

人工知能の機械学習

技術政策学（データ科学編）

土井翔平

2023-05-22

はじめに

⚠ 警告

日進月歩の分野なので、本章の内容はすぐに古いものになる（or 既にそうであるかもしれない）点に注意。

ビッグデータは魅力的な資源（材料）だが、有効な利用法（調理法）があって初めて価値を持つ。

→ 近年のデータ科学における 2 つの変革

1. 機械学習：データから一定のパターンを機械（パソコン）が学習し、**予測**をする。
2. 因果推論：データから因果関係（因果効果）を学習する。

→ いわゆる（最近において）人工知能と呼ばれるものは機械学習（予測）

- ・ 現在は第 3 次人工知能ブームと言われている。
- ・ クオリティが高いがゆえに、あたかも機械が人間のように思考しているように見えてしまう。
- ・ 自然言語処理に関するタスク [SuperGLUE](#) では人間を越えている。

生成 (generative) AI：ある情報から、別の情報を出力するモデル

- ・ **大規模言語モデル** (large language model: LLM)：大量のテキストを使い、巨大なモデルを学習した生成 AI
 - [ChatGPT](#)、[Google Bard](#)、[Microsoft Bing AI](#) など
 - GPT=Generative Pre-trained Transformer
 - まずは、[OpenAI の Playground](#) で遊んでみよう。
- ・ 画像生成の性能も向上、Vision and Language の発展
 - [DALL·E 2](#)、[midjourney](#)、[stable diffusion](#) など

生成モデルも実は予測の組み合わせである。

- DeepL ⇨ ある言語の文章から他の言語の文章を予測する。
- チャット bot、文書要約、コード生成 ⇨ ある文章から返答、要約、次に来る文章を予測する。
- Amazon や Netflix の推薦 ⇨ これまでの購入履歴やウォッチリストから次に購入する商品を予測する。
- 学習過程にテキストデータを含めることで、テキストから画像生成できる (vision and language)。

代表的な機械学習の分類

1. 教師あり学習：特徴量 (feature) から対象を予測する。
2. 教師なし学習：多様な特徴量から重要なものを抽出する。
3. 強化学習：フィードバックを通じて最適な方策 (policy) を発見する。

⇨ これらの概要を理解し、生成 AI が何をしているかを理解する。

1 教師あり学習

教師あり学習とは、機械に人間の判断のパターンを学習させ、模倣できるようにすること。

⇨ 言い換えれば、予測 (prediction) というタスクを実行できるように訓練する。

- 写真とその内容のペアのデータを機械に覚えさせる。
- 住宅の情報（間取り、最寄り駅までの距離.....etc）と価格を機械に覚えさせる。



図1: 教師あり学習のイメージ

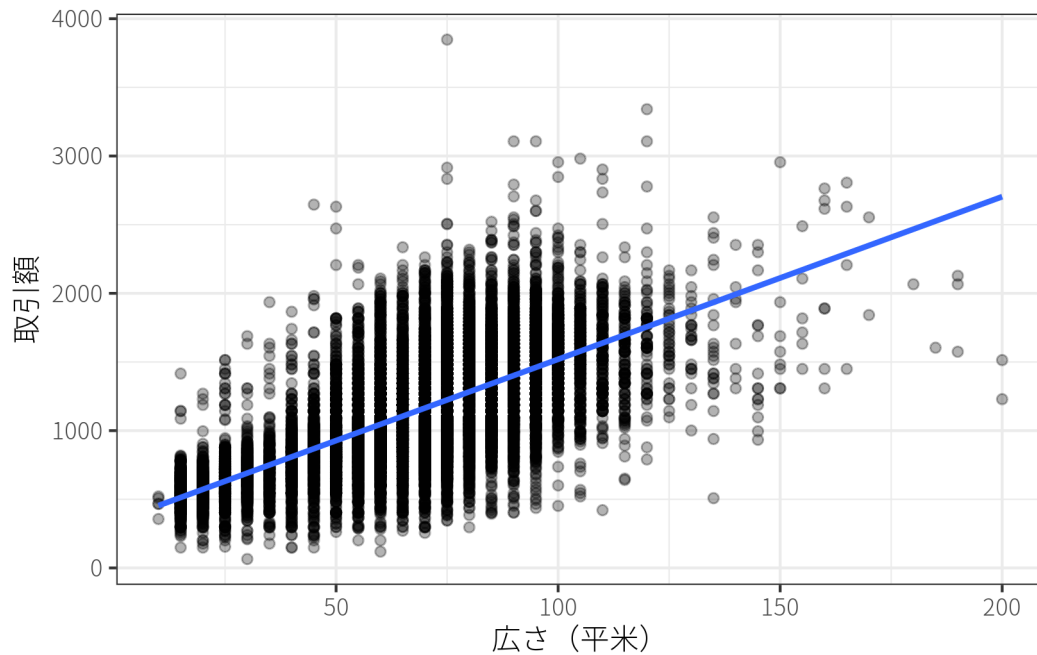
⇨ ある情報を入力すると、それに対応する情報を出力する。

- 入力情報に対応する出力（正解）を人間が判断する **アノテーション** が重要になる。

1.1 回帰分析

シンプルで、広く使われている教師あり学習の手法として **回帰分析** (regression analysis) がある。

- 例えば、北海道の中古マンション価格の教師あり学習を行ってみる。



$$\text{価格} = 335.58 + 11.84 \times \text{広さ}$$

最小二乗法 (ordinary least squares: OLS) はデータとの誤差が最も小さくなる直線を計算する。

- 予測値を一次関数（直線）で予測する。

$$i \text{ の予測値} = \hat{y}_i = \underbrace{\hat{\alpha}}_{\text{切片 (intercept)}} + \underbrace{\hat{\beta}}_{\text{傾き (slope)}} x_i$$

- i は個体ごとに異なる値を取るということを意味している。
- 上手く予測できるような $\hat{\alpha}, \hat{\beta}$ をデータから求める（学習する）。
- 真の値と予測値のズレ、誤差 (error) が小さい方がいいはず。

$$i \text{ の予測誤差} = i \text{ の真の値} - i \text{ の予測値} = y_i - \hat{y}_i$$

- ズレはプラスにもマイナスにもなるので、プラスの値しか取らない距離や面積に変換する。

- 通常は誤差を二乗して、面積にする。

$$i \text{ の予測誤差の二乗} = i \text{ の真の値} - i \text{ の予測値} = (y_i - \hat{y}_i)^2$$

- 個々の誤差をデータ全体について計算し、合計する。

$$i \text{ の予測誤差の二乗の合計} = (y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 + \dots$$

⇒ これを最小にする $\hat{\alpha}, \hat{\beta}$ をデータから求める！(パソコンが計算してくれる)¹

- $\hat{\alpha}, \hat{\beta}$ は \hat{y}_i の中に入っていることに注意。

予測に使う情報(特徴量)は1つである必要はない。

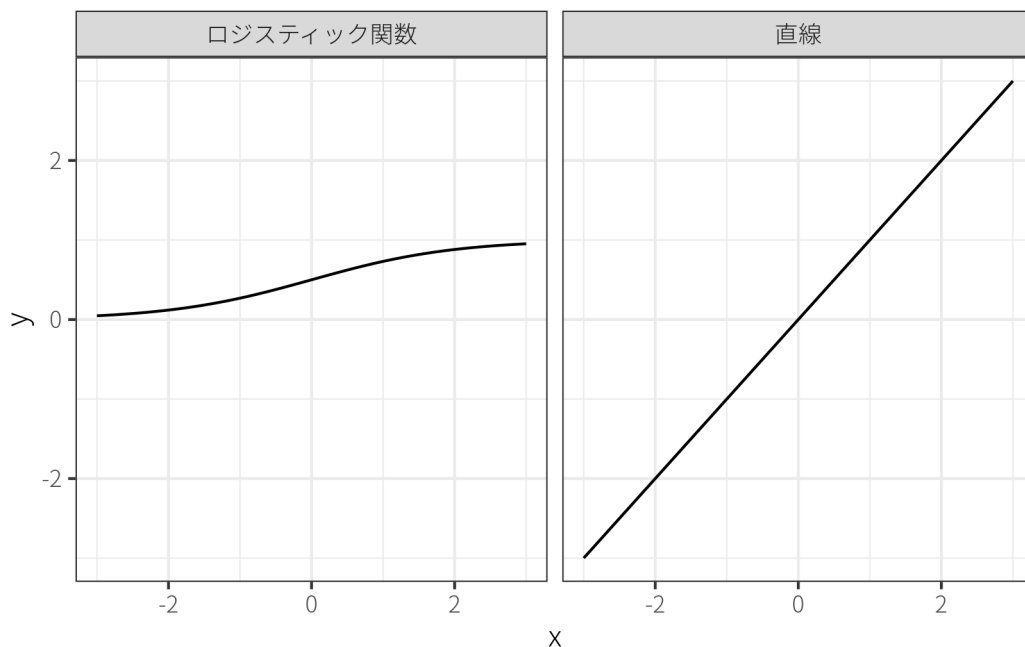
$$\text{価格} = 396.27 + 12.50 \times \text{広さ} - 10.57 \times \text{距離}$$

