

1-1 社会で起きている変化

東京大学 数理・情報教育研究センター
久野遼平

2022年10月21日

概要

- データサイエンスやAIとはどういう分野なのでしょうか？なぜ社会の関心を集めているのでしょうか？
- 本節では社会におけるデータ・AIの利活用例を幅広く学ぶことで、社会で起きている変化を知り、データサイエンスやAIを学ぶことの意義を理解することを目標とします
- 特にAIを活用した新しいビジネスやサービスは、複数の技術が組み合わせられて実現していることに注目して下さい
- 後半の方では社会科学や自然科学など学問の分野での変化についても簡単に紹介します

1-1-1 データサイエンス入門

データサイエンスとは

- データを有効活用し、数理モデリングや計算技術と適用ドメインの専門知識を結合することで新たな知識を生み出し、その活用のシナリオを導き出すことです
- なぜ注目を集めているのでしょうか？

古くからあります

- 例えば17世紀の天文学者であるヨハネス・ケプラーは
 - (1) データの有効活用：ティコ・ブラーエが長年観測した惑星の運動データを用い
 - (2) 数理モデル・計算技術・適用ドメインの専門知識：天文学の知識を活用し、手計算で惑星運動を分析し
 - 計算機はおろか消しゴムすらありません
 - (3) 活用のシナリオ：仮説検証や知識発見などを通じてその後の天文学に大きな影響を与えました
 - ケプラーの法則はニュートンの万有引力へと繋がっていきます
- なぜ今注目を集めているのでしょうか？

The image shows a page from Johannes Kepler's 'Tabularum Rudolphinarum' (1627), which is a detailed astronomical table for Mars. The title at the top is 'Tabularum Rudolphinarum Tabula Aequationum MARTIS'. The table contains numerous columns of numerical data, likely representing planetary positions, distances, and other astronomical measurements. The text is in Latin and the table is a complex grid of numbers and some descriptive text.

出典：
[Kepler1627]
<https://www.loc.gov/item/49038330/>

技術のブレークスルー：理論

- 統計学・機械学習・最適化技術の進歩により今まで最適化できなかった問題も「効率的に解ける」ようになりました。データの利活用の際に必須の統計や学習の技術も飛躍的に進歩しました。
- 根本技術の土台には数理の蓄積があります
 - 縁の下での力持ち
- AIと将棋棋士が熱戦を繰り広げられるようになったのもそうした技術の進歩による所が大きいです
 - 電王戦 (<https://www.shogi.or.jp/match/denou/>)
 - 後述する深層学習・強化学習と関連します

技術のブレークスルー：ハードウェア

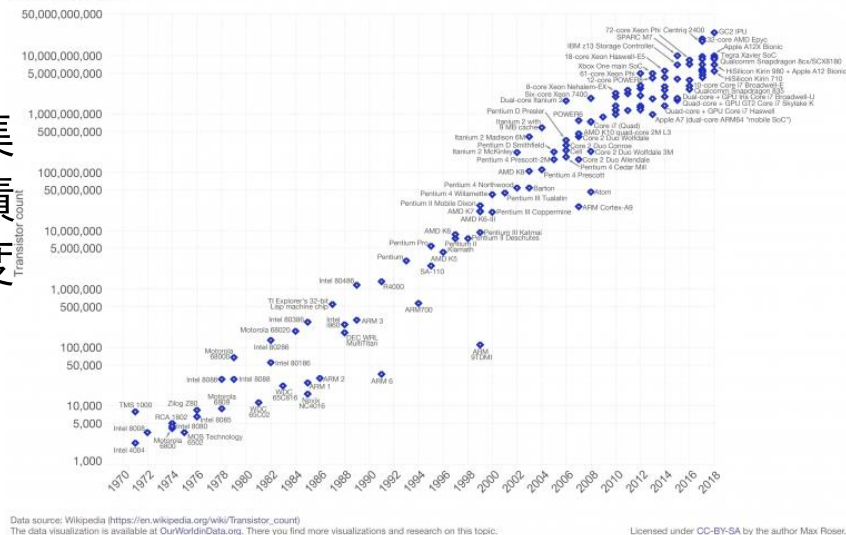
- 計算機の進化
 - 処理性能の向上→CPU（ムーアの法則）、GPUの進歩
 - 記憶装置の進化→メモリ、HDD、SSD
 - 情報通信技術の革新→インターネット、クラウド、エッジ

