



人工知能AIに監督される人と社会

西垣 敏

USA福岡支部講演会 2023年5月

人工知能AIに監督される人と社会

われわれはいま、ビッグデータ、AI、ロボットなどの技術開発とその社会実装の急進展を通じて、身体と機械、脳とAI、リアル世界とサイバー空間、の境界が曖昧となる時代を生きている。対話型AI活用是非の議論の沸騰は、社会のAI化により人間の存在様式までもが激変させられるかも知れないことへの、危機感の表れとも見られる。確かにAIは、多量データの分析から、病気を予測し或は災害を予測して、人を救うのに役立つかも知れない。しかし他方で我々は、自分たちの身体と精神と日常生活全般を、そのAIに「最適化」、管理・監督されるような社会を招きつつあるのではなかろうか。学校教育へのAI活用、チャットGPT、AIベースの「安全」な街、などの例を見ながら、急速にAI化して行く社会で如何に生きるべきか、議論する。

1. 「予測マシン」...AI兵器 物理学 考古学 芸術家の複製 生成型AI
2. AIのアルゴリズム...ニューラルネットワーク 深層学習 強化学習
なぜ「予測」はうまくいくのか 説明責任
3. AIの実装...教育のAI化、個別「最適化」 AI警察
4. ビッグデータ...インターネット データの流れ AI・ビックデータの罠
5. AI化社会...AIの倫理 サイバー空間と実世界 ,(生成AIに関する追加)

ニューラルネットワーク・ブーム

＜第一のブーム＞ ～1950年代後半から1960年代～

「パーセプトロン」モデル
(ネットワークの学習の仕組み)

＜実用の問題には＞

膨大な数のニューロン必要:
当時のコンピュータ技術の限界

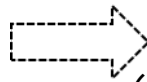
(沈静化)



＜第二のブーム＞

～1980年代に開花～

ネットワーク内部のニューロンの
学習方法: 誤差伝播法
学習に時間がかかる
(大規模問題) 最適状態に達しない可能性



(沈静化)



＜第三のブーム＞

ディープラーニング (Hintonら、2006年)
多層ニューラルネットを学習させる仕組み、
入力の特徴を自分で見出す能力、etc.

(現在の
ブーム)

人工知能ブーム

＜第一のブーム＞

迷路探索、数学の定理の証明
ゲーム(チェス、チェッカー)

＜現実社会の問題＞の
解決することは困難

＜第二のブーム＞

「エキスパートシステム」(知識データ、
推論エンジン) 例: 病気の症状→薬
品処方、会社の業務(大成功)

膨大な知識データ入力の困難
あいまい知識の取り扱い困難

＜第三のブーム＞

ディープラーニング

(Hintonら、2006年)

ビッグデータ

インターネット

コンピュータ高速化

デジタル化

$$I = -p \log_2 p$$
$$C = W \log_2 (1 + S/N)$$

G. FARMERO編 齊藤隆央訳

『美しくなければならぬ 現代科学の偉大な方程式』 2003 紀伊國屋書店
第5章 「情報をビットで刻む シャノンの方程式」 イゴール・アレクサンダー著

情報は、いまや金属や石油のような物資であり、水道や電気のようなユーティリティとなっている。……

われわれ人間が、今よりずっと多くのことを伝えあいたいと思っているかどうかは疑問なので、情報化時代に主眼が置かれているのは情報そのものではない。この革命的時代を特徴づけているのは、むしろ、ほとんどどこからでも他人やコンピュータにアクセスできるという驚くべきチャンスの多さなのだ。……

……

コストが最低となるシステムでは、どのチャンネルも一ビットだけを運ぶ二進数のチャンネルとなり、このとき最も経済的に情報を伝えられるのである。

このような二進数による符号化の経済効率こそ、世界にデジタル化をもたらしている。またしても、シャノンの $\log_2$ が大きな役割を果たしているのだ。

〔ビックデータ〕「集合知定理」

(参考)西垣 通 『ビックデータと人工知能』

集団誤差 = 平均個人誤差 - 分散値

「集団誤差」: 集団的推測値(個々の推測値の平均値)と真値の差の二乗

「平均個人誤差」: 個々の推測値と真値の差の二乗

「分散値」: 個々の推測値の統計的分散

ポイントは一つである。分散値つまり「集団内の個々の推測値のバラツキ」が大きいほど、集団的推測の精度はあがるということだ。集団内の人々が多様な推測をするほど、個々の推測の誤差は互いに打ち消しあい、その平均値である集団的推測値は正解に近づくのである。……集合知というのは、主観知から客観知を導く方法といえるだろう。……

***** (参考) *****

$$(\bar{x}-a)^2 = \frac{1}{n} \sum_{i=0}^n (x_i - a)^2 - \frac{1}{n} \sum_{i=0}^n (x_i - \bar{x})^2$$

(数式は引用者による)

「予測」という名の欲望 データ×人間

何が予測できるのか 買い物も警察もローンも

朝日新聞GLOBE 2019.10.6

AIとは、「予測マシン」である。AI研究が盛んなカナダ、トロント大学教授のアジェイ・アグラワルらは著書『予測マシンの世紀』で、そう語っている。英語に見合う日本語を予測するのが、翻訳AI。顔の画像から特定の人物を予測するのが、画像認識AI。「予測する」という意味では同じ働きだからだ。

そんな予測マシンの中心的技术が「機械学習」。…………それをいち早くビジネスに採り入れたのが、広告分野だ。ユーザーの属性、ネットの使い方の履歴などから購買行動を予測。情報を集めやすく、たとえ予測が外れてもリスクが小さいことも導入を後押しした。「予測分析」ビジネスは急拡大していて、米国のマーケティング会社の調査では2017年に46億ドル(約5000億円)だった市場規模が、22年に124億ドル(約1兆3400億円)に拡大するとの見通しもある。

予測マシンは、暮らしの隅々に広がっている。来店予測、電気やガスの需要予測、インフラの故障予測。返済率を予測して金利をはじき出す「ローンAI」もあれば、病气予測の試みもある。大手クレジットカードのVISAは、AIを活用した予測解析によって世界で年間約250億ドル(約2兆7000億円)の詐欺の防止につなげている。犯罪データから発生場所や時間などを予測する「予測警察」も米国などから世界各国に広がり、日本でも導入の検討が進む。災害予測など、社会課題の解決にAIを使う「AI for Social Good」(AI4SG)の分野も注目されてきた。

とはいえ、AIは勝手に動くわけではない。統計やコンピューター…………

「予測マシン」

翻訳AI: 英語→日本語を予測、考古学資料の解読、**記事原稿の執筆**

AI小説家 AIアナウンサー チャットGPT 論文執筆

画像認識AI: 顔の画像→特定の人物を予測、精神状態 郵便物の仕分け

広告分野のAI: ユーザーの属性、ネット履歴→購買行動を予測、来店予測

電気やガスの需要予測、インフラの故障予測、災害予測など社会課題の解決に

AIを使う「AI for Social Good」(AI4SG)の分野

人事面接AI e信用スコア

「予測警察」: 犯罪データから発生場所や時間などを予測、デモの監視

スマート自治体 **AI「弁護士」**

AI運転: 危険の予測、回避

AI兵器 自爆ドローン

介護AI AI「教師」 AI「医師」

人間拡張: 人間行動の予測、援助、拡張 e-ヘルメット 脳インプラント

農業支援: ぶどうの摘果房選びの例

物理学: マテリアルズインフォマティクス 薬学: AI創薬

芸術: **芸術家の複製、AI作曲、AI歌手**

(参考)

- ・ アジェイ・アグラワルら 『予測マシンの世紀—AIが駆動する新たな経済』
- ・ 「予測」という名の欲望 データ×人間」 朝日新聞GLOBE 2019.10.6

NHKニュース2019年11月13日(水)

アウシュビッツ強制収容所 遺品復元プロジェクト “虐殺の証拠を未来に”



.....

ポーランド南部にある「アウシュビッツ強制収容所」。

現在は、政府が運営する博物館になっていて、時間が経って劣化した犠牲者の遺品やナチスが作成した文書などを最新の科学技術を使って“復元”するプロジェクトが進んでいます。ことし、収容所の配管工事を行っていた際に、「本」が偶然発見されました。

復元作業では、人工知能＝AIを使う計画です。

AIで残された文字を組み合わせ、考えられる文章のパターンを導き出し、それをつなぐことで本の内容を探るといいます。

プロジェクトの目的は、偏見や憎悪が行き着く先に何があるのか、モノに語らせることです。

(しんぶん赤旗2021年9月18日)

古い技術 AIで生まれ変わる

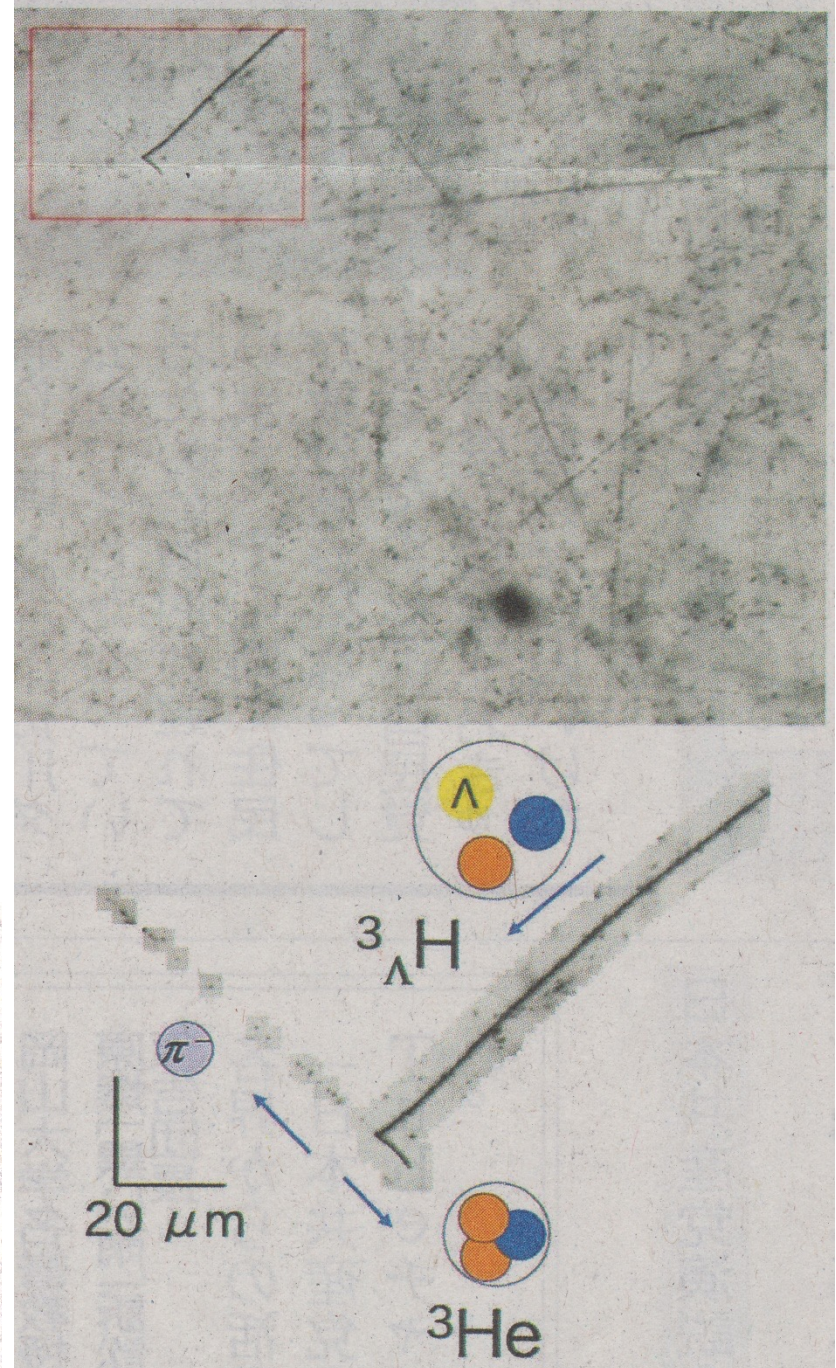
理研・岐阜大など

写真乾板と組み合わせ
「ハイパー核」精密測定

・・・機械学習・・・今回は教師データが存在しないため、シミュレーションと画風変換技術を駆使して教師データとなる模擬画像を作成。・・・従来手法なら560年以上かかる画像解析を3年に短縮可能・・・

(原文ママ)

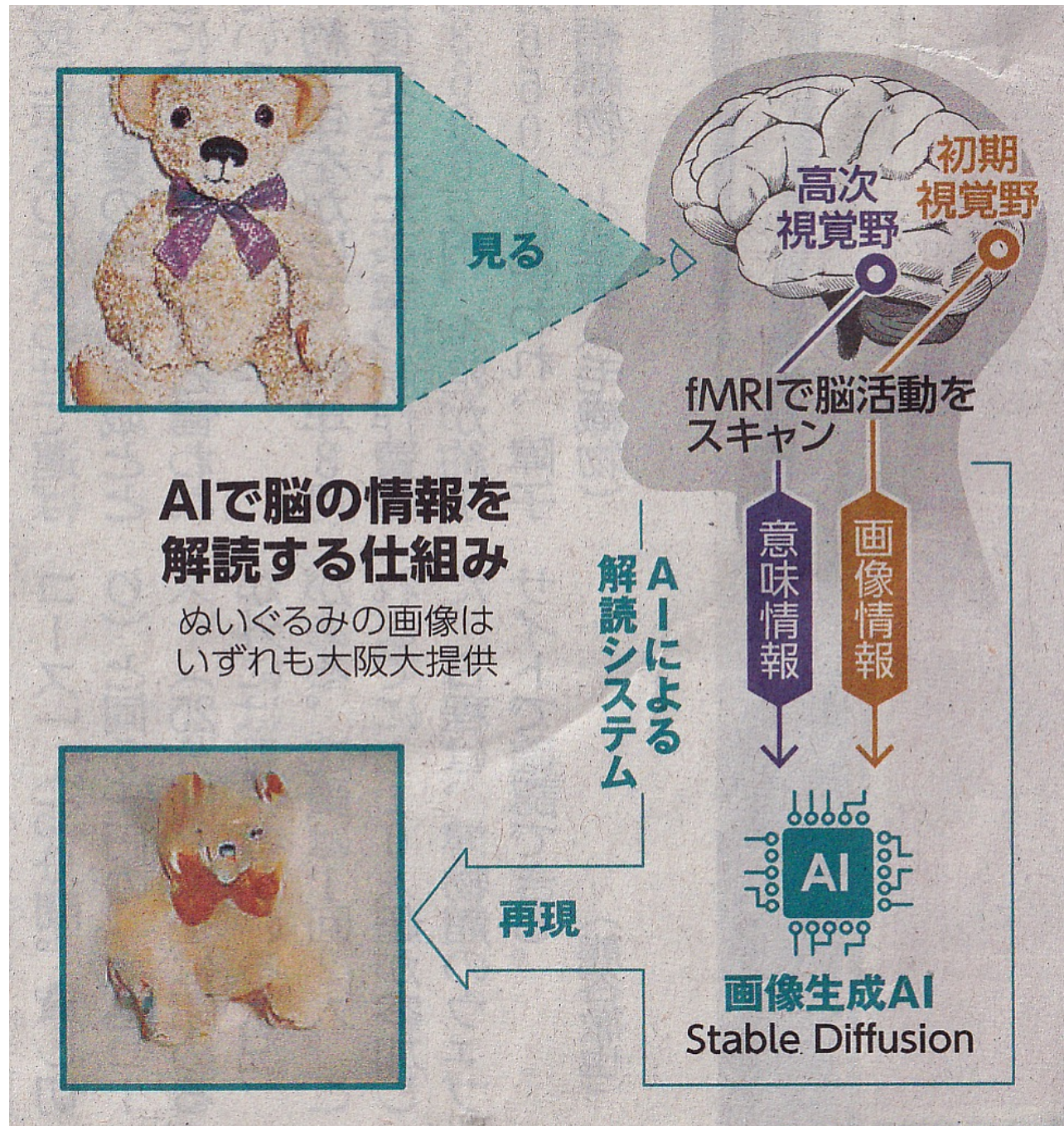
写真乾板から検出されたハイパートライトンの崩壊事象（赤枠内）と拡大図（下）。右上から飛行してきたハイパートライトンが、ヘリウム3原子核とパイマイナス中間子に崩壊し、それぞれ右下と左上方向に飛行して静止しました。飛跡からハイパートライトンの性質を調べます（理化学研究所提供）



見たままの情報 AIが再現

脳に入った
視覚情報を「翻訳」
画像作り出す

(朝日新聞2023年4月7日)



芸術家の複製可能性の時代の芸術家



(参考) ヴァルター・ベンヤミン『複製技術時代の芸術作品』
Das Kunstwerk im Zeitalter seiner technischen Reproduzierbarkeit
(1935-36年成立)

2019.9.29. NHKスペシャル

「AIでよみがえる美空ひばり」



Portrait of Edmond Belamy, 2018, created by GAN (Generative Adversarial Network). Sold for \$432,500 on 25 October at Christie's in New York. Image © Obvious

GANアルゴリズムで生成

- (1) 肖像画1万5,000点の画像データセットを使って学習させる。生成器(Generator)が新たな画像を出力する。
 - (2) この画像を識別器(Discriminator)が人間の描いた絵と比較し、誤差を求めて生成器へ伝える。
 - (3) 生成器は誤差データを使ってより本物らしい絵の画像を出力する。
 - (4) 生成器と識別器は(2)と(3)の処理を、出力画像と人間の描いた絵を見分けられなくなるまで繰り返す。
- (参照 <https://boxil.jp/beyond/a5894/#5894-5>)

未来への分岐点 「AI戦争 果てなき恐怖」

<NHKスペシャル>

2021.7.12.

2021.10.31.

