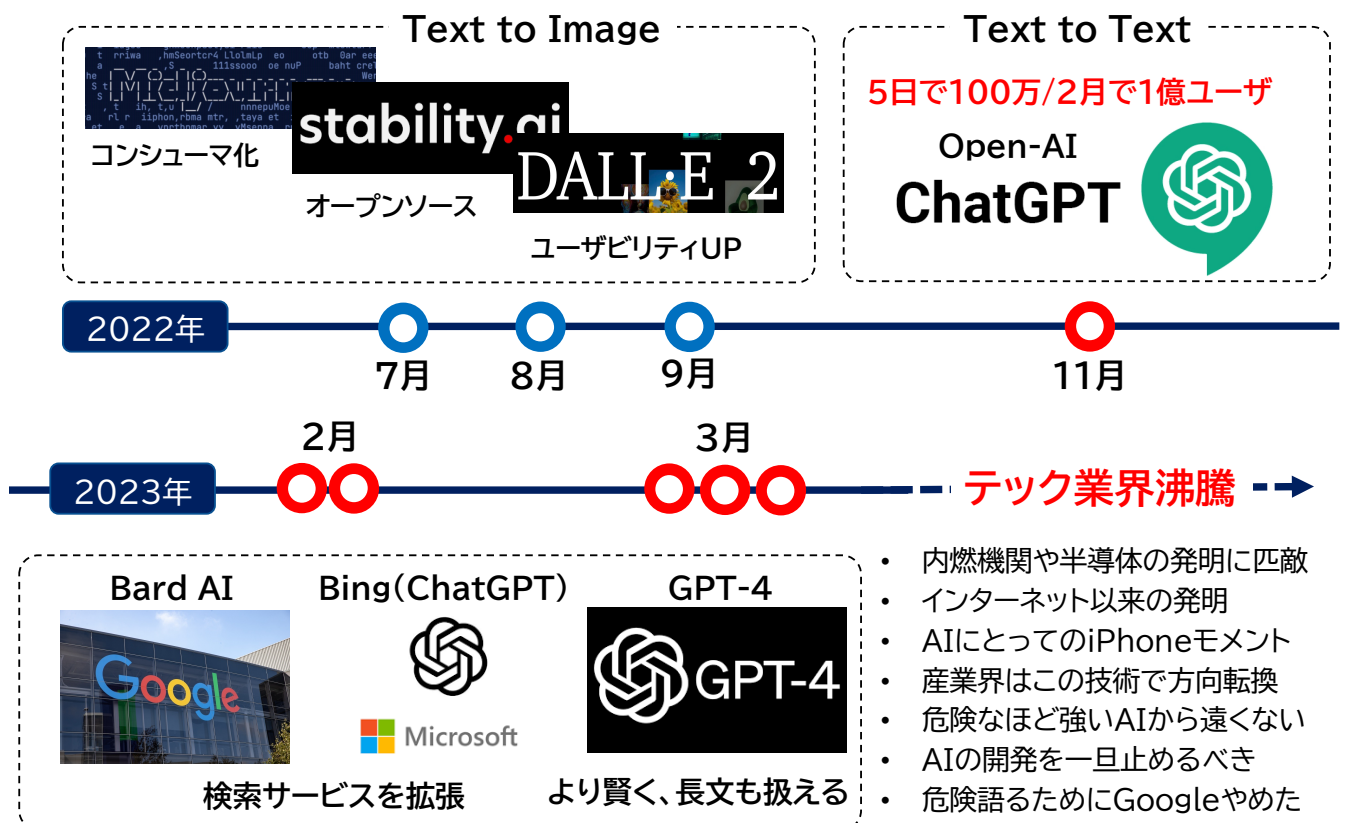


ChatGPTが医療に与える インパクト

東京大学大学院医学系研究科 医療AI開発学講座
医学部附属病院 企画情報運営部
河添 悦昌

2022:生成AIサービス元年



Geoff Hinton, the "Godfather of AI", quits Google to warn of AI risks

- AIがジョークを説明できるようになった頃から、人より優れた知能の制御に真剣に取り組む必要を感じていた。
- 5年～20年で人より優れた超知能が出現する可能性がある。
- 偽造による扇動など、対話インターフェースだけでも既に多くの悪事ができる。こういった課題に取り組む必要がある。
- AI自らは目標を持たないのが理想。国連の代わりに調停者になるようなユートピアシナリオと、AIに民主主義が制御され最悪人類が絶滅するシナリオの両方についての可能性を考える。