ムーアの法則[OurWorldData]

Moore's Law – The number of transistors on integrated circuit chips (1971-2018)
Moore's law describes the empirical regularity that the number of transistors on integrated circuits doubles approximately every two years. This advancement is important as other aspects of technological progress – such as processing speed or the price of electronic products – are linked to Moore's law.



集積度



年

GPUの実物

CPUとGPUはどちらも計算処理を行う計算機のいわば頭脳ですが、GPUはCPUに比べ簡単な計算を並列で大量に実行するのに向いています。

図の出典：

<https://ja.wikipedia.org/wiki/NVIDIA#/media/%E3%83%95%E3%82%A1%E3%82%A4%E3%83%AB:NVIDIA-GTX-1070-FoundersEdition-FL.jpg>

データ量の増加

- 記録の電子化
 - 紙媒体で記録していた時はデータを分析しようと思っても不可能でした。ペーパーレス化が進むとデータの有効活用に対するニーズが高まりました。
- ICT技術（インターネット）の普及
 - 人と人が繋がりがやすくなった結果、民泊などシェアリングエコノミーと呼ばれる新しいサービスも多く生まれました
 - シェアリングエコノミーとは、個人が保有する資産をインターネットを介して他者にも利用してもらう社会のことです

ペーパーレス化の例：

NACCS(輸出入・港湾関連情報処理システム)：国際貿易に関連する通関手続きは非常に面倒です。日本では比較的早い時期から電子手続きに移行しています

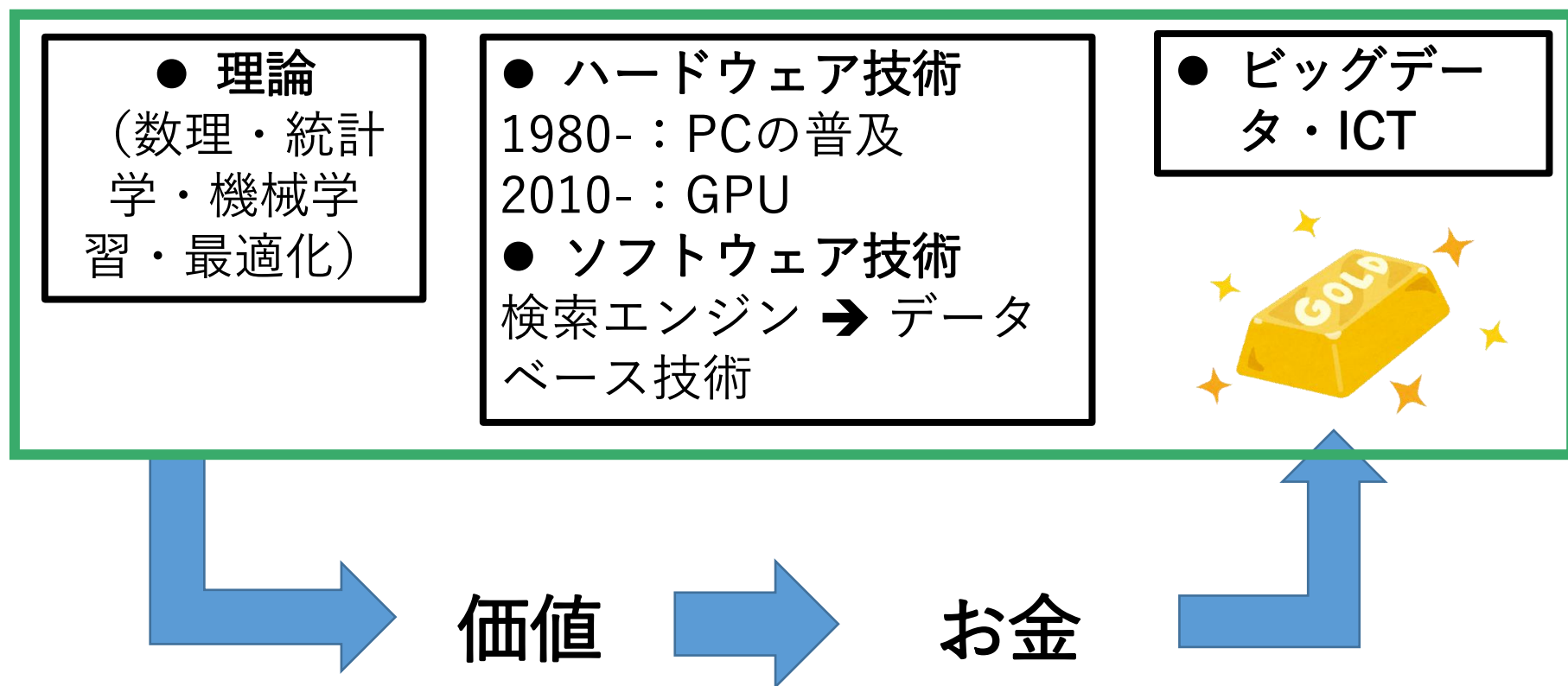
(<https://www.naccs.jp/>)

