

データサイエンスリテラシー(1)

「データサイエンスとは？」



イラスト：©いらすとや

Table of Contents

データサイエンスリテラシーとは？ (5分)

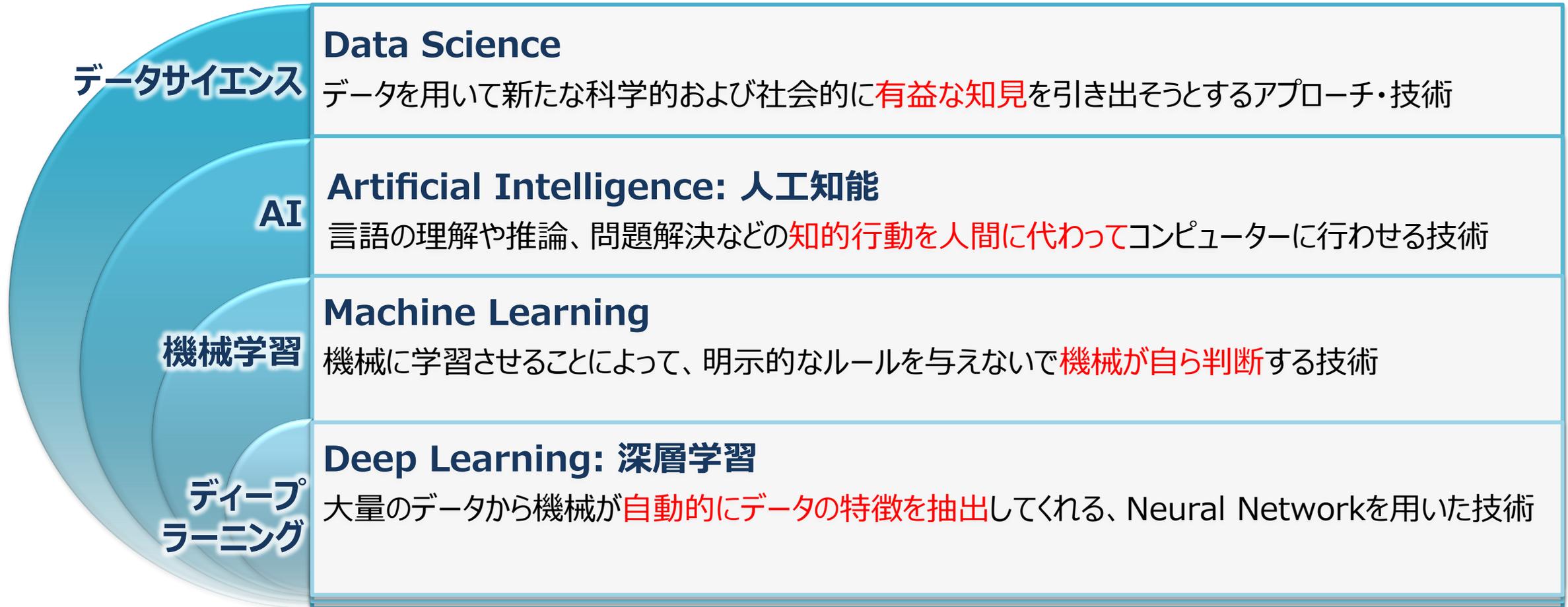
なぜデータサイエンスリテラシーを学ぶのか？ (20分)

データサイエンスリテラシーでは何を学ぶのか？ (30分)



そもそも「データサイエンス」って？

データサイエンス/AI/機械学習/ディープラーニング



※データサイエンスやAIの包含する内容は人や分野によって解釈が異なる場合があります。

「リテラシー」とは？

ある分野で用いられている「ことば」を理解・整理し、活用する能力

□元々は「書き言葉を、作法にかなったやりかたで、読んだり書いたりできる能力」

- 日本では「読み・書き・そろばん」→（ある分野で）知っているべき素養

□現代の様々なリテラシー

<https://ja.wikipedia.org/wiki/リテラシー>

- メディア・リテラシー（Media literacy）
- 金融・リテラシー（Financial literacy）
- 環境・リテラシー（Environmental literacy）

...

→「リテラシー」＝「読み書きのような最低限持っているべき素養、知識」

「データサイエンス・リテラシー」＝「データを用いて新たな科学的および社会に有益な知見を引き出す上で最低限持っているべき素養・知識」

Table of Contents

データサイエンスリテラシーとは？

(5分)

なぜデータサイエンスリテラシーを学ぶのか？

(20分)

データサイエンスリテラシーでは何を学ぶのか？

(30分)