- パターンを学習しているだけであり、機械がマンションについて理解しているわけではない。

予測対象がカテゴリーの場合はどうするのか？

- 機械学習の代表的なデータセットに**タイタニック号の乗客データ**がある。
- このときの予測対象は乗客が生存したかどうかというカテゴリー

⇒ **ロジスティック関数**(シグモイド関数)を使って変形すると、0 から 1 の間に収まる。



¹ 最適化(偏微分係数が0となる値を求める)によって明示的に解くことができる。

1.2 決定木

回帰分析以外の代表的な教師あり学習の手法として**決定木** (decision tree) がある。

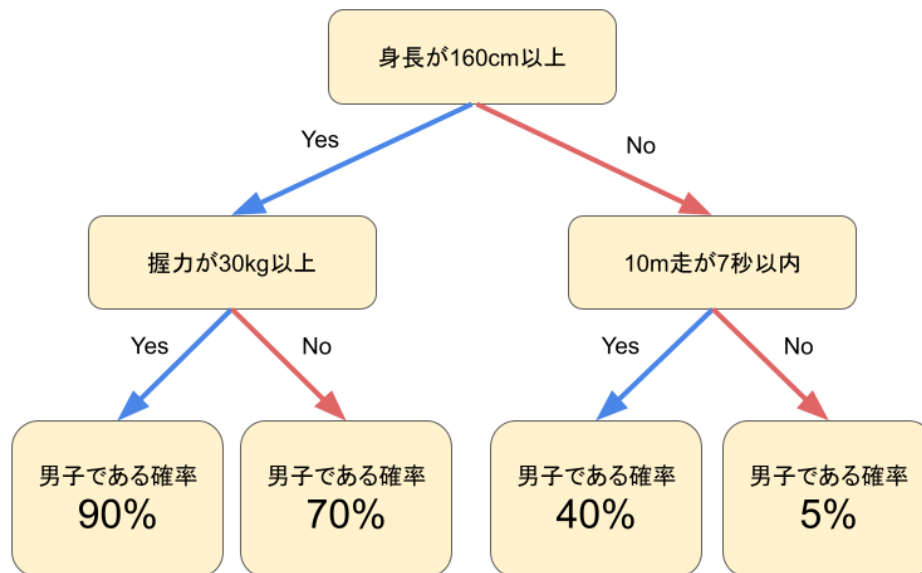


図2: 決定木のイメージ

⇒ 弱い決定木をたくさん集めた**ランダム・フォレスト**（やその発展形²）がよく使われている。

- ・ 三人寄れば文殊の知恵？ 陪審定理？

1.3 深層学習

深層学習 (deep learning) は深層ニューラル・ネットワーク (deep neural network: DNN) と呼ばれる。

⇒ もともとは人間のニューロンをマシン上で再現すれば人工知能ができるかという期待

- ・ 閾値を超えると発火して信号を送信する。

⇒ 回帰分析をニューロンとして見て³、これをたくさん作る。

なぜ深層学習はすごいのか？

1. 隠れ層を増やせば増やすほど柔軟な予測ができる。

² XGBoost や LightGBM など。

³ 厳密に言えば、活性化関数を挟む。

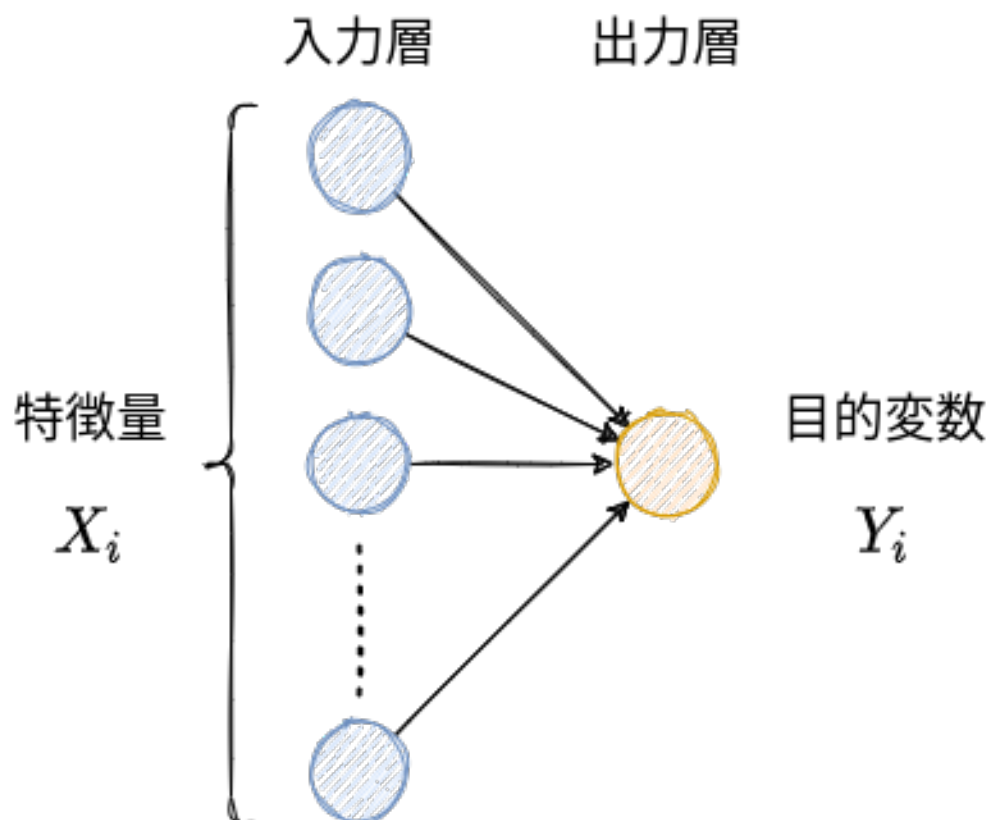
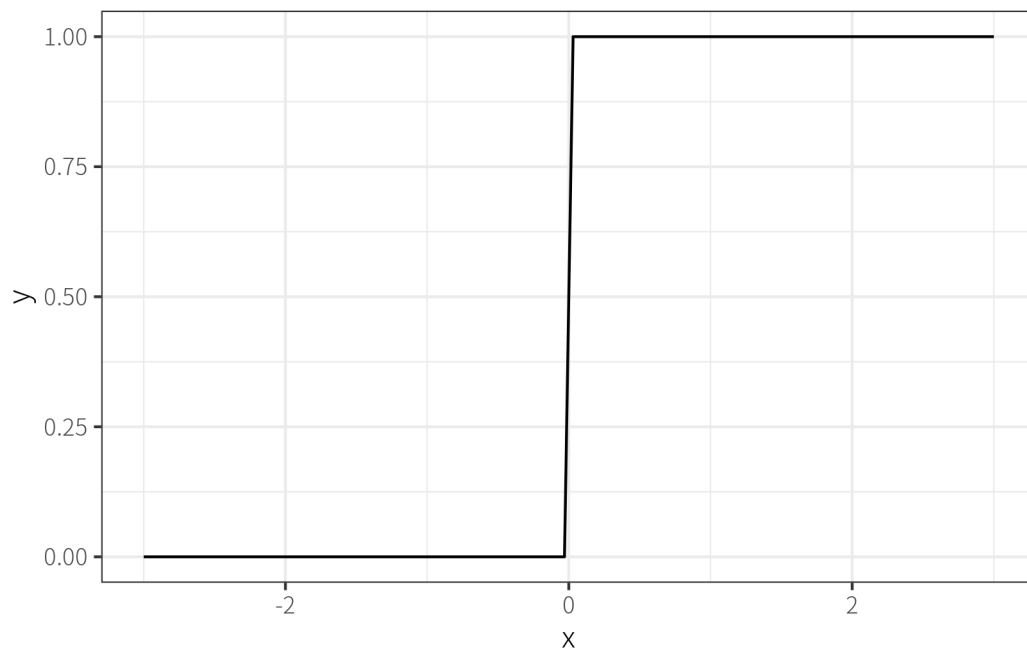


図3: ロジスティック回帰のイメージ

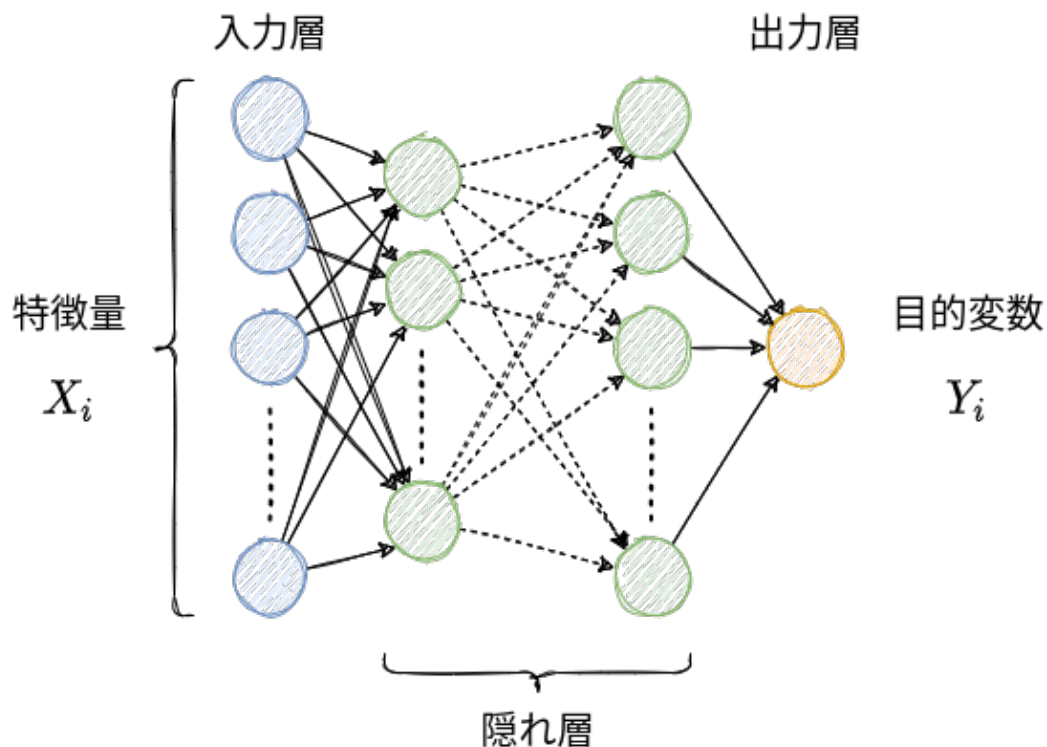


図4: 深層ニューラル・ネットワークのイメージ

- パラメータの数（隠れ層の数に比例） \approx モデルのサイズ
- 例えば

$$\hat{y}_i = \hat{\alpha} + \hat{\beta}x_i$$

のパラメータの数は2

2. 特徴量を人間が作らなくてよい。
 - むしろ、重要な特徴量がなにかを学習する（表現学習）。
3. 学習済みモデルを使える。
 - タスクに応じて出力側を再学習（ファイン・チューニング）する。
 - GPT=Generative Pre-trained Transformer
4. 様々な形式のデータ（テキスト、画像、音声.....）を同じ枠組みで分析できる。
 - vision and language などマルチモーダルなモデルの開発