2019年において：

- Youtubeでは毎分約450万本のビデオが再生されました
- Uberでは毎分約1万人がタクシーを利用しました
- Googleでは毎分約450万件の検索が行われました（出典[Domo]）

全てが組み合わさった結果

- 価値が生まれるようになり好循環が誕生



1-1-2 応用分野（ビジネス）

画像生成

- テキストを入力とし画像を新規生成
 - OpenAIのDallE[Ramesh2021] (<https://openai.com/blog/dall-e/>)
- 生成した画像に対してさらに修正箇所を指定し
 - DALL・E2 (<https://openai.com/dall-e-2/>)
 - Glide (<https://gpt3demo.com/apps/openai-glide>)
- 複数の技術：自然言語処理 × 画像生成（マルチモーダルの例）
 - 生成技術の進歩はフェイクニュースやディープフェイクなど新たな問題も生んでいます

入力  「アボカドのような椅子」 「③の位置にフラミンゴを追加」

機械が生成した画像  <https://openai.com/blog/dall-e/> で実例を見ることができます <https://openai.com/dall-e-2/> で実例を見ることができます

IoTと自動制御

- センサーからリアルタイムでデータを読み取り自動で複雑な制御を行うこともあります

- IoT (Internet of Things) :
モノのインターネット

- 例えば熟練した農家の栽培技術のAIでの再現を目標に、IoT機器と機械学習の技術で上手な水やりを学習し、甘いトマトの栽培に成功した研究者もいます

(<https://www.jst.go.jp/kisoken/jyonetsu/interview/h29/mineno.html>)



出典：2020 科学技術振興機構(JST)
峰野博史、多様な環境に自律順応できる水分ストレス高精度予測基盤技術の確立

- 自動運転技術の開発が盛んです
(<https://www.tesla.com/autopilot>)