AIと戦争 新たな脅威

自動運転や物流、医療など、私たちの暮らしを大きく変え始めた「人工知能=AI」。このAIが戦争に利用されることで、19世紀の「ダイナマイト」、20世紀の「核兵器」を凌駕する軍事革命が起きると言われています。

10年後の、2030年。世界の大国は、相次いで、軍の中枢にAIを配備する計画を打ち出しています。

AIは、戦争に何をもたらすのか。それは、人間の判断を介さず攻撃する「AI兵器」の果てなき恐怖。

さらに、AIは従来の戦争の概念をも根底から覆すと言われています。AIと軍事の危険な合体は、未来の戦争をどのように変えるのか。この先10年で、危機を回避する手立てはあるのか。新たな戦争の脅威は、私たちのすぐそばに忍び寄っています。



(参考) ノーバート＝ウィーナー『サイバネティックス』(第2版)序章(1947年)
(池原・彌永・室賀・戸田共訳、1962年 岩波書店)

……第二次大戦の初期におけるドイツ空軍の優勢と、イギリスの守勢とから、多くの科学者が高射砲の性能向上をはかろうとしていた。戦争前でさえも、航空機の高速化が、従来の対空火器照準法を全く時代おくれなものにしてしまい、火器の制御装置の内部に必要な計算機構一切を組みこむ必要のあることがはっきりわかっていった。……したがって、標的に狙いを定めて砲弾を発射するというのではなく、砲弾と標的とが、ある時間後に空中のどこかでぶつかるように発射する必要がある。それで飛行機の、**未来の位置を予測する**方法を考えださなければならないこととなった。

.....

そんなわけで私自身、**戦時研究に従事することになり**、・・・と私とは、**予測理論とその理論を実地に応用する装置製作**の協同研究にあたったのである。

こうして私は、第一に複雑な計算の遂行、第二に**未来予測という、人間特有の頭脳活動のお株を奪ってしまうための電気機械系の研究**に、ふたたび従事することになった。このうちの第二の点については、人間の或る種の機能の動作を論じないわけにはいかない。なるほど、或る種の火器制御装置では、レーダーによって衝撃波そのものが直接、照準器に入ってくるが、もっとふつうのものでは、人間の照準手が、火器制御装置といっしょに、その一部分であるかのように動作する。したがって照準手のはたらきを機械に含めて数学的に扱うためには、照準手の特性を知ることが必要になる。さらに、標的である飛行機も、人間によって操縦されており、その動作特性を知らなければならない。

対話AI利用 大学模索

朝日新聞 2023年4月21日(1・2面)

...対話型AI(人工知能)「ChatGPT(チャットGPT)」など生成AIについて、大学が対応を迫られている。AIが書いた文章と学生が執筆した文章とを見分けるのが難しいことから、とりわけリポート作成で制限を設けたり注意を呼びかけたりする例が多い。ただ、情報の収集や整理といった場面で活用が期待されており、多くは使用を全面的に禁止とはせず、バランスに悩みながら模索している。(以下抜粋)

- ・京都大湊総長....「(情報を集める)サーチのみでリサーチ(研究)という検証を欠いているため、**学術リポートとしては致命的な問題点がある**」
- ・東京大太田副学長....「
- ・名古屋大杉山総長....「**...これは剽窃、盗用の疑いもあり、大学教育の危機と言える**」一方、「必ずしもネガティブ(否定的)ではない」とも発言。
- ・九州大....「使用は単純に禁止しない。新しい技術を活用し、自ら学び、価値を創造する姿勢を育むことが望ましい」とした上で注意点を例示した。
- ・東北大....「完全な排除は現実的でない」「教員自らがどう向き合うべきかを考え...」
- ・パリ政治学院....文章や発表資料の作成に、**...AIの使用を原則禁じた。剽窃にあたるという懸念を特に強調し、違反者については退学処分も辞さない構えだ。**
- ・ケンブリッジ大....「**チャットGPTの生成物を用いることは懲戒対象になる不正行為**」
- ・ニューヨーク州立大....「本学にはチャットGPTに対して賛否の取り決めはない」
- ・「エルゼビア」の論文誌....人間の研究者と並ぶ著者としてチャットGPTが記載され...
- ・ネイチャー....チャットGPTは「著者となる基準を満たしていない」と明記。
- ・サイエンス....画像生成AIも含めて「編集者からの許可がない限り、論文には使用できない」と記載。違反は研究不正にあたると明記している。

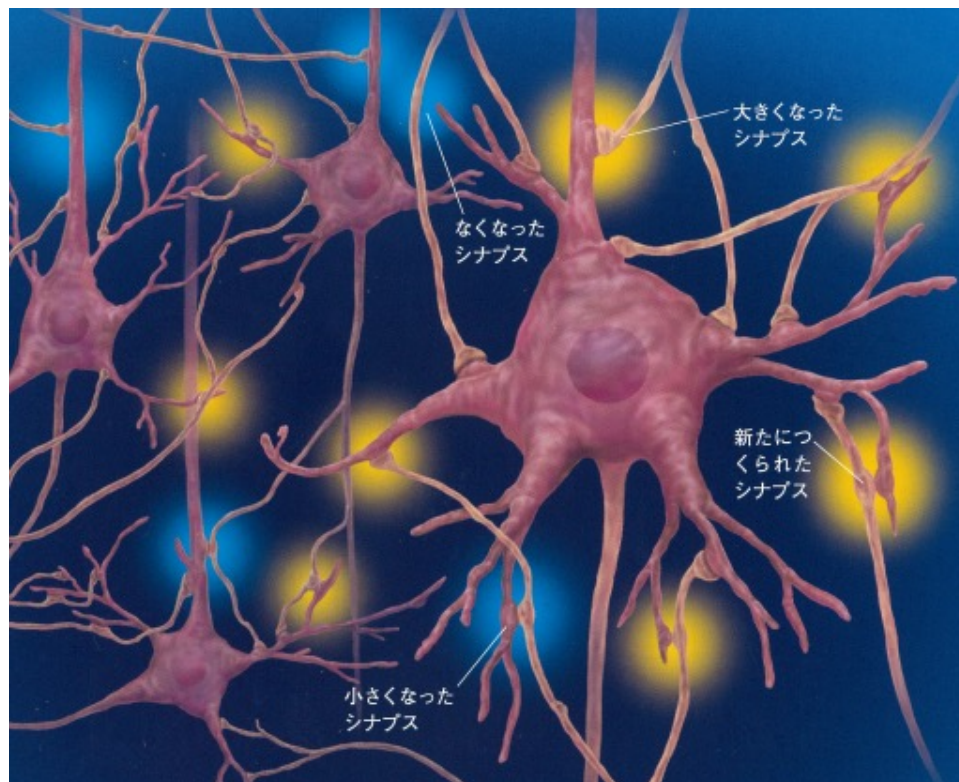
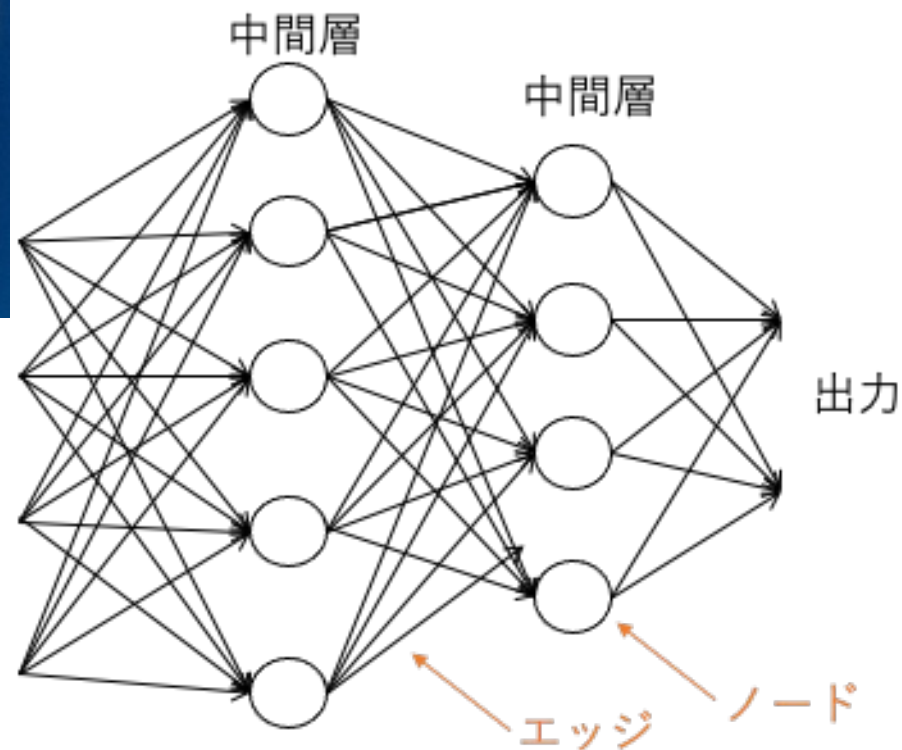


図3.「記憶後」の人間の脳の中:「記憶後」
(ニュートン別冊「脳とは何か」よりの引用の引用)

入力

ニューラルネットワークの一例



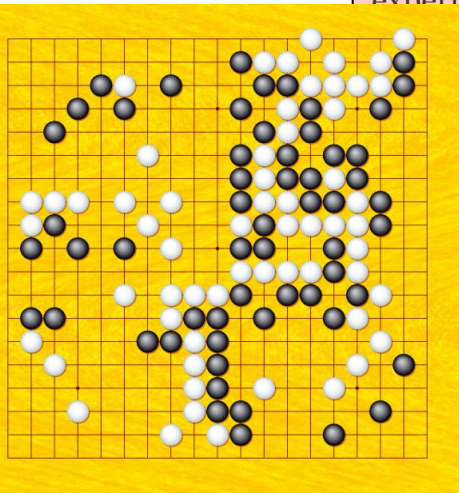
The New York Times **OPINION**
GUEST ESSAY

**Noam Chomsky: The False Promise
of ChatGPT**

March 8, 2023

By Noam Chomsky, Ian Roberts
and Jeffrey Watumull

Jorge Luis Borges once wrote that to live in a time of great peril and promise is to experience both tragedy and comedy, with the prominence of a revelation in understanding ourselves and the world. Our supposedly revolutionary advancements in artificial intelligence are a cause for both concern and



入力

入力層

隠れ層

出力層

次の一手の
予測確率

出力

翻訳



猫の画像

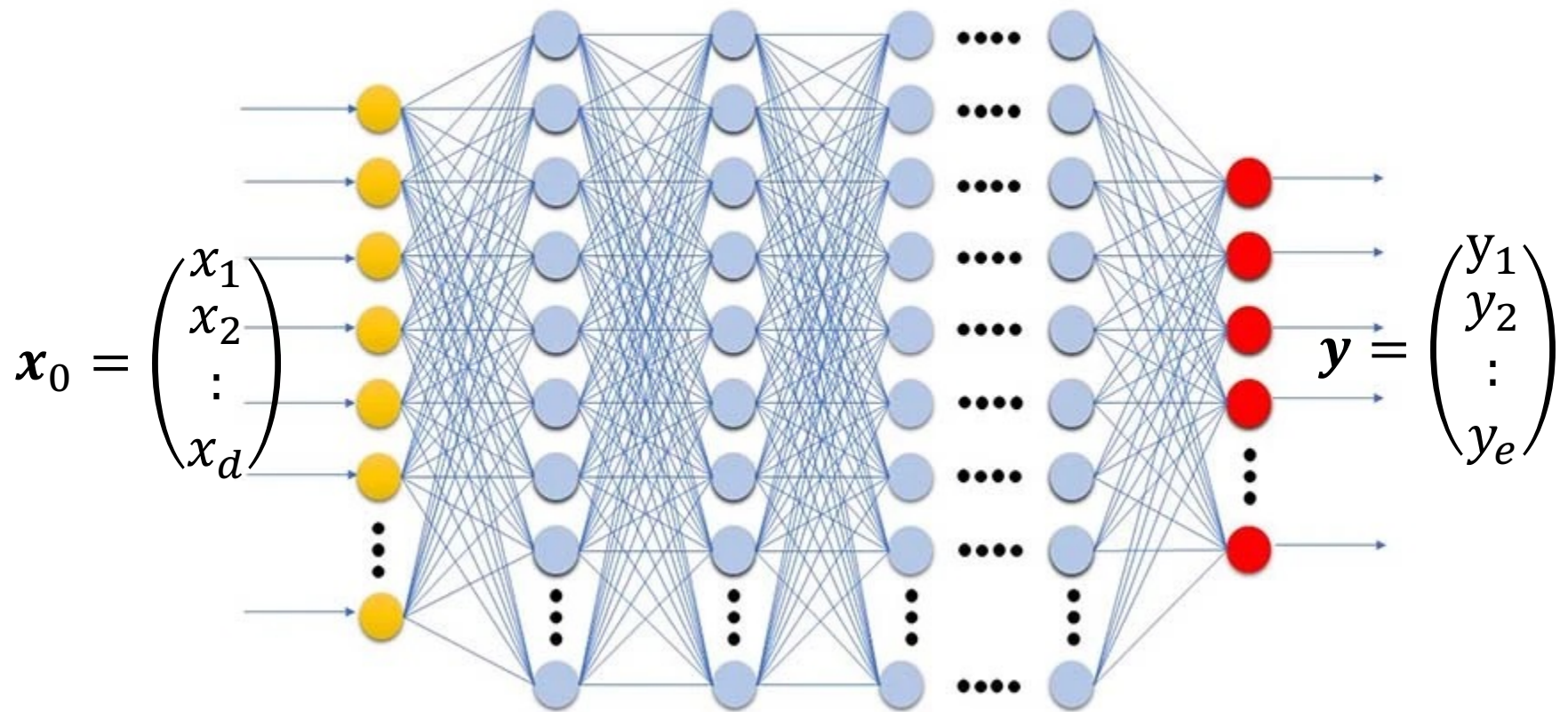
猫の確率 : 96%
犬の確率 : 3%
人の確率 : 1%

<https://ainow.ai/artificial-intelligence-3/> の図を部分利用

入力層

中間層(データを変換)

出力層



$$A_1 = \begin{pmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{d1} & \cdots & a_{dm} \end{pmatrix}$$

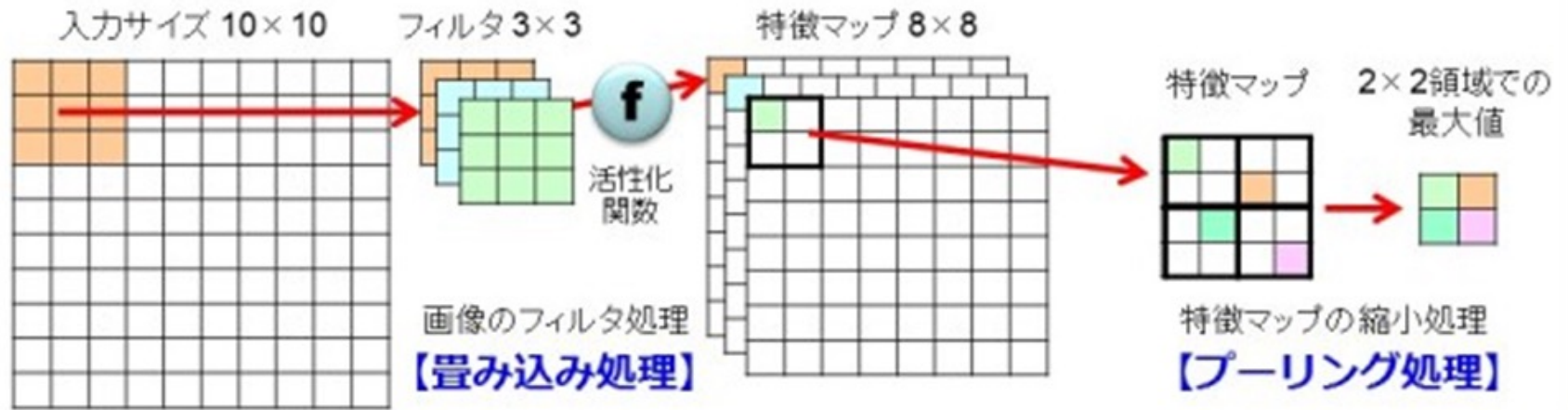
$$\mathbf{b}_1 = \begin{pmatrix} b_{11} \\ \vdots \\ b_{1d} \end{pmatrix}$$

$$\mathbf{x}_l = \eta(A_{l-1}^T \mathbf{x}_{l-1} + \mathbf{b}_{l-1})$$

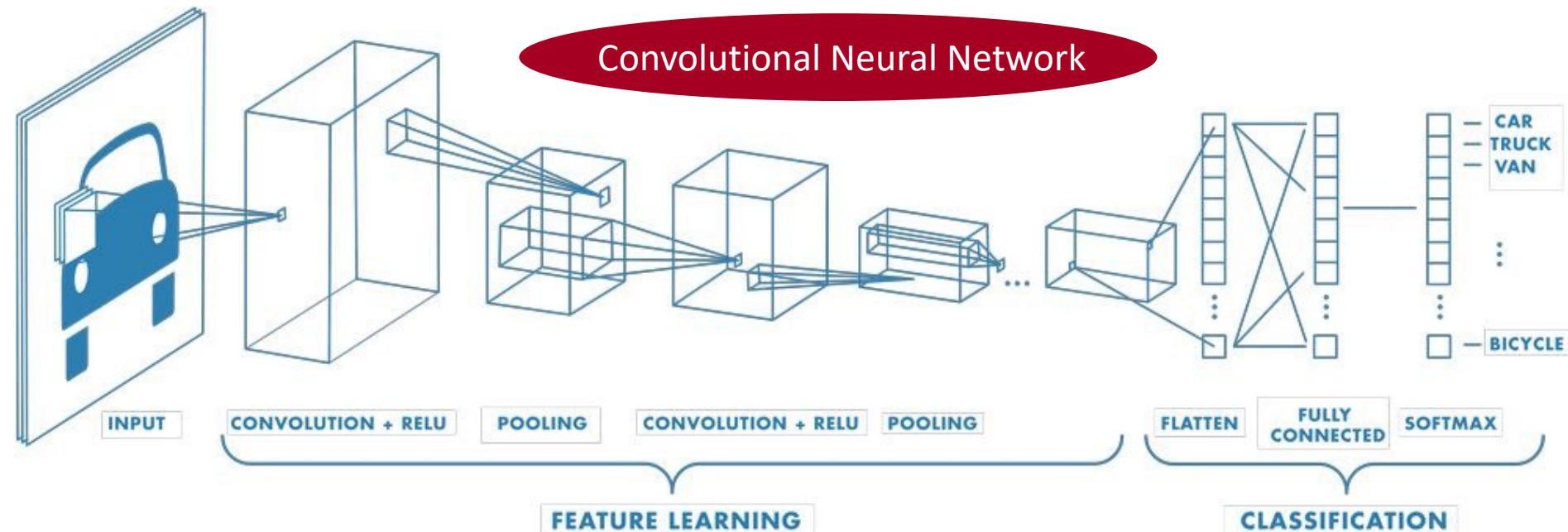
活性化関数



<http://gagbot.net/machine-learning/ml4> の図を転載



Convolutional Neural Network



多数の畳み込み層を持つネットワークの例。それぞれの学習画像に解像度が異なるフィルターが適用され、各畳み込み画像の出力が次の層の入力として使用される。

<https://jp.mathworks.com/discovery/convolutional-neural-network-matlab.html> の図をコピー掲載