Table of Contents

なぜデータサイエンスリテラシーを学ぶのか？

●社会で起きている変化

- 人手不足と長時間労働
- 社会課題の複雑化・高度化

●科学技術の進展

- IoTデバイス等によるデータ増大
- 計算機の処理性能の向上
- AIの非連続的進化

●データサイエンス・AIによる社会課題の解決



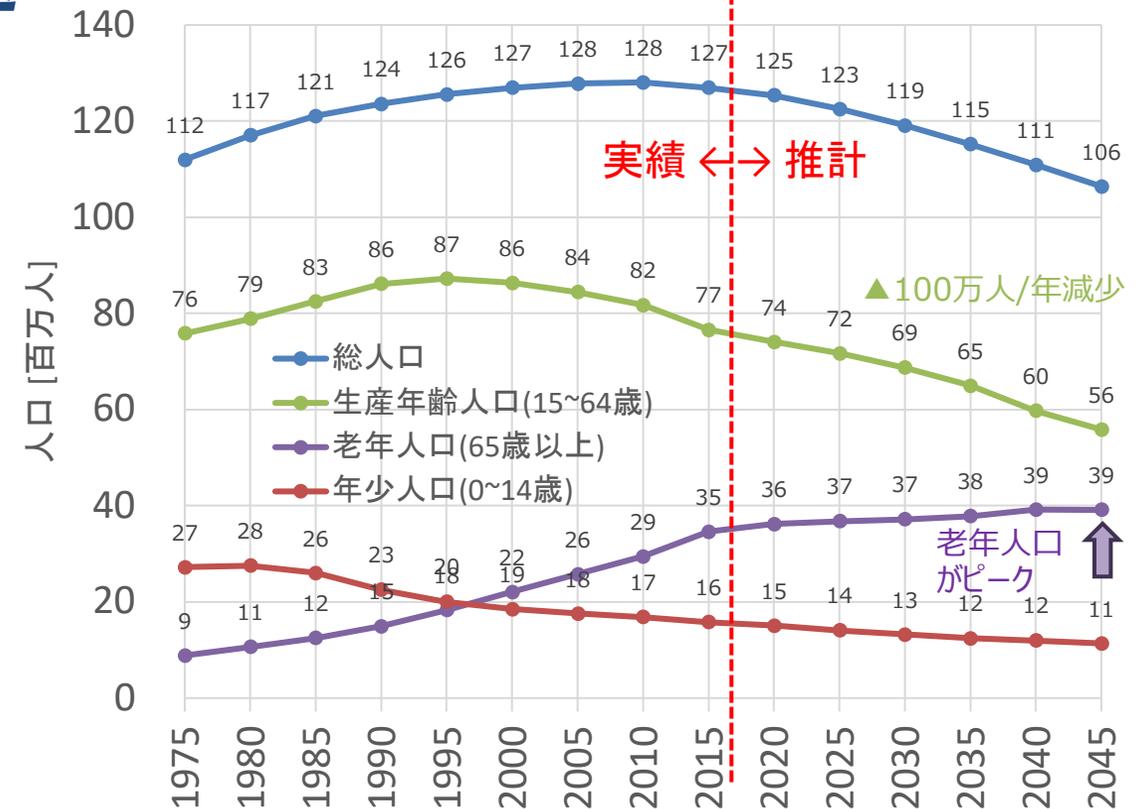
社会で起きている変化

人手不足と長時間労働

労働人口の減少・熟練者の引退

- 1人が複数の業務を管理・実施
 - 型通りの業務の自動化
- 熟練者のノウハウ引き継ぎ
 - ノウハウの文書化、数値データ化

総人口及び年齢3区分人口の推移



「年齢3区分別人口：実績値および推計値」（国立社会保障・人口問題研究所）

(http://www.ipss.go.jp/pp-zenkoku/j/zenkoku2017/db_zenkoku2017/g_tables/pp29gg0103data.htm)

を加工して作成。(©IPSS)

社会で起きている変化

人手不足と長時間労働

□ 時間あたり労働生産性の低さ

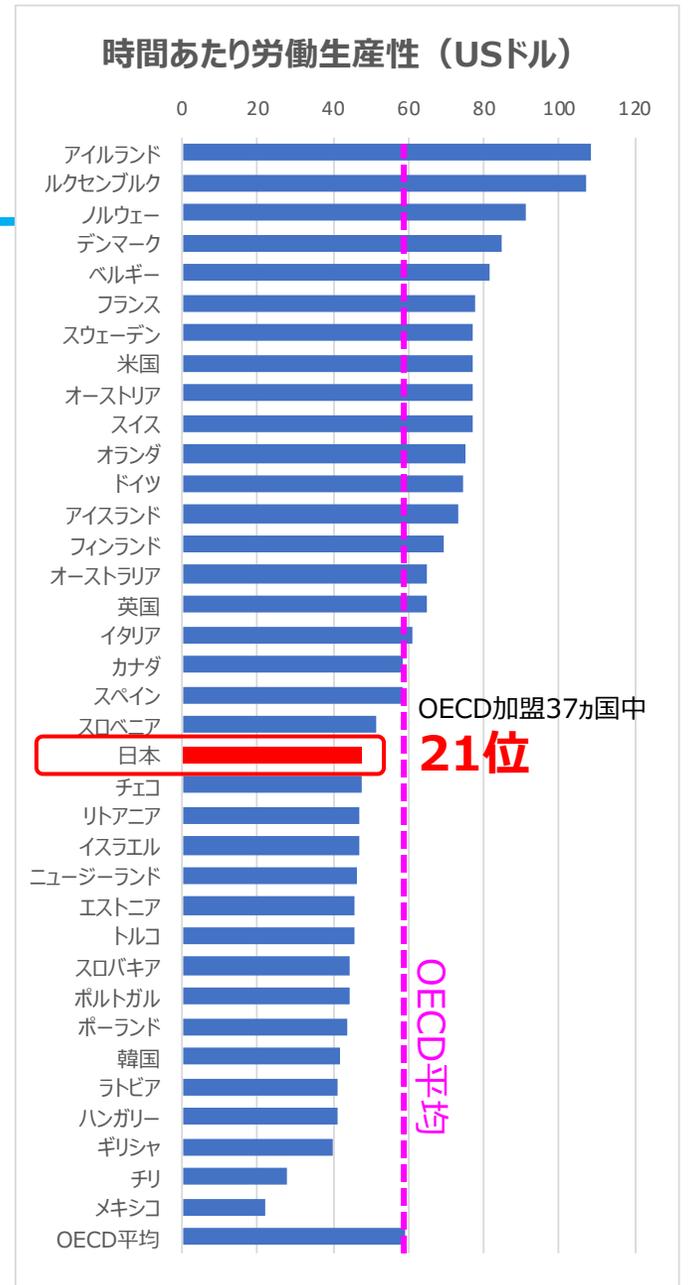
- 日本の時間あたり労働生産性は46.8ドルで、OECD加盟36カ国中21位

「労働生産性の国際比較 2020」、公益財団法人・日本生産性本部

- ロボット化・機械化によって、1人当たりの労働時間を削減し、長時間労働を是正
- 成果に影響のない業務や作業の適切な分担（最適化）によって働き方を改善



ヒトが体力的に困難だった単純作業を機械に置き換え



出典：「時間あたり労働生産性の国際比較」(公財)日本生産性本部・生産性データベース

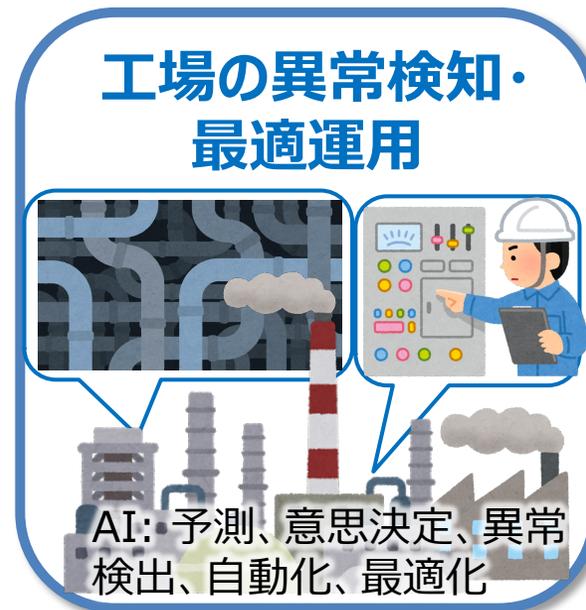
©JAPAN PRODUCTIVITY CENTER

社会で起きている変化

社会課題の複雑化・高度化

□ 複雑システムの最適運用

- ヒトには複雑に絡み合った複数の要因の関係は理解できない！
 - Ex. 違うメーカーが作った装置を組み合わせた製造装置の運用
 - Ex. コンビニ弁当の需給予測
(イベント、季節、天気、立地、etc.)



