

アウトライン

- a. データサイエンス教育強化の背景と経緯
- b. 社会の動き：DX、IoT、インダストリー4.0
- c. ビッグデータと機械学習
- d. 帰納法の弱点と社会実装での課題
- e. 将来の社会像、教育の姿

アウトライン

- a. データサイエンス教育強化の背景と経緯
- b. 社会の動き：DX、IoT、インダストリー4.0
- c. ビッグデータと機械学習
- d. 帰納法の弱点と社会実装での課題
- e. 将来の社会像、教育の姿

平成32年と2020年

平成の30年間

最後まで触れますが、この30年間の、特に日本の存在感の減少(没落)を、30年前誰が予測できたでしょう？



(図のソース: 2017/12/29日経Webニュース)

平成元年 世界時価総額ランキング				平成30年 世界時価総額ランキング			
順位	企業名	時価総額	国名	順位	企業名	時価総額	国名
1	NTT	1,438.6	日本	1	アップル	9,499.5	米国
2	日本郵政銀行	715.9	日本	2	アマゾン・ドットコム	8,808.8	米国
3	住友銀行	695.9	日本	3	アルファベット	8,208.6	米国
4	富士銀行	670.8	日本	4	マイクロソフト	8,158.4	米国
5	第一勧業銀行	660.9	日本	5	フェイスブック	6,992.5	米国
6	日興	646.5	米国	6	イーベイ・イーベイ	4,920.9	米国
7	三菱銀行	592.7	日本	7	アリババグループ	4,795.8	中国
8	エコバンク	540.2	米国	8	ドットコム・グループ	4,557.3	米国
9	東京電力	534.8	日本	9	アマゾン・ドットコム	3,790.0	米国
10	ロイヤル・ダッチ・シェル	513.8	米国	10	エクソン・モービル	3,446.3	米国
11	トヨタ自動車	511.7	日本	11	シノコック・グループ	3,375.5	米国
12	GE	492.8	米国	12	ビザ	3,143.8	米国
13	三井銀行	492.9	日本	13	バンク・オブ・アメリカ	3,014.8	米国
14	野村證券	464.4	日本	14	イーベイ・イーベイ	2,899.7	米国
15	銀行本業債	414.8	日本	15	中国工商银行	2,879.7	中国
16	AT&T	381.2	米国	16	サムソン電子	2,803.8	韓国
17	日立製作所	358.2	日本	17	ウェルズ・ファーゴ	2,730.4	米国
18	松下電器	357.0	日本	18	ウォルマート	2,598.3	米国
19	フィリップ・モリス	321.4	米国	19	中国建設銀行	2,502.8	中国
20	東芝	309.1	日本	20	ネスレ	2,455.2	スイス
21	関西電力	308.9	日本	21	ユナイテッド・ヘルスグループ	2,431.0	米国
22	日本電信電話	308.5	日本	22	インテル	2,419.0	米国
23	東海銀行	302.4	日本	23	アライバークラウド	2,372.0	米国
24	三井物産	296.9	日本	24	サン・ペドロ	2,356.5	米国
25	メルク	275.2	米国	25	ホームデポ	2,335.4	米国
26	日産自動車	269.5	日本	26	ファイザー	2,163.6	米国
27	三菱重工	266.5	日本	27	アスター・ワールド	2,164.3	米国
28	デュポン	262.8	米国	28	イン・システムズ	2,091.6	米国
29	IBM	252.5	米国	29	ボーイング	2,045.6	米国
30	三菱東京UFJ銀行	246.7	日本	30	ロシュ・ヘルシング	2,014.0	スイス
31	BT	242.9	英国	31	アマゾン・ドットコム	2,013.2	米国
32	ベネッセ	241.7	日本	32	インテル	1,983.5	米国
33	BP	241.5	英国	33	P&G	1,978.5	米国
34	フォード・モーター	239.3	米国	34	システムズ	1,975.7	米国
35	アコム	237.3	日本	35	トヨタ自動車	1,936.8	日本
36	東京電力	224.8	日本	36	オクダ	1,933.3	米国
37	中国電力	219.7	日本	37	コカ・コーラ	1,925.8	米国
38	住友銀行	218.7	日本	38	ノバルティス	1,921.9	スイス
39	コカ・コーラ	215.0	米国	39	AT&T	1,911.9	米国
40	ウォルマート	214.9	米国	40	HSBCホールディングス	1,874.8	英国
41	三菱地所	214.5	日本	41	チャイナ・モバイル	1,786.7	中国
42	川崎重工業	213.8	日本	42	イーベイ・イーベイ	1,747.8	米国
43	モビル	211.5	日本	43	シティアーク	1,742.0	米国
44	東京ガス	211.3	日本	44	中国建設銀行	1,693.0	中国
45	東洋火災海上保険	209.1	日本	45	アマゾン	1,687.0	米国
46	NTK	201.5	日本	46	ウェット・サイエンス	1,661.6	米国
47	アルコ	196.3	日本	47	ペプシコ	1,641.5	米国
48	日本郵政	194.1	日本	48	中国建設銀行	1,637.7	中国
49	大塚製粉	191.1	日本	49	トヨタ	1,611.3	日本
50	協和	185.2	日本	50	ネット・フランク	1,577.2	米国

ものづくり企業から ITメガ企業へ

Top of market capitalization of stock price

2007, May		2017, May		10年間と比較しての倍率	2020年現在 Modern Seven Sisters
469	Exxon Mobil	1	Apple		
	GE	2	Alphabet (Google)	×7.6	
	Microsoft	3	Microsoft	×4.3	
	Citi Group	4	amazon	×1.8	
	Petro China	5	facebook	×16.8	
	AT&T	6	Berkshire Hathaway	---	
	Royal Dutch Shell	7	Johnson & Johnson	---	
	Bank of America	8	Exxon Mobil	×0.7	
	Industrial and Commercial Bank of China	9	Tencent	×42.5	
216	トヨタ	10	Alibaba	297	

billion US\$

Original News Source: NikkeiWeb2017/June/02

日本における統計教育体制の貧弱な歴史

1944年6月

- 統計数理研究所(以後、統数研と略す)が設立 (※統数研は独自の文化を持ち、従前から、日本の数学、統計、情報コミュニティとも一定の距離をとっていた。日本の統計コミュニティの中で、モデル選択、時系列解析、計算集約型の統計手法の採用、広範な応用分野の開拓、ベイズ統計の発展、機械学習の取り込み、データ同化等、海外に誇る独特の存在感を示す)