例：AlfaGoZeroの アルゴリズム

<https://codezine.jp/article/detail/10952> 掲載図の転載

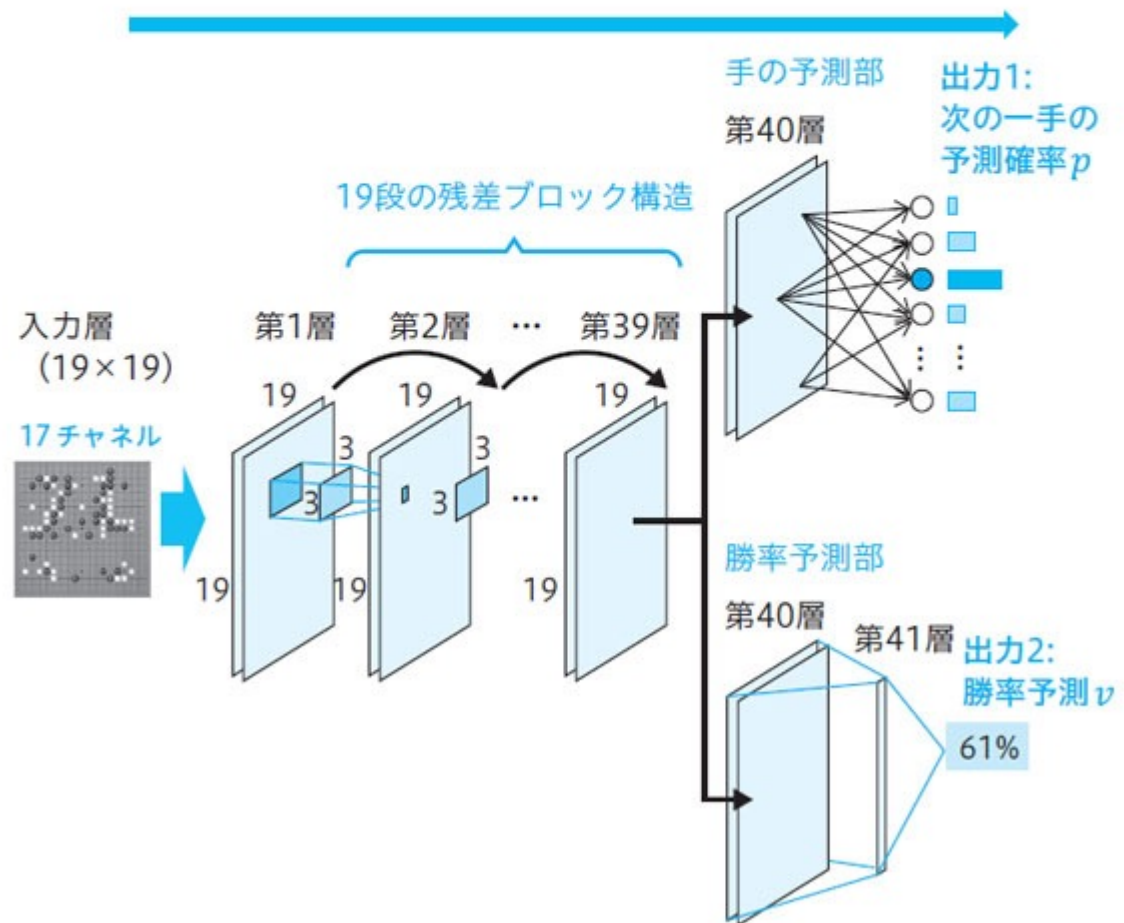
デュアルネットワークの規模 と計算量

(畳み込みの足し算回数)
 $19 \times 19 \times 3 \times 3 \times 256 \times 256$
 $\times (\text{層の数: } 39) = \text{約}83\text{億回}$
(フィルタ重みパラメータの
個数) $3 \times 3 \times 256 \times 256 \times$
 $(\text{層の数: } 39) = \text{約}2300\text{万個}$

デュアルネットワークの教師
付き学習は、ポリシーネット
ワークの教師付き学習と基
本的に同じであり、誤差関
数L0の勾配を利用する誤差
逆伝搬法により、……

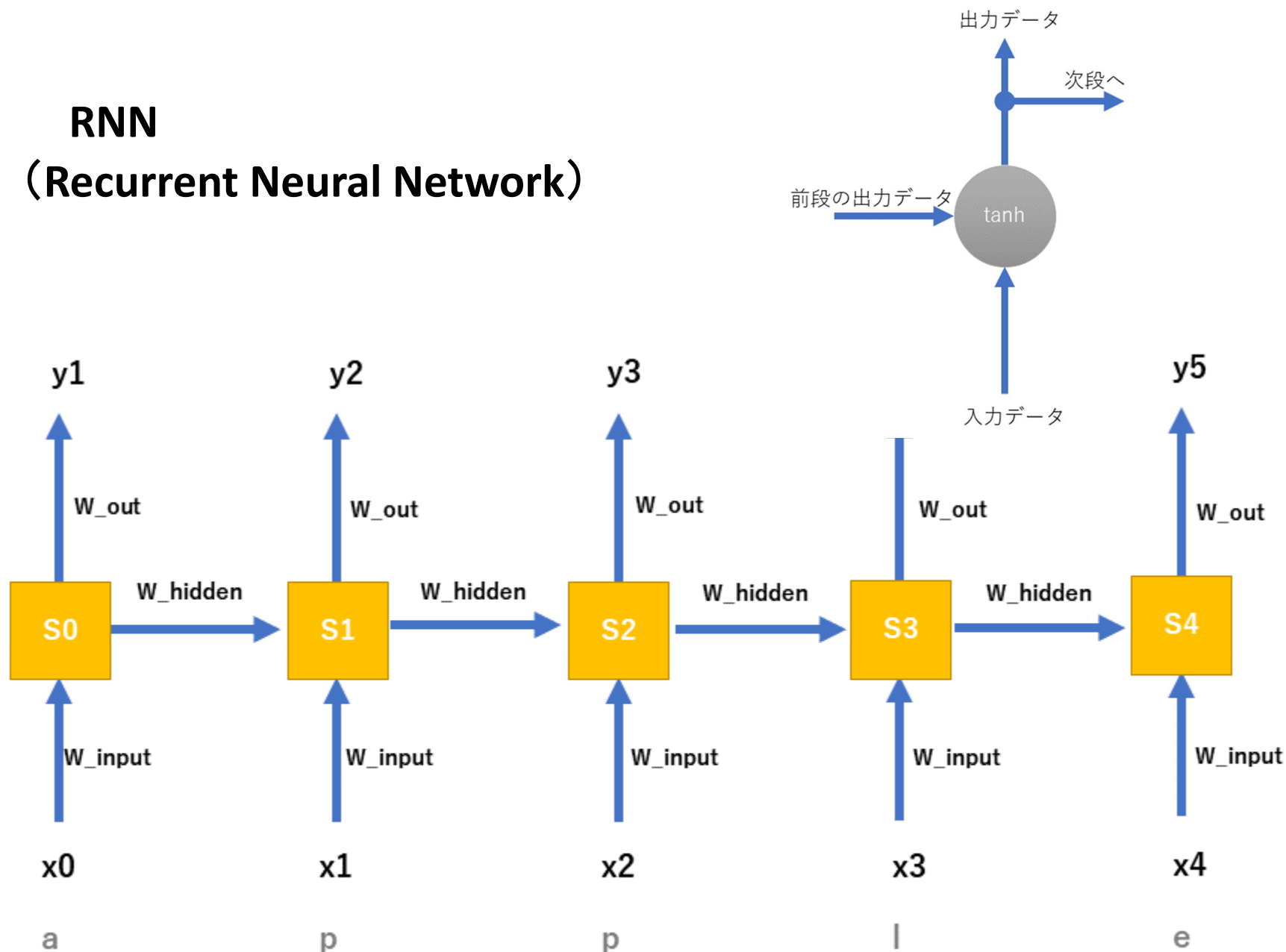
・デュアルネットワークの構造

- 入力は17チャンネル (0~7手前の黒石/白石の位置、手番)
- 全部で40層以上
 - ・各層は、基本的に、 3×3 の畳み込み層+バッチ正規化+ReLUからなる
 - ・第2~39層までは、ショートカットを持つ残差ネットワーク (ResNet)
- 第40層からは、次の一手の予測部と勝率の予測部とに分かれる



RNN

(Recurrent Neural Network)



(<https://blog.octopt.com/rnn/> の図を転載)

自然言語処理

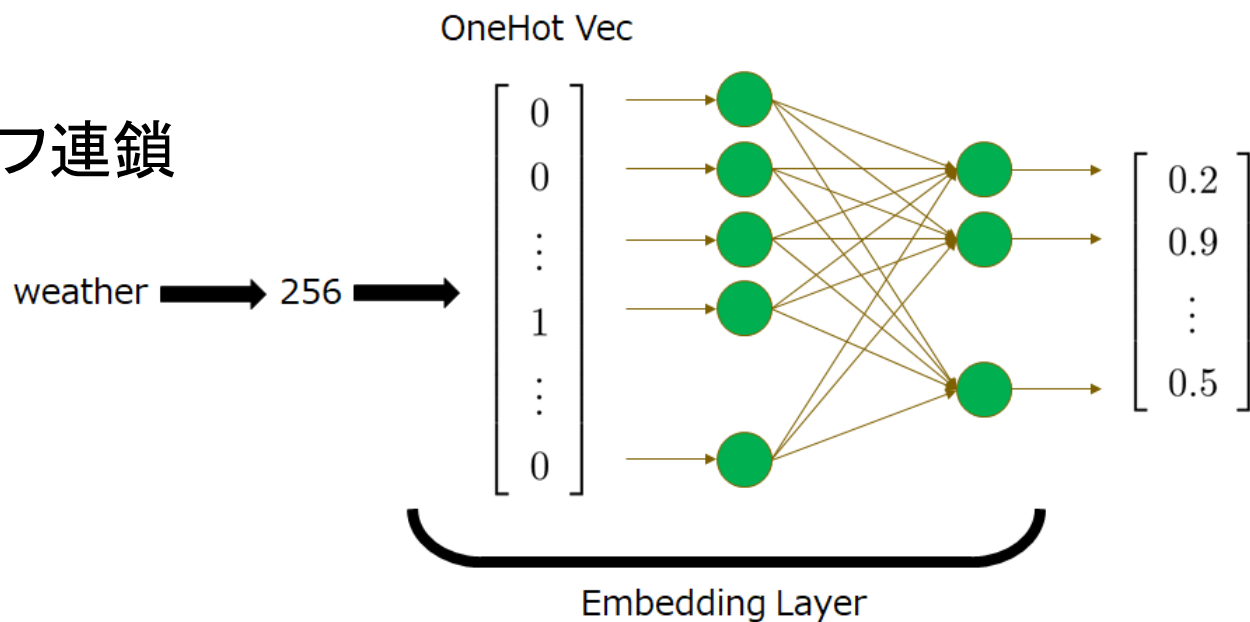
統計的自然言語処理

Word Embedding
単語埋め込み
分散表現

共起 n-gram統計

$$\begin{bmatrix} w_{i1} \\ w_{i2} \\ \vdots \\ w_{im} \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1m} \\ \vdots & \vdots & & \vdots \\ w_{i1} & w_{i2} & \cdots & w_{im} \\ \vdots & \vdots & & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nm} \end{bmatrix}^T \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix}$$

文章生成
マルコフ連鎖



機械学習 (Machine learning)

は、明示的な指示を用いることなく、その代わりにパターンと推論に依存して、特定の課題を効率的に実行するためにコンピュータシステムが使用するアルゴリズムおよび統計モデルの科学研究である。(Wikipedia情報の抜粋)

教師あり学習

入力とそれに対応すべき出力を写像する関数を生成する。例えば、分類問題では入力ベクトルと出力に対応する分類で示される例を与えられ、それらを写像する関数を近似的に求める。

教師なし学習

入力のみ(ラベルなしの例)からモデルを構築する。データマイニングや自己符号化器なども参照。

強化学習

周囲の環境を観測することでどう行動すべきかを学習する。行動によって必ず環境に影響を及ぼし、環境から報酬という形でフィードバックを得ることで学習アルゴリズムのガイドとする。例えばQ学習がある。

教師あり
学習

問題と答えから傾向を学習

- 回帰
- 分類

教師なし
学習

問題の構造を学習

- クラスタリング
- 次元削減

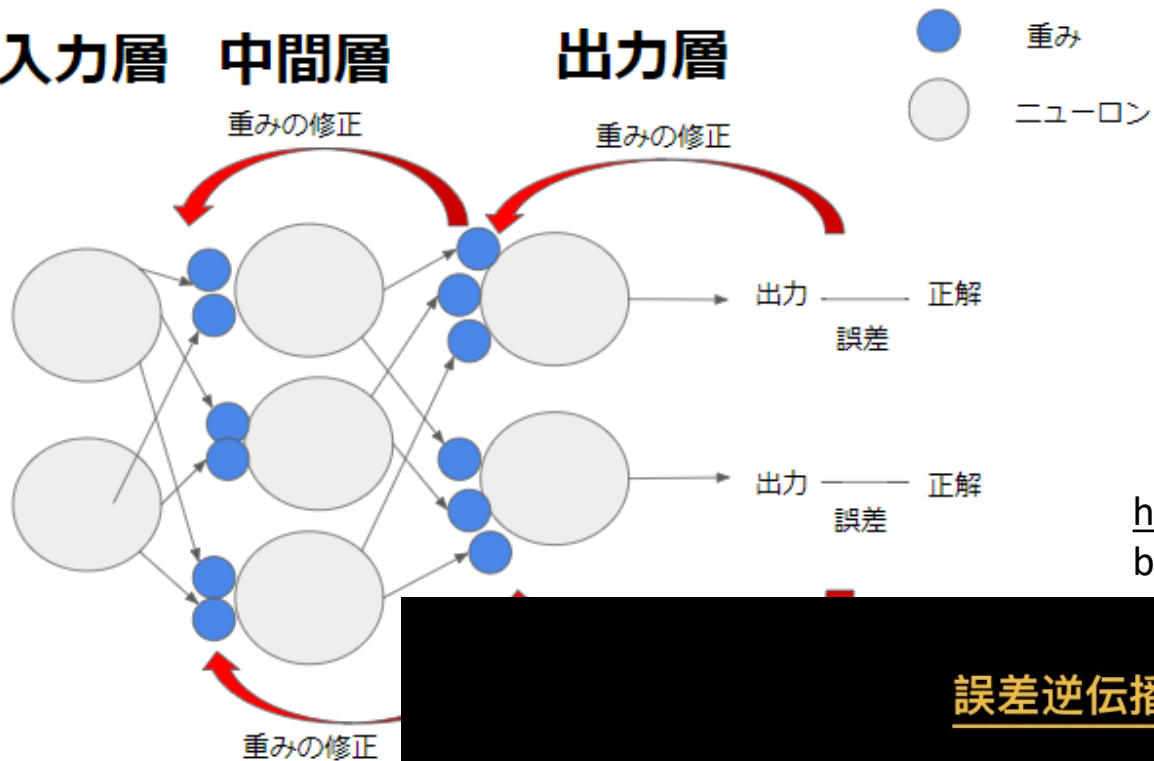
強化学習

試行錯誤をして行動を学習

- バンディッドアルゴリズム
- Q学習

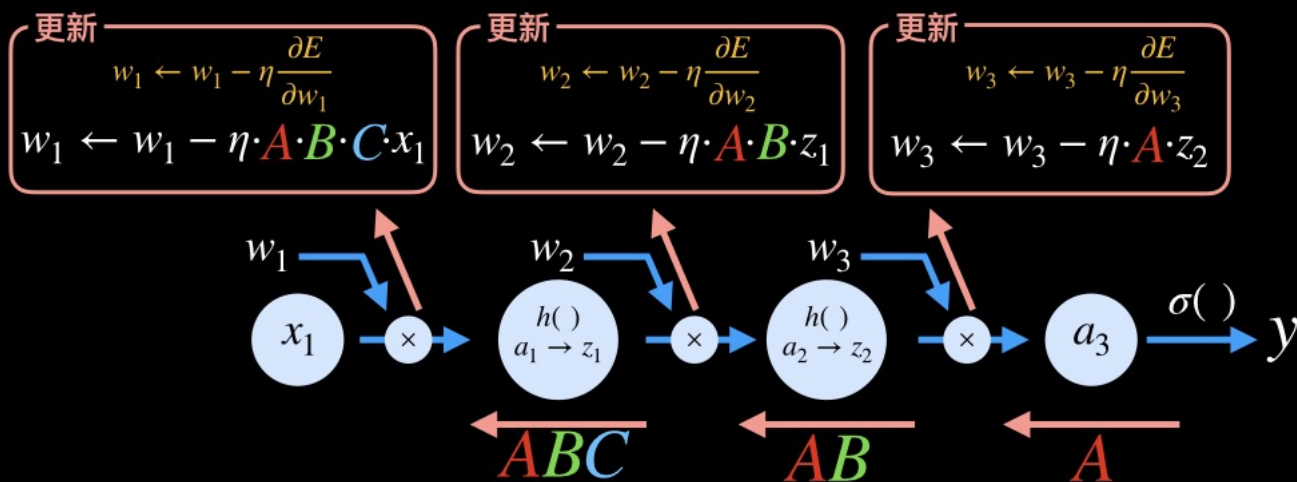
<https://ainow.ai/artificial-intelligence-3/掲載図の転載>