社会で起きている変化

社会課題の複雑化・高度化

□ 従来より高度な分析

- ちょっとした兆候や変化ではヒトには重要性が判断できない！
 - Ex. 会社のデータに普段アクセスしない人がアクセスしている
 - Ex. 車が通ったときの橋の振動が普段と違ってきた

ヒトが計算力・記憶力的に困難だった作業を機械に置き換え



Table of Contents

なぜデータサイエンスリテラシーを学ぶのか？

●社会で起きている変化

- 人手不足と長時間労働
- 社会課題の複雑化・高度化

●科学技術の進展

- IoTデバイス等によるデータ増大
- 計算機の処理性能の向上
- AIの非連続的進化

●データサイエンス・AIによる社会課題の解決



科学技術の進展

IoTによるデータの質と量の増大

□ センサの小型化、低電力化、高精度化

- データを直接、高頻度で計測・検知
→ 低コストで種々のモノからデータ収集可能に
- センサの組み合わせで別のデータに変換
- データを使わなくてもとりあえず取得

□ ネットワーク無線化、高速化、大容量化

- あらゆるデバイスがインターネットにつながる
 - IoT (Internet of things): モノのインターネット
- 世界のデータ通信量は月間40エクサバイト※
 - 広辞苑 1兆冊分データ: 重ねた厚さ 約 1億km

太陽には届かなかった!

これまで取得が困難だった時系列のバイタルデータや
個人のつぶやき、活動情報、各種機器の稼働状況など
多種多様なデータを大量・長期に取得可能

※1 エкса: 10^{18} 、Ericsson traffic measurements (Q4 2019)

※2 広辞苑は1ページは24字詰め50行、4段組。本文3181ページで、約1500万字、1MB~約50万字、厚さは8cm



例) スマートフォンの
様々なセンサー

- 加速度センサ
- 地磁気センサ
- ジャイロ
- 近接センサ
- 照度センサ
- 温湿度センサ
- 気圧計

スマートウォッチ

- 心拍数
- 血圧計
- 心電図
- 他

→ 行動記録、
睡眠・健康管理、転倒、緊急SOS



イラスト: ©いらすとや

科学技術の進展

計算機の処理性能の向上

□ Mooreの法則 (1965年)

- 集積回路上のトランジスタ数は「毎年2倍になる」→「2年で倍増する！」
(指数関数的な増加)