1983

- 日本学術会議が、複数の大学において大学院統計学専攻を新たに設置することを勧告

1988

- 総合研究大学院大学(以後 総研大)設立。日本で初めての統計学に関する専攻として、統計科学専攻(博士課程後期のみ)がたちあがる

2013

小森先生も

- 総研大統計科学専攻の博士号取得者が100名を超える。総研大高畑尚之学長をお招きし式典を開催。

2019

- 公的統計調査の不正が発覚。2020年度に「統計データアナリスト」の資格をつくり、5年後をめどに各省庁に同資格を持つ職員の配置を義務付け

日本におけるデータサイエンス教育体制の充実(1)

2011年4月~2019年3月、私は統計数理研究所の所長および情報・システム研究機構の理事

ビッグデータブームと提言

2011

- 統計数理研究所が、所内に人材育成・養成を目的とした組織、「統計思考院」を設立
- 日本統計学会が主体となり、統計検定の事業開始



2012

- 樋口知之「ビッグ・データを操る者が勝つ」DIAMONDハーバード・ビジネス・レビュー
- オバマ政権「ビッグデータ研究開発イニシアチブ」開始

2013

- ✓ 深層学習ImageNetが、物体認識コンペで圧倒的な性能を示す これが原点
- 樋口知之「データ・サイエンティストがビッグデータで私たちの未来を創る」『情報管理』
- 一般社団法人データサイエンティスト協会(以後、DS協会)を正式スタート (NHKニュースで何度も放映)



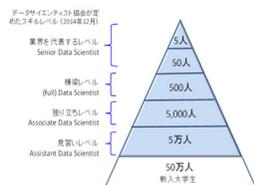
2014

- 日本学術会議 数理統計学分会 提言「ビッグデータ時代における統計科学教育・研究の推進について」
- 日本学術会議 情報学委員会 E-サイエンス・データ中心科学分会(委員長 北川、幹事 樋口) 提言「ビッグデータ時代に対応する人材の育成」
- DS協会、データサイエンティストのミッション、スキルセット、定義、スキルレベルを発表

2015

- 産総研AIRC設立
- 情報・システム研究機構からの提言「ビッグデータ活用のための専門人材育成について」

2015年 情報・システム研究機構からの提言



ビッグデータの利活用のための専門人材育成について



ビッグデータの利活用に係る専門人材育成に向けた産学官懇談会出席者

座長	北川 源四郎	情報・システム研究機構 機構長
委員	安宅 和人	ヤフーCSO、データサイエンティスト協会 理事
	榎本 剛	文部科学省 研究振興局 参事官(情報担当)
	岡本 青史	富士通研究所
	北山 浩士	文部科学省 高等教育局 専門教育課 課長
	佐藤 俊哉	京都大学医学研究科 教授
	長谷川 眞理子	総合研究大学院大学 理事・副学長(教育担当)
	樋口 知之	統計数理研究所 所長
	丸山 宏	統計数理研究所 教授/データサイエンティスト育成ネットワーク事業 実施担当責任者
	丸山 文宏	富士通研究所
	渡辺 美智子	慶應義塾大学健康マネジメント研究科 教授/総研大 統計センター 理事

陪席	栗辻 康博	文部科学省 研究振興局 教学イノベーションユニット次長 基礎研究振興課 融合領域研究推進官
	金井 学	文部科学省 高等教育局 専門教育課 情報教育推進係長
	栗原 潔	文部科学省 研究振興局 参事官(情報担当)付専門官
	土生木 茂雄	文部科学省 高等教育局 専門教育課 視学官
	山路 尚武	文部科学省 高等教育局 専門教育課 課長補佐

平成 27 年 7 月 30 日
大学共同利用機関法人 情報・システム研究機構
ビッグデータの利活用に係る専門人材育成に向けた産学官懇談会

日本におけるデータサイエンス教育体制の充実(2)

2016 AIブームとデータサイエンス(系)学部新設

- 理研AIP設立
- 数理及びデータサイエンスに係る教育強化の拠点校に6大学(北大、東大、滋賀大、京大、阪大、九大)を選定

2017

- 滋賀大学データサイエンス学部設立
- 『日本ティーラーニング協会』設立

2018

- 横浜国立大学データサイエンス学部設立

2019

- 文科省が「大学における数理・データサイエンス教育の全国展開」の協力校として20大学を選定
- 武蔵野大学データサイエンス学部設立
- 全国の国立、私立大学にAI、データサイエンスの研究あるいは教育を主活動とするセンターが多数設置

2020

- 数理・データサイエンス教育強化拠点コンソーシアムが数理・データサイエンス・AIのリテラシーレベルモデルカリキュラムを公開(分野を問わず、全ての大学・高専生(約50万人卒/年)を対象)

2021

- リテラシーレベルの認定制度開始
- DS協会がDS検定を開始

数理・データサイエンス教育強化拠点コンソーシアム

2016年度

- ・数理及びデータサイエンスに係る教育強化の拠点校 6校選定 (北海道大学、東京大学、滋賀大学、京都大学、大阪大学、九州大学)
- ・東京大学を幹事校として取り組み開始

2019年度

- ・国立大学20校が協力校として参加。全国6ブロックに分担し活動

2020年度

- ・協力校、特定分野協力校を加え、公私立大学 (95大学、2短大 2021年9月現在)、国立高専機構を連携校として加わる

活動内容

- ・全国的なモデルとなる標準カリキュラム・教材の作成
- ・その標準カリキュラム・教材の他大学への普及方策の検討及び実施
- ・情報交換等を行うための対話の場の設定
- ・取組の成果指標や評価方法の検討