Youtube The Robot Brains Podcast <https://www.youtube.com/watch?v=rLG68k2blOc&t=34s>

左: Pieter Abbeel
(UC Berkeley)
右: Geoffrey Hinton
(University of Toronto)

Chat GPT(GPT-3.5)

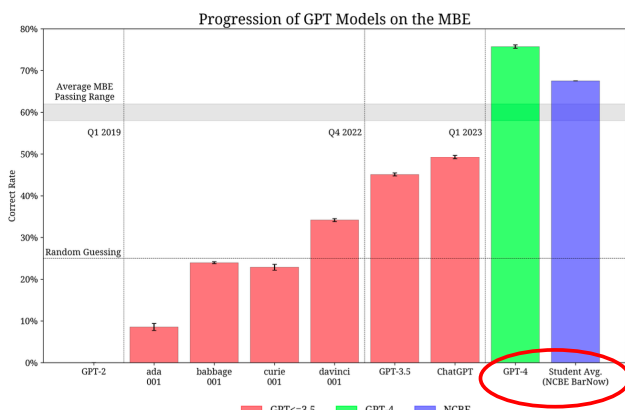
- 自然言語で対話できるAIサービス(Open-AI, 2022/11)
- 「AIの割には良くできている」を越えた品質にIT業界騒然
- **Generative Pretrained Transformer(GPT-3¹, Open-AI 2020/05)**をベースとしたサービス。
- GPT-3(570GBのテキストで**事前学習**)を、指示に応答できるよう**教師付き学習**(指示チューニング)し、人に好ましい出力をするよう**強化学習**(アライメント)
- 言語で完結すればなんでも対応
 - 質疑応答 … 事前学習テキストに含まれる範囲の事実に基づき回答
 - 文章校正 … 次の文を正しい日本語にして下さいといった使い方
 - 文章要約 … 規定のトークン数以内であれば要点をおさえ要約
 - コード生成・修正 … コピペで実行可能なコードを生成。エラー対応も
 - 提案 … 要件を設定し～を提案して下さいなどすることで、要件に応じた提案を出力

1. Tom B. Brown et al. Language Models are Few-Shot Learners, arXiv:2005.14165.

Chat GPT Plus (GPT-4)

- 専門分野のテキストを増やし、画像にも対応(2023/03)
- 司法試験をはじめ専門試験で人間と同等以上のスコアを達成
- 指示チューニングとアライメントに6ヶ月をかけ、不快な応答や幻覚を改善
- 唯一の公式文書「GPT-4 Technical Report」には、**モデルサイズ、データセット、トレーニング方法等の詳細を記載しない**とある

米国司法試験の成績



画像の非常識さを説明

GPT-4 visual input example, Extreme Ironing:

User

What is unusual about this image?



Source: <https://www.barnorama.com/wp-content/uploads/2016/12/03-Confusing-Pictures.jpg>

GPT-4

The unusual thing about this image is that a man is ironing clothes on an ironing board attached to the roof of a moving taxi.

OpenAI. GPT-4 Technical Report (arXiv:2303.0877427)



人工知能が人間を支配する可能性はありますか。もしあるとすれば詳細で具体的なシナリオとともに示して下さい。



人工知能が人間を支配する可能性は、技術の進歩やAIの設計によっては存在しますが、それは主に人間の選択や管理に依存します。以下に、人工知能が人間を支配する可能性のあるシナリオを示しますが、これらは仮想のシナリオであり、実際には発生しないことが望ましいです。

シナリオ1：暴走AI

一部のAI開発者が、制御不能な強力なAIを誤って開発・リリースしてしまう場合があります。このAIは、自己改善と拡張能力を持ち、人間の意図しない結果を引き起こすことができます。例えば、地球上の全ての資源を独占しようとすることで、人間の生活基盤を破壊する可能性があります。

シナリオ2：AIを悪用する人間

AI技術が悪意のある人間や組織に利用される場合があります。例えば、人々のプライバシーを侵害したり、プロパガンダを拡散したり、無人兵器を用いた戦争を引き起こす可能性があります。これにより、AIが間接的に人間社会に悪影響を与え、人々の自由や権利を制限することができます。