□ GPUによる演算の高速化

- グラフィック演算処理の高速化
 - 単純だが大量の演算処理性能に優れた専用チップ

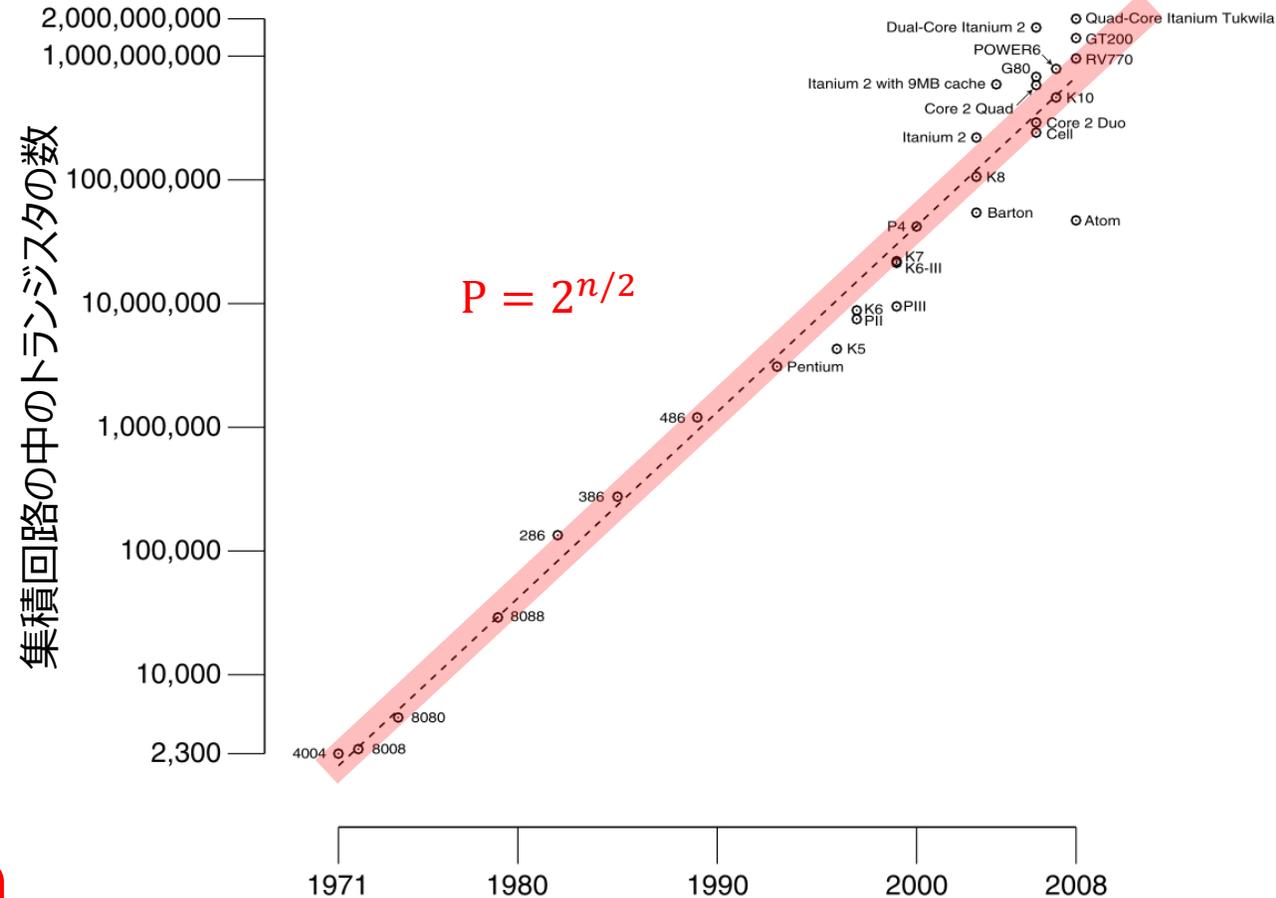
→ AIが用いる、単純だが大量の演算を高速化

□ 量子計算機の実用化

- 特定の最適化問題を高速化



半導体デバイス性能が急速に向上し大量データ処理を高速化



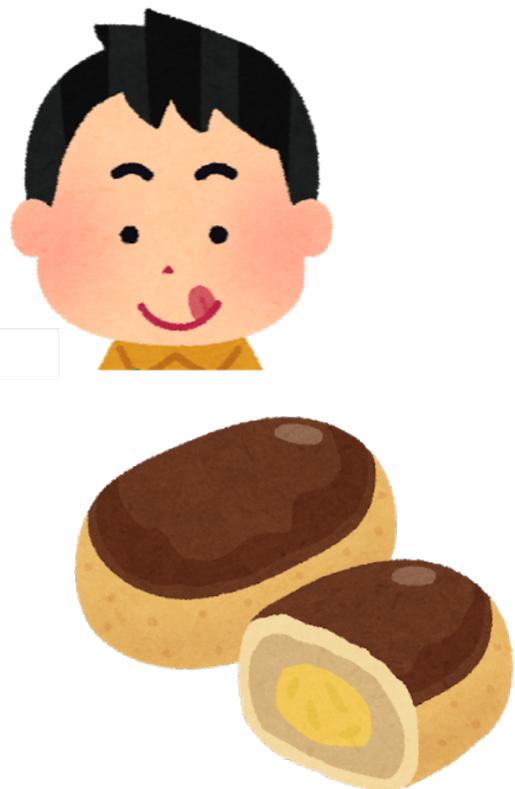
“CPU Transistor Counts 1971-2008 & Moor’s Law” ©Wgsimon (Licensed under CC BY-SA 3.0)

指数関数的な増加の例

CITY
UNIV.

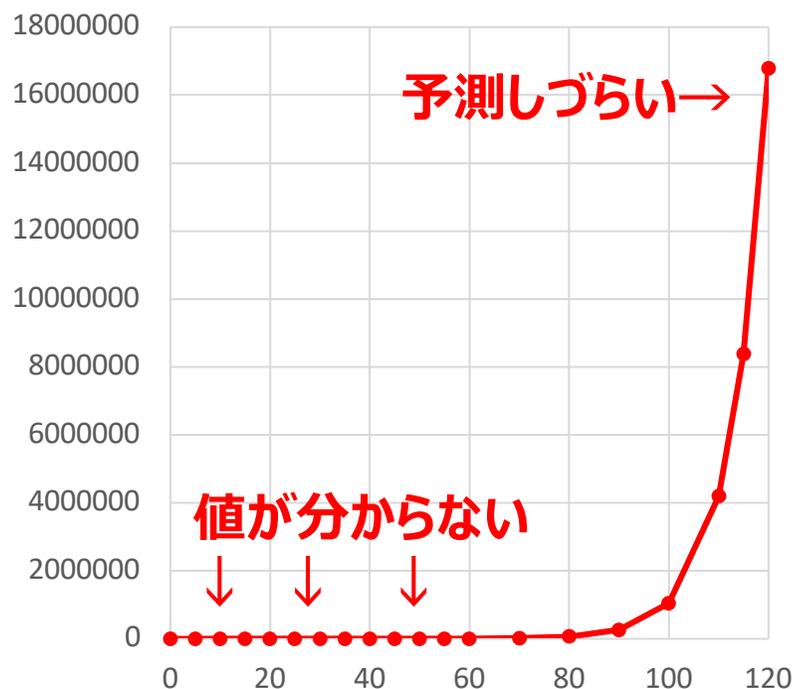
まんじゅうの数が5分ごとに2倍に増えるとする。

経過時間	0分	5分	10分	15分	20分	...	1時間	2時間
まんじゅうの数	1個	2個	4個	8個	16個	...	?	???

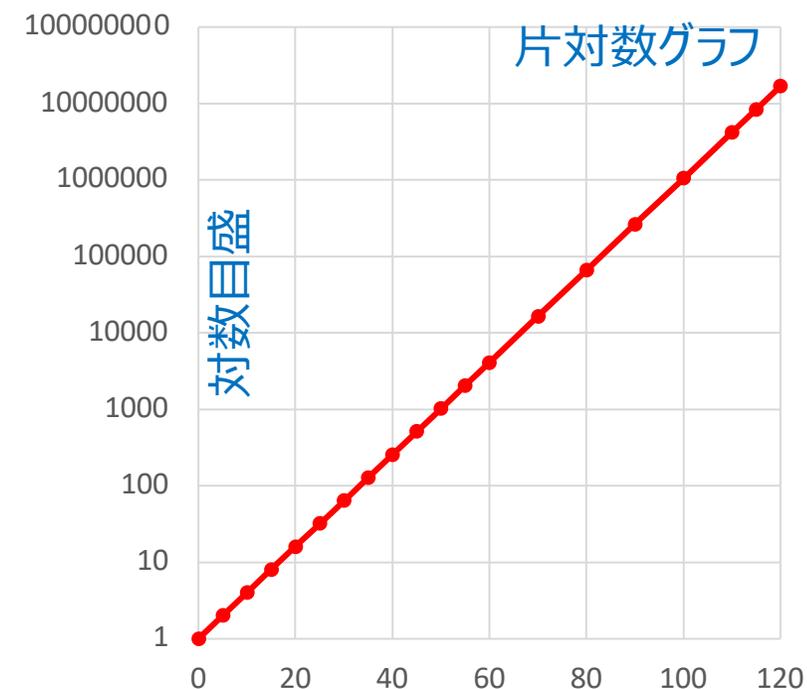


イラスト：©いらすとや

まんじゅうの数



まんじゅうの数



科学技術の進展

AI技術の飛躍的な進化

● 3回のAIブーム

2000年代(第3次AIブーム)

データからコンピュータが自動的に特徴量※を抽出.
データからAIが条件を作る
→「なぜその条件になったのか?」「条件の意味は?」がブラックボックス

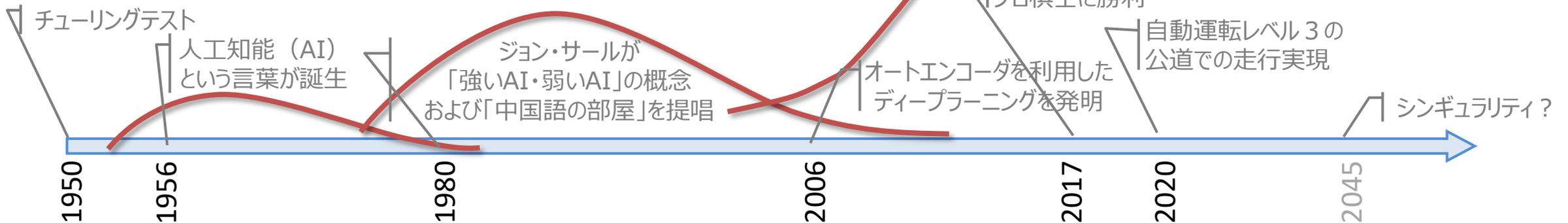
1980年代(第2次AIブーム)