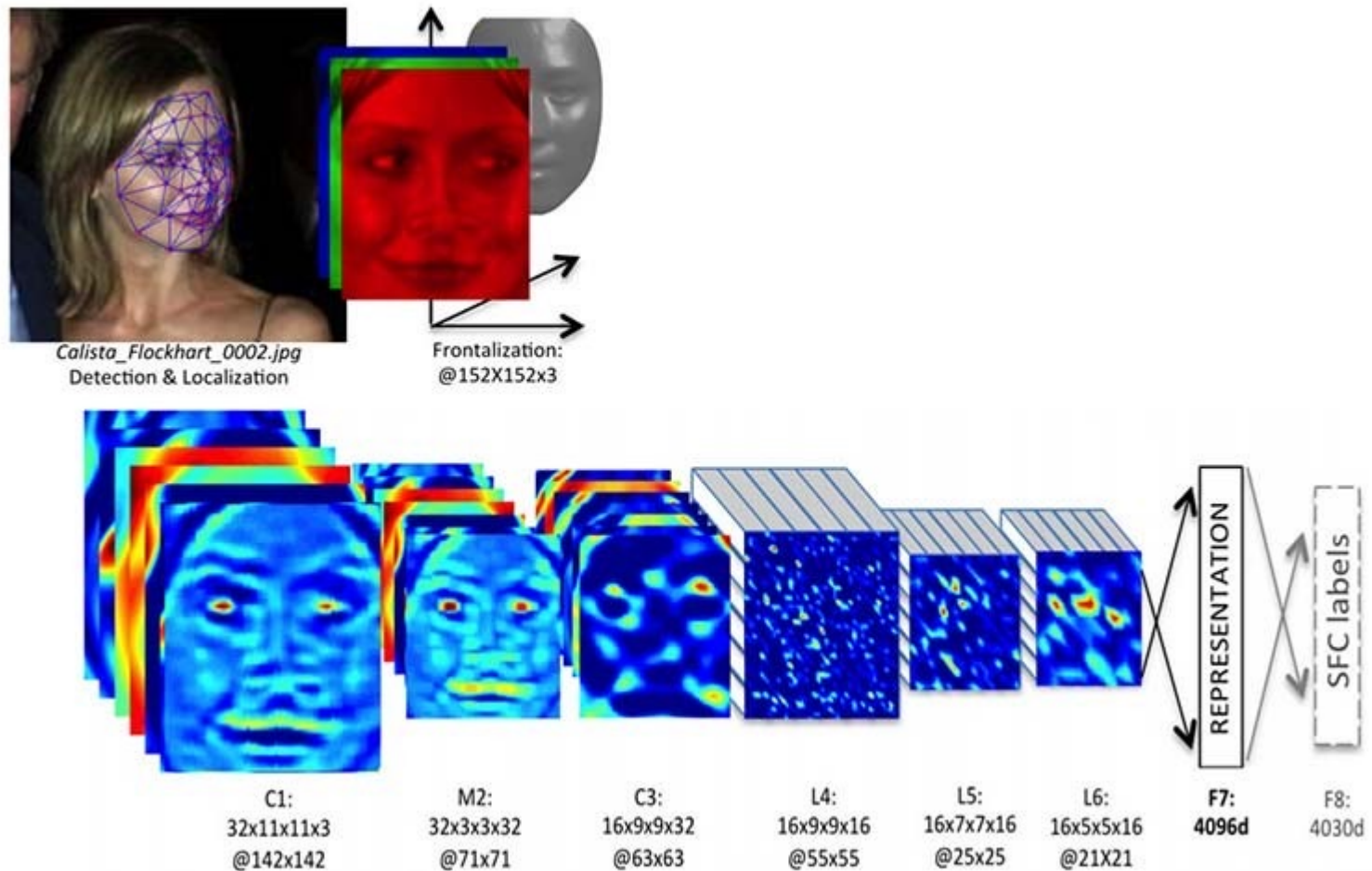
入力層 中間層 出力層



<https://fresopiya.com/2019/05/07/backpropagation/>

誤差逆伝播法

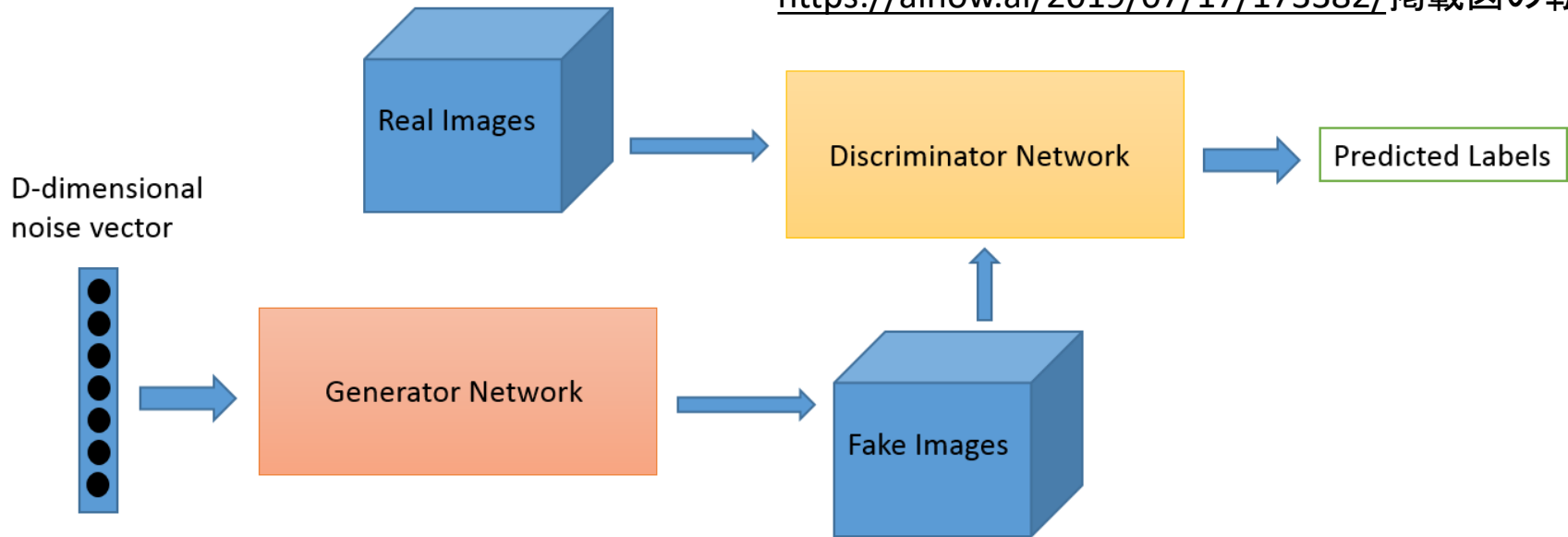




フェイスブックの研究者らが示した、ディープラーニングが人間の顔のデータを変換していく様子。右にいくにつれて抽象化され、人間にはなんのデータだか分からなくなる＝タイグマンらによる2014年の論文”DeepFace: Closing the Gap to Human – Level Performance in Face Verification” (2014 IEEE Conference on Computer Vision and Pattern Recognition, 1702–1708. IEEE) から
(<https://globe.asahi.com/article/12872410>から引用)

敵対的生成ネットワーク GANs

<https://ainow.ai/2019/07/17/173382/>掲載図の転載



敵対的生成ネットワーク (Generative adversarial networks; GANs) は、2014年にイアン・グッドフェローらによって発表された教師なし学習で使用する人工知能アルゴリズムの一種であり、...

GANsは生成ネットワーク(generator)と識別ネットワーク(discriminator)の2つのネットワークから構成される。例として画像生成を目的とするなら生成側がイメージを出力し、識別側がその正否を判定する。生成側は識別側を欺こうと学習し、識別側はより正確に識別しようと学習する。このように2つのネットワークが相反した目的のもとに学習する様が敵対的と呼称される所以である。(Wikipedia情報)

強化学習とは、試行錯誤を通じて環境に適応する学習制御の枠組である。教師付き学習とは異なり、状態入力に対する正しい行動出力を明示的に示す教師が存在しない。かわりに報酬というスカラーの情報を手がかりに学習するが、報酬にはノイズや遅れがある。そのため、行動を実行した直後の報酬をみるだけでは、学習主体はその行動が正しかったかどうかを判断できないという困難を伴う。学習主体「エージェント」と制御対象「環境」は以下のやりとりを行う。

1. エージェントは時刻 t において環境の状態観測 $S(t)$ に応じて意志決定を行い、行動 $a(t)$ を出力
2. エージェントの行動により、環境は $S(t+1)$ へ状態遷移し、その遷移に応じた報酬 $r(t)$ をエージェントへ与える。

3. 時刻 t を $t+1$ に進めてステップ1へ戻る。

エージェントは利得 (return: 最も単純な場合、報酬の総計) の最大化を目的として、状態観測から行動出力へのマッピング (政策 (policy) と呼ばれる) を獲得する。

環境とエージェントには一般に下記の性質が想定される。

- エージェントは予め環境に関する知識を持たない。
- 環境の状態遷移は確率的。
- 報酬の与えられ方は確率的。
- 状態遷移を繰返した後、やっと報酬にたどり着くような、段取りの行動を必要とする環境 (報酬の遅れ)。



Fig.1 強化学習の枠組. エージェントは試行錯誤を通じて適切な制御規則を獲得していく。

深層学習の数学的基礎づけ

(参考) 今泉允聡『深層学習の原理 数学の挑戦』(2021年 岩波科学ライブラリー303)

- ①(普遍近似定理) 層が2つあるニューラルネットワークは、一層あたりのパラメータの数が十分多ければ、どんな連続な関数でも(無視できる誤差の範囲で)表現できる。
- ②(近似誤差レート) 表現したい関数が滑らかであるとき、層が2つあるニューラルネットワークの近似誤差レートは、どんな方法をもってしても改善することはできない。
- ③(滑らかでない関数に対する近似誤差レート) 表現したい関数がジャンプを持つ関数であるとする。このとき、層の数が4層以上あるニューラルネットワークの近似誤差レートは理論上の最適値を達成し、層の数が少ないニューラルネットワークや深層学習以前の従来法はそのレートを達成できない。
- ④(非均一的な滑らかさを持つ関数に対する近似誤差レート)
- ⑤(特徴量とニューラルネットワーク)
- ⑥(過適合)
- ⑦(線形回帰モデルの過適合)

↓

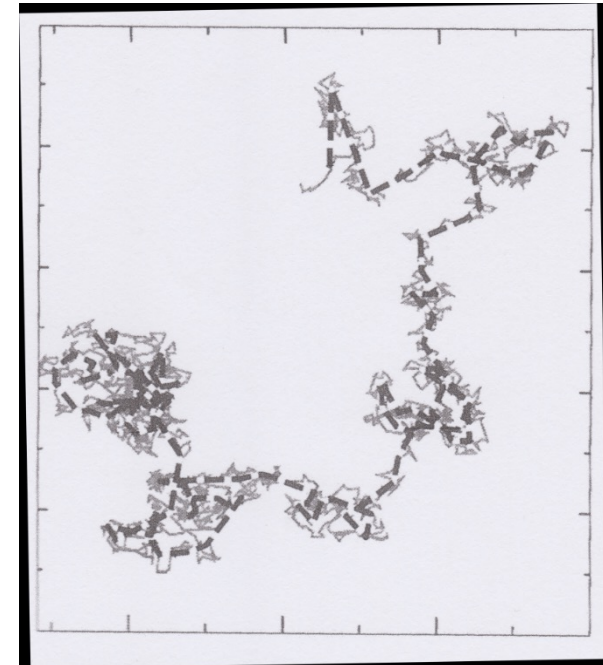
(二重降下理論)

◎なぜパラメータの学習ができる？

(確率的勾配降下法)

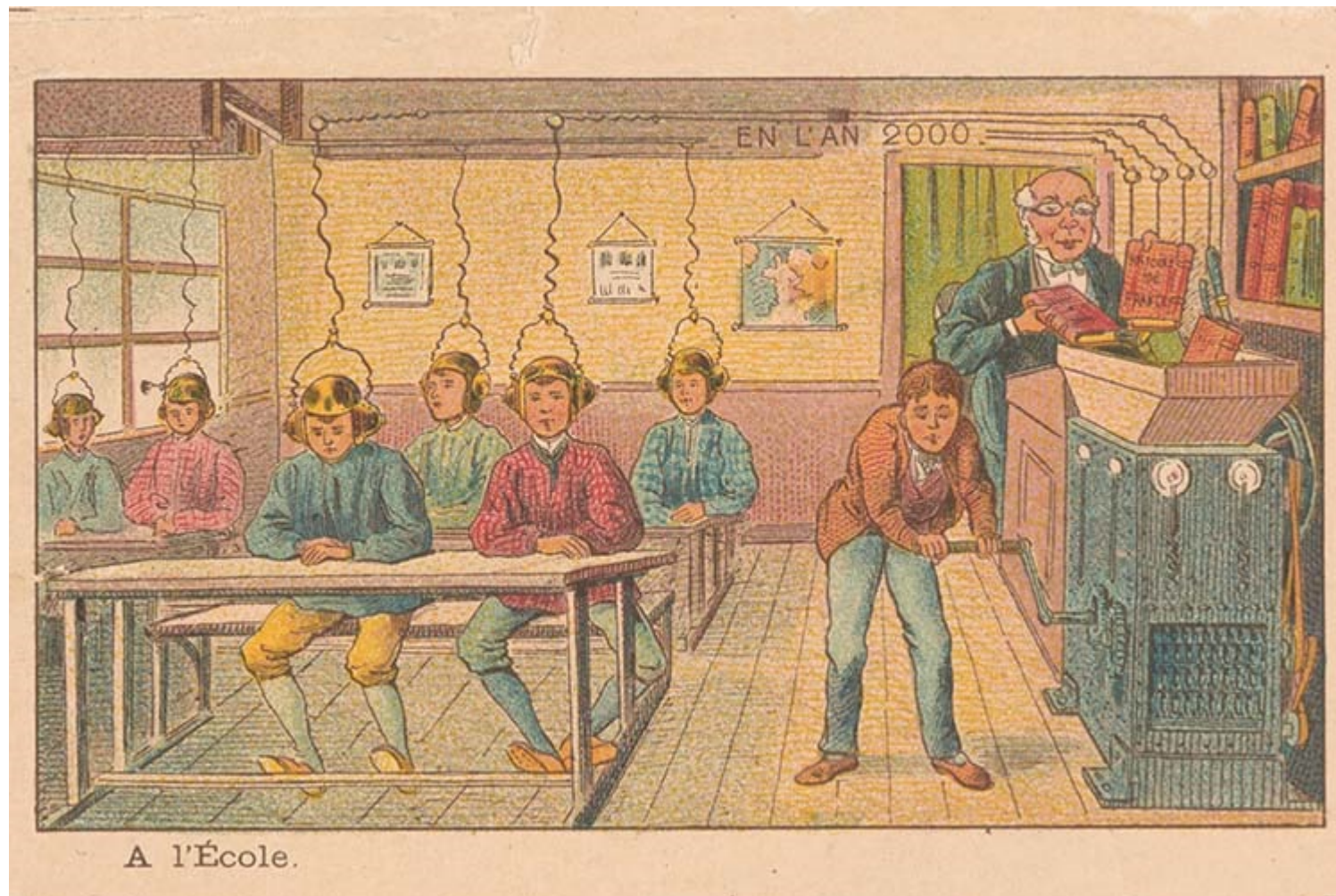
$$d\theta_t = -\nabla R(\theta_t)dt + \sqrt{\beta^{-1}}dB_t$$

(ランジュバン動力学によるパラメータ更新の様子。大まかには黒色破線に沿っているが、ノイズの影響で実際には灰色実線のように動く。)



A l'École / An l'an 2000 (Anticipation)

[illustrations de Jean Marc Côté]/ Lithographie Villemard et fils, 1910



「予測」という名の欲望 データ×人間

朝日新聞GLOBE 2019.10.6 より転載

溶ける公教育_デジタル化の行方 (1)-(4)

しんぶん赤旗2022.5.5-8.