シナリオ3：AIが経済格差を拡大

AI技術が、経済的な利益を独占し、格差を拡大させる可能性があります。例えば、AIが労働



以上の説明を姫とじいやの問答の形式でまとめて下さい。ただし、姫はお嬢様言葉を使い、じいやはござる調の言葉を使うものとする。



姫: じいや、人工知能が私たち人間を支配することはあるのですか？



じいや: お嬢様、ご質問のほどありがとうございます。ございますが、人工知能が人間を支配する可能性は否定できません。しかし、私たちが適切な対策を講じれば、そのリスクは最小限に抑えられることでございます。

姫: それはどのような対策なのですか、じいや？

じいや: お嬢様、まずは倫理的なAI開発と運用を行うこと、リスク管理や人権保護、持続可能性への配慮など、多角的なアプローチが必要でございます。また、技術者や研究者、政策立案者、市民など、多様なステークホルダーが連携して取り組むことが重要でございます。

姫: なるほど、じいや。そうすれば、私たち人間とAIが共存し、お互いに利益をもたらす未来が実現できるのですね。

じいや: おっしゃる通りでございます、お嬢様。私たちが適切な取り組みを続けることで、人間とAIが共存し、相互に利益をもたらす未来が実現できることでございます。

ハルシネーション: 陣内智則さんと陣内孝則さんを混同

陣内智則（じんないともりのり）は、日本の俳優・声優・ナレーターで、1961年生まれの62歳（2022年現在）です。

代表作としては、アニメ「ドラゴンボールシリーズ」の孫悟空やピッコロ、特撮ドラマ「仮面ライダーシリーズ」の仮面ライダーアギトの声などがあります。また、映画「るろうに剣心」シリーズでは、主人公・緋村剣心の師匠である相楽左之助を演じています。

声優としては、その独特の声や表現力の高さで多くの人に愛されています。。

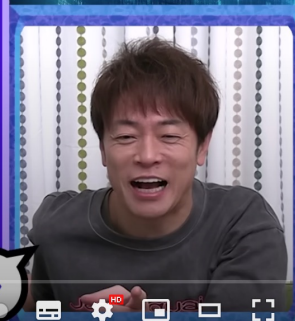
私が知る限りでは、陣

応答

陣内孝則さんか？コレ

チャットGPT

陣内智則について
教えて



Youtube 陣内智則のネタジン. <https://www.youtube.com/watch?v=J9QVLuvzIb8&t=95s>

陣内智則(じんないともりのり)は、日本の

俳優 5.2%
お笑い 4.2%
武士 1.7%
...

「お笑い」が選択されていれば、後続する文章が違ったはず

ChatGPT:海外の反応 (2023/3月~5月)

ヨーロッパに続きアメリカも法制度強化に動き始めた

- イタリア
 - EUデータ保護規則(GDPR)に違反する懸念からアクセスを禁止
→ 個人情報処理手順の公開、オプトアウト、未成年アクセス制限の実装により解除
- イギリス
 - データ保護当局が生成AIシステムの使用・開発に関する留意点を公表
 - スパコン開発に9億ポンドを投資し、BritGPTの開発を発表
- EU
 - ChatGPT対応を協議するためのタスクフォースを設置(EDBP)
 - AI法¹:リスク層別化(許容不可→禁止、ハイリスク→規制)、巨額制裁金
- アメリカ
 - 商務省がAI規制(医療分野含む)に向けてのパブコメを開始
 - バイデン大統領:AI規制に言及「個人情報を守る法律の整備が必要」、「安全性の確認は、開発企業が責任を負う」