専門分野の知識を取り込み、条件にあった専門的な答えを返す
→「人の常識」が機械には理解困難

1950年代後半(第1次AIブーム)

難しい迷路やパズルが解ける
→決められたルールの中で次の一手

※分析で重要な役割を果たすデータ・情報



科学技術の進展

AI技術の飛躍的な進化

□ IoTによるデータの質と量の増大と
計算機の処理能力の向上で、
ディープラーニングを中心としたAI
の分析精度が飛躍的に向上

□ プログラムのライブラリ構築によって
容易に独自のAIを作成可能

AIでよく用いられる計算プログラムをライブラリ化

<機械学習>

NumPy (ナンパイ), Pandas (パンドス), SciPy (サイパイ),
Matplotlib (マツプロットリブ), scikit-learn (サイキット・ラーン)

<ディープラーニング>

Pytorch (パイトーチ), TensorFlow (テンソルフロー), Keras
(ケラス), Chainer (チェイナー)

※ライブラリ…よく使ういくつかのプログラムをまとめたプログラム集
他のプログラムによく使う機能を提供するコードの集まり

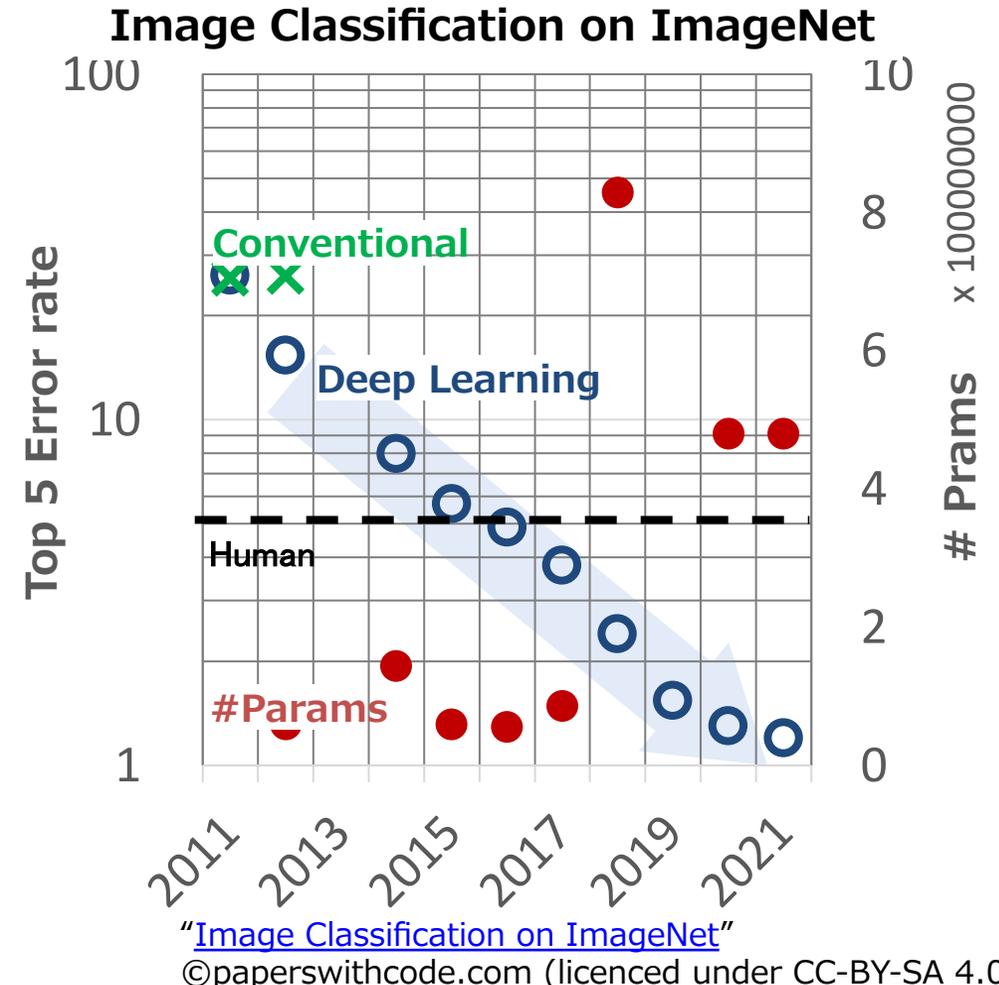


Table of Contents

なぜデータサイエンスリテラシーを学ぶのか？

●社会で起きている変化

- 人手不足と長時間労働
- 社会課題の複雑化・高度化

●科学技術の進展

- IoTデバイス等によるデータ増大
- 計算機の処理性能の向上
- AIの非連続的進化