出典: Cityscape <https://www.cityscapes-dataset.com/examples/#fine-annotations>

- KOMTRAX [KOMTRAX]
 - コマツ機械稼働管理システム

マーケティング：購買、販売、サービス

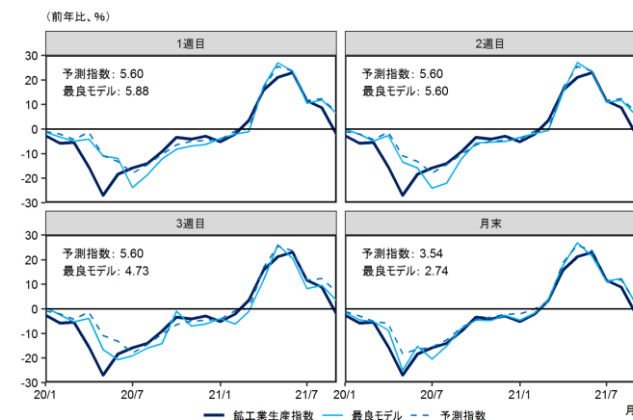
- ウェブログやGPSデータなどは購買、販売、調達、サービスのあらゆる場面で利活用されています
- 例えばAmazonのおすすめは皆さんも一度は利用したことがあると思います
- ブログウォッチャーの例でも見ましたが、GPSデータに基づいた人口統計は商圈分析や人間行動の分析にも役立てられています
(<https://mobaku.jp/>)
 - こうしたGPSデータを用いるとどのような属性の人がある時・ある場所にいたのかがわかるので、マーケティングに役立ちます

1-1-3 応用分野（社会科学）

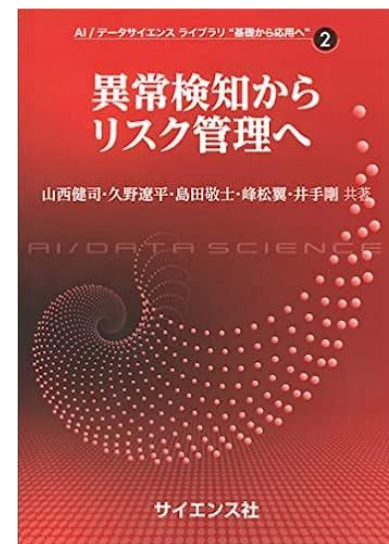
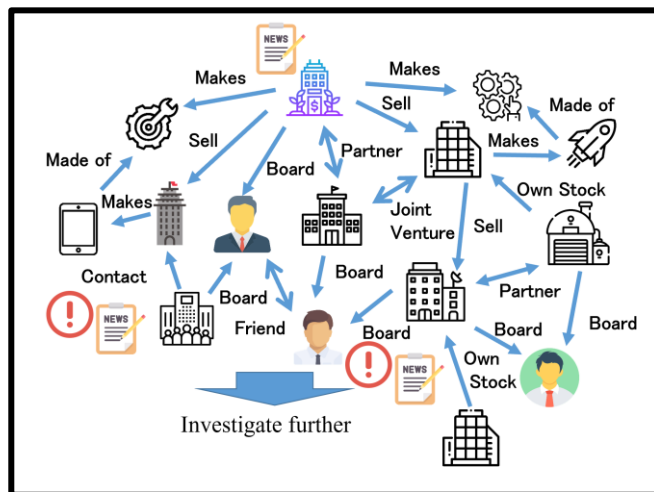
経済学・ファイナンス

- 経済学・ファイナンスでは従来利用されなかったようなデータ（オルタナティブデータ）の活用が盛んです
 - 下記以外にもPOS、GPS、衛星写真などが活用されています
- 日本銀行（[古川,箕浦,久野,八木2022]）ではGPSと交通量データを用いマクロ経済指標のナウキャストリングを試みています（左下）
 - 内閣府や経済産業省も同様の試み：データ駆動型の公共政策（EBPM）
- 複数のデータベースをネットワーク上に格納し探索することで、株価に影響を与えるニュースが報道されるリスクを予測する試み（真中）
 - 機械学習×ファイナンスは活発（例：[山西,久野,島田,峰松,井手2022]）