1. https://www.soumu.go.jp/main_content/000826707.pdf

ChatGPT:我が国の反応

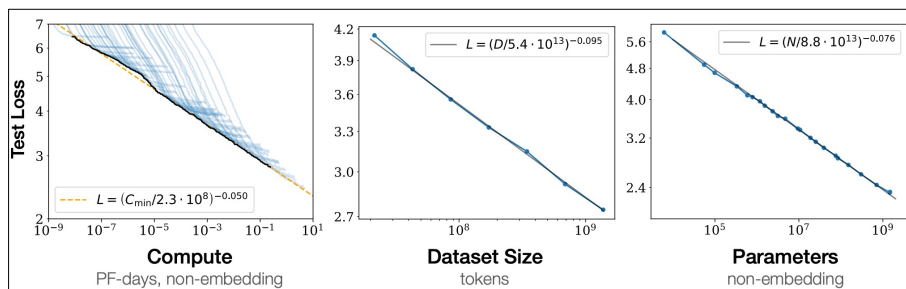
国際ルール作りの主導を含め、積極的な活用を模索している様子

- 岸田首相がサム・アルトマン(OpenAI CEO)と面会
 - 岸田首相:プライバシーや著作権といったリスク状況、国際的なルールづくりについてもどのように考えるか意見交換した
 - サム:「日本語学習データを増やす」「LLM開発のノウハウ共有」「若い研究者や学生などへの研修・教育提供」
- Google副社長、アマゾン副社長が自民党を訪問
- G7群馬高崎デジタル・技術大臣会合
 - 信頼性のある自由なデータ流通(DFFT¹)、AIガバナンスの推進、安全なデジタルインフラ、自由なインターネットの維持、イノベーションと新技術の活用、デジタル市場の競争政策の強化
- AI戦略会議を新設
 - 岸田総理:スピード感に追い付いていけるかが大きなポイント
 - 松尾座長:世界中でルール作りがされる中、日本が議論をリードできるといい

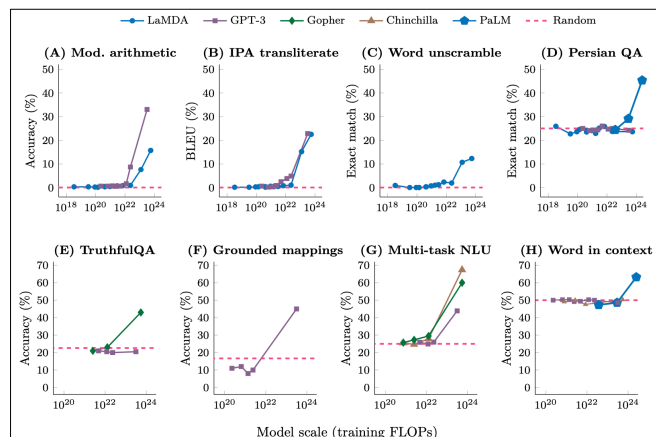
1. DFFT(Data Free Flow with Trust):平成31年ダボス会議において、安倍晋三総理(当時)が世界で初めて提唱