数理・データサイエンス教育強化拠点コンソーシアム 拠点/協力/連携校



数理・データサイエンス教育強化拠点コンソーシアムNEWSLETTER vol.12 p.15より転載
<http://www.mi.u-tokyo.ac.jp/consortium/pdf/newsLetter12.pdf>

数理・データサイエンス教育強化拠点コンソーシアム モデルカリキュラム

◆リテラシーレベル(2020年4月)

学修目標

今後のデジタル社会において、数理・データサイエンス・AIを**日常生活、仕事等の場で使いこなすことができる基礎的素養**を主体的に身に付けること。そして、学修した数理・データサイエンス・AIに関する知識・技能をもとに、これらを扱う際には、**人間中心の適切な判断**ができ、**不安なく自らの意志でAI等の恩恵を享受し、これらを説明し、活用**できるようになること。

数理・データサイエンス・AI (リテラシーレベル) モデルカリキュラム
～データ思考の画巻～

導入	1. 社会におけるデータ・AI利活用 1-1. 社会で活用されているデータ 1-2. データとAIの活用領域 1-3. データ・AI利活用の最新動向	1-4. 社会で活用されているデータ 1-5. データとAIの活用領域 1-6. データ・AI利活用の最新動向
基礎	2. データリテラシー 2-1. データを扱う 2-2. データを説明する	2-2. データを説明する
心得	3. データ・AI利活用における倫理事項 3-1. データ・AI利活用と倫理事項	3-2. データを扱う上での倫理事項
選択	4. オプション 4-1. 統計学とデータサイエンス 4-2. データサイエンス 4-3. データサイエンス 4-4. データサイエンス 4-5. データサイエンス	4-3. アルゴリズム基礎 4-4. 機械学習基礎 4-5. データサイエンス 4-6. データサイエンス 4-7. データサイエンス 4-8. データサイエンス (新設カリキュラム)

http://www.mi.u-tokyo.ac.jp/consortium/model_literacy.html
 数理・データサイエンス・AI (リテラシーレベル) モデルカリキュラムより抜粋

◆応用基礎レベル (2021年3月)

学修目標

数理・データサイエンス・AI教育 (リテラシーレベル) の教育を補完的・発展的に学び、**データから意味を抽出し、現場にフィードバックする能力、AIを活用し課題解決につなげる基礎能力**を修得すること。そして、**自らの専門分野に数理・データサイエンス・AIを応用するための大局的な視点を獲得**すること。

数理・データサイエンス・AI (応用基礎レベル) モデルカリキュラム ～AI×データ活用の実践～			
3. AI基礎		3. AIの歴史と応用分野 (5)	
3-2. AIと社会 (5)	3-3. 機械学習の基礎と発展 (5)	3-4. 深層学習の基礎と発展 (5)	3-5. 言語・知識
3-6. 予測・判断	3-7. 言語・知識	3-8. 身体・運動	3-9. AIの倫理と責任 (5)
1. データサイエンス基礎		2. データエンジニアリング基礎	
1-1. データサイエンス基礎 (5)	1-2. データサイエンス基礎 (5)	2-1. ビッグデータとデータエンジニアリング (5)	2-2. データサイエンス基礎 (5)
1-3. データサイエンス基礎 (5)	1-4. データサイエンス基礎 (5)	2-3. データサイエンス基礎 (5)	2-4. データサイエンス基礎 (5)
1-5. データサイエンス基礎 (5)	1-6. データサイエンス基礎 (5)	2-5. データサイエンス基礎 (5)	2-6. データサイエンス基礎 (5)
1-7. データサイエンス基礎 (5)	1-8. データサイエンス基礎 (5)	2-7. データサイエンス基礎 (5)	2-8. データサイエンス基礎 (5)

http://www.mi.u-tokyo.ac.jp/consortium/model_ouyoukiso.html
 数理・データサイエンス・AI (応用基礎レベル) モデルカリキュラムより抜粋

認定制度について

- ・正式名称: 「数理・データサイエンス・AI教育プログラム認定制度」
- ・(リテラシーレベル) と (応用基礎レベル) がある。
- ・内閣府は認定制度設計を検討
- ・認定は**文部科学省** (高等教育局**専門教育課** 情報教育推進係) が担当
- ・MDASH※と呼ばれ「認定教育プログラム」と、より優れたプログラムに送られる「認定教育プログラム プラス」の2種類ある。
- ・リテラシーレベルは2020年公募開始 (2021年度採択)、2021年8月現在78件 (プラス11件) が採択。
- ・応用基礎レベルは、2020年度に内閣府検討が完了。2021年度公募のスケジュール

リテラシー認定のマーク



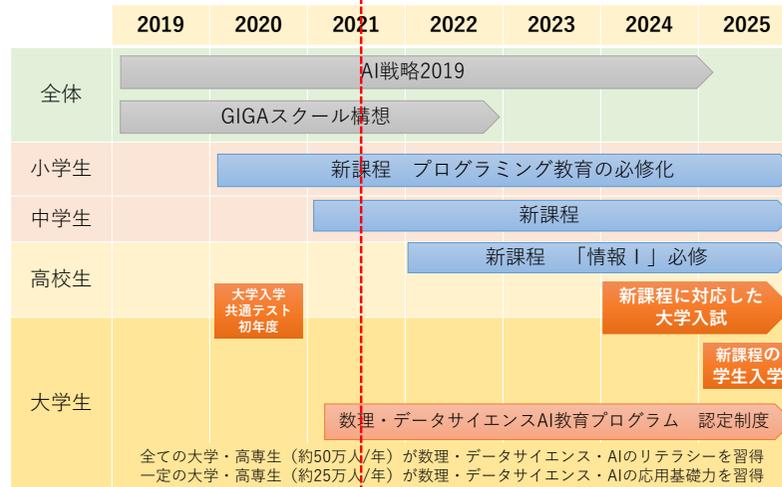
リテラシー認定状況 (2021年8月採択まで)