図6. ナウキャストリングモデルの推計結果（鉱工業）



(注1) 予測指数は、1週目～3週目時点では翌月予測指数、月末時点では当月予測指数を表す。
(注2) 図中の数字は、RMSEの値。



https://www.boj.or.jp/research/wps_rev/wps_2022/wp22j14.htm/

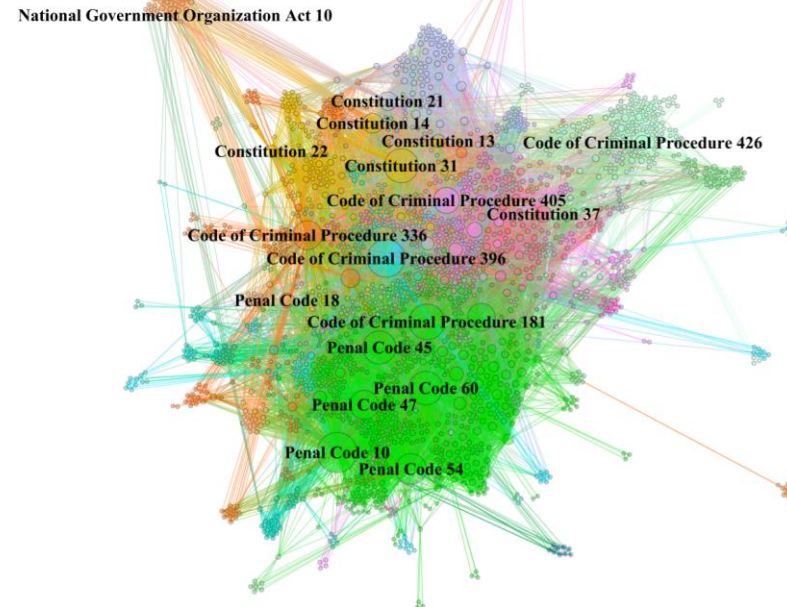
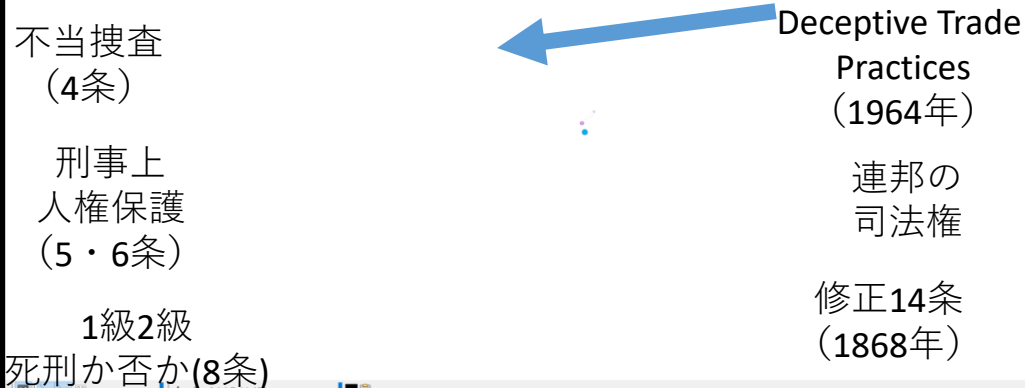
[Hisano et al.2020]

法律とデータサイエンス・A I

- 法体系の全体像をとらえることは非常に困難。テキストとネットワークマイニングの技術を用い法体系を数量的に分析する試みも
 - 左下：[FreeLaw]より著者作成
 - 右下：[Kondo,Yoshida,Hisano2022]より
- 契約書を締結の度に読み直すのは非常に大変です。そのため契約書の自動分析ツールを提供する企業もあります
 - eBrevia：契約書の自動分析

日本の法・判例構造

米国判例引用（18c～現代）



文学と音楽

- データサイエンスとAIの活用範囲は文学や音楽にも及びます
- 例えば古文書の崩し文字を読めるようになるには専門的な訓練を必要とします。そこで広く一般層にも古文書を読んでもらえるように機械学習を用い、崩し文字の「翻刻」（書き換え）に挑戦している研究者もいます[カラーヌワット2020]

(https://about.google/intl/ALL_jp/stories/tensorflow-ai-japanese-culture/?utm_source=google&utm_medium=hpp&utm_campaign=Japan)