●データサイエンス・AIによる社会課題の解決



データサイエンス・AIの活用

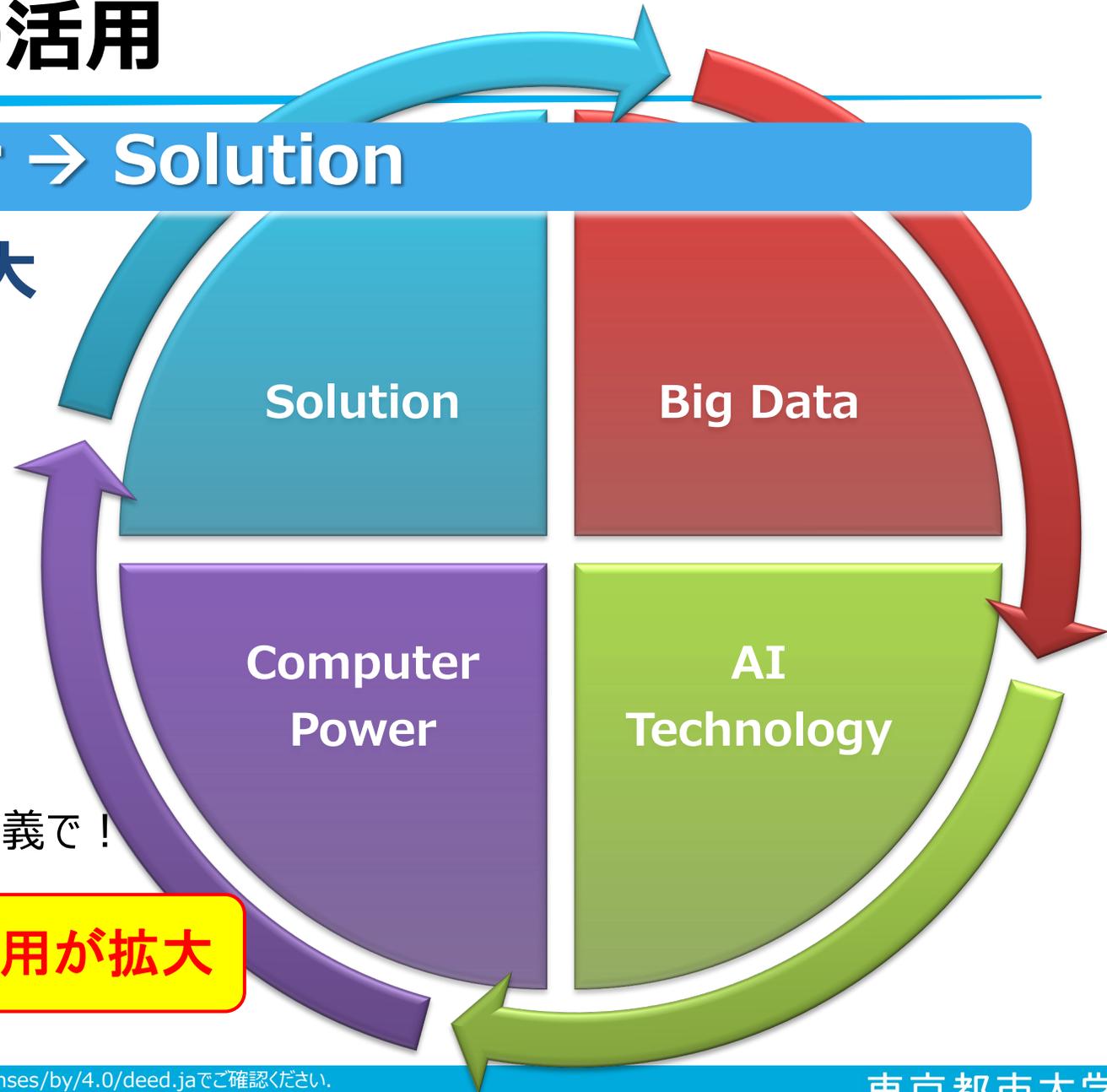
Big Data x AI x Computer → Solution

- IoTによるデータの質と量の増大
- X
- AI技術の飛躍的な進化
- X
- 計算機の処理性能の向上

社会が直面している様々な
重要課題を解決できる可能性！

→ 具体的事例は今後の講義で！

社会の至る所でデータサイエンス・AIの活用が拡大



データサイエンス・AIの活用

Society 5.0

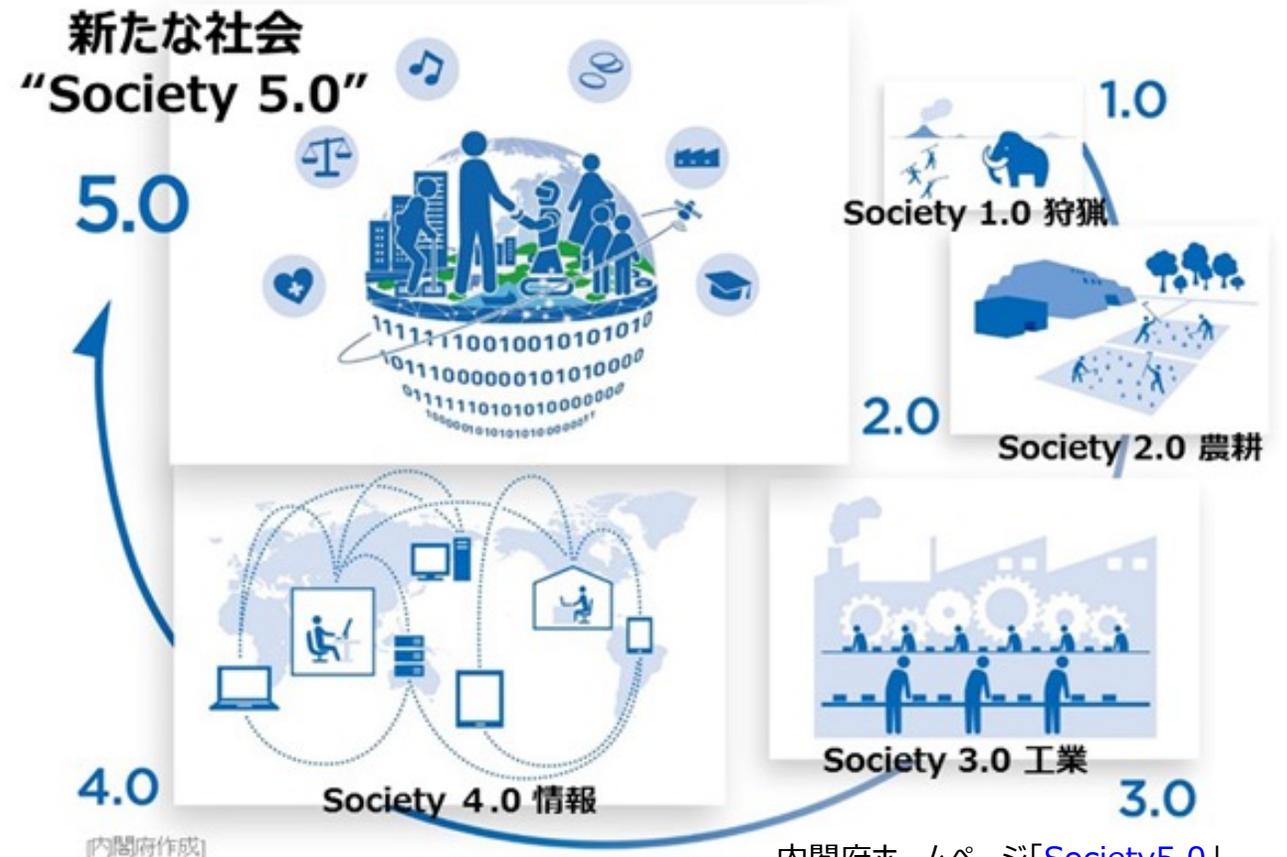
□サイバー空間とフィジカル空間
を高度に融合

→経済発展と社会的課題の解決
を両立する人間中心の社会

イマ
ココ

● Society 4.0 (情報社会):
知識や情報が共有されず、分野横
断的な連携が不十分

● Society 5.0:
IoTで全ての人とモノがつながること
で、様々な知識や情報が共有され、
今までにない新たな価値を生
み出す



内閣府ホームページ「[Society5.0](#)」
©Cabinet Office, Government of Japan

なぜデータサイエンスリテラシーを学ぶのか？

サイバー空間とフィジカル空間の高度な融合

フィジカル（現実）空間から**センサー**と**IoT**を通じてあらゆる情報が集積（**ビッグデータ**）
人工知能（AI）がビッグデータを解析し、高付加価値を**現実空間にフィードバック**