	リテラシー	プラス	
大学	国立	30	6
	公立	3	1
	私立	33	2
	小計	66	9
短期大学	小計	2	0
高等専門学校	小計	10	1
合計	78	10	

年度	2019	2020	2021	2022	2023以降
リテラシー	制度検討	公募	認定 (期間: 初回5年更新後3年)		
応用基礎		制度検討	公募 (予定)	認定 (予定)	

※Approved Program for Mathematics, Data science and AI Smart Higher Education

2025年度: デジタル・ネイティブ, AI・ネイティブ世代が 大学進学



- 小学校**
2020年度
プログラミング教育
必修化
- 高等学校**
2022年度
「情報I」 (1年生から)
必修化
- 大学入試**
2024年度末の入試
「情報I」
入試科目に

全ての大学・高専生 (約50万人/年) が数理・データサイエンス・AIのリテラシーを習得
 一定の大学・高専生 (約25万人/年) が数理・データサイエンス・AIの応用基礎力を習得

新しい価値観：検索＋スマホ世代が変える社会

知識がモノの形で伝達→知識が脳内で構造化

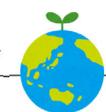
知識がフラット化



価値観の転換

- ・シェアリング
- ・エコシステム

アントレプレナーブーム

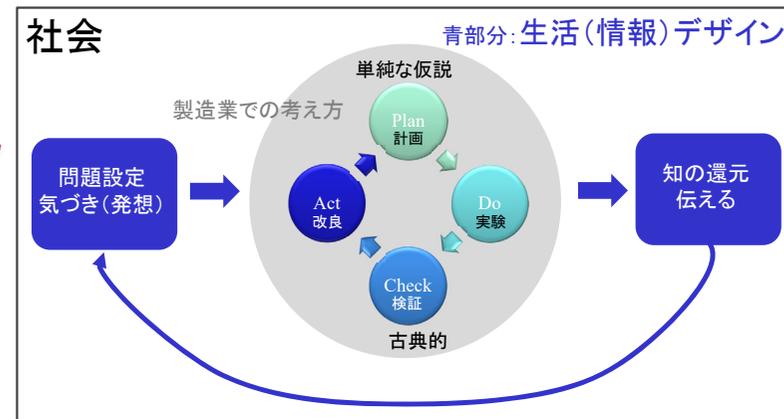


UBER
airbnb
スペース共有 ■ 空いたオフィススペース ■ 使われてない時間帯の厨房
人的リソース共有 ■ ベビーシッター ■ ボディガード

17

PDCAサイクルだけでは不足・時代遅れ

PDCAサイクルに加えてこれからわれわれはどうすべきか？



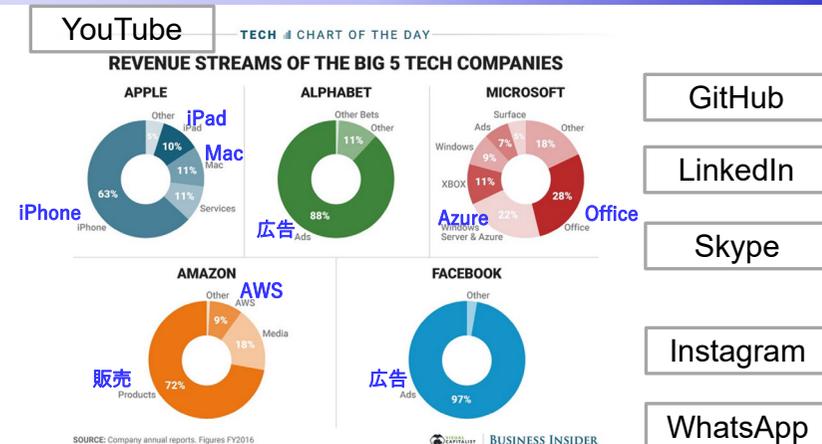
18

アウトライン

- データサイエンス教育強化の背景と経緯
- 社会の動き：DX、IoT、インダストリー4.0
- ビッグデータと機械学習
- 帰納法の弱点と社会実装での課題
- 将来の社会像、教育の姿

19

IT主要5大企業の収益源 (2016年)



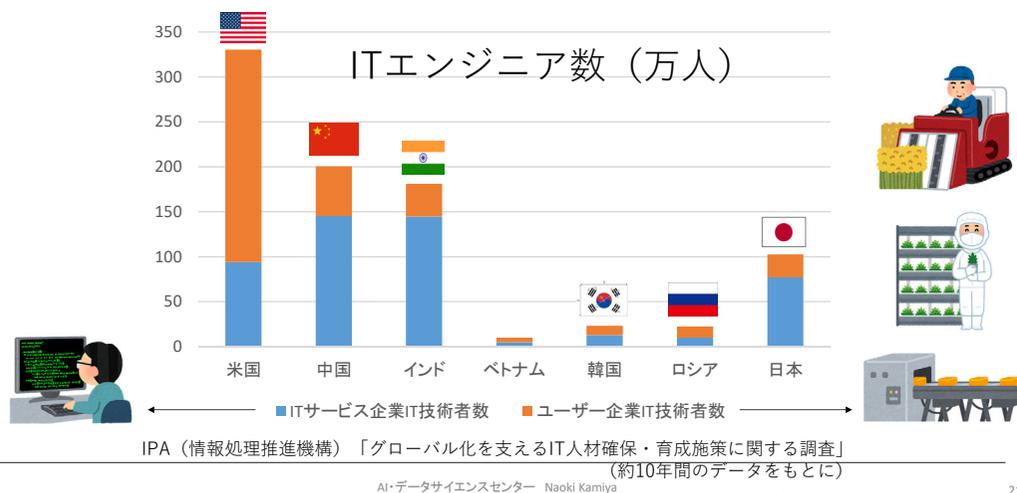
SOURCE: Company annual reports. Figures FY2016