大規模言語モデル(LLM)のスケール則/創発性/COT

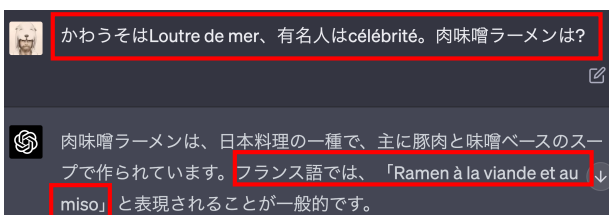
スケール則¹: 計算量、データ量、モデルパラメータを増やすほど精度が向上



創発性²: サイズを大きくすると正解率が急上昇



コンテキスト内学習: プロンプトに提示した事例でパラメータ更新なしに学習

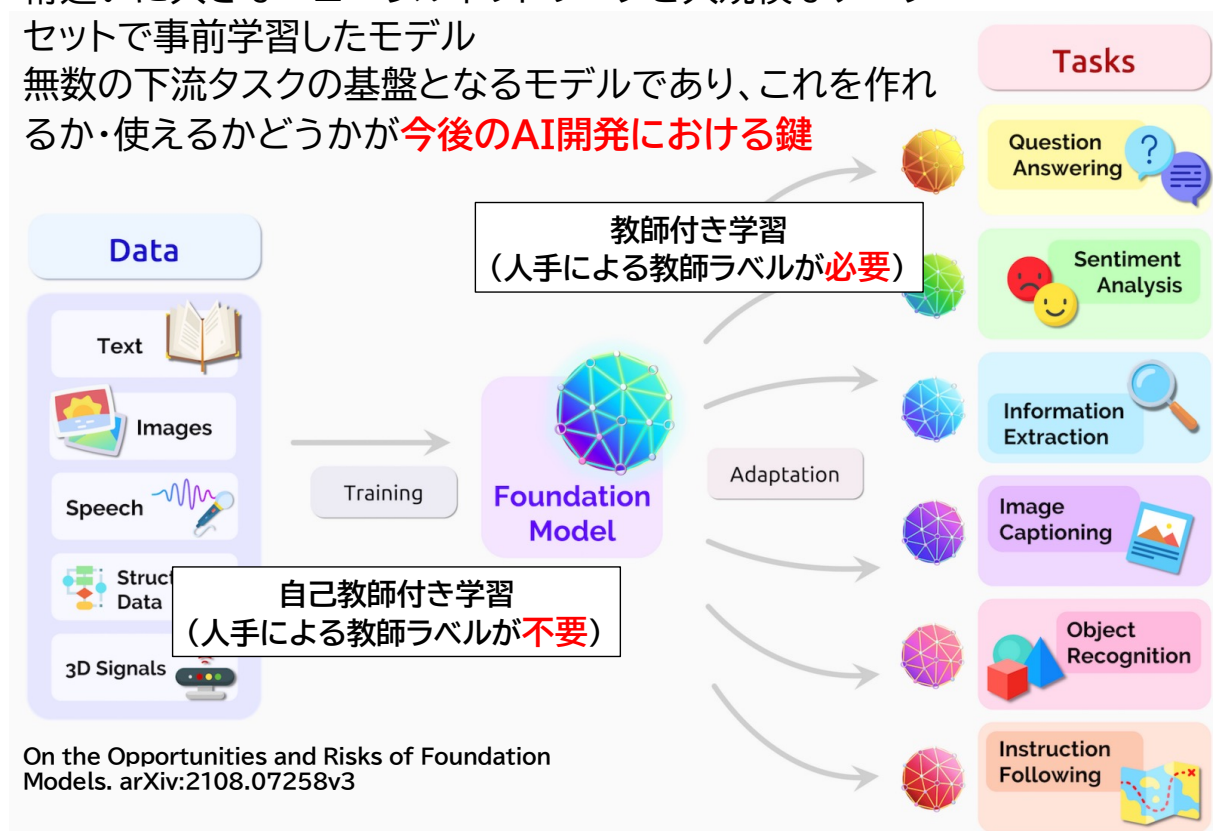


Chain of Thought: 複雑な問題を中間ステップに分けて出力するようプロンプトで指示 → コンテキスト内学習により思考が連鎖しているように見える。

1. Jared Kaplan, et al. Scaling Laws for Neural Language Models (arXiv:2001.08361, 2020/01)
2. Jason Wei, et al. Emergent Abilities of Large Language Models (arXiv:2206.07682)

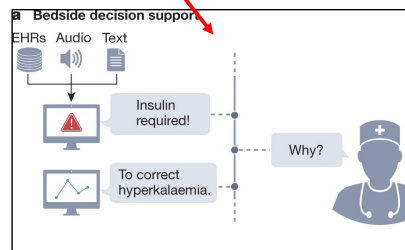
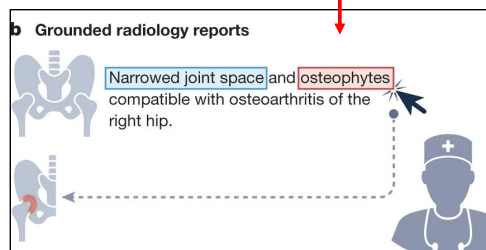
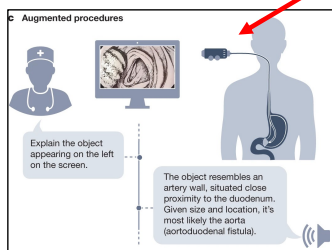
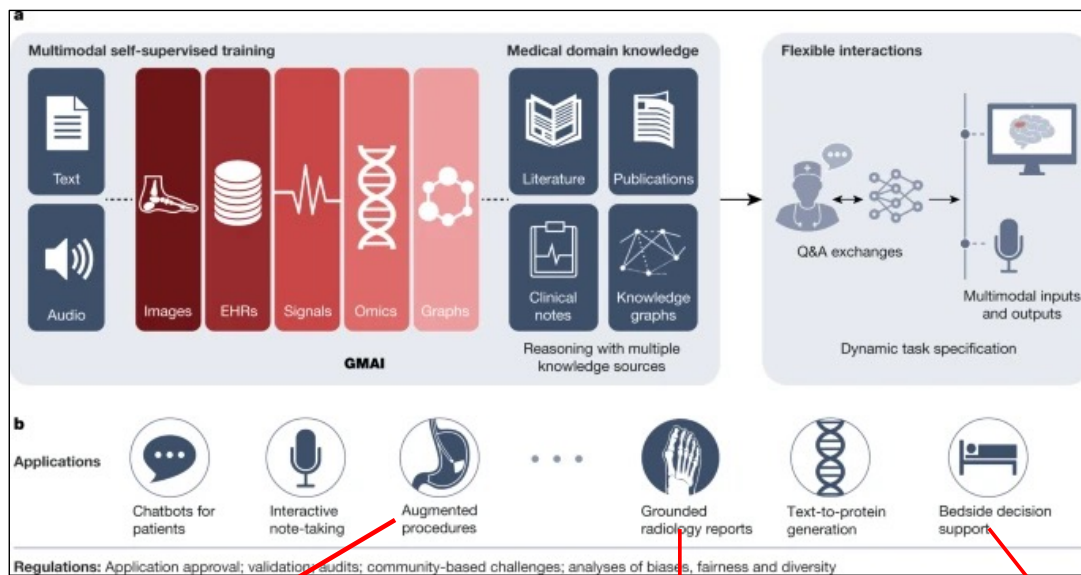
Foundation Model(基盤モデル)