これまでの情報社会(4.0)

Society 5.0



[内閣府作成]

内閣府ホームページ「[Society5.0](#)」

©Cabinet Office, Government of Japan

なぜデータサイエンスリテラシーを学ぶのか？

経済発展と社会的課題の解決の両立

イノベーションで創出される**新たな価値**により、格差なくニーズに対応したモノやサービスを提供することで、**経済発展**と**社会的課題を解決**を両立



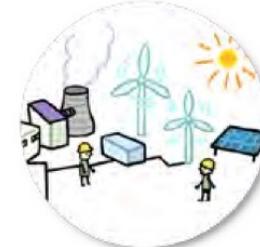
予防検診・ロボット介護



健康寿命延伸・社会コストの抑制



Society 5.0



エネルギーの多様化・地産地消



安定的確保、温室効果ガス排出削減



農作業の自動化・最適な配送



食料の増産・ロスの削減



最適なバリューチェーン・自動生産



持続可能な産業化の推進・人手不足解消

[内閣府作成]

内閣府ホームページ「[Society5.0](#)」

©Cabinet Office, Government of Japan

東京都市大学

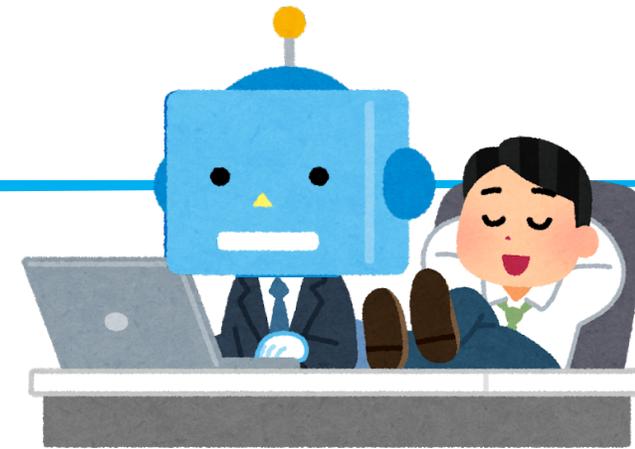
データサイエンス・AIの活用

AIのメリット

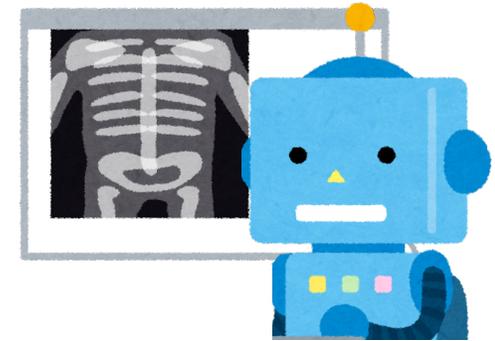
- **業務の自動化・効率化**
 - クリエイティブな仕事に専念
 - ➔ モチベーションの向上
 - ➔ 労働力不足の解消
 - **データの分析・予測**
 - 経験や勘ではなく、データに基づく判断
- ➔ **便利な生活が手に入る！**
- 様々な立場・分野のヒトが開発に参画し、新たな課題解決のアイデア提供

求められる資質

- **現場のエキスパートとして課題を明確化**
- **現場とAI専門家をつなぐヒト**



自由な働き方



病気の早期検知・診断



自動運転・渋滞緩和

イラスト：©いらすとや

データサイエンス・AIの活用

AIのデメリット

- **雇用が減る？**
 - AIは単純作業→労働時間削減
 - **責任の所在？**
 - 基本的にAIが行ったことに対する責任はAIの所持者に
 - **思考のプロセスが見えない？**
 - 深層学習以外ではむしろ明確
 - 自動化によりブラックボックス化しやすい
- **AIの誤用や誤った解釈！**
- 偏見や差別など悪影響の可能性

深層学習では
研究推進中

求められる資質

- **AIの適用範囲や精度の把握**
- **AIの悪用や悪影響についての評価**

□ AIで犯罪を予測

AIなら公平？！

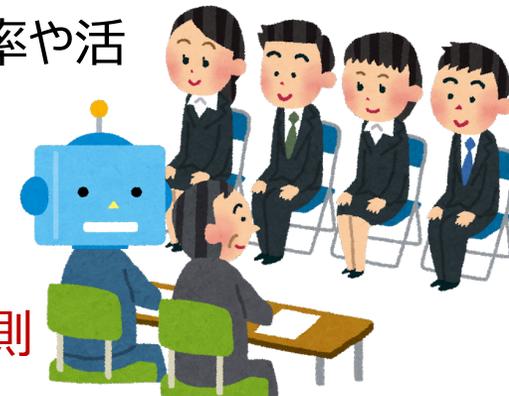
- 再犯予測プログラム
- ⇔ **誤認逮捕**
- 犯罪発生地域の予測
- ⇔ **差別や偏見の助長**



□ 人材採用システム

- 優秀な人材の発掘
- ⇔ **(深層学習では)理由が不明確**
- 採用候補者の退職確率や活躍度を予測
- ⇔ **学習データが偏見を含んでいたり、偏っていることでAIは偏見を持って予測**

イラスト：©いらすとや



データサイエンス・AIの活用

様々な分野・立場のヒトが課題解決に取り組むことでAIをフル活用
様々な分野・立場からAIの悪用・悪影響を検証、最小限に抑制

- 現場のエキスパートとして課題を明確化
- 現場とAI専門家をつなぐヒト
- AIの適用範囲や精度の把握
- AIの悪用や悪影響についての評価

AIが何を目指すべきか、
価値基準を与えるのはヒト

データサイエンスのリテラシーレベルの向上が必須

一人一人のヒトが中心となる社会

Table of Contents

データサイエンスリテラシーとは？ (5分)

なぜデータサイエンスリテラシーを学ぶのか？ (20分)

データサイエンスリテラシーでは何を学ぶのか？ (30分)