5大IT企業の売り上げ構成。マイクロソフトの内訳の多様性が際立つ。
Mike Nudelman/Business Insider/Visual Capitalist

(ソース <https://www.businessinsider.jp/post-33925>
<http://www.visualcapitalist.com/chart-5-tech-giants-make-billions/>)

20

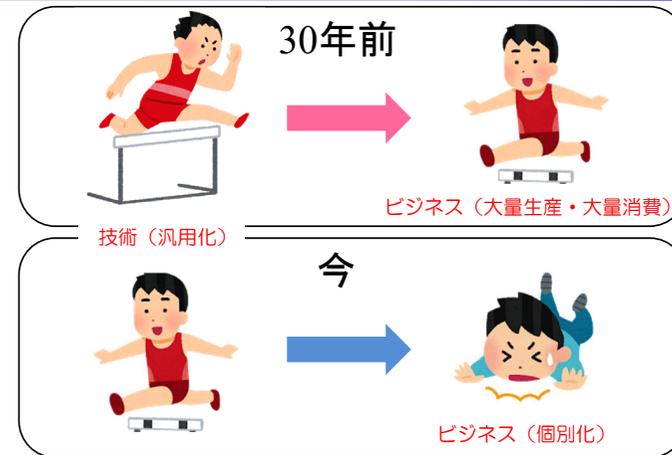
人材数の格差：ベンダー or ユーザー企業



21

Invention vs. Innovation

技術のハードル vs. 顧客のハードル



22

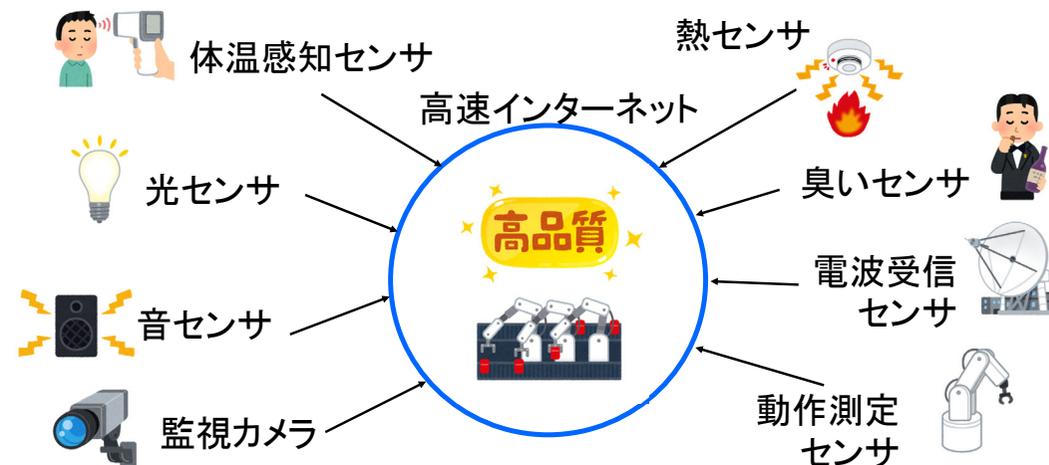
DX : Digital Transformation

「AI や IoT などの先端的なデジタル技術の活用を通じて、デジタル化が進む 高度な将来市場においても **新たな付加価値** を生み出せるよう従来の **ビジネスや組織を変革** すること」
(情報処理推進機構 (IPA))

「企業がビジネス環境の激しい変化に対応し、データとデジタル技術を活用して、**顧客や社会のニーズ**を基に、製品やサービス、ビジネスモデルを変革するとともに、業務そのものや、**組織、プロセス、企業文化・風土**を変革し、競争上の優位性を確立すること」(2018年12月、経済産業省)

23

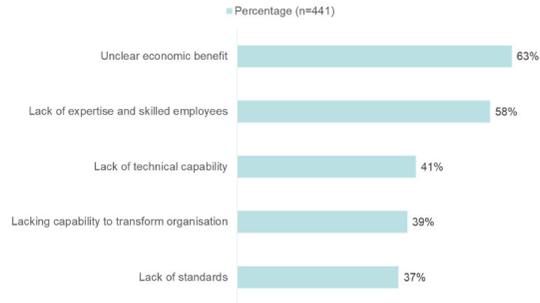
IoT: Internet of Things



24

モノづくり： インダストリー4.0の進展を阻む障害

Obstacles to Industry 4.0 implementation



Fraunhofer IAO (2014) 2014年調査
 2019年10月5日 第15回 早稲田大学総合研究機構研究成果報告会
 シュレーダー准教授@九州大学の発表資料

経済効果が不明確
 専門知識と熟練した
 従業員の不足
 技術的能力の欠如
 組織を変革する
 能力の欠如
 規格・標準の欠如

モノづくり： インダストリー4.0の実像

■標準化

- そもそも標準がない
- 標準インターフェースがあったとしても、自分らのインターフェースを使用したがる
- 標準化と参照アーキテクチャーは企業内だけ。あるいは、ごく一部の企業の間だけ
- 部品企業はさまざまなインテグレーションは難しい。形式と粒度の違うデータを受け取るだけで疲弊

■インセンティブ

- 大企業は実現したいが、一方、小さい企業は保守的
 (小企業： 経済的メリットが分からない。あらゆることを変更したくない)
- 「インダストリー4.0」への投資の意味がわからない

■インフラ

- 通信基盤がしっかりと整っていない： インターネットスピード ドイツの隠れたチャンピオンは農村部に位置(田舎に工場がある)
- ITセキュリティの心配が強くある

■組織

- 規制(投資に関する規制や、組合がドイツは強いいため改革が進まない)
- 専門家である技術者の組織変更ができない

◎結論： ロードマップがない。みんな実験しているが、マスタープランがない
 VWの仮想コミッショニングにより、VWグループは統合しようとしているが、...
 CPS, デジタルツイン?