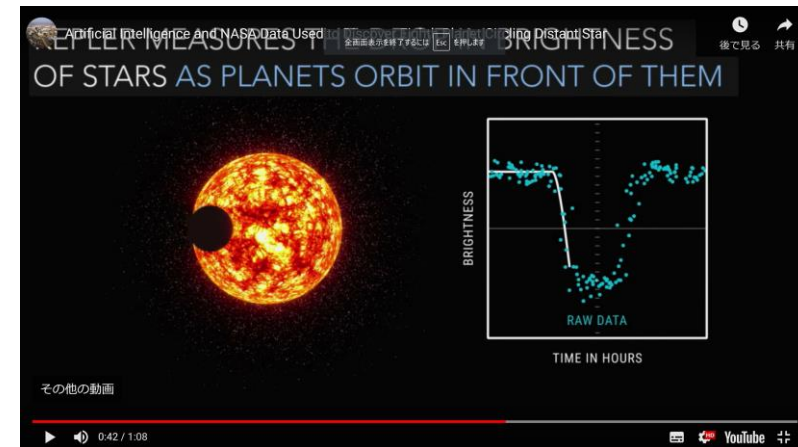
- 他にもGoogleは与えられたメロディに対してバッハ風の伴奏を自動生成するツールを公開し話題になりました

(<https://www.google.com/doodles/celebrating-johann-sebastian-bach>)

1-1-4 応用分野（自然科学）

物理学

- 物理学では探査機から送られてくる大量のデータや望遠鏡で撮影された画像をAIで分析することで新たな発見が生まれています
- 例えばGoogle AIの研究者はNASAのケプラー探査機から送られてくる膨大な輝度のデータを深層学習で分析することによって、太陽系外惑星を発見する試みを行っています
(<https://www.nasa.gov/press-release/artificial-intelligence-nasa-data-used-to-discover-eighth-planet-circling-distant-star>)
- 日本ではマルチスペクトル観測装置のMuSCATを駆使し同様の研究を行っています (<https://gendai.ismedia.jp/articles/-/71410>)
 - 地球と似た惑星が見つかる日もそう遠くないのかもしれませんが



Video credit: NASA/Ames Research Center

生物学

- 生物学では遺伝子データや画像分析技術などを用いて様々なデータサイエンスが行われています
- 日本のライフサイエンス統合データベースセンターでは、ライフサイエンス分野におけるデータベースの統合化に取り組んでいます
(<https://ds.rois.ac.jp/center1/>)
- DeepMindでは遺伝子データと深層学習を用いたタンパク質構造を予想する試みに取り組んでいます
 - AlphaFoldと呼ばれています[Senior et al.2020]
(<https://deepmind.com/blog/article/AlphaFold-Using-AI-for-scientific-discovery>)

化学

- 化学でもビッグデータの利用は盛んです
- 例えば化学反応予測の分野では、既に判明している化学反応のデータをネットワークとして整理し、未知の反応の予測をリンク予測問題として定義することで、新しい化学反応を探す試みが行われています[Segler,Waller2017]
 - 同研究ではおよそ1400万の分子を対象にしたネットワークが活用されました。ノードを分子や反応とし、エッジを化学反応における役割として定義しています。
- 日本の物質材料研究機構でも情報技術と化学実験を融合することで物質・材料の研究開発を革新する試みが行われています(<https://www.nims.go.jp/research/MaDIS/index.html>)

まとめ

- データサイエンスとは
 - データを有効活用し、数理モデリングや計算技術と適用ドメインの専門知識を結合することで新たな知識を生み出し、その活用のシナリオを導き出すこと
- なぜ注目を集めているか
 - 計算技術、ビッグデータ、情報通信技術などの進歩により急激にできることが広がったことがその背景にあります
- なぜ多くの人が学ぶ必要があるのか
 - 社会にとって望ましい技術と社会の融合を実現するには、一部の技術者だけでなく、多くの人を巻き込んだ対話の結果として、新しい社会を作る必要があります
 - そのためには多くの人が基本的な理解をもっている必要があります

1-1-5 参考文献

参考文献

[Aker] <https://www.akersolutions.com/>

[Brundage et al.2018] M. Brundage et al., "The Malicious Use of Artificial Intelligence : Forecasting, Prevention, and Mitigation", Feb 2018.