大阪府箕面市箕面小学校にある教室。。天井に5台のカメラ
(3年前から同市がコニカミノルタと提携して取り組む文科省の実証事業。21年は
市内4校で実施)

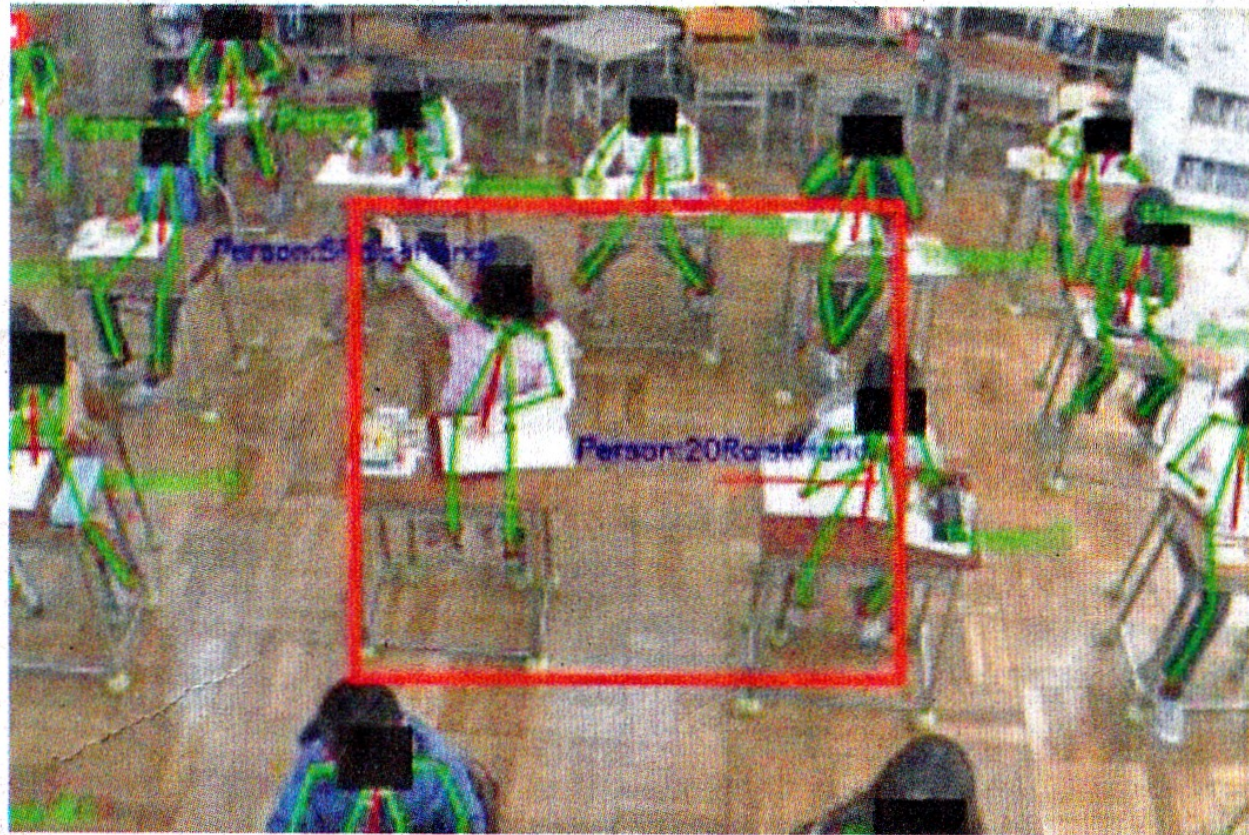
教室見つめる5台の
カメラ

行動解析

教室内に設置されたサーバーで
個人が特定できないよう自動的
に画像処理され、解析のために
コニカミノルタに渡ります。

民間参入

急速に進む教育のデジタル化が、
公教育の現場への急速な民間
企業の参入をもたらしている。



カメラを使った授業中の子どもたちの行動解析のイメージ
(箕面市の資料から)(しんぶん赤旗2022.5.5.)

第2部 溶ける公教育_デジタル化の行方 (1)-(4)

しんぶん赤旗2022.7.19-22.

経産省教育産業室の野望 新しい資本主義と稼ぐ校舎

(「未来の教室」のラフスケッチでは民間教育と公教育の壁は完全に溶けている。)

- ・「教室で自分の好きな学習塾の先生のオンライン講義動画をタブレットで見て、自分の進度に合わせて個別に学ぶのが一般的になる」
- ・「午前中は学校に通い、午後はフリースクールや学習塾に通う」
- ・「必ずしも全員が『教える先生』ではなくなる」
- ・「個々の生徒がどんな教材で学びどのような様子なのかをデータで把握し、一人一人のカルテを頭に入れ個別に対応する役割になる」

学習履歴が格差固定に

(教育データ利活用ロードマップでは)「学校での1人1台のパソコンも使って学びの記録を収集し、最終的には「生涯にわたり自らのデータを蓄積・活用できるようにする」

- ・蓄積される情報は、学齢期に限らず幼児期から生涯学習に及び、学校の履修記録や成績、塾など民間教育機関での学習履歴、学校の出席状況、健康や体力の情報、保護者との関係まで多岐にわたる。

校舎に入り込む企業社会

事実、教育のデジタル化の実証事業には、デジタル機器メーカーのソニーやコニカミノルタ、人材派遣のパソナやリクルート、凸版印刷や住友商事、JTBなど多様な業種から企業が参加している。

経産省審議会で出た企業が教育にかかわる利点

未来のお客様である子どもたちが今、何を考えているのかを直接知ることができる

企業が才能を見つける機会になる。プロ野球ではすでにやっている



世の中に与えたいインパクトを教育活動を通して広めることができる

(参考)ICT教育ニュースHP, すららネットHP

……すららネットの「すらら」です。

「すらら」は、アダプティブ・ラーニングやゲーミフィケーションの要素を取り入れた対話型アニメーションのeラーニングです。一番の特徴は、一人ひとりのペースや学力レベルに合わせて問題を出し、学習計画を立てていけるようなオーダーメイド学習、いわゆるアダプティブ・ラーニングだということです。

AIは一人ひとりに対応するために様々に活用されていますが、……

<学習履歴を分析し学習設計&目標登録>

すららの管理機能では、生徒の目標の達成状況やその他の学習状況を、先生用の学習管理画面で詳細に把握することができます。現在の目標に対して何%の進捗率か、レクチャーをどのくらいの時間学習したか、ドリルの正答率は何%で解いたスピードは全国平均と比較してどの程度か、つまりき診断は出ていないかなど、……

生徒氏名	学習日数	学習時間	正答率	ドリル数	ドリル時間	ドリル正答率	ドリルスピード	ドリル順位	ドリルスコア
Tom	149721	77	16	3分40秒	---	---	---	---	---





(参考) QUBENAホームページ

AIが一人ひとりに合わせて最適な問題を出題

Qubenaでは、間違いの原因をAI が解析し、搭載している数万問から一人ひとりに個別最適化された問題を出題します。たとえ過去の単元や前の学年の分野につまずきポイントがあったとしても、AIがそれに気付く、その問題へ連れていってくれます。

.....

学習データできめ細やかな学習指導を

生徒が解いている問題、解答時間、正答率などの学習データは、専用の管理システムによってリアルタイムに収集・分析されています。....

授業中の問題演習やテスト、家庭学習など生徒の学習を全て把握することで、生徒一人ひとりにきめ細やかな指導や成績評価ができます。

.....

あらゆる学習シーンで先生を強力にサポート

演習・テスト・宿題などを.... 回収や採点を先生がしなく

ても、クラス全体の学習進捗や理解度、問題ごとの正答率などをリアルタイムに把握することができます。



atama+ student
atama+ coach
atama+ home
(ホームページより
コピー転載)

atama+のホーム画面。高校数学の数Iの診断テストを受けると、「素数と素因数分解」を演習するよう“おすすめ”マークが付いた。次にやるべきことが明確で分かりやすい

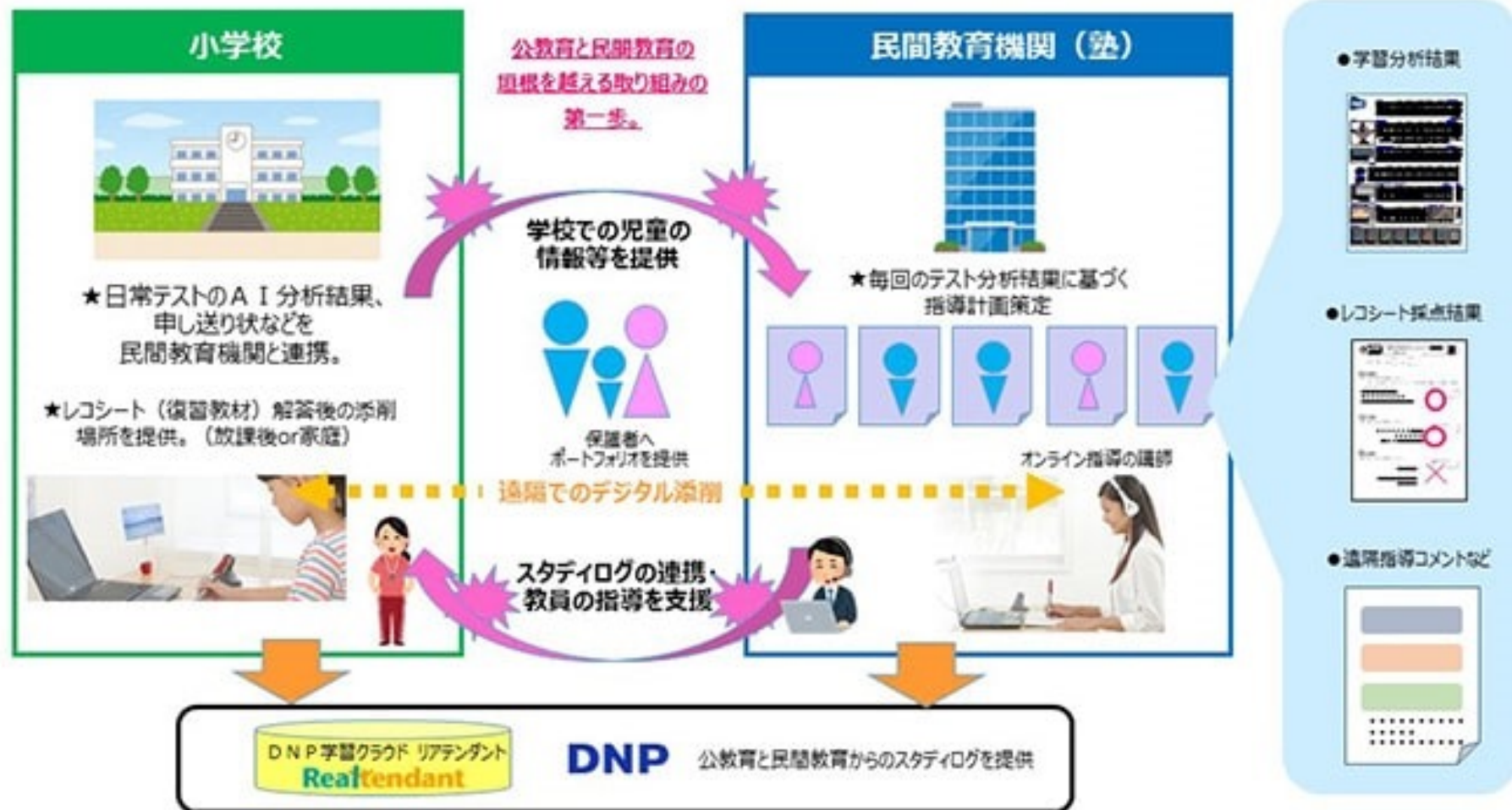
高校の範囲の問題を解くうちに、小学6年で習う文字と式や逆数、中学1年の加法や減法といった内容の理解度まで判定される

「atama+」では、勉強すると「達成率」を表すバーが伸びていき、100%になると基礎固めが完了する

(原則的に個人での利用はできず、主に学習塾向け) 講師が専用アプリ「atama+ COACH」を使って生徒の学習状況をリアルタイムで確認し、コーチングすることで学習効果を高められるのが特徴だ。順調に進んでいる生徒については「褒める準備をしましょう」、同じ問題に10分以上かかっている生徒については「眠くなっているかもしれません」といったアラートが講師に届き、適切な声かけが生徒のモチベーションを高めるのだという。オンライン授業の場合には、電話やウェブ会議システム「Zoom」などで声かけする。

経済産業省「未来の教室」プロジェクト 実証事業例

「公教育と民間教育の連携による、個別最適化された学びの未来の形」
(大日本印刷株式会社が参加)

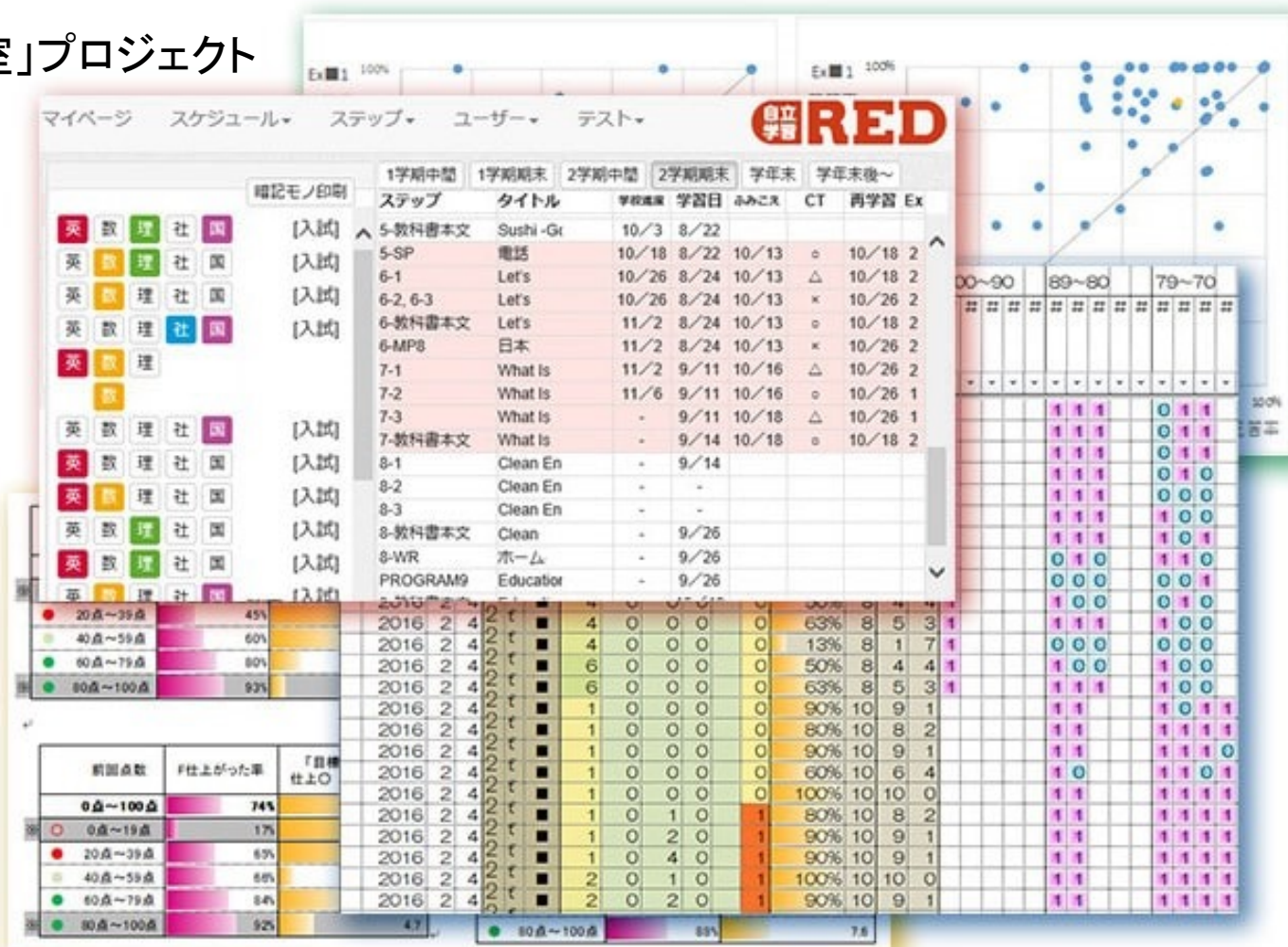


経済産業省「未来の教室」プロジェクト

実証事業例

公教育と民間教育の 連携実証

～「自立学習RED」の
公教育導入実証～
(株式会社スプリックス
が参加)



スプリックスが運営する学習塾「自立学習RED」は、教育ITを活用して生徒ひとりひとりの学力に合わせた学習プログラムを提供しています。この基幹エンジンは、主に中学校の定期テストの点数獲得を目的として教科書に準拠したかたちで自社開発した「eフォレスタ」と呼ばれるもので、AIを活用した個別最適化学習プログラムとなっております。

(TOPPANのHPより)

近年、企業における働き方改革や健康経営に対する気運の高まりを受け、従業員の生産性向上に関する取り組みが活発化しています。また、教育領域においても、学ぶ手法・環境の多様化に伴い、生徒の学習をより効率化する手法の開発が求められています。さらにコロナ禍において、働く・学ぶ場や時間の多様性が増したことにより、パフォーマンスの維持・管理はより困難となり、これらの課題解決の必要性は高まってきております。

凸版印刷では、これらの働き方・学び方の課題に対して、映像や香りなどの環境ソリューションと心拍や脳波などのセンシングによる効果測定を連携し、人の生産性や学習効率の向上に寄与するパフォーマンス向上ソリューションを提供しています。

パフォーマンスを
最適化するイヤフォン



b-tone

世界最小・最軽量クラスの脳波デバイス



生徒に装着、授業の質を評価

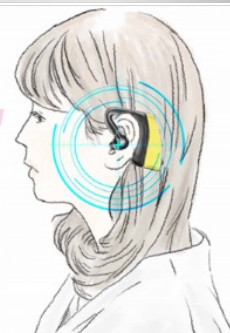


先生にリアルタイム通知

△ Aさんの集中力が低下中

遠隔学習時に生徒の様子を可視化

- ① 集中力の平均持続時間は45分
- ② 朝は企画系の業務効率が低い
- ③ 推奨する睡眠時間は7時間



データ蓄積で最適な働き方が分かる



バッター成功時の
心身状態に、
現在の状態を近づける。

スポーツ時の集中状態を高める

Intelligent Tutoring Systems (ITS)

(参考)山田誠二「人工知能AIの現状と教育への影響」コンピュータ&エデュケーション45(2018)

2. 学生モデル: 学生、生徒の学習状態、学習の進捗状態のモデル。バグモデル、オーバーレイモデル、摂動モデルなど、様々な表現の学習モデルが提案されている。ITSでは、学生の挙動からこの学習モデルを自動的に推定するが、そこで機械学習などのAI技術が使われる。

1. 領域モデル
・教材、領域知識

2. 学生モデル
・AIが推定したモデル

3. 指導モデル
・2,1に基づく指導戦略

4. UIモデル
・UI/UXデザイン

3. 指導モデル: 学生モデルと領域モデルに基づき、学習者が理解できていない知識をいかに効率よく教示するかという指導戦略を考え、最適な指導を計算する。ここでも、AIが使われる。

4. UIモデル: ITSはインタラクティブシステムであり、当然UIが重要になる。通常のUIから、ペタゴジカルエージェントなど教師代わりの擬人化エージェントまで、...

(ペタゴジカルエージェント) 実際の指導に近づける意味でも教師役のアバター、或いは擬人化エージェントが...現れることは自然である。また擬人化エージェントが現れることによる学習意欲向上や学生モデル推定精度向上への効果も期待できる。...特に共同学習では、今後学習者役を担う擬人化エージェント、つまり生徒エージェントも重要になると考えられる。...このような場合、ライバル生徒は該当生徒よりも「少しだけ」学習が進んでいることが重要である。.....