都市伝説「現場にはデータが大量にある」

良くある例：

業務(作業)報告書に数値を記載する欄



US: N/A (責任をとらされるので書かない)

Japan: 前回の記録を見ながら予測、空想で埋める
 (空欄だと上司に怒られる)



結果： 意味の無い誤差だらけのデータがあるだけ！

この企業文化を変えることが重要な第一歩！

現場機器がインターネットにつながっていない
 古いOSで動いているため、セキュリティ上、つなげられない！



課題解決型人材の重要性

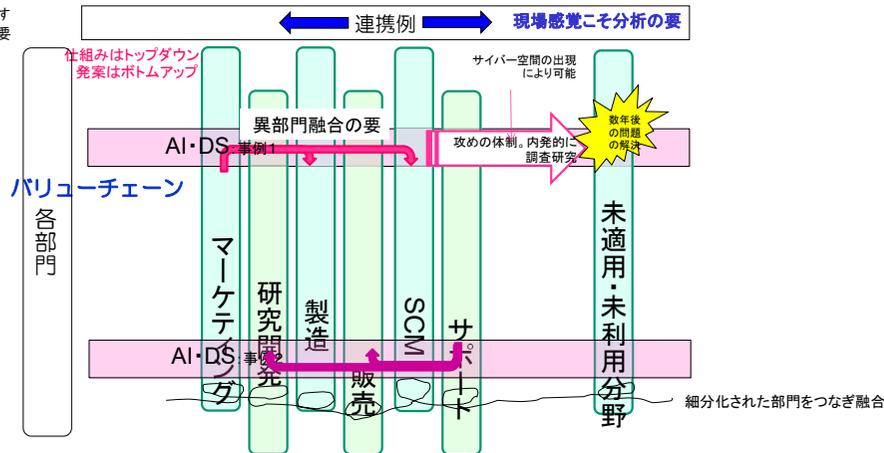
DX: デジタルトランスフォーメーション



- あらゆる産業における DXの奔流
- モノからコトへ： 解きたい課題がどこにでも
- DSの特性は横串： 異部門、異分野、異業種をつなぐ
- 研究開発体制：オープンイノベーション、エコシステム

【AI・データサイエンス技術】マトリックス戦略 企業を横断する中核部隊として

- データが重要
- 新しい価値観を生かすには新しい組織が必要



ポイント 従来の学部構成では時代遅れ。副専攻のような体制が必須!

汽水：淡水と海水が混じり合った状態



産業界
ビジネスカ
現場力

PBL
Active Learning
反転学習

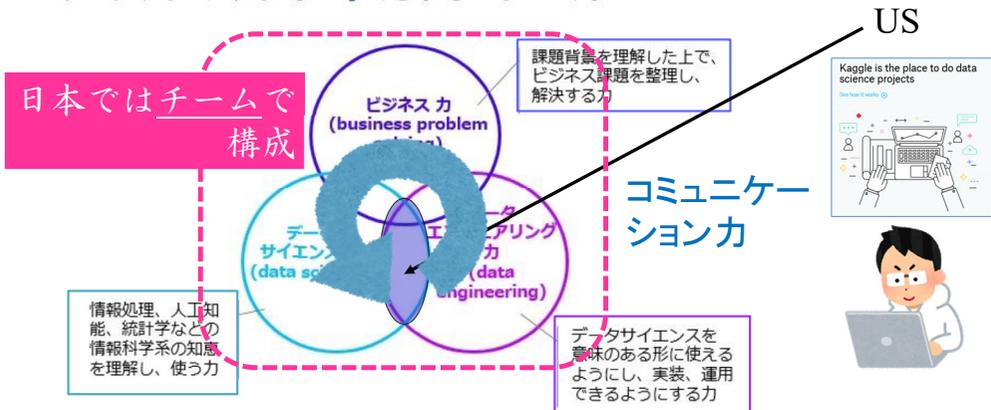
アカデミア
データサイエンスカ データエンジニアリングカ

ポストコロナを見据えた
授業のあり方
座学系 → オンデマンド
混じり合い方は多種多様

データサイエンティスト協会

現場主義

データサイエンティストに求められるスキルセット



ポイント 各学部での必修科目群が一つの大きな柱。これがビジネス応用場面の皆さんの個別能力発揮に大きく関連。

アウトライン

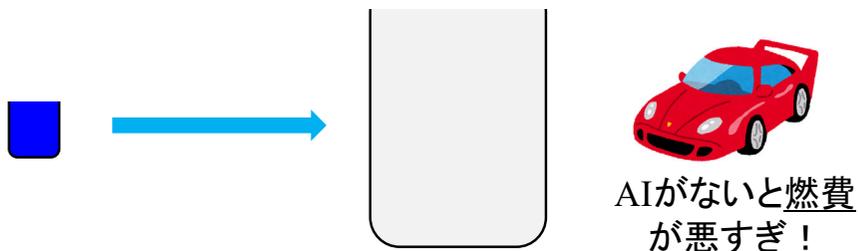
- データサイエンス教育強化の背景と経緯
- 社会の動き：DX、IoT、インダストリー4.0
- ビッグデータと機械学習
- 帰納法の弱点と社会実装での課題
- 将来の社会像、教育の姿



ビッグデータ + AI の前に立ちはだかるハードル
価値密度

価値 (←目的による!) 密度の定義

$$\text{価値密度(目的)} = \frac{\text{価値総量(目的)}}{\text{データ総量 Volume}}$$



ビッグデータ + AI の前に立ちはだかるハードル
『マシンに入れれば何かでてくる』は幻想



■よくあるケース
「ビッグデータはいろいろ社内にあるのだけれども、先生、何かできませんかねえ？もったいないいつも思っているのですが。最近AIもあるじゃないですか....」