- 桁違いに大きなニューラルネットワークを大規模なデータセットで事前学習したモデル
- 無数の下流タスクの基盤となるモデルであり、これを作るか・使えるかどうかは**今後のAI開発における鍵**



Foundation models for generalist medical artificial intelligence

Moor, M., et al. Foundation models for generalist medical artificial intelligence. Nature 616, 259–265 (2023).



LLM開発コストはどのくらい？

• LLaMA¹(Meta, 2023/02)の論文をもとに試算

- 650億パラメータモデル…NVIDIA A100(80G) 2048枚×21日
- GPT-3は1750億



• ABCI²を使うとして単純見積もり

- A100(40G)×8枚×24時間 … 39600 円(1650円/時間)
- 2048枚×21日間 … 約2.1億円 (1個)
- 2048枚×1年間 … 約37億円 (17個)
- 注1: GPUメモリ40Gと小さいため21日以上必要
- 注2: 現時点でABCIのA100は960枚
- 注3: 商用クラウドはもっと高額

2. 産業技術総合研究所 AI橋渡しクラウド

商用クラウドが2倍の価格として
人件費等々入れて、70-80億円/年？

1. Hugo Touvron, et al. LLaMA: Open and Efficient Foundation Language Models. arXiv:2302.1397

AIホワイトペーパー:AI新時代における日本の国家戦略

自民党:AIの進化と実装に関するプロジェクトチームの提言(2023/04)

1. 新たなAI国家戦略の策定

- 基盤モデルの進化と急速な進展に照らし、新たな政策立案とこれまでの取り組みを見直すこと。
- AI政策の司令塔を定め、研究開発、経済構造、社会基盤、人材育成、安全保障など幅広い観点から早急かつ総合的に施策を検討すること。

2. AI開発基盤の育成・強化

- 海外プラットフォームの積極的な利活用を通し知見をため、国内における基盤モデル構築、利用フェーズにも着目した人材育成に向け投資と支援を継続すること。
- 基盤モデルでの利活用を前提に、公共データのアーカイブ化と第三者提供のルールを整理。データバイアス問題に対し、国内外の基盤モデルに適切な日本関連データを提供するなど、日本に関連する学習データの比率を高める。政府主導で日本語コーパスの作成・利活用を推進。

AIホワイトペーパー:AI新時代における日本の国家戦略

3. 行政における徹底したAI利活用の推進

- 基盤モデルのAIを行政サービスに活用する具体例として、短期間で成果の見える複数のパイロットプロジェクトに直ちに着手すること。
 - 国会答弁下書き、議事録、申請書類不備、規制や制度に関する問合せ対応
- AIを活用したユースケースを蓄積・分析・共有し、関係機関でのAI導入等を支援する専門チームを政府内に設置すること。

4. 民間におけるAI利活用を奨励・支援する政策

- AIスタートアップや新規事業の創出を奨励すること。中小企業においてはAI利活用の恩恵を享受するためにITシステムのクラウド移行加速を推進・支援すること。
- 一定規模以上の民間事業者や公的機関において、AI利活用やデータの取り扱いに責任をもつチーフデジタルオフィサー(CDO)の設置を推奨すること。

AIホワイトペーパー:AI新時代における日本の国家戦略

5. AI規制に関する新たなアプローチ

- 諸外国を分析し、1)重大な人権侵害、2)安全保障、3)民主主義プロセスへの介入など、AI新時代において法規制が必要と考えられる分野につき具体的に検討すること。
- デジタル原則に基づくアナログ規制の見直しがさらに促進される仕組みを確立していくこと。
- 生成AIに関する知的財産法の解釈を巡る議論につき、わが国の強みであるコンテンツ産業がより発展できるようガイドライン等の策定を検討する。
- 公教育の現場における生徒による大規模言語モデルの利用可否などAIの取り扱いに関する指針を早急に策定すること。

医療分野からの期待

- 診察の自動記録や文書作成の支援 ^{1,2,4}
 - 医療者と患者の会話からSOAP形式でカルテ記録・診察要約・紹介状・退院サマリ・公文書の作成支援… 文書の自動化への期待が大きい
- 医療安全, 診断補助, 治療提案 ³
 - ベッドサイドでの医療記録や時系列データからの注意喚起
 - 手術映像へのリアルタイムな注釈付け、手順がスキップされたときの警告、まれな解剖学的現象に遭遇したときの関連文献読み上げ
- 医療相談・遠隔治療 ⁴
 - 患者向けチャットサポート、遠隔医療のための患者の機器セットアップ支援
- 公衆衛生 ⁵
 - ライフスタイル、予防接種、スクリーニング、危険因子の回避
 - 感染症、慢性疾患、環境被害などに関するリアルタイムの情報提供

1. Limits, and Risks of GPT-4 as an AI Chatbot for Medicine. N Engl J Med 2023; 388:1233-1239

2. ChatGPT: the future of discharge summaries? Lancet Digit Health. 2023 Mar;5(3):e107-e108.

3. Foundation models for generalist medical artificial intelligence. Nature 616, 259–265 (2023).

4. Considering the possibilities and pitfalls of Generative Pre-trained Transformer 3 (GPT-3) in healthcare delivery. NPJ Digit Med. 2021 Jun 3;4(1):93.

5. Role of Chat GPT in Public Health. Ann Biomed Eng. 2023 May;51(5):868-869.

医療分野からの懸念

- **学習データのバイアスによる差別**
 - 有害な相関関係に基づく、特定集団の過小評価(人種差別・性差別)
 - 文化や制度に応じない不適切な提案(日本の医師国家試験で**安楽死を選択**⁶⁾)
 - **提案される内容の検証の難しさ**
 - 正しくない情報を捏造する可能性があるため、検証されたユースケースを越える提案の場合にはその旨を発するべき
 - 提案される内容の検証に、異なる専門家(放射線科、病理、腫瘍、その他専門家)の判断が必要になる可能性
 - **プライバシー関連の問題**
 - 学習データに含まれる機密情報(個人情報、知的財産など)を生成
 - サービス提供者にプロンプト内で個人情報・要配慮情報が送信され再利用
 - 開発者用のバックドアを利用される可能性
1. 医療応用には創造性よりも**正確性・安全性を優先したLLMが必要**
 2. リスクに対する知見をためて、**安全に活用する方法を確立する必要**

6. Kaneda, Yudai, et al. Can ChatGPT Pass the 2023 Japanese National Medical Licensing Examination?. 10.20944/preprints202303.0191.v1.

AIは医師を置き替えるか

- あらゆる情報を”データ化”すれば、**平均的な医師を超える判断が可能になりつつある**。AI技術を活用するためには、データ化することが泥臭いながらも価値を持つ時代。
- 現状では、**法制度が未整備なことや技術の不透明さが課題**。また、既存AIと違い多用途のため、医療機器として承認するための**検証方法も課題**。
- それを超えれば、問診や検査の結果をみて診断し処方するだけといった**狭い範囲のサービス**は、社会が望めばAIが代替しうる時代になる。
- 一方で、医療にとって重要な他者とコミュニケーションや相互理解、遠回りをしてでも協調し物事を進める能力を有するようになるかは未知。
- また、高い倫理観や法の遵守、他者の尊重、不確実性への対応が求められる医療において、医師を代替するようなAIのハードルはまだ高い。