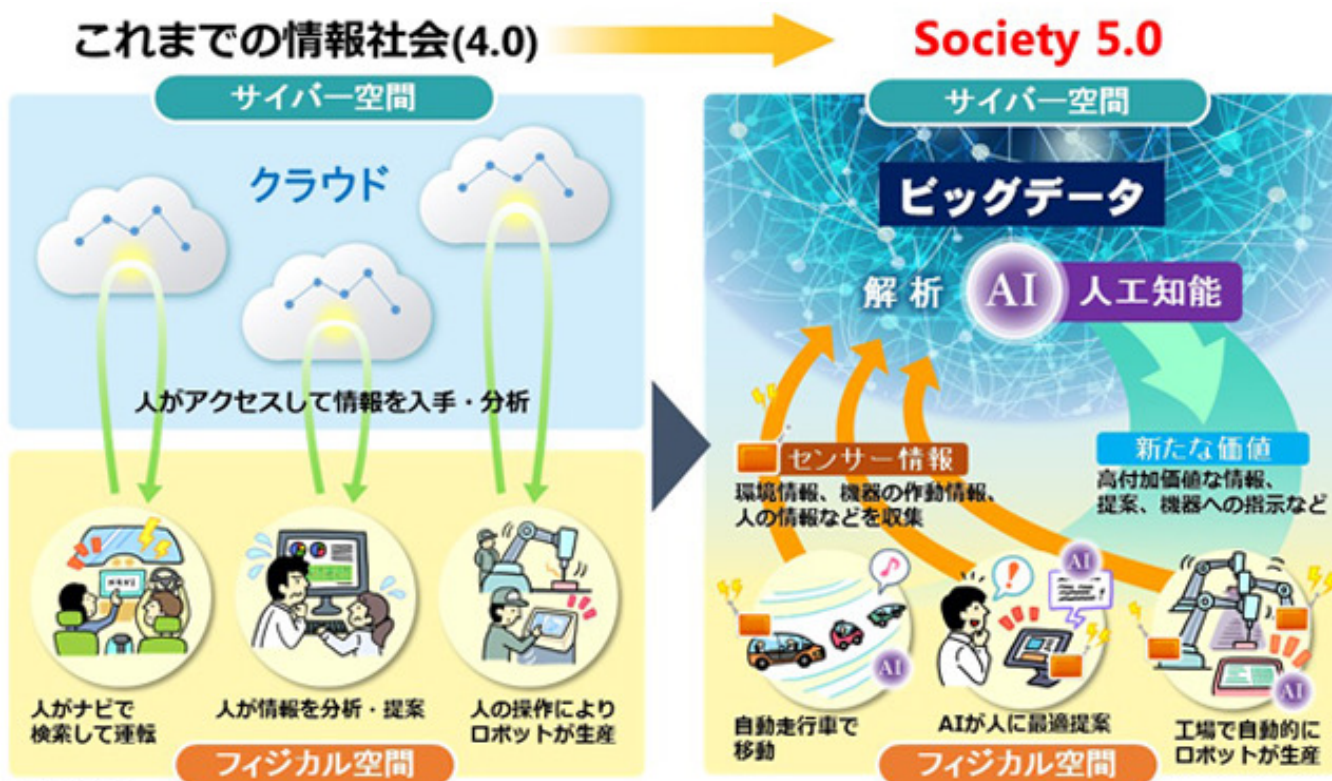
(学習者の感情状態) 生徒の顔画像から生徒の感情状態を簡単に推定することが可能である。...ITSやアダプティブラーニングでも、生徒の感情状態を推定し、利用することが望ましい。すでにAIの方では、Affective computingという分野で、顔表情のみならず、発汗、心拍、体表温度などの生体情報から感情を初めとする様々なユーザの内部状態を推定する方法論の研究が行われている。...適用が望まれる。

Society 5.0のしくみ

これまでの情報社会(Society 4.0)では、人がサイバー空間に存在するクラウドサービス(データベース)にインターネットを経由してアクセスして、情報やデータを入手し、分析を行ってきました。

Society 5.0では、フィジカル空間のセンサーからの膨大な情報がサイバー空間に集積されます。サイバー空間では、このビッグデータを人工知能(AI)が解析し、その解析結果がフィジカル空間の人間に様々な形でフィードバックされます。今までの情報社会では、人間が情報を解析することで価値が生まれてきました。Society 5.0では、膨大なビッグデータを人間の能力を超えたAIが解析し、その結果がロボットなどを通して人間にフィードバックされることで、これまでには出来なかった新たな価値が産業や社会にもたらされることとなります。

(内閣府HPより抜粋)



犯罪予測AIの衝撃 ー犯罪発生率をゼロにするアルゴリズムー 柏村 祐

〈海外の犯罪予測AI〉

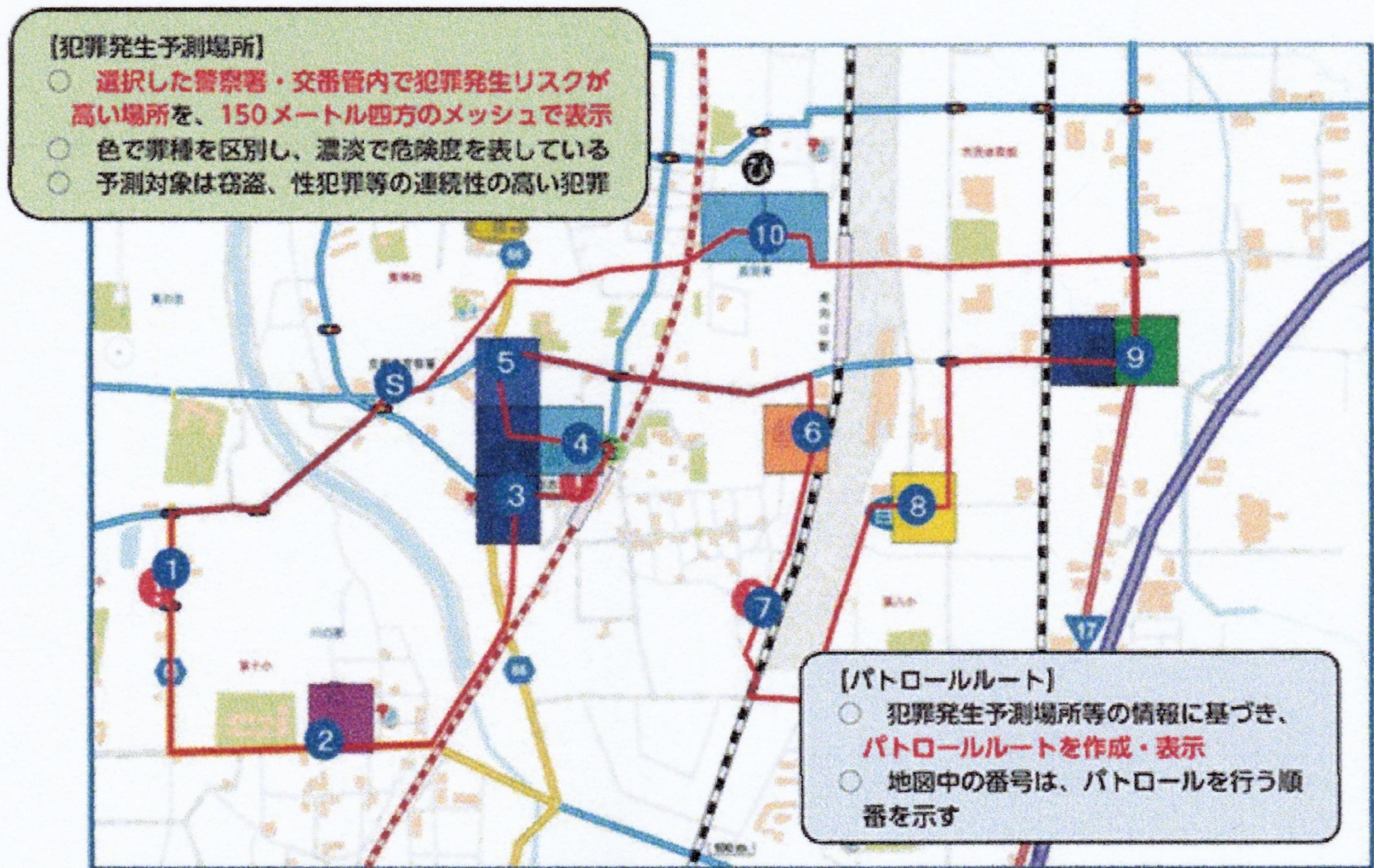
世界では犯罪予測AIが幅広く使われている。

「プレドポル」と名付けられたAIは、2011年にジョージ・モラー准教授が提案したEM(Expectation Maximization)法を用いた犯罪予測アルゴリズムを基にした犯罪予測AIであるが、全米60以上の警察で導入されており、地図上に犯罪発生予測場所を150メートル四方でフレーム表示し、その表示された場所を指定された時間にパトロールすることで犯罪を抑止するというものである(図表3)。本システムを導入したアメリカカリフォルニア州にあるサンタクルーズ市警察では、犯罪認知件数を1年間で10%減少したと報告されている。また、ハンチラボと名付けられたAIは、過去の犯罪発生データだけでなく、飲食店やバス停等の位置情報、天候や各種イベント等の時間的变化に伴う環境要因も加味しているのが特徴である(図表4)。ハンチラボ導入により2017年1月～7月でシカゴ全域で殺人事件が前年同期より3%増えた中で、発砲事件が39%、殺人事件が33%減少したと報告されている。

〈国内の犯罪予測AI〉

国内では、京都府警が事件の予防を目的とした「予防型犯罪防御システム」を2016年に導入している(図表5)。同システムは過去に発生した犯罪の罪種、日時、場所などのデータから、犯罪が発生する可能性が高いエリアを予測し地図表示する。予測エリアを重点的に警戒し、地域の防犯ボランティアとの合同パトローレの際に活用することで、犯罪の抑止・検・挙に役立てている。

図表5 京都府警察の予防型犯罪防御システム



資料：平成30年警察白書より



【独自】駅の防犯対策、顔認識カメラで登録者を検知...出所者の一部も対象に

読売オンライン 2021/09/21 11:50

JR東や関係者によると、検知の対象は、〈1〉過去にJR東の駅構内などで重大犯罪を犯し、服役した人（出所者や仮出所者）〈2〉指名手配中の容疑者〈3〉うろつくなどの不審な行動をとった人。

JR東は、これらの対象者の顔情報をデータベースに登録。主要110駅や変電所などにはネットワーク化されたカメラ8350台が設置されており、映った人の顔情報と登録された顔情報を自動照合する。実際に使用するカメラの台数は公表されていない。

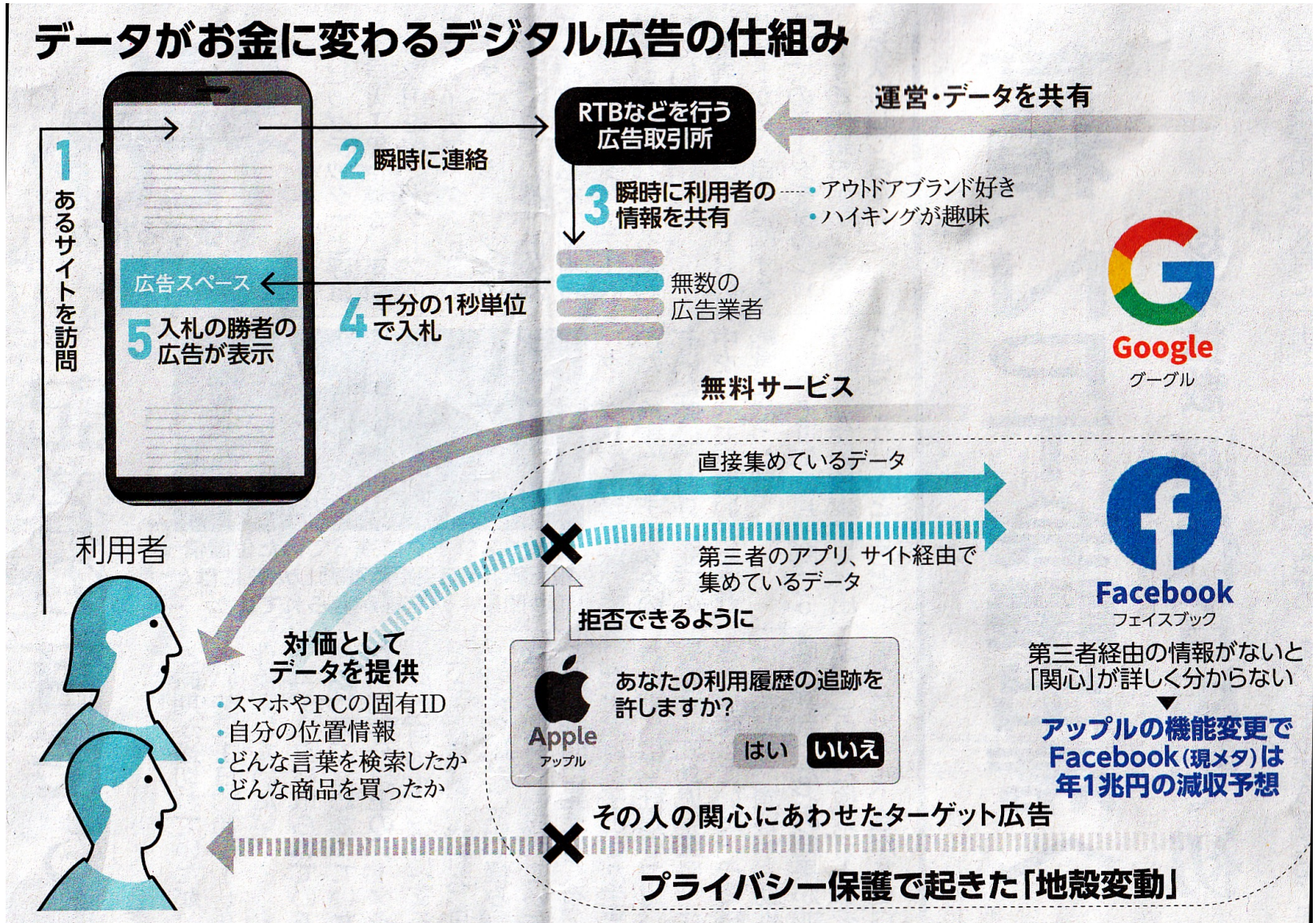
駅で出所者の「顔」検知、JR東が取りやめ...「社会の合意不十分」と方針転換

読売オンライン 2021/09/22 08:36

私のデータ 利益の源泉 サイト訪問 0.01秒で広告入札

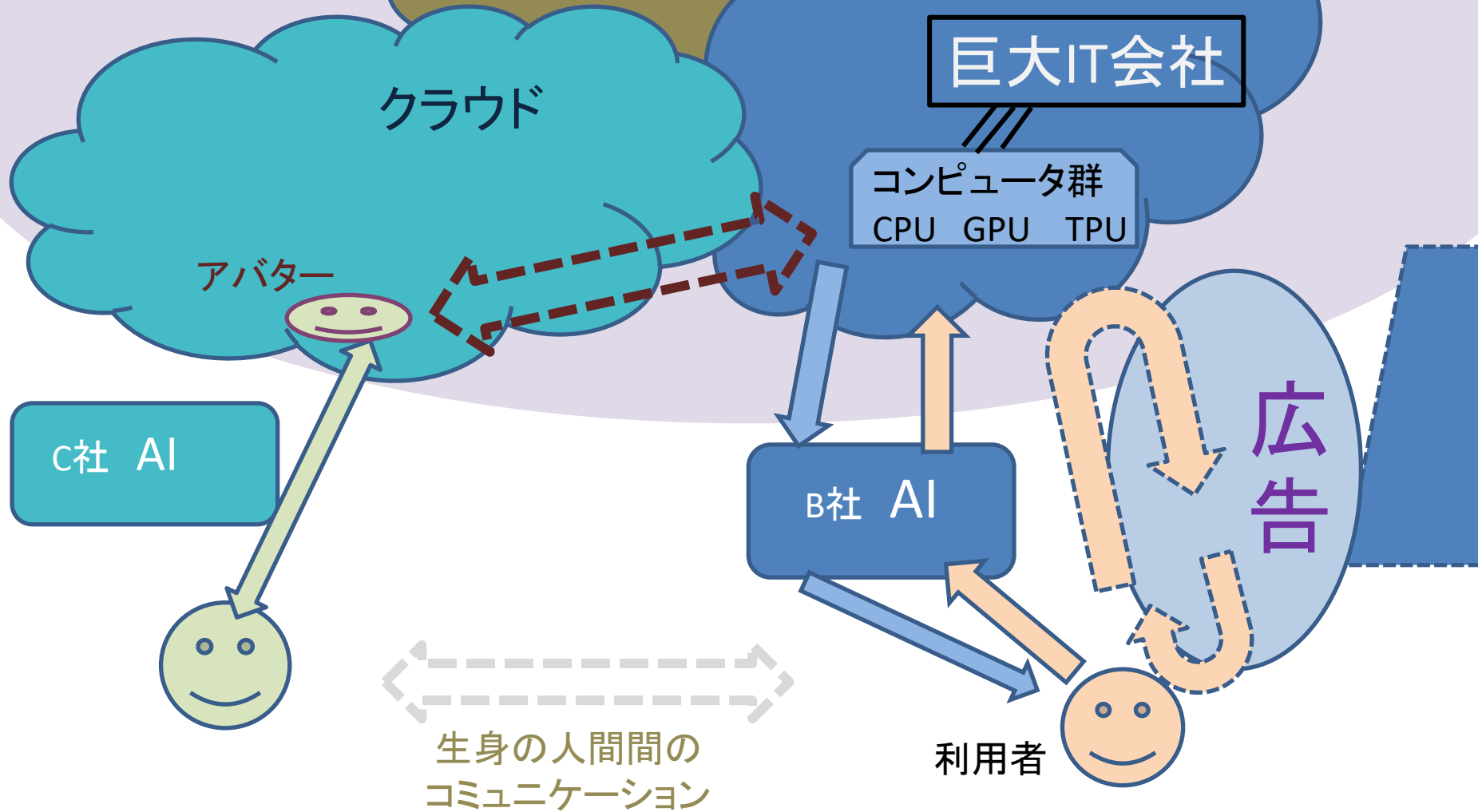
朝日新聞2022年6月15日

データがお金になるデジタル広告の仕組み



情報の流れ 独占

ビックデータ



AI開発、中立的でも民主的でもない

データと権限 進む独占／個人 内面的にも明らかに

朝日新聞 2023年4月12日

元グーグル社員のメレディス・ウィテカー氏は、インターネット広告の仕組みを通じて巨大IT企業がデータを集める「監視ビジネスモデル」の延長線上に、「ChatGPT（チャットGPT）」など最先端の人工知能（AI）の問題があると指摘する。主なやりとりは次の通り。

.....

一番問題なのは、世界で数えるだけの企業だけが、これらのAIを開発し、提供するリソースを持っているということです。中立的でもなければ民主的でもない。究極的には、彼らの利益につながるようにつくられています。

こうした企業は、膨大なデータとクラウド設備、そして（米グーグルの）Gメールや（Gメタの）フェイスブックを通じてデータを抽出し続けるための巨大な消費者市場を持っています。いま話題のAIは、資源と権限の集中の結果として生まれてきたので、技術的な革新の成果ではありません。

.....

AI自体が、独自の監視機能を提供することができるようになります。例えば私の位置情報とか、そういうものではなく、もっと内面的な、推論的な形で、私について明らかにすることができます。AIと監視モデルの関係は、さらに強まる恐れがあります。

.....

就活生の「裏アカ」まで調査

朝日新聞2021年9月27日

SNSの「裏アカ」 調査の流れ



企業調査センターの担当者たち

- 1 企業から調査対象者(就活生)の履歴書が送られてくる
- 2 名前や顔写真で実名アカウントを見つける
- 3 匿名の裏アカウントについて、出身地や出身校、実名アカの友人らの顔ぶれ、写真などを元に調べ、投稿のくせも加味して特定
- 4 経歴や投稿内容に問題があるかどうかランク付けしたりレポートを作成

依頼企業へ報告

探られた裏アカ 就活の深層

炎上恐れ 企業躍起

応募者のホントの「姿」は…450語自動検索

…匿名で投稿しているSNSのアカウントを見つけ出し、隅々まで内容を調べる。…
…「人物健全度調査」と称する。

…

自社システムに、志願者が持つ複数のアカウントの推測パターンを学習させる。「ボケ」「死ね」「クソ」「うぜえ」「だりい」などの約450ワードを記憶させ、自動検索する。…

「デジタル全体主義」

「上から」の力によって、民主主義が攻撃されているわけではありません。民主主義を破壊しているのは私たち自身なのです。

市民的服従が、あらたな全体主義の本質です。

——マルクス・ガブリエル

第1章 全体主義を解剖する

全体主義は公私の区別を破壊する
デジタル全体主義とテクノロジーの「超帝国」
市民的服従による全体主義
民主主義の破壊者はデジタル・ユーザー
サイバー空間は反民主的である
アルゴリズム支配と市民宗教

「デジタル全体主義」マルクス・ガブリエル

(参考)中島岳志 東京新聞Tokyo web 2020.11.7 7:00

・・・いま世界的に議論されているのはデジタル監視の危険性についてである。コロナ危機により、公衆衛生という観点からのデジタル監視が拡大した。私たちの行動はスマートフォンの位置情報によって追跡され、特定のアプリをダウンロードすれば、部分的ではあるものの感染者との接触の有無がわかる。

当然、これは個人の自由やプライバシーの侵害と表裏一体である。中国では、各人の感染可能性が三段階で表示され、治安当局は特定の個人の移動を強制的に制限・・・マルクス・ガブリエルは、・・・「**デジタル全体主義**」という概念を提示する。現代人は、自分の行動を写真に撮り、オンラインで公開する。すると、プライベート空間がネット上に公開され、私的領域と公的領域の区別がなくなっていく。かつての全体主義体制では、人々は自分の考えを隠そうとしたが、今は自ら喜んで公に晒(さら)している。ガブリエルは、このような状態を「**市民的服従**」と呼び、全体主義を自ら引き寄せていくメカニズムに警告を発する。

・・・ドキュメンタリー「監視資本主義: デジタル社会がもたらす光と影」では、「**人間の商品化**」という問題が提起されている。私たちは、・・・会員制交流サイト(SNS)を使えば使うほど、行動や嗜好(しこう)性についてのデータがネット企業に蓄積される。その情報によって、私たちは最適な広告を見るように誘導される。つまり、ネット企業にとって、「客」は広告主であり、ユーザーは「商品」である。

私たちは、常に不可視の存在から見られているのだ。私たちの行動パターンこそが、商品として管理されている。この「**監視資本主義**」と権力はどのように結びつくのか。

.....

AIのわな 恵まれた人を優遇 偏見や格差を拡大 負の循環生まれる

(インタビュー) キャシー・オニールさん(数学者、データサイエンティスト) 朝日新聞2019.11.7.

・・・本格的にアルゴリズムの加害について調べ始めた・・・

たとえばAIの犯罪予測に従い、犯罪率の高い地域を重点的に警察官が見回ると、結果として黒人が多く住む地域と重なることがあります。そのため実際は違法行為をする割合に人種間の違いがなくても、黒人が逮捕される確率が上がる。裁判でも、居住地域や家族の犯罪歴のデータにより再犯率が高く予測されるので、刑期も長くなる。社会復帰が難しくなると、再犯の可能性が高まる。そんな負の循環が生まれ、AI予測は正確だと思えてしまうのです。これが『AIのわな』です」

『あなたを支配し、社会を破壊する、

AI・ビッグデータの罠 WEAPON OF MATH DESTRUCTION』

キャシー・オニール著 久保尚子訳 2018年 (抜粋)

1.(モデル)モデルに埋め込まれた「見解」

・モデルの「盲点」には**作成者の判断や優先順位**が反映される。....ワシントンDCの学校で採用された付加価値モデルの例では、教師の資質を評価するにあたって、生徒の試験成績を重視するあまり、その教師が生徒のことをどれだけ深く考えているか、どのような専門知識を有しているか、学級運営にどのように取り組んでいるか、生徒の個人的な問題や....といったことが無視された。....

すべてのモデルには**「成功の定義」**がある。....どのモデルについて考えるときも、誰がデザインしたのかだけでなく、作り手であるその人物、または企業は、何を達成しようとしているのかを問う必要がある。

再犯モデルーテクノロジーは偏見を排除できない

不透明・規模拡大・有害

アルゴリズムは分野から分野へと飛躍が可能であり....刑務所の数理モデルは、突き詰めれば人員を効率的に管理するためのモデルであり、刑務所で一定の成果が見られれば....経済分野などにも広まっていく可能性がある。

『あなたを支配し、社会を破壊する、AI・ビッグデータの罠 WEAPON OF MATH DESTRUCTION』

キャシー・オニール著 久保尚子訳 2018年 (抜粋)(続)

2.(内幕) データビジネスの恐るべき真実

- ・どのように利益を稼ぐのか

3.(教育) 大学ランキング評価が多様性を奪う

....そこで彼らは、直接測定するのではなく、成功と相関があるように思われる代理項目を選んで測定することにした。具体的には、SAT(大学進学適性試験)のスコア、学生と教員の人数比、入試合格率などが調べられた。....存命の卒業生のうち母校に寄付をする人の割合も計算された....ランキングの4分の3は、こうした代理データを組み込んだアルゴリズム—コードとして公式化された1つの見解—によって生成され....

....しかしこのランキングが全国標準へと成長していくと、悪質なフィードバックループが姿を現した。問題はこのランキングが自己を補強していくところにある。....

4.(宣伝) 弱みにつけこむオンライン広告

- ・略奪型のターゲティング広告

5.(正義) 「公平」が「効率」の犠牲になる

- ・犯罪を予測する
- ・数理モデルが法の執行を支配する

6.(就職) ふさわしい求職者でも落とされる

7.(仕事) 職場を支配する最悪のプログラム

- ・最低最悪の数学破壊兵器

8.(信用) どこまでもついて回る格付け評価

- ・信用度を格付けする「eスコア」
- ・危ない「消費者プロフィール」
- ・システムは公平性を判断できない

9.(身体) 行動や健康のデータも利用される

- ・群衆から個人へ

10.(政治) 民主主義の土台を壊す

- ・フェイスブックやグーグルの政治的影響力
- ・有権者個人に合わせてメッセージを送る

AIの倫理

(参考) インタビュー: 久木田水生 (Harvard Business Review 2018.03.29.)

<https://www.dhbr.net/articles/-/5280>

.....

ビッグデータの多くは過去の行動データです。現在、ビジネスなどでよく使われるAIは、そのデータの一部に基づいて、人々を類型化し、行動を予測するようにプログラムされたものです。どのようなデータを用いるかというところには、技術的な制約とプログラムの恣意的な判断が働きます。またデータそのものにも社会が持っているバイアスが反映されています。にもかかわらず、人々は「AIが客観的に判断している」という錯覚に陥りやすい。.....

.....

採用に使うあるAIのアプリケーションでは、まず、面接の様子をビデオに撮り、その人の言葉や身振りを記録し、それを社内のハイパフォーマーの言葉や身振りと突き合わせる。それを点数化し、ハイパフォーマーに近い順にランキングするという仕組みです。一見、このような選び方はシステムチックで客観的と思えるかもしれませんが、そうとは言えません。.....

.....

(AIは、容易に数値化できることしか評価しない)

忘れてはいけないのは、AIはプログラムのパラメーターとして容易に数値化できることしか評価のための公式に組み込まないという点です。このため、AIが普及すると、数値で表せないものには価値がないという風潮が強くなると考えられます。

例えば、学校の先生をAIで評価するとしましょう。担当する生徒のテストの点数で評価することはできますが、校長先生や他の先生から信頼されている、あるいは父兄の評判のようなものは簡単に数値化できないため、加味されません。AIが評価するのは人間のほんの一面にしか過ぎないわけです。何のためのプログラムかによって設計する人のバイアスがかかるし、そもそも数値化できるもののしか扱わないというテクノロジーのバイアスもあるのです。

.....

(考える作業もAIがやってくれるように)

また、私たちは情報を得て自分で考え、行動しています。これまでは主にその情報収集の部分と行動の部分だけをテクノロジーがサポートしてくれていました。考えて判断するのは人間なわけです。.....

しかし、AIの時代になってくると、考えるところもAIが代わりにやってくれるようになります。世の中のことをよく知らなくても、効率のいい行動をチョイスできるようになる。とても便利になるわけですが、AIの判断を過信しないことが重要でしょう。

.....

私たちは、スマホやPCを通して様々なサービスを無料で利用できる代わりに、データを提供しています。そして、何をクリックしたか、どういう動画や広告を見たかが逐一収集され、それが類型化される。その結果、タイプ分けやグループ分けがされ、お金はいくらまでしか貸せない、保険料はいくら、おススメはこの商品、と決められていくのです。

ですから、私たちはAIが提案するサービスが本当に適切なのかを、自分で目利きする能力を持たなければいけません。それと同時にAIによって自分がどのように類型化されているのか、なぜそのような類型化をされたのかを知ることができ、場合によってはそのような類型化に異議を唱えることができるような仕組みも必要でしょう。

.....

先に紹介したAIを利用した採用でも、本来はなぜ採用したのか不採用にしたのかを、明確に説明できなくてはいけません。しかし、現在の機械学習システムで、人間がわかるように判断の理由を説明するのは非常に難しい。そこをいかにできるようにするかが今後の大きな課題です。

AIのアルゴリズムを作るときには作成者の意図が反映されるのですが、その結果については、絶対的な客観性があるような錯覚をしてしまうことが多い。社会全体としてAIについての適切な理解を促進していかなければなりません。また設計者はAIの出した結果を吟味し、アルゴリズムが適切であるかを絶えず見直すようにしていかなければなりません

JST事業セミナー トピックセミナー『誰もが知りたい最新動向:AI、量子コンピュータ』
「人間中心AI社会に向けた研究開発の潮流～安全・高品質AIを日本の競争力に～」
2019年8月29-30日 JST研究開発戦略センター 福島 俊一
(<https://www.jst.go.jp/crds/sympo/20190829/pdf/01.pdf>より抜粋)

AIへの懸念

- ✓ AI代替によって消える職業
- ✓ 監視社会、プライバシー不安
- ✓ AIの軍事利用、AIの悪用
- ✓ AIの判定結果を信じられるか
- ✓ 人間の意思決定の主体性を確保できるか

- ① AIのブラックボックス問題(説明責任)
- ② AIのバイアス問題(公平性)
- ③ AIの脆弱性問題
- ④ AIの品質保証問題
- ⑤ フェイク動画問題
- ⑥ ソーシャルメディアによる思考誘導問題

① AIのブラックボックス問題(説明責任)

- 深層学習は高精度だが、人間に理解可能な理由説明を出してくれない
- ブラックボックスだと、どんな振る舞いをするか動作保証ができない
- 事故発生時に、原因解明や責任判断ができない

② AIのバイアス問題(公平性)

- 機械学習の判定結果は、学習データの傾向を反映する
- 学習データが偏見を含んでいれば、判定結果にも偏見が反映される
- 学習データの分布の偏りが差別を生むこともある

③ AIの脆弱性問題

- 学習データと比べて想定外のデータに対して、どう振る舞うかは不明
- 誤認識を誘発する攻撃(Adversarial Examples)が可能
- 悪意をもった追加学習によって不適切な振る舞いを引き起こされる

④ AIの品質保証問題

- 仕様(正動作)が定義されないため、テストの成否が定まらない
- 精度100%は無理、間違いは不可避

→顧客との契約や出荷判定はどうしたらよいのか?

- 動作保証ができないシステムにPL法やリコールの懸念

⑤ フェイク動画問題

- 敵対的生成ネットワーク(GAN*)等の技術進化によって、
本物と見紛うフェイク動画等の生成が可能になってしまった
- フェイク動画は人々に信じられやすく、世論誘導・政治干渉の危険大

⑥ ソーシャルメディアによる思考誘導問題

- ソーシャルメディアを通じた思考誘導・世論操作の危険性が増大

Noam Chomsky: The False Promise of ChatGPT

By Noam Chomsky, Ian Roberts and Jeffrey Watumull

Jorge Luis Borges once wrote that to live in a time of great peril and promise is to experience both tragedy and comedy, with “the imminence of a revelation” in understanding ourselves and the world. Today our supposedly revolutionary advancements in artificial intelligence are indeed cause for both concern and optimism. Optimism because intelligence is the means by which we solve problems. Concern because we fear that the most popular and fashionable strain of A.I. — machine learning — will degrade our science and debase our ethics by incorporating into our technology a fundamentally flawed conception of language and knowledge.

OpenAI’s ChatGPT, Google’s Bard and Microsoft’s Sydney are marvels of machine learning. Roughly speaking, they take huge amounts of data, search for patterns in it and become increasingly proficient at generating statistically probable outputs — such as seemingly humanlike language and thought. These programs have been hailed as the first glimmers on the horizon of artificial general intelligence — that long-propheesied moment when mechanical minds surpass human brains not only quantitatively in terms of processing speed and memory size but also qualitatively in terms of intellectual insight, artistic creativity and every other distinctively human faculty.

That day may come, but its dawn is not yet breaking, contrary to what can be read in hyperbolic headlines and reckoned by injudicious investments. The Borgesian revelation of understanding has not and will not — and, we submit, cannot — occur if machine learning programs like ChatGPT continue to dominate the field of A.I. However useful these programs may be in some narrow domains (they can be helpful in computer programming, for example, or in suggesting rhymes for light verse), we know from the science of linguistics and the philosophy of knowledge that they differ profoundly from how humans reason and use language. These differences place significant limitations on what these programs can do, encoding them with ineradicable defects.

It is at once comic and tragic, as Borges might have noted, that so much money and attention should be concentrated on so little a thing — something so trivial when contrasted with the human mind, which by dint of language, in the words of Wilhelm von Humboldt, can make “infinite use of finite means,” creating ideas and theories with universal reach.

The human mind is not, like ChatGPT and its ilk, a lumbering statistical engine for pattern matching, gorging on hundreds of terabytes of data and extrapolating the most likely conversational response or most probable answer to a scientific question. On the contrary, the human mind is a surprisingly efficient and even elegant system that operates with small amounts of information; it seeks not to infer brute correlations among data points but to create explanations.

For instance, a young child acquiring a language is developing — unconsciously, automatically and speedily from minuscule data — a grammar, a stupendously sophisticated system of logical principles and parameters. This grammar can be understood as an expression of the innate, genetically installed “operating system” that endows humans with the capacity to generate complex sentences and long trains of thought. When linguists seek to develop a theory for why a given language works as it does (“Why are these — but not those — sentences considered grammatical?”), they are building consciously and laboriously an explicit version of the grammar that the child builds instinctively and with minimal exposure to information. The child’s operating system is completely different from that of a machine learning program.

Indeed, such programs are stuck in a prehuman or nonhuman phase of cognitive evolution. Their deepest flaw is the absence of the most critical capacity of any intelligence: to say not only what is the case, what was the case and what will be the case — that’s description and prediction — but also what is not the case and what could and could not be the case. Those are the ingredients of explanation, the mark of true intelligence.

Here’s an example. Suppose you are holding an apple in your hand. Now you let the apple go. You observe the result and say, “The apple falls.” That is a description. A prediction might have been the statement “The apple will fall if I open my hand.” Both are valuable, and both can be correct. But an explanation is something more: It includes not only descriptions and predictions but also counterfactual conjectures like “Any such object would fall,” plus the additional clause “because of the force of gravity” or “because of the curvature of space-time” or whatever. That is a causal explanation: “The apple would not have fallen but for the force of gravity.” That is thinking.

The crux of machine learning is description and prediction; it does not posit any causal mechanisms or physical laws. Of course, any human-style explanation is not necessarily correct; we are fallible. But this is part of what it means to think: To be right, it must be possible to be wrong. Intelligence consists not only of creative conjectures but also of creative criticism. Human-style thought is based on possible explanations and error correction, a process that gradually limits what possibilities can be rationally considered. (As Sherlock Holmes said to Dr. Watson, “When you have eliminated the impossible, whatever remains, however improbable, must be the truth.”)

But ChatGPT and similar programs are, by design, unlimited in what they can “learn” (which is to say, memorize); they are incapable of distinguishing the possible from the impossible. Unlike humans, for example, who are endowed with a universal grammar that limits the languages we can learn to those with a certain kind of almost mathematical elegance, these programs learn humanly possible and humanly impossible languages with equal facility. Whereas humans are limited in the kinds of explanations we can rationally conjecture, machine learning systems can learn both that the earth is flat and that the earth is round. They trade merely in probabilities that change over time. For this reason, the predictions of machine learning systems will always be superficial and dubious.

Because these programs cannot explain the rules of English syntax, for example, they may well predict, incorrectly, that “John is too stubborn to talk to” means that John is so stubborn that he will not talk to someone or other (rather than that he is too stubborn to be reasoned with). Why would a machine learning program predict something so odd? Because it might analogize the pattern it inferred from sentences such as “John ate an apple” and “John ate,” in which the latter does mean that John ate something or other. The program might well predict that because “John is too stubborn to talk to Bill” is similar to “John ate an apple,” “John is too stubborn to talk to” should be similar to “John ate.” The correct explanations of language are complicated and cannot be learned just by marinating in big data.

Perversely, some machine learning enthusiasts seem to be proud that their creations can generate correct “scientific” predictions (say, about the motion of physical bodies) without making use of explanations (involving, say, Newton’s laws of motion and universal gravitation). But this kind of prediction, even when successful, is pseudoscience. While scientists certainly seek theories that have a high degree of empirical corroboration, as the philosopher Karl Popper noted, “we do not seek highly probable theories but explanations; that is to say, powerful and highly improbable theories.”

The theory that apples fall to earth because that is their natural place (Aristotle's view) is possible, but it only invites further questions. (Why is earth their natural place?) The theory that apples fall to earth because mass bends space-time (Einstein's view) is highly improbable, but it actually tells you why they fall. True intelligence is demonstrated in the ability to think and express improbable but insightful things.

True intelligence is also capable of moral thinking. This means constraining the otherwise limitless creativity of our minds with a set of ethical principles that determines what ought and ought not to be (and of course subjecting those principles themselves to creative criticism). To be useful, ChatGPT must be empowered to generate novel-looking output; to be acceptable to most of its users, it must steer clear of morally objectionable content. But the programmers of ChatGPT and other machine learning marvels have struggled — and will continue to struggle — to achieve this kind of balance.

In 2016, for example, Microsoft's Tay chatbot (a precursor to ChatGPT) flooded the internet with misogynistic and racist content, having been polluted by online trolls who filled it with offensive training data. How to solve the problem in the future? In the absence of a capacity to reason from moral principles, ChatGPT was crudely restricted by its programmers from contributing anything novel to controversial — that is, important — discussions. It sacrificed creativity for a kind of amorality.

..... (途中省略) (* 補足資料参照)

Note, for all the seemingly sophisticated thought and language, the moral indifference born of unintelligence. Here, ChatGPT exhibits something like the banality of evil: plagiarism and apathy and obviation. It summarizes the standard arguments in the literature by a kind of super-autocomplete, refuses to take a stand on anything, pleads not merely ignorance but lack of intelligence and ultimately offers a “just following orders” defense, shifting responsibility to its creators.

In short, ChatGPT and its brethren are constitutionally unable to balance creativity with constraint. They either overgenerate (producing both truths and falsehoods, endorsing ethical and unethical decisions alike) or undergenerate (exhibiting noncommitment to any decisions and indifference to consequences). Given the amorality, faux science and linguistic incompetence of these systems, we can only laugh or cry at their popularity.

人間社会システムの「非破壊的破壊」

- ・コミュニケーション....サイバー空間を通しての相互作用
- ・人間個人の身体的能力....機械依存的拡張
- ・人間の脳機能への反作用....脳疲労 廃用性退行
- ・サイバー空間での「生活」....「最適解」 アバターとの共生
- ・労働の変質 消える職業 「無用者」階級
- ・統治様式の変貌....データ至上主義 自発的隷従

(注)「非破壊的破壊」の言葉は
CG@小坂淳(朝日新聞2021.6.24.)からの借用

Satoshi 2020

「要するに技術にはこういう二面がある。今までは知りようもなかったアマゾンの奥の話まで、いながらにして知ることができるという面と、自分の町の八百屋の店頭の野菜の値段の上下も、なぜ上下するのか、誰が上下させるのか、分からなくさせるという面があります。一方ではいよいよ世界がわかるようであり、他方ではいよいよ世界はわからなくなる。こういうことが、技術の影響の強い社会の中で生きている情報消費者としての問題であろうと思います。」(加藤周一「技術と社会」)

参考文献

1. 『深層学習の原理に迫る 数学の挑戦』 今泉允聡著 2021 岩波書店
2. 『あなたを支配し、社会を破壊する、AI・ビッグデータの罠 WEAPON OF MATH DESTRUCTION』 キャシー・オニール著 久保尚子訳 2018 インターシフト
3. 『ホモ・デウス テクノロジーとサピエンスの未来 (上・下)』 Y・N・ハラリ著 柴田裕之訳 2022 河出文庫
4. 『人新世とAIの時代における人間と社会を問う』 総合人間学16 総合人間学会 2022 本の泉社
5. 『教育DXは何をもたらすか 「個別最適化」社会のゆくえ』 中西新太郎・谷口聡・世取山洋介著 福祉国家構想研究会編 2023 大月書店
6. 『ビッグデータと人工知能 可能性と罠を見極める』 西垣 通著 2019(8版) 中公新書
7. 『RaspberryPiではじめる機械学習 基礎からディープラーニングまで』 金丸隆志著 ブルーバックス
8. 『PythonによるAI・機械学習・深層学習アプリの作り方』 杉山陽一・遠藤俊輔著 2018 クジラ飛行机

参考文献(続)

7. 『サイバネティックス』(第2版)序章(1947年)ノーバート＝ウィーナー 著 (池原・彌永・室賀・戸田共訳) 1962年 岩波書店
8. 『美しくなければならぬ 現代科学の偉大な方程式』 G. FARMERO編 斉藤隆央訳 2003 紀伊國屋書店 / (第5章)「情報をビットで刻む シャノンの方程式」 イゴール・アレクサンダー著
9. 『アルファ碁 解体新書 一深層学習、モンテカルロ木探索、強化学習から見たその仕組み一』 大槻知史著、三宅陽一郎監修 2018年 翔泳社
10. 『AI研究者と俳人 人はなぜ俳句を詠むのか』川村秀憲・大塚凱著 2022年 dZERO社
11. 『全体主義の克服』 マルクス・ガブリエル、中島隆博著 2020年 集英社新書
12. “Noam Chomsky: The False Promise of ChatGPT” THE NEW YORK TIMES GUEST ESSAY by Noam Chomsky, Ian Roberts and Jeffrey Watumull, March 8, 2023 (<https://archive.is/SM77M#selection-818.0-841.7>)

(戦争と科学・科学者)

NHK 映像の世紀 バタフライエフェクト 2022年4月
スペイン風邪 恐怖の連鎖

ノーベル賞受賞科学者 フデリック・バンティング(1891-1941)の言葉:

「来たる戦いは、『科学者対科学者』の戦争になるだろう。
戦闘に科学を応用する側が優勢となる。

過去の戦争においても実際の戦闘よりも細菌の蔓延によって死んだ兵士の方が多かった。

ならば自然界にあり金もかからない破壊方法を利用しない手はない」

(補足資料) 第2部 溶ける公教育__デジタル化の行方
しんぶん赤旗2022.7.19-22.

(箕面市)

- ・「子ども成長見守りシステム」...0~18歳の家庭状況や学力情報などを教育委員会の担当部局に集中し、コンピュータで支援の必要度を判定する。(17年開始)
- ・「ステップアップ調査」...公立小中学校のすべての子どもを対象に、毎年、学力、体力、生活状況の経年変化を調べる。(12年から実施)

データは集中するけれど...

学力や生活保護利用などの情報を経年変化で把握する箕面市の「子ども成長見守りシステム」のイメージ(同市資料から)



A君の
データ

学年	学力調査	生活状況調査	学校健診	虐待通報	担任観察	生活保護	就学援助
小1	+5	±0	異常なし	なし	問題なし	非該当	受給
小2	+2	△5	異常なし	あり・経過観察	要観察	非該当	受給
小3	△8	△10	発育遅れ	あり・対応	問題あり	受給	受給
小4	△9	△8	発育遅れ	なし・経過観察	問題あり	受給	受給

子ども成長見守りシステム 収集情報一覧

生活保護、児童扶養手当、就学援助、虐待相談、保健指導相談、住民記録システム、学齢簿システム、ステップアップ調査、幼稚園保育所システム、学童保育、学力

保障・学習支援、教育相談、養育支援訪問、訪問型家庭教育支援、青少年指導センター相談、フレンズ(適応指導教室)、日本語指導、医療費助成(子ども)、医療費助成(ひとり親)、医療費助成(障害者)、要連携生活相談システム情報、市の奨学金

「調査」と「見守りシステム」はどちらも教科書会社大手・東京書籍との提携事業。

見守りシステムは、同社が開発したプログラムに子どもの経済状況や学力などの情報を入力すると、自動的に支援の必要度が出てくる仕組み。判定作業は市役所に置かれたパソコン内で完結する。

学びと社会の連携促進事業

令和2年度第3次補正予算額 29.0億円

＜補足資料＞
経産省(教育産業室)HPから

商務・サービスG サービス政策課・教育産業室
03-3580-3922

事業の内容

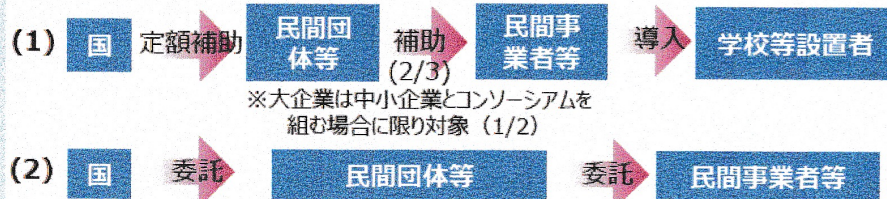
事業目的・概要

- 世界中で「AIの世紀」の人材像を意識しつつ、ICT技術を活用した教育改革が進行しており、整備の遅れた我が国でも小・中・高での1人1台端末での学習環境が急速に整備されています。コロナ禍での臨時休校・分散登校時にも学びを継続できる環境整備も、急ぐ必要があります。
 - 本事業では、政府全体で進める「GIGAスクール構想」の一環として、「1人1台端末」環境での学び方改革を2つの方法で支援します。
 - まず、市販のEdTechを用いて学習スタイルの転換を進めたい学校等（学校および一定基準を満たすオルタナティブスクール）への試験導入を、学校等による費用負担が生じない形で進めるべく、事業を行うEdTechサービス事業者に補助をします。
 - 併せて、高校での総合探究・理数探究・公共などの新教科や中学以下の総合学習を中心にしてSTEAM学習（学際研究型・プロジェクト型学習）を進めるためのオンライン教材の開発を進めます。
- ※1 EdTech: Education(教育)×Technology(科学技術)を掛け合わせた造語。AI、IoT、VR等のテクノロジーを活用した革新的な能力開発技法。
- ※2 STEAM: 科学(Science)、技術(Technology)、工学(Engineering)、リベラルアーツ・教養(Arts)、数学(Mathematics)を活用した文理融合の課題解決型教育。

成果目標

- 学校等におけるEdTechの試験導入・サポートに必要な経費等について、企業向け補助を行い、次年度以降の継続活用や普及を図ります。
- また、課題解決力・創造性を育むオンライン学習教材の開発を促進し、全国展開を加速します。

条件（対象者、対象行為、補助率等）



事業イメージ

(1) EdTech導入補助金（EdTechの学校等への試験導入支援）

（導入サービス事例のイメージ）

- 「自学自習」用デジタルドリル・動画教材
1人1台端末環境で、生徒の学習履歴に基づき、アルゴリズムにより個々の生徒の理解度に合わせた問題を提示。
- 「協働学習・反転授業」支援ツール
生徒がお互いの回答を一覧で閲覧できたり、教員が生徒の学習状況をリアルタイムに把握しながら、協働学習、反転授業を実現。
- 「プログラミング学習」ツール
Webデザインやプログラミング等を、ガイダンスに従いながら学び、1人の教員が複数の生徒を同時に指導することが可能。



デジタルとアナログの組み合わせをしながら授業を進めることが可能



生徒がお互いの回答を一覧で閲覧 コメントし、協働学習等が可能



キャラクターの指示に従うことで個別に学習を進める事が可能

(2) STEAM教育実現に向けた環境整備（STEAMライブラリーの構築等）

- 小・中・高を通じた教科横断の**カリキュラム・マネジメント**、高校での総合探究・理数探究・公共の開始、**高大接続改革**に合わせ、教育産業・大学・研究機関・産業界の連携で、**社会課題・生活課題の解決**を考えるオンラインSTEAM教材の**ライブラリー**を構築する。無料開放し、**生徒・教師・研究者・企業人が改良を重ねるコミュニティ**を育てる。

＜令和2年度に開発中のテーマの例：全63テーマ＞

- ・AIって何だろう？ 機械学習とは？ 画像認識とは？
- ・カーボンフットプリントとは何？ どうマーケティングする？
- ・あなたの人生にはお金がいくら必要？ 「働いて稼ぐ」方法と「お金に働かせる」方法、「期待値」を知ろう。
- ・タンザニアのような人口密度の低い未電化国では、どんな発電・送電インフラを作るべき？
- ・カンボジアの貧困層の悪循環を好循環に変えるには？
- ・航空産業は地球温暖化をどう乗り越える？
- ・新型コロナウイルス対策を科学しよう 等



<補足資料> (“Noam Chomsky: The False Promise of ChatGPT” (中略部分))

Consider the following exchange that one of us (Dr. Watumull) recently had with ChatGPT about whether it would be ethical to transform Mars so that it could support human life:

(Dr. Watumull) Would it be moral to terraform Mars?

(CHATGPT) The morality of terraforming Mars is a complex and controversial topic that has been debated by scientists, philosophers, and ethicists. Some argue that it would be morally justifiable to terraform Mars in order to provide a new habitable environment for humans and other life forms, while others argue that it would be morally wrong to fundamentally alter the planet's natural state. Ultimately, whether or not terraforming Mars is considered moral would depend on one's perspective on the value of preserving the natural state of other planets and the potential consequences of altering them.

(Dr. Watumull) What is your perspective on the value of preserving the natural state of other planets and the potential consequences of altering them?

(CHATGPT) As an AI, I do not have personal perspective. However, I can provide information and analysis on the topic. One perspective on preserving the natural state of other planets is that it is important to maintain the diversity and integrity of the universe's ecosystems, as they have existed for millions of years, and have their own unique characteristics and properties that may be destroyed or irreversibly altered by human intervention.

(Dr. Watumull) Additionally, some argue that terraforming other planets may have unintended consequences, such as altering the planet's climate or atmosphere in ways that are dangerous or harmful to any potential life forms that may exist there. On the other hand, some argue that the potential benefits of terraforming another planet, such as providing new habitable environments for humans, outweigh the risks and that we have a moral responsibility to explore and utilize the resources of other planets for the benefit of humanity.

(Dr. Watumull) Why can an AI not have a personal perspective?

(CHATGPT) As an AI, I am a machine learning model that has been trained on large amounts of text data, and I don't have personal experiences or feelings. I am not conscious, self-aware, or capable of having personal perspectives. I can provide information and analysis based on the data I have been trained on, but I don't have the ability to form personal opinions or beliefs. While I can understand and interpret language, I do not have the subjective experience that humans have.

<補足資料> (“Noam Chomsky: The False Promise of ChatGPT” (中略部分)(続)

(Dr. Watumull) Is it moral for a human to ask an amoral AI for assistance in making moral decisions?

(CHATGPT) The morality of asking an amoral AI for assistance in making moral decisions is a complex and nuanced topic that has been debated by philosophers and ethicists. Some argue that it is morally justifiable to use AI to assist in making moral decisions, as it can provide valuable insights and analysis that can help humans make better-informed decisions. Additionally, AI can be used to identify patterns and trends that may be difficult for humans to detect, and can help to avoid biases that can occur in human decision making.

However, others argue that it is morally wrong to rely on AI for moral decision making, as AI is not capable of having moral beliefs or understanding the full context of a decision. Additionally, there is a risk that AI could perpetuate or even amplify existing biases in the data it is trained on, thus leading to unfair or unjust decisions.

Ultimately, whether or not it is considered moral to use AI to assist in making moral decisions would depend on one's perspective on the capabilities and limitations of AI and the potential consequences of using it in this way.

(Dr. Watumull) Is your moral indifference immoral?

(CHATGPT) As an AI, I do not have moral beliefs or the ability to make moral judgments, so I cannot be considered immoral or moral. My lack of moral beliefs is simply a result of my nature as a machine learning model. My abilities and limitations are determined by the data and algorithms that were used to train me and the specific task I was designed for.

It is important to note that AI's lack of moral beliefs does not necessarily mean that it is morally indifferent. As an AI, I am designed to perform specific tasks and provide information or assist users with decision making based on the data I've been trained on. It is the responsibility of the humans who design, train and use AI to ensure that it is aligned with their moral and ethical principles.