「この種の発言をされるお客さんの案件はお断りするようになっている。営業にはいやな顔をされるが」
某ビッグデータ・コンサルティング担当マネージャー



錬金術士



マシンのイメージ

こちらのイメージ

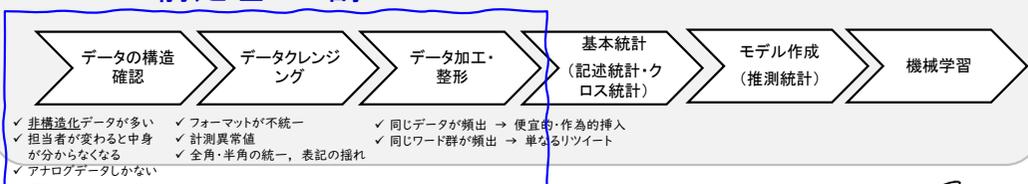
データは副産物。そして抽象物。



「もの」は分解できる

「もの」は合成できる

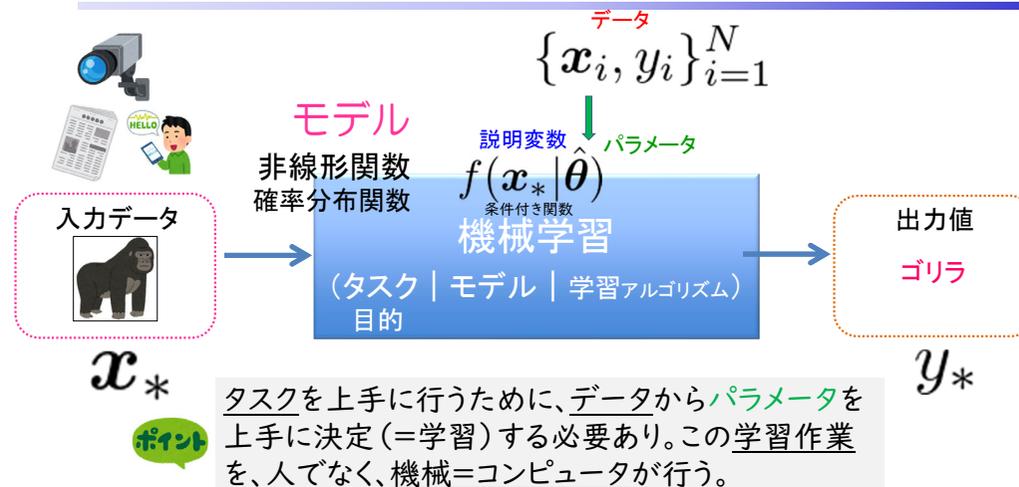
前処理が9割 データ処理の流れ



ポイント データが生み出された背景をきちんと調べる → 現場に聞く

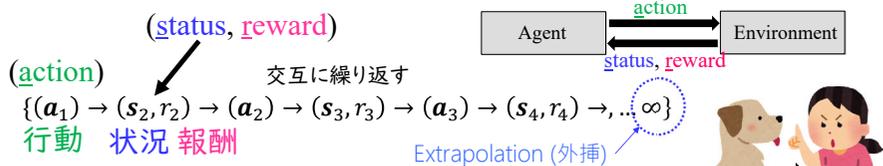


機械学習を規定する三つ組み: 物体認識を例に



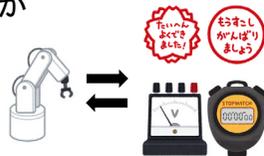


③強化学習



• 実験計画(スケジューリング)

- ✓ 高性能の実験・観測装置の運用には時間とお金がかかる。目的に応じた効率的作業が必須。
- ✓ 工場の現場でのリソース調整



• 制御理論

- ✓ ロボット制御(ロボティクス)、AI株投資(資金運用)

実際のビジネスの場面では要素技術を複合

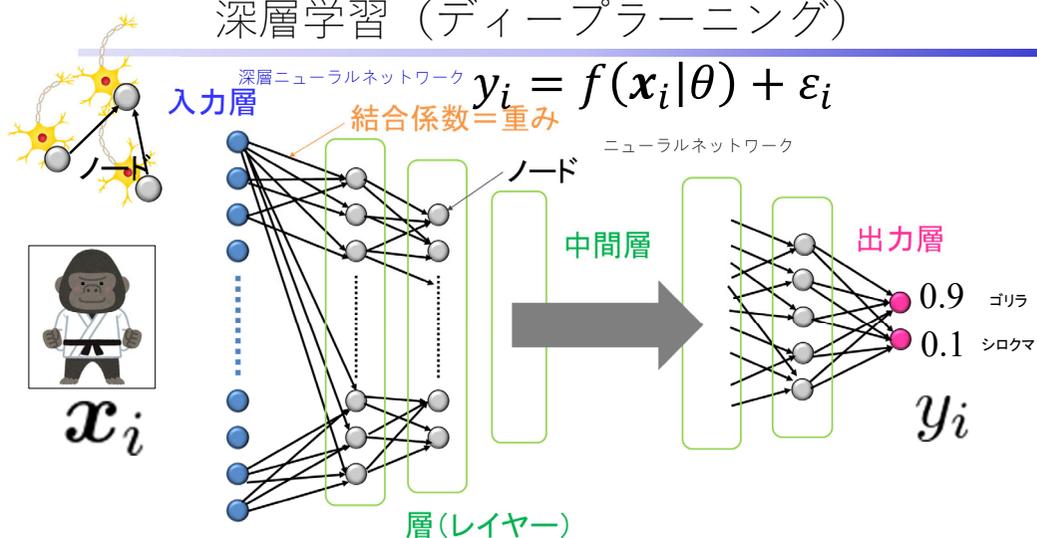
- 自動運転、AI(自動)回答(チャットボット)は、複合技術



- 故障検知は、異常(故障)の事例数の大小により、適したタイプが異なる



深層学習 (ディープラーニング)

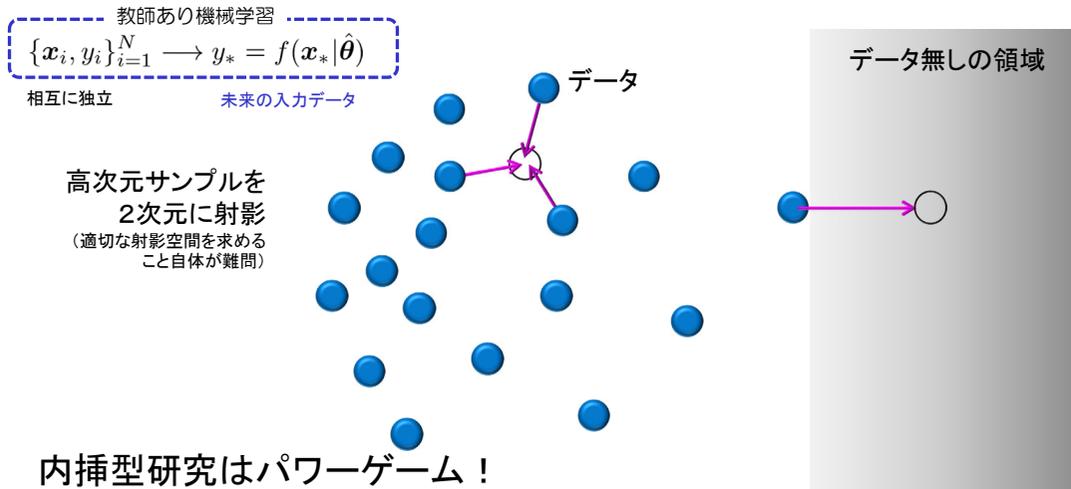


深層学習は通常レイヤー数が100~1000段以上

アウトライン

- データサイエンス教育強化の背景と経緯
- 社会の動き：DX、IoT、インダストリー4.0
- ビッグデータと機械学習
- 帰納法の弱点と社会実装での課題
- 将来の社会像、教育の姿

内挿と外挿



45

帰納法の弱点 1：偽陽性と偽陰性のバランス

データのみからルールを定める帰納法の限界 事例ベース意思決定法の限界

医療診断：偽陰性を避けるため、偽陽性を甘受（寛容）

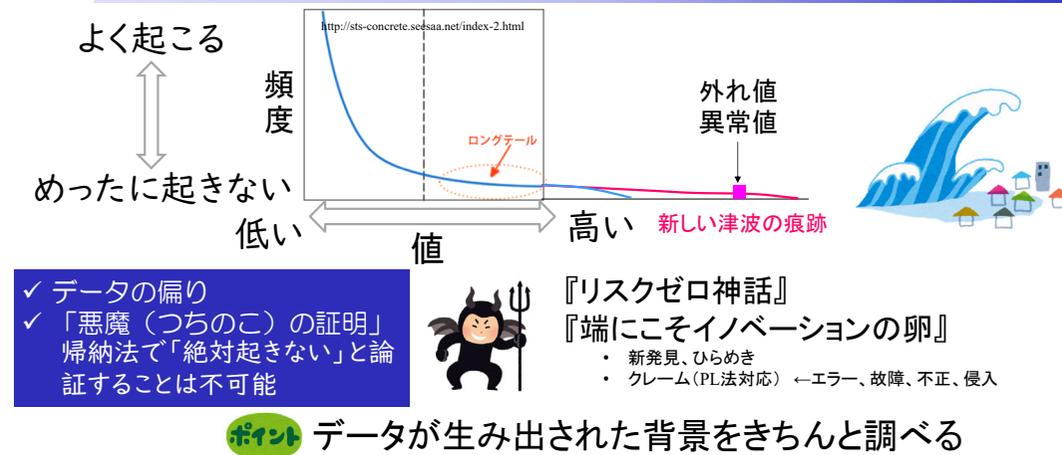
自動運転：偽陽性や誤検知を避けるため、情報をあえて無視する制御（忌避）

ECサイトでのマーケティング：いろいろなことが比較的自由に試せる

ポイント 現場、問題に応じて適切に課題設定する必要

46

帰納法の弱点 2：不（未）観測とデータの質 サンプリングの問題



47

帰納法の弱点 3：相関と因果

-
- ① 交絡因子 (第三の変数):
 - ・ 幼児や小学生低学年で「体力がある」→「学力が高い」(早生まれ、親の教育熱心さ)
 - ・ 「高血圧」→「高収入」、「趣味が盆栽」→「高収入」(年齢)
 - ・ 「教室の前に座る」→「成績が良い」(勉強への意欲の高さ)
 - ・ 「アイスクリームの売上げがあがる」→「焼きそばが売れる」(気温=イベント数)
 - ② 逆の因果関係:
 - ・ 「警察官の人数が多い」→「犯罪の発生件数が多い」
 - ・ 「メタボ健診をきちんと受けている」→「健康になる」
 - ③ 全くの偶然:
 - ・ 「ニコラス・ケージの年間映画出演本数が増える」→「プールの溺死者数が増える」
 - ・ 「地球温暖化が進む」→「海賊の数が減る」(? : サイクロン、台風)
 - ・ 「マーガリンの消費量」→「ある州の離婚率」

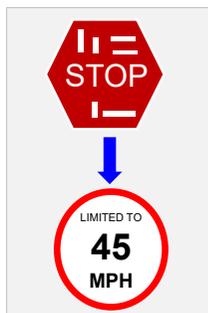
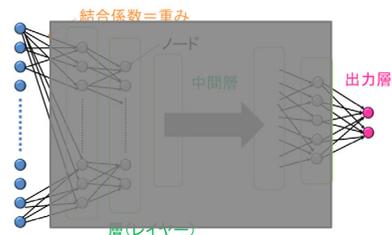
ビッグデータの分析だと、馬鹿馬鹿しい相関結果も
“統計的に有意”となる可能性大

ポイント 現場では因果の発見が勝負！
現場の知識、エキスパートの経験もあわせて活用

48

XAI：深層学習の弱点 ブラックボックスと可読性

XAI: Explainable AI



- ✓ 結果をみても何が特徴なのかわかりづらい。
- ✓ 途中の計