[Buchanan and Miller2017] Ben Buchanan and Taylor Miller, "Machine Learning for Policymakers", Cyber Security Project, Belfer Center, June 2017.

[BuzzFeed] <https://www.youtube.com/watch?v=cQ54GDm1eL0>

[Caltech2019] Biology and Big Data,
<https://magazine.caltech.edu/post/biology-and-big-data>

[Copernicus 1543] Nicolaus Copernicus, "De revolutionibus orbium coelestium"

[DeepMind]DeepMind, "DeepMind AI Reduces Google Data Centre Cooling Bill by 40%",<https://deepmind.com/blog/article/deepmind-ai-reduces-google-data-centre-cooling-bill-40>

[Domo] <https://www.domo.com/learn/data-never-sleeps-7>

[GoogleBach] <https://www.google.com/doodles/celebrating-johann-sebastian-bach>

[GoogleAutoML] <https://cloud.google.com/automl>

参考文献

- [Hisano et al.2020] Ryohei Hisano, Didier Sornette, Takayuki Mizuno ,
“Prediction of ESG compliance using a heterogeneous information network. J
Big Data 7, 22 (2020). <https://doi.org/10.1186/s40537-020-00295-9>.
- [Jung] Jung, A., “Playing Mario with Deep Reinforcement Learning”,
<https://github.com/aleju/mario-ai>
- [Kahneman2013] Kahneman, D, "Thinking, Fast and Slow", FSG Adult; 1st
edition, 2013.
- [Kepler1627] Kepler Johannes, “Tabulae Rudolphinae”, 1627.
- [KOMTRAX] <https://sanki.komatsu/komtrax/>
- [Kondo2022] Ryoma Kondo, Takahiro Yoshida, Ryohei Hisano, "Masked
prediction and interdependence network of the law using data from large-
scale Japanese court judgments", Artificial Intelligence and Law,
<https://doi.org/10.1007/s10506-022-09336-5>, In press, 2022.
- [Krakauer et al.2011] David C. Krakauer, Jessica C. Flack, Simon Dedeo,
Doyne Farmer and Daniel Rockmore, “Intelligent data analysis of intelligent
systems”, IDA 2010: Advances in Intelligent Data Analysis IX pp 8-17.
- [McKinsey2018] McKinsey Analytics, "Analytics comes of age", Jan 2018.
- [MicrosoftAutoML] [https://docs.microsoft.com/ja-jp/azure/machine-
learning/concept-automated-ml](https://docs.microsoft.com/ja-jp/azure/machine-learning/concept-automated-ml)

参考文献

- [MIT2019] Karen Hao, "AI thinks this flood photo is a toilet. Fixing that could improve disaster response. A new data set aims to teach computer vision systems to recognize images from disasters.",
<https://www.technologyreview.com/s/614246/ai-image-recognition-improves-disaster-response/>
- [NACCS] NACCS（輸出入・港湾関連情報処理センター株式会社）
<https://www.naccs.jp/aboutnaccs/aboutnaccs.html>
- [Nasa2017] <https://www.nasa.gov/press-release/artificial-intelligence-nasa-data-used-to-discover-eighth-planet-circling-distant-star>
- [OpenAI] <https://talktotransformer.com/>
- [OurWorldData] <https://ourworldindata.org/technological-progress>
- [ProPublica2016] by Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, "Machine Bias There's software used across the country to predict future criminals. And it's biased against blacks.", ProPublica, May 23, 2016.
- [Rahwan2017] Iyad Rahwan, "Society-in-the-Loop: Programming the Algorithmic Social Contract", Ethics and Information Technology, 2017.