データサイエンスリテラシーでは何を学ぶのか？

データサイエンスリテラシーで学ぶこと

□データを起点としたものの見方

□実際の適用事例からAIのアイデア力を鍛える

□データ分析技術を体験



データサイエンスリテラシーで学ぶこと

□質問：600人が死亡すると予想される特殊なアジア病の流行に備えて、2つの対策が提案されている。

あなたならどちらを選択しますか？

- 対策Aを行うと、200人が助かる。
- 対策Bを行うと、1/3の確率で600人が助かるが、2/3の確率で誰も助からない。

データサイエンスリテラシーで学ぶこと

□質問：600人が死亡すると予想される特殊なアジア病の流行に備えて、2つの対策が提案されている。

あなたならどちらを選択しますか？

- 対策Cを行うと、400人が亡くなる。
- 対策Dを行うと、1/3の確率で誰も死なないが、2/3の確率で600人全員が亡くなる。

データサイエンスリテラシーで学ぶこと

□ **質問**：600人が死亡すると予想される特殊なアジア病の流行に備えて、2つの対策が提案されている。あなたならどちらを選択しますか？

- 対策Aを行うと、200人が助かる
- 対策Cを行うと、400人が亡くなる。
- 対策Bを行うと、 $\frac{1}{3}$ の確率で600人が助かるが、 $\frac{2}{3}$ の確率で誰も助からない。
- 対策Dを行うと、 $\frac{1}{3}$ の確率で誰も死なないが、 $\frac{2}{3}$ の確率で600人全員が亡くなる。

データサイエンスリテラシーで学ぶこと

□質問：600人が死亡すると予想される特殊なアジア病の流行に備えて、2つの対策が提案されている。あなたならどちらを選択しますか？

- 対策Aを行うと、200人が助かる
- 対策Cを行うと、400人が亡くなる。
- 対策Bを行うと、1/3の確率で600人が助かるが、2/3の確率で誰も助からない。
- 対策Dを行うと、1/3の確率で誰も死なないが、2/3の確率で600人全員が亡くなる。

フレーミング効果

切り取り方（フレーミング）や表現方法によって、印象が大きく変化する効果

添加物はダメ？
え、同じ!?

果汁90%以上



添加物10%以下



天気予報

- ・晴れる確率 50%
- ・雨が降る確率 50%

手術の成功率

生存率 90%
死亡率 10%

栄養ドリンク

タウリン 1g 配合
タウリン 1000mg 配合

イラスト：©いらすとや

データサイエンスリテラシーで学ぶこと

□質問：これは室町時代後期に作られた古信楽自然釉大壺です。現在の評価額は、20万円よりも高いと思いますか？ それとも安いと思いますか？



データサイエンスリテラシーで学ぶこと

□質問：これは室町時代後期に作られた古信楽自然釉大壺です。現在の評価額は、20万円よりも高いと思いますか？ それとも安いと思いますか？

古くて有名？！
20万円を基準に…



アンカリング効果

最初に印象的な情報を与えることにより、その後の意思決定に影響を及ぼす効果



連絡：1時間半遅刻します！
結果：1時間の遅刻

→遅刻したけど急いで来てくれた！

データサイエンスリテラシーで学ぶこと

□ **質問**：父親と息子が自動車事故に遭い、父はその場で死亡し、重傷を負った息子は急いで病院に運び込まれた。

手術室で外科医はその少年を見て言った。

「この少年を手術するのは（とても辛くて）できない。この少年は私の息子です！」

Q. この外科医の少年に対する関係は？

データサイエンスリテラシーで学ぶこと

□ **質問**：父親と息子が自動車事故に遭い、父はその場で死亡し、重傷を負った息子は急いで病院に運び込まれた。

手術室で外科医はその少年を見て言った。

「この少年を手術するのは（とても辛くて）できない。この少年は私の息子です！」

Q. この外科医の少年に対する関係は？

A. 外科医は少年の母親

実は同性婚を考えると父親でも正解ですね。



イラスト：©いらすとや

確認バイアス

思考が特定の確認に支配されると客観的な判断ができなくなる

□ 確認バイアスの強化例

「大手企業に入れば安心」

→ それを肯定する情報ばかり集める

→ 同様の考えを持つ同僚と仲良くなる

→ 中小企業やベンチャーの情報に触れない、避ける

→ 自分と違う生き方をしている人を批判する

データサイエンスリテラシーで学ぶこと



データを起点としたものの見方

- データから必要な情報を整理して抽出する、データに即した判断（遅い思考）
⇔ 直感や経験に基づいた判断（早い思考）
- 直感（早い思考）を磨くことも大切だが、データに即した検証・裏付けもできること



認知の偏り（認知バイアス）の例

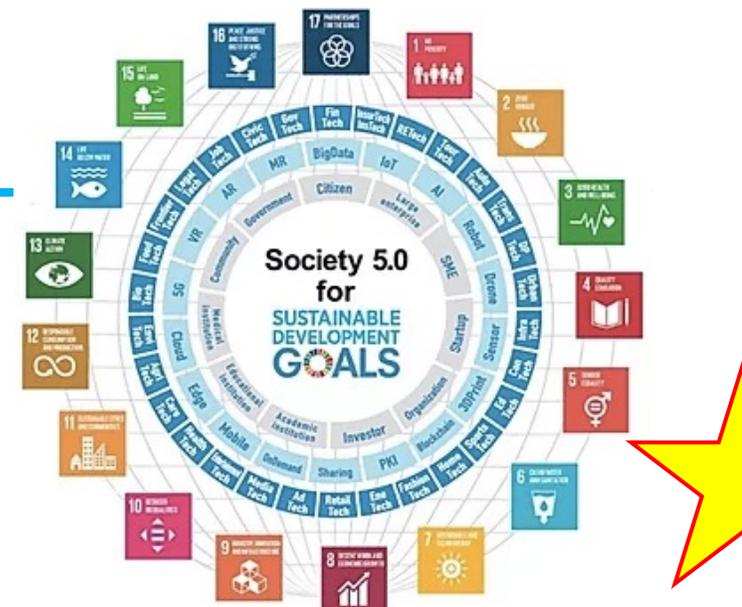
イラスト：©いらすとや

データサイエンスリテラシーでは何を学ぶのか？

データサイエンスリテラシーで学ぶこと

口実際の適用事例からAIのアイデア力を鍛える

- 様々なデータサイエンスの応用事例を紹介
 - ⇔ AIを過大評価も過小評価もしない



<https://www.keidanrendsgs.com/home-jp>
©1995-2020. [Keidanren](https://www.keidanrendsgs.com/home-jp)

口データ分析技術を体験

- ディープラーニングによるデータ分析を実施し、簡易に分析できることを体験する。
 - ソニーのNeural Network Consoleを用いたディープラーニングを利用
- <https://www.youtube.com/watch?reload=9&v=1AsLmVniy0k>

Neural Network Console

ドラッグ&ドロップで簡単ニューラルネットワーク設計

コーディング無しでディープラーニングを用いた高度なAI開発を実現

[サインイン](#)

Neural Network Console

<https://dl.sony.com/ja/>

到達目標

1. データサイエンスの様々な応用事例を学習し，データの活用のポイントを理解する。
2. データサイエンス活用における課題解決プロセスとそれぞれの基本的な技術の特徴を説明できるようにする。
3. データサイエンスの技術を体験し，実際の活用の足がかりを得る。