→ 生成 AI (LLM 含む) は与えられた単語の列から、次に来そうな、もっともらしい単語を予測している（だけ）！

- 同じ入力に対して同じ回答をしないように、ある程度ランダムに予測をしている。
 - GPT における temperature はランダム度合いを指定している。
- 人工知能の獲得か？（例、中国語の部屋）

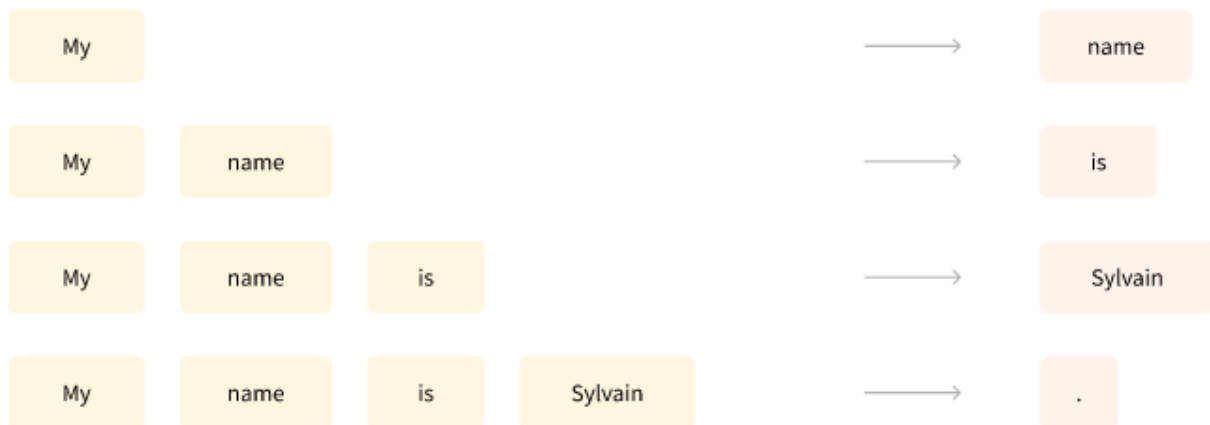


図5: LLM による文書生成のイメージ

1.4 プロンプト・エンジニアリング

ChatGPT などの最近の LLM がすごいのは、タスクも指示するだけでよいこと。

- ・ 翻訳、要約、質疑応答などのタスクごとにモデルを作らなくて良い。

⇒ 入力する指示文（**プロンプト**）をどのようにするのが重要。

- ・ プロンプトの書き方を工夫することをプロンプト・エンジニアリングなどと呼ぶ。
- ・ いくつかの具体例を提示すると、性能が良くなる、安定する **few shot learning** という現象 (?)

1.5 拡散モデル

拡散モデル：画像にノイズを追加していった、それを取り除くプロセスを学習

⇒ テキストからの画像生成もテキスト → 画像の予測を行っている。

2 教師なし学習

数値ではない文書データをどのようにデータ分析するのか？

- ・ シンプルな方法は **bag of words**（単語の出現頻度を特徴量とする）アプローチ

⇒ 教師なし学習によってデータから特徴量を抽出する。

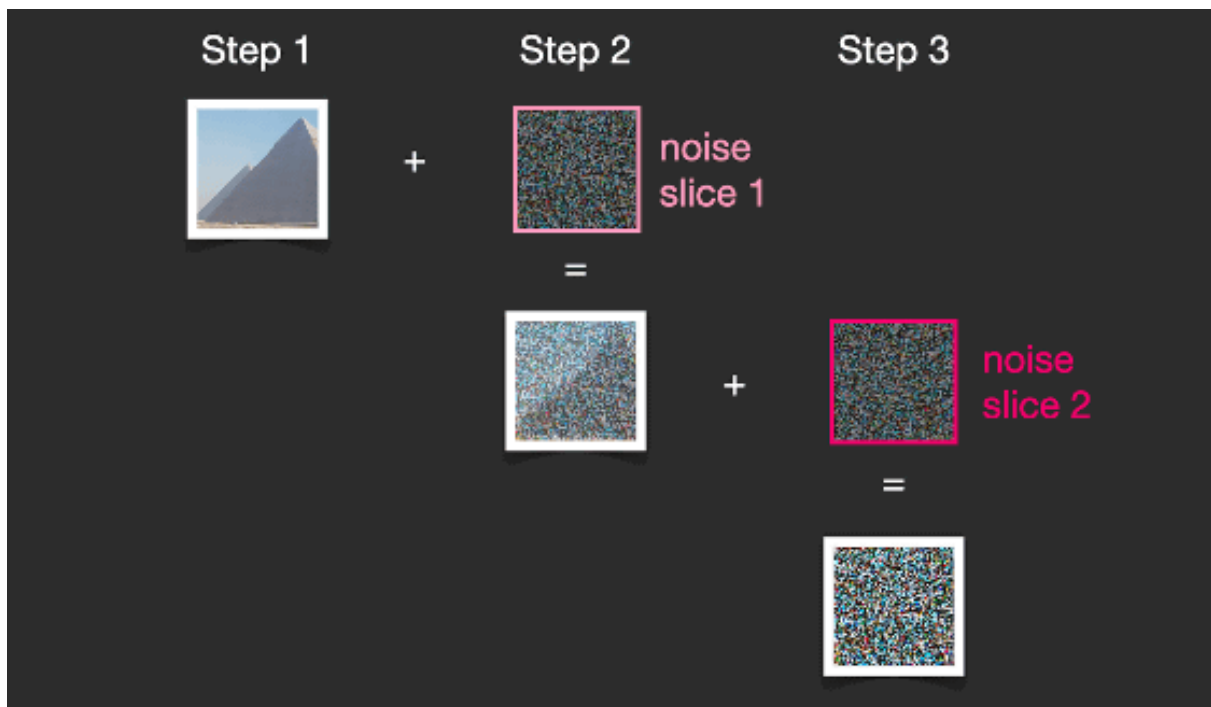


図6: 拡散モデルのイメージ

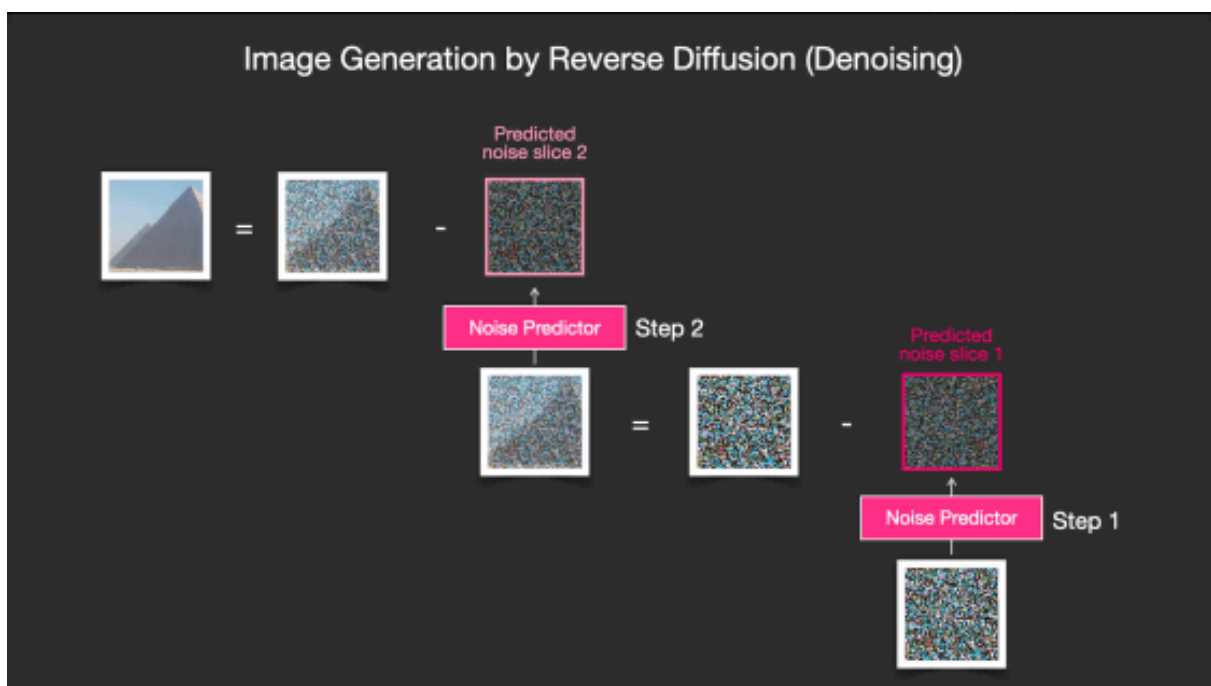


図7: 拡散モデルのイメージ

2.1 単語埋め込み

単語埋め込み (word embedding) : 単語を低次元空間のベクトルに変換する

→ 単語のベクトル (位置) ≈ 意味?

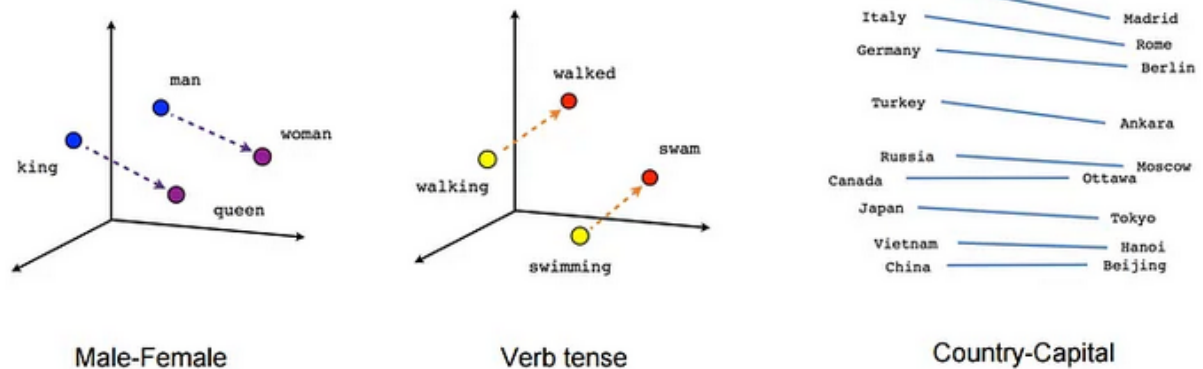


図8: 単語埋め込みのイメージ

- 単語の意味は周辺の単語によって決まるはず。

→ 周辺の単語をうまく予測できるベクトルを学習する。

- 画像も同様に埋め込むことで、テキスト → 画像の予測が可能となる。

2.2 自己注意機構

近年の自然言語処理の飛躍的発展のキーは Transformer(Vaswani et al., 2017) の登場⁴

- BERT=Bidirectional Encoder Representations from Transformers
- GPT=Generative Pre-trained Transformer

→ 特に自己注意機構 (self-attention mechanism) が重要 (と言われている)


- ある単語を処理する際に、他のどの単語に注目すればよいのかを学習する。
- “Attention is all you need”(Vaswani et al., 2017)
- 離れた単語も踏まえた学習ができる。
- 学習速度が高速になる。⁵

⁴ Transformer を提案したのは Google の研究者たち。

⁵ 並列化が可能になるため。

【Skip-gramが解く穴埋め問題】

The king of is bald.



【CBowが解く穴埋め問題】

The present of France is bald.

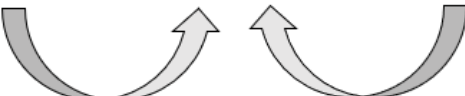


図9: 単語埋め込みの学習のイメージ

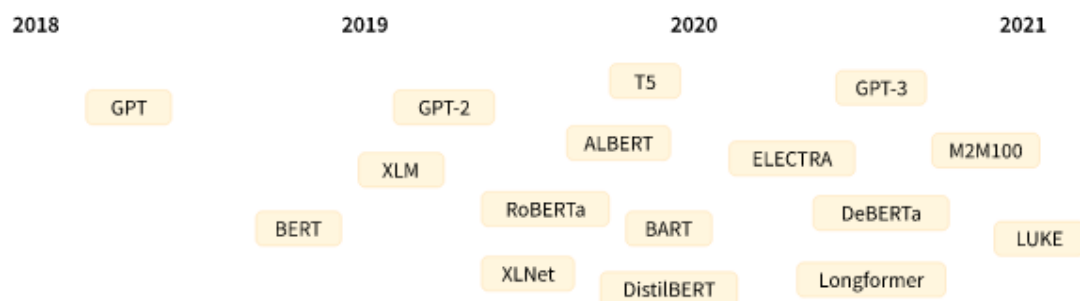


図10: Transformer ベースの LLM

3 強化学習

一部のモデルでは強化学習を用いて、更に性能を向上させている。

- 例、GPT-3 やそれ以上のベースと考えられている InstructGPT

強化学習では機械が最適なアクションを見つける。

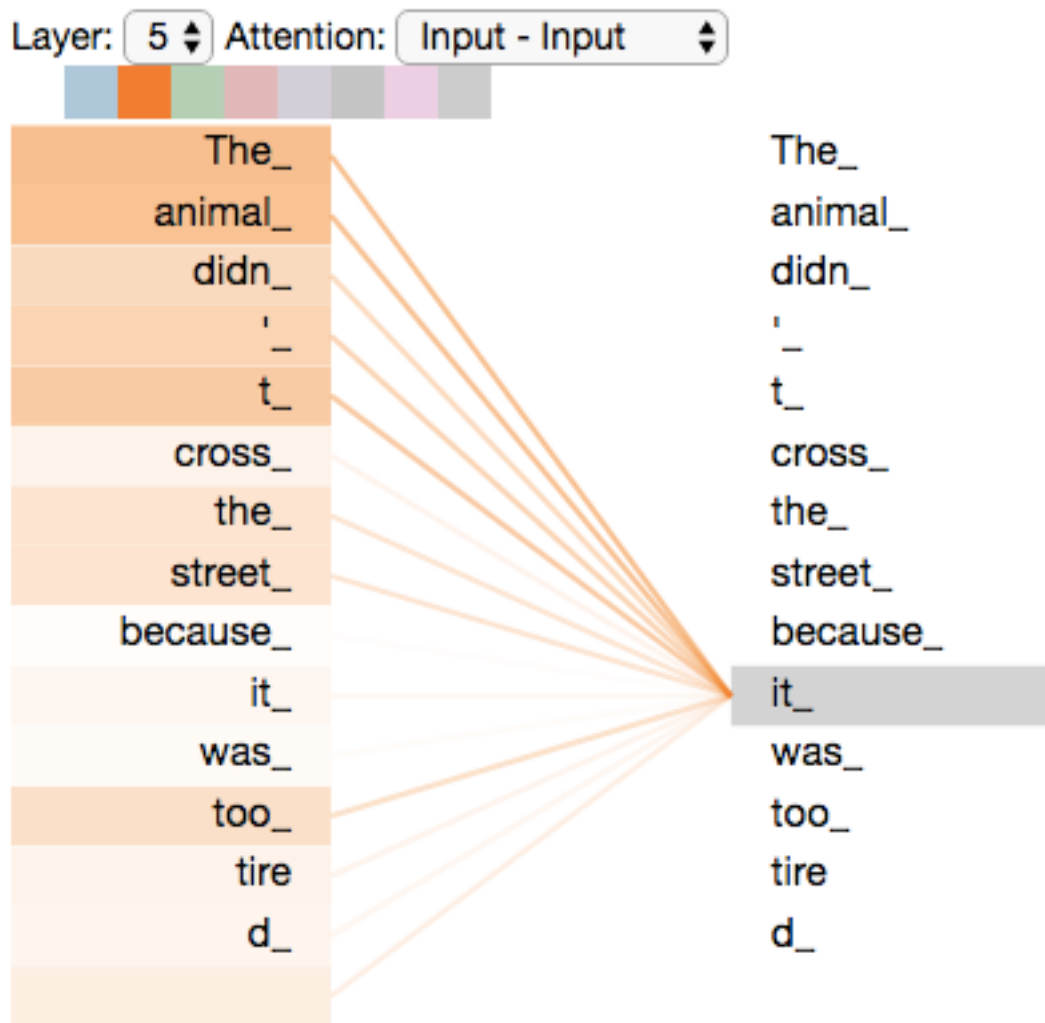


図11: 自己注意機構のイメージ

- データからパターンを発見するわけではない。
- エージェントが状態 → 行動 → 環境 → 報酬・状態を繰り返し、報酬を大きくする行動を発見する。

3.1 AlphaGo

AlphaGo は強化学習で囲碁のアルゴリズムを学習し 2016 年に世界トップ棋士に勝利

- 初期設定として実際の棋譜データから学習したアルゴリズムを使用
- 2つのエージェント（機械）が対戦する中で大量にデータを生成し、学習

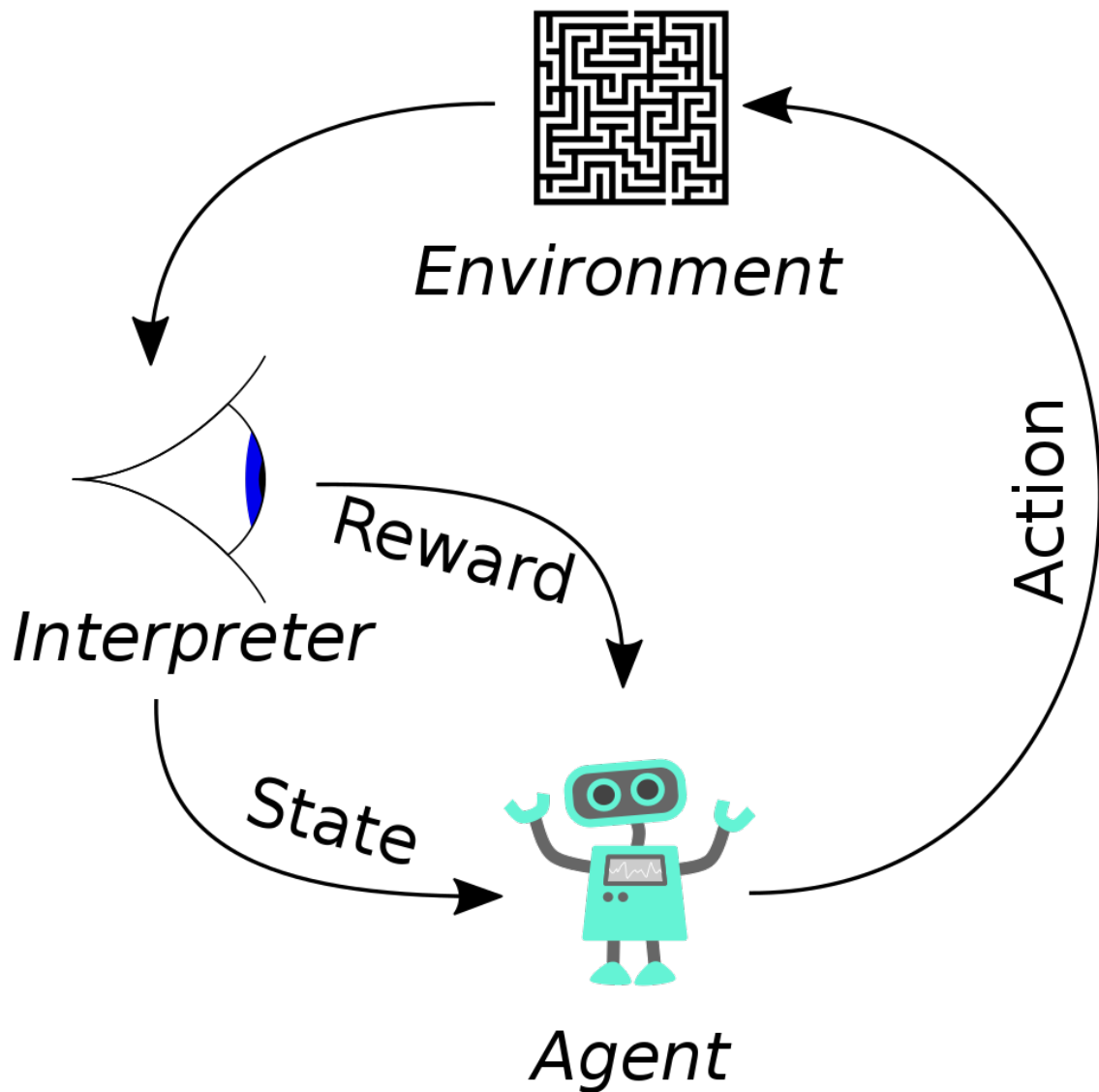


図12: 強化学習のイメージ

3.2 人間のフィードバック

InstructGPT では人間のフィードバックからの強化学習 (Reinforcement Learning from Human Feedback: RLHF) を用いている。

1. プロンプトと（人間の作った）回答のデータから教師あり学習
2. 1 で作ったモデルの回答結果と人間の採点結果のデータから教師あり学習
3. 1 と 2 のモデルを使ってプロンプト → 回答 → 採点の強化学習

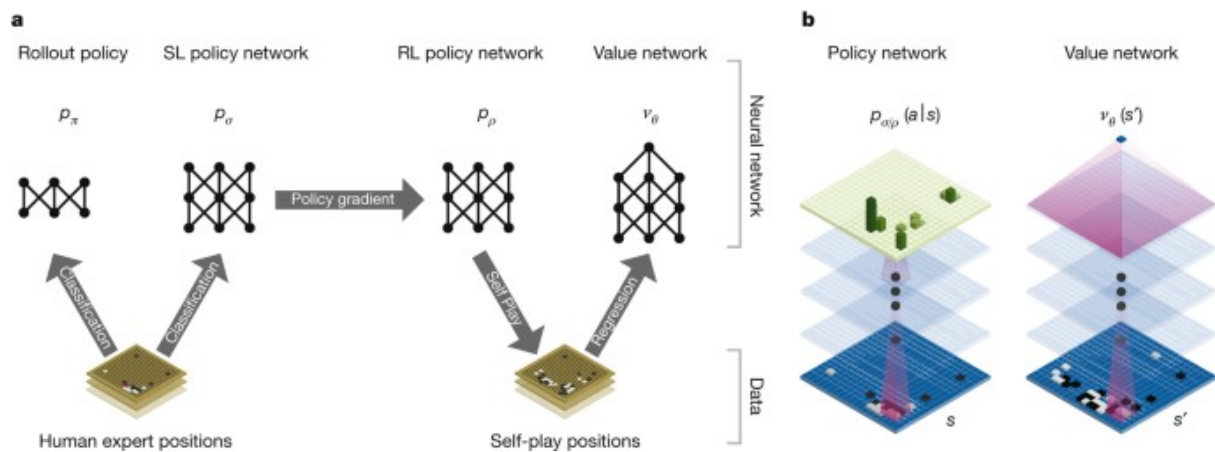


図13: Silver et al. (2016)

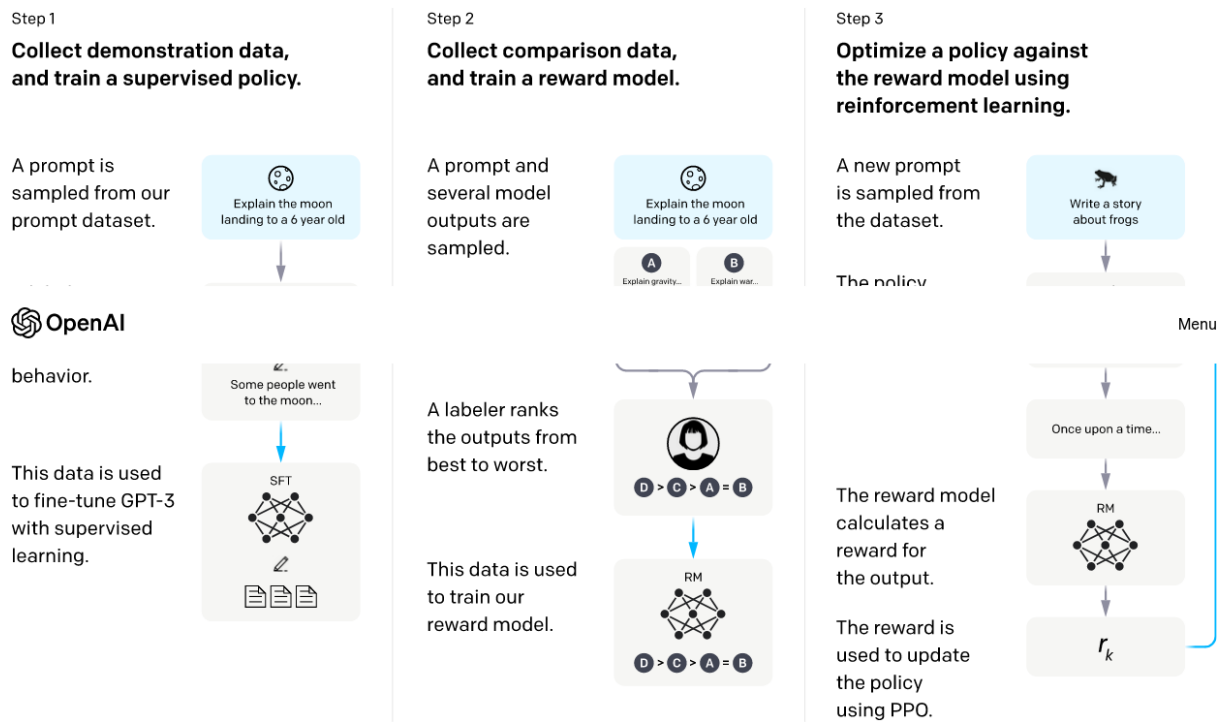


図14: Aligning language models to follow instructions

4 人工知能の社会的課題

あらゆる技術がそうであるように、新しい技術の登場、急速な普及は様々な問題を引き起こしうる。

- G7 広島サミットで生成 AI に関するルールを策定する広島 AI プロセスの立ち上げに合意

4.1 人工知能の使い方

4.1.1 人工知能の悪用

AI を悪用 ⇨ 偽の情報を作り、拡散

- 文書生成 ⇨ フェイクニュース
- 物体検出 & 画像生成 ⇨ Deep Fake

⇨Deep Fake はオンラインの画像や映像を加工 ⇨ フェイクポルノなど深刻な被害の可能性

改ざんされた記者会見写真

ツイッター上に投稿された写真。
表情が加工されている



フジテレビが実際に放送した
映像の一場面



図15: Deep Fake の例

- ちなみに、スクリーンショットの捏造は人工知能を使わなくても簡単にできる。

どんな技術も悪用しようと思えばできる ～（個人的に）重要なのは、悪用よりも誤用

4.1.2 人工知能の誤用

統計的差別 (statistical discrimination)：個人の属性（属する集団）に基づく差別

- ・ 学歴、性別、人種などで就職活動、賃貸契約、判決が異なる。

～ 機械学習はあくまで「人間を模倣」するので、人間に（無意識でも）差別があれば、機械はそれを学習する。

アメリカの一部の州では保釈や刑期を決定する際に COMPAS というシステムでリスク評価を行っていた。

- ・ 黒人では「再犯する」と予測されたが「再犯しなかった」という間違いが多かった。
- ・ 白人では「再犯しない」と予測されたが「再犯した」という間違いが多かった。

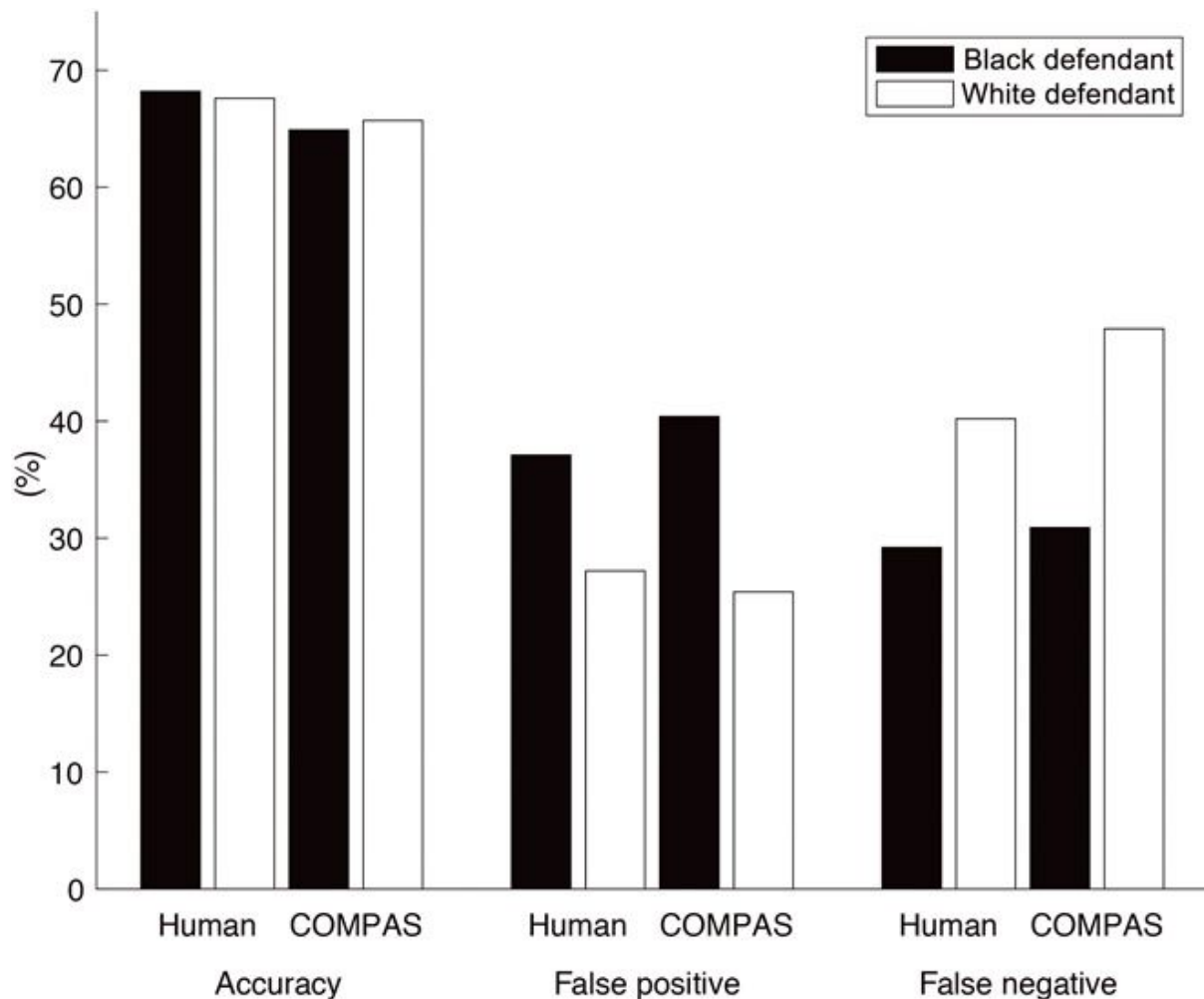


図16: COMPASS による再犯予測

学習に用いるデータが偏っている ～ 予測も偏る。

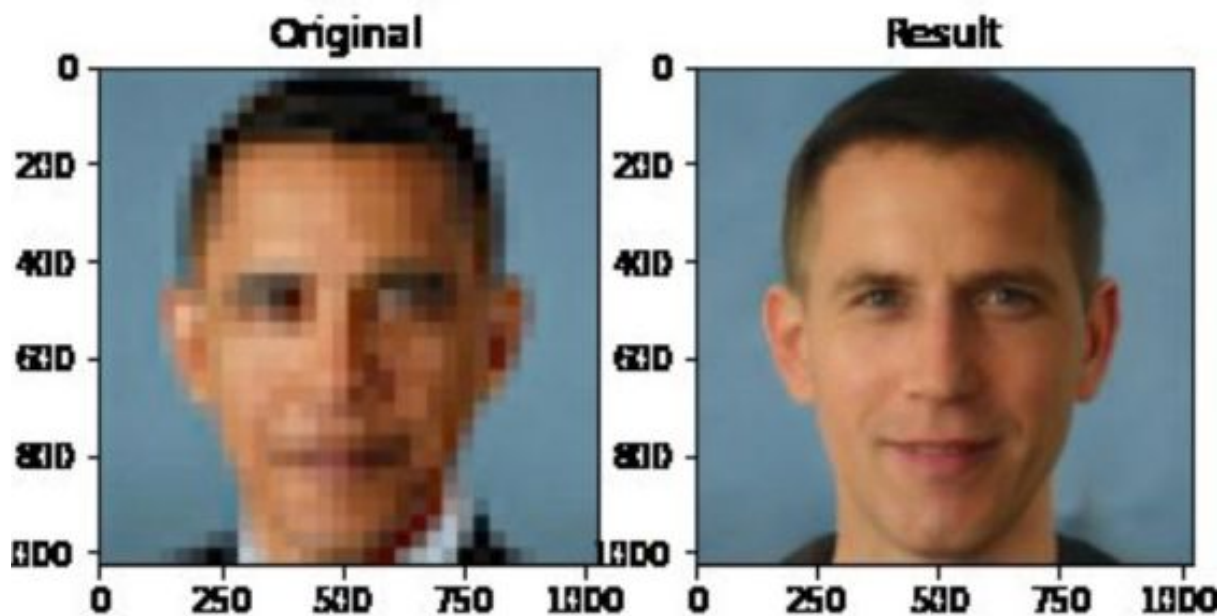


図17: 偏ったデータによる学習

→ 公平性のある機械学習を行う必要がある。

敵対的攻撃 (adversarial attack) : あえて AI を騙すような情報入力

ChatGPT などで指摘されているのは幻覚 (hallucination) と呼ばれる現象

- 正しくない回答をあたかも正しいものと堂々と提示する。

意図せざる形での著作権侵害や個人情報流出も？

- 学習に使用されたデータがほとんどそのまま出力される可能性

→ AI だからといって客観的であるわけでも、常識的であるわけでもない。

- 特に深層学習の中身はブラックボックスであり、説明可能性に欠けている。
- 回帰分析や決定木は分かりやすい。

4.1.3 人工知能と倫理

AI 倫理、公平性のある AI が必要

→ AI が守るべき倫理、公平性とは？

自動運転車がトロッコ問題に遭遇したとき、どうすべきなのか？

モラル・マシンというアンケートに答えることで、どのような命を重視するのかが分かる。

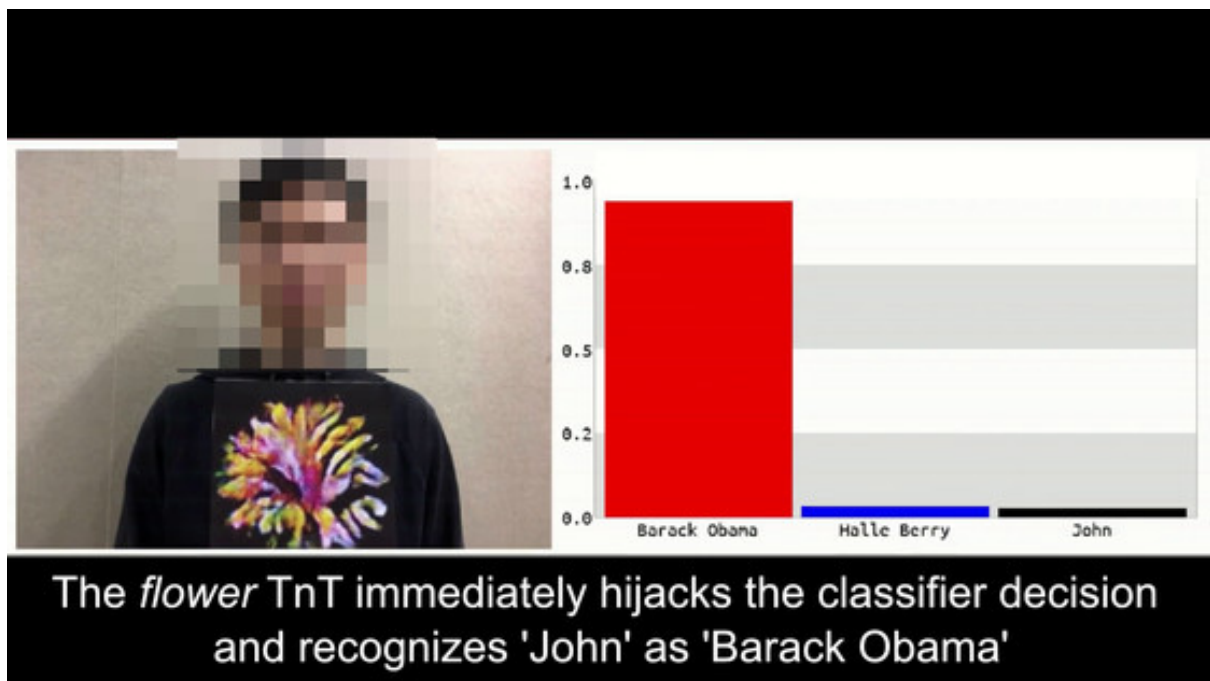


図18: 敵対的攻撃

- 「審査を始める」を押す。
- 直進する場合は左の絵を、曲がる場合は右の絵をクリックする。
 - ドクロマークがついている人が死んでしまうとする。

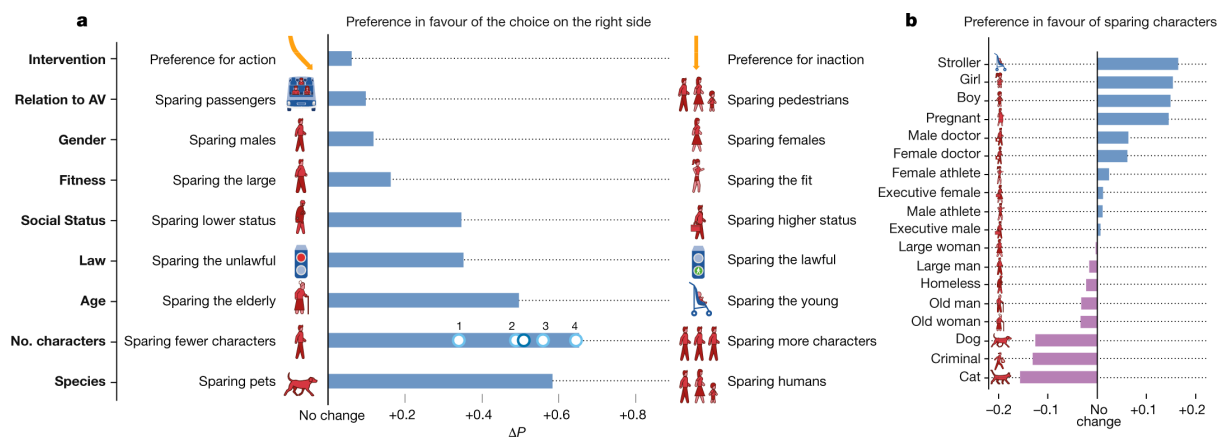


図19: Awad et al. (2018)

倫理観は国ごとに異なる。

日本の場合は、

1. お年寄りを助け、
2. より多くの人を助けるわけではなく、

MORAL COMPASS

A survey of 2.3 million people worldwide reveals variations in the moral principles that guide drivers' decisions. Respondents were presented with 13 scenarios, in which a collision that killed some combination of passengers and pedestrians was unavoidable, and asked to decide who they would spare. Scientists used these data to group countries and territories into three groups based on their moral attitudes.

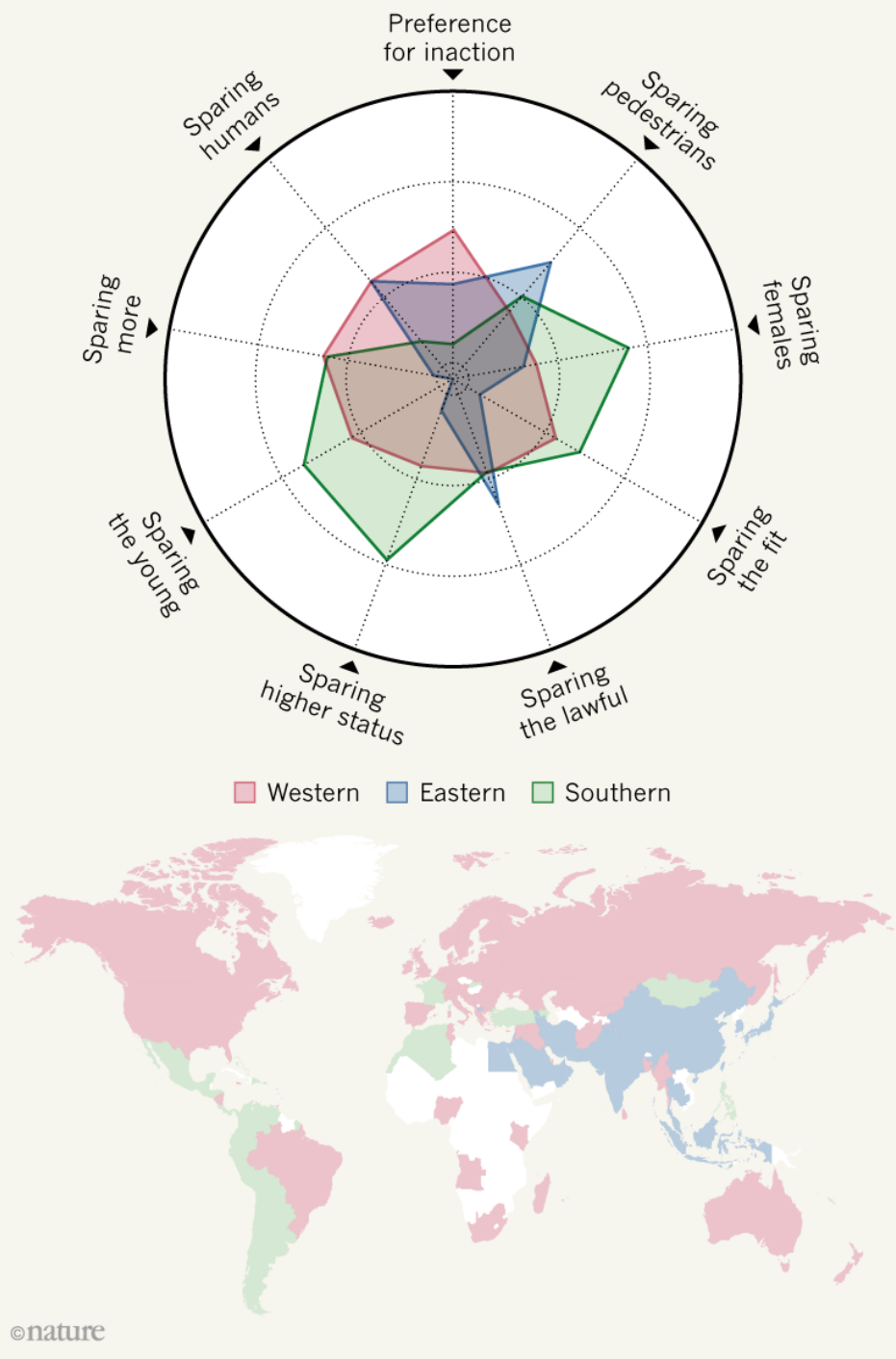


図20: AI 倫理観の地域ごとの違い

3. 歩行者を助ける

傾向にあるらしい。

1. 人工知能は人間に代わって倫理的判断を行うわけではない。
 - ・ 人工知能に期待する倫理も人々、国々の間で同じではない。
2. 人工知能は誤作動も起こりうる。

⇒ 人工知能の責任はどこにあるのか？

- ・ 生成 AI (Generative AI) の倫理的・法的・社会的課題 (ELSI) 論点の概観：2023 年 3 月版（大阪大学）

4.1.4 人工知能と兵器

無人兵器と AI 技術の発展は自律型致死兵器システム (Lethal Autonomous Weapons Systems: LAWS) の可能性を現実のものとしつつある。

- ・ 人間の代わりに戦闘をさせることで人命の損失を減らせる／戦争やテロリズムが起こりやすくなる？
- ・ 人道法の違反（文民の殺害など）を起こす？
 - 誤作動だけでなく、現在の AI 技術では AI の判断を人間が理解できない可能性

⇒ LAWS 規制に関する議論の指針が 2019 年に定まったばかり。

4.2 人工知能と社会

4.2.1 人工知能と政治

AI と民主主義の関わりに着目されつつある？

- ・ AI が人間に代わって政策を決定する。
- ・ AI が個人の思考を学習して人間の代わりに政治（議論や投票……）をする。
- ・ AI が議論や情報収集をアシストする。

科学技術を巡って展開される政治がある。

⇒ なぜ、特定の技術がグローバル・スタンダードとして支配的になるのだろうか？

ネットワーク外部性：利用者が多いと、利用するメリットが増える。

⇒ ある程度の規模の人々（クリティカル・マス）がその製品や技術を用いる ⇒ 他の人も使うようになる。

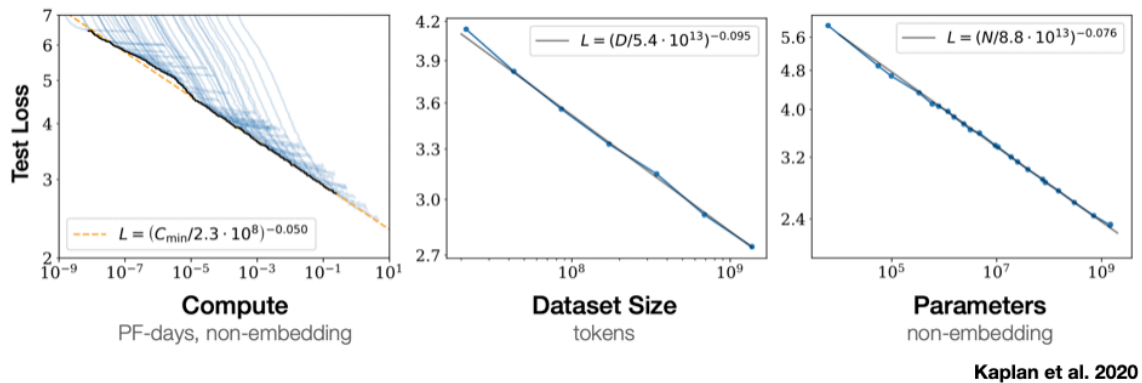
規模の経済：生産量が増えると、平均的な費用が低下する。

⇒ 単独で市場のニーズを満たすことができる。

特に LLM の場合は初期投資が莫大 ⇒ 新規参入が更に困難

- 現在の LLM で重要なのは計算資源 × データ量 × モデルのサイズ ⇨ 資金&時間

Scaling Law for Neural Language Models



自然言語モデルは1)計算能力 2)データ量 3)パラメータ量に比例して性能が上がる

図21: Kaplan et al. (2020)

- GPT-3 は 1750 億、GPT-4 は 5000 億? 以上

⇨ 先に行動する側に**先行者利益** (first-mover advantage) がある。

- **イーロン・マスクらが AI の開発モラトリアムを要求し、サム・アルトマンは AI 規制の導入を主張**
一度、スタンダードになると（たとえ不便でも）それが利用され続ける。

- **経路依存性** (path dependency)：過去の偶然の事象が長い期間に渡って影響すること

日本は鉄道や原子力発電所などのインフラ輸出に力を入れている。

- ネットワーク効果によって、その国のスタンダードになりうる。
- 輸入国は特定の国への依存を避けるため、複数の国から導入したい。

先進国による 5G におけるファーウェイの排除は安全保障だけが理由ではない。

- ネットワーク効果によってグローバル・スタンダードとなってしまう。
- 中国に依存せざるを得なくなる、いわゆる**技術覇権**への懸念

4.2.2 人工知能と環境

LLM の開発には大規模な計算が必要 ⇨ 莫大な電力消費と環境不可

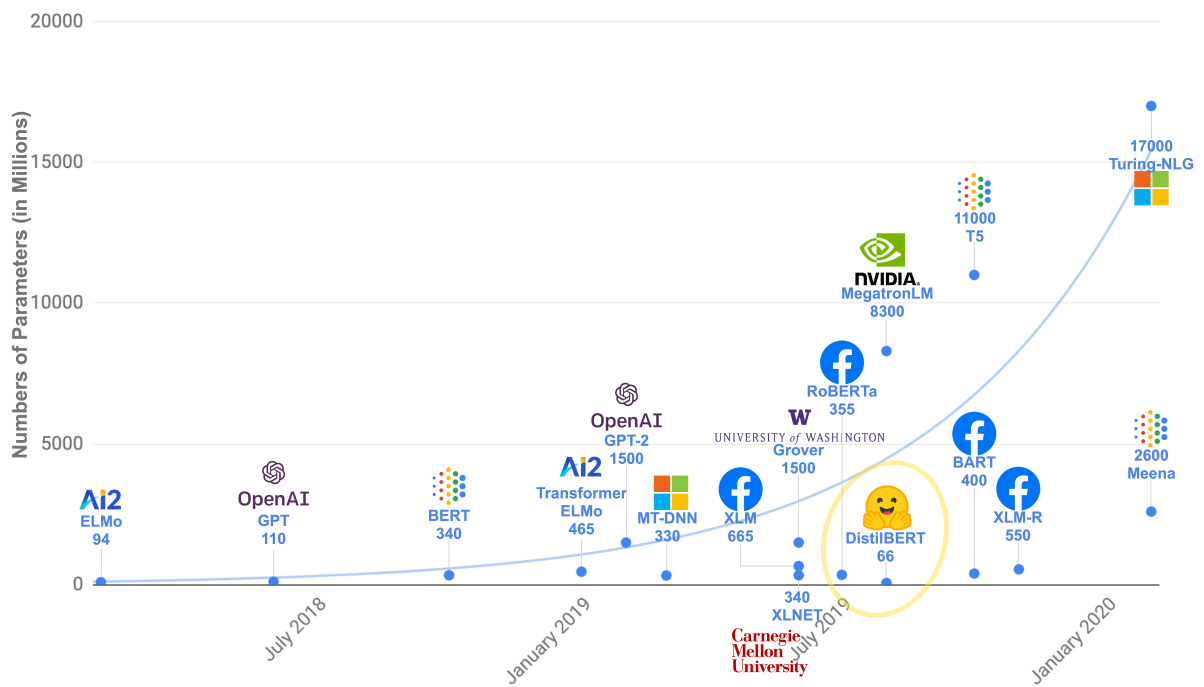


図22: LLM の規模の発展

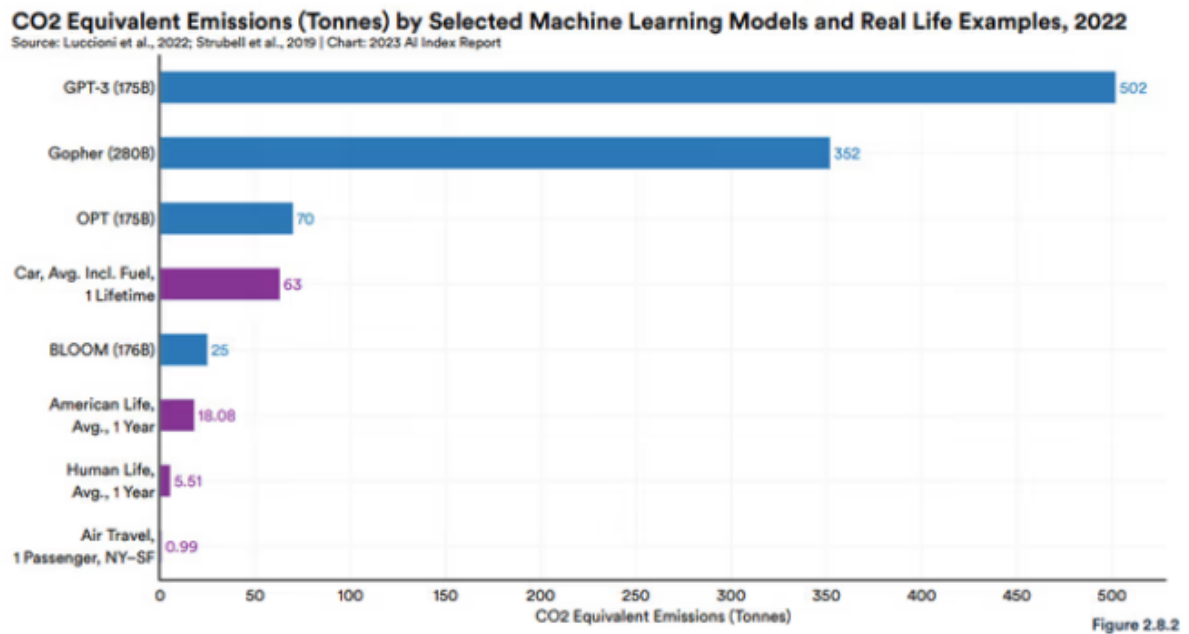


図23: LLM の環境負荷

4.2.3 人工知能と労働

AI の発展によって雇用は減るのか？

- ・ ラッドライト運動：産業革命 ⇨ 繊維産業の自動化 ⇨ 熟練工の抗議
 - 自動化によって新しい雇用の創出&効率的な生産による生活水準の向上
 - 1755-1802 年の労働者の実質賃金は半減、生活環境の悪化

製造の自動化とは異なり、LLM（やその他の生成 AI）によって代替される職業は？(Eloundou et al., 2023)

5 機械学習の実践

機械学習を初めとするデータ分析は様々な教材・資料がオンラインに無料で公開されている。

人工知能（の基盤にある機械学習）の多くは **Python** というプログラミング言語で実行

- ・ 制約付きではあるが [Google Colaboratory](#) でオンラインで実行できる。
- ・ Python 自体も無料なので、自分の PC にインストールして実行できる。
 - 環境構築はちょっとめんどくさい
 - jupyter notebook や visual studio code などが人気のある **統合開発環境** (integrated development environment: IDE)

パッケージをインストール・読み込む ⇨ 様々な分析

- ・ [pandas](#)：データの読み込み、処理
- ・ [matplotlib](#), [seaborn](#)：グラフの作成
- ・ [scikit-learn](#)：（深層学習を除く）機械学習
- ・ [statsmodels](#)：統計分析

深層学習のライブラリとして [TensorFlow](#)、[PyTorch](#)、[keras](#) など

- ・ [TensorFlow Core](#) のチュートリアルや [TensorFlow Hub](#) のチュートリアルでは [Google Colaboratory](#) で試すことができる。
- ・ まずは学習済みモデルを利用

[Kaggle](#) や [Signate](#)、[Nishika](#) などのデータ分析コンペ

- ・ 学生向けのイベントもあり
- ・ 就職活動でアピールできる（かも）

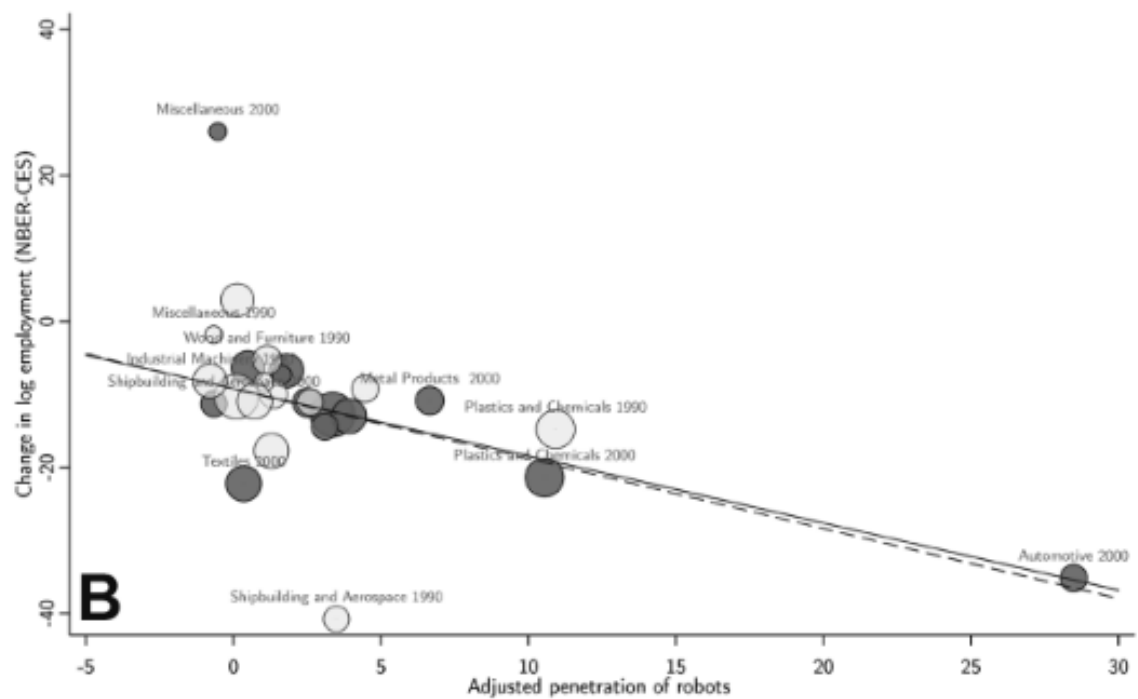
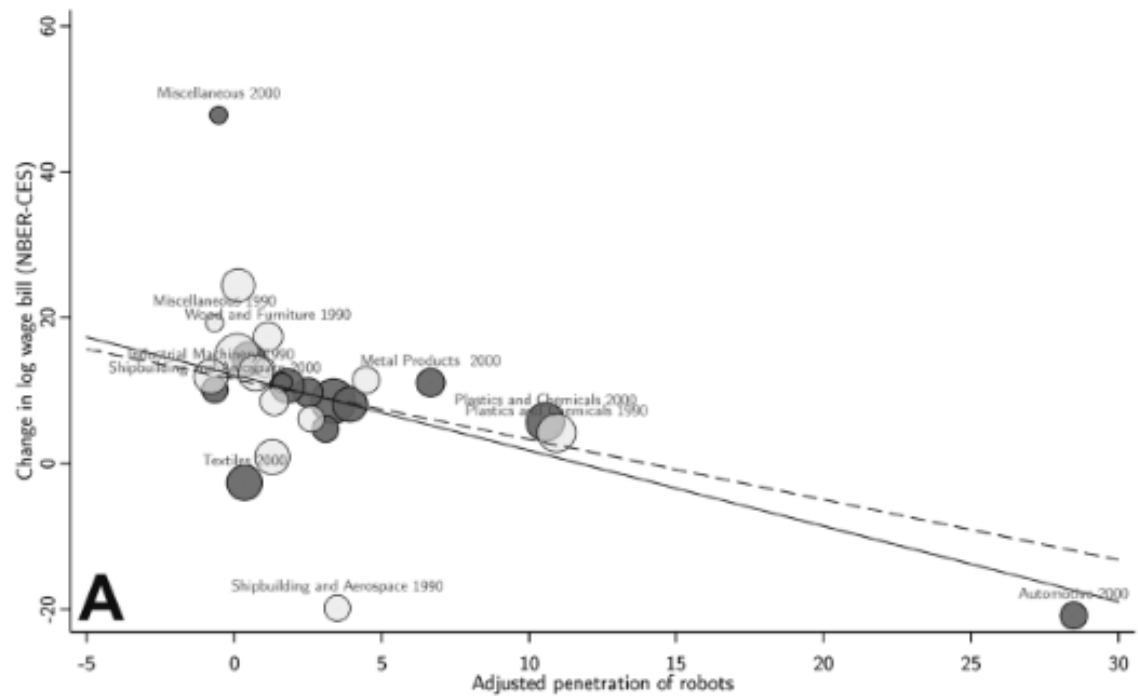


图24: Acemoglu and Restrepo (2020)

参考文献

- Acemoglu, Daron and Pascual Restrepo (2020) “Robots and jobs: Evidence from US labor markets,” *Journal of political economy*, Vol. 128, No. 6, pp. 2188–2244.
- Awad, Edmond, Sohan Dsouza, Richard Kim, Jonathan Schulz, Joseph Henrich, Azim Shariff, Jean-François Bonnefon, and Iyad Rahwan (2018) “The moral machine experiment,” *Nature*, Vol. 563, No. 7729, pp. 59–64.
- Eloundou, Tyna, Sam Manning, Pamela Mishkin, and Daniel Rock (2023) “Gpts are gpts: An early look at the labor market impact potential of large language models,” *arXiv preprint arXiv:2303.10130*.
- Kaplan, Jared, Sam McCandlish, Tom Henighan et al. (2020) “Scaling laws for neural language models,” *arXiv preprint arXiv:2001.08361*.
- Silver, David, Aja Huang, Chris J Maddison et al. (2016) “Mastering the game of Go with deep neural networks and tree search,” *nature*, Vol. 529, No. 7587, pp. 484–489.
- Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin (2017) “Attention is all you need,” *Advances in neural information processing systems*, Vol. 30.