参考文献

[Sakaji et al.2017] Hiroki Sakaji, Risa Muro, Hiroyuki Sakai, Jason Bennett, Kiyoshi Izumi, Discovery of Rare Causal Knowledge from Financial Statement Summaries, The 2017 IEEE Symposium on Computational Intelligence for Financial Engineering and Economics (CIFEr), pp.602-608, Hawaii, USA, November, 2017.

[Segler,Waller2017] Segler, M. H. S. and Waller, M. P., "Modelling Chemical Reasoning to Predict and Invent Reactions.", Chem. – A Eur. J. 2017, 23 (25), 6118–6128.

[Song et al.2018]Kyungwoo Song, Wonsung Lee, Il-Chul Moon, "Neural Ideal Point Estimation Network", Thirty-Second AAAI Conference on Artificial Intelligence, 2018.

[小川|2017] 小川 安英、「リクルートにおけるデータ活用」、第4回 日銀 FinTechフォーラム、2017年11月1日、
https://www.boj.or.jp/announcements/release_2017/rel171030a.htm/

参考文献

[カラーヌワット2020] 機械学習で日本の過去を紐解く文学研究者 さがそう。
TensorFlow のストーリー。

https://about.google/intl/ALL_jp/stories/tensorflow-ai-japanese-culture/?utm_source=google&utm_medium=hpp&utm_campaign=Japan

[河合2018] 河合祐子、「FinTech（フィンテック）－現状とこれから（金融高度化セミナー＜3月5日、大阪＞）」、2018年 3月14日、

https://www.boj.or.jp/announcements/release_2018/rel180314b.pdf

[高口2019] 高口太朗、「LINEアプリの機能改善を支える、データサイエンスチームの知られざる裏側」、LINE DEVELOPER DAY 2019,2019年11月20日～2019年11月21日、<https://logmi.jp/tech/articles/322207>

[内閣府] 内閣府地方創生推進事務局、「「スーパーシティ」構想について」
<https://www.kantei.go.jp/jp/singi/tiiki/kokusentoc/supercity/openlabo/supercitycontents.html>

[野村総合研究所2018] 「AI×言葉（合言葉）によるチャットボットの取り組み」第6回 FinTechフォーラム～チャットボットの進化と金融ビジネスへの活用～、2018年8月10日、
https://www.boj.or.jp/announcements/release_2018/rel180810d.htm/

参考文献

[古川2022] 古川角歩, 箕浦征郎, 久野遼平, 八木智之, 「オルタナティブデータと機械学習アプローチを用いた鋳工業生産ナウキャスティングモデルの構築」, 日本銀行ワーキングペーパーシリーズ 2022年7月

[毎日新聞2020年2月4日] スーパーシティ構想へ特区 政府法案再提出 AI、ビッグデータ活用、毎日新聞

[峰野] 峰野博史、多様な環境に自律順応できる水分ストレス高精度予測基盤技術の確立、

<https://www.jst.go.jp/kisoken/jyonetsu/interview/h29/mineno.html>

[山西2022] 山西, 健司, 久野, 遼平, 島田 敬士, 峰松 翼, 井手剛, 異常検知からリスク管理へ, サイエンス社 2022年8月 (ISBN: 9784781915463)

[山本2018] 山本 龍彦 (編集)、AIと憲法、2018/8/25

[ライフネット生命2018] 岩田慎一、「ライフネット生命のLINE活用術」、第6回 FinTechフォーラム～チャットボットの進化と金融ビジネスへの活用～、2018年8月10日、

https://www.boj.or.jp/announcements/release_2018/rel180810d.htm/

[渡辺2018] 特許第6395307号 (P6395307) 「物価指数推定装置、物価指数推定プログラム及び方法」登録日:2018/09/07. 出願人：国立大学法人東京大学. 発明者：渡辺努、渡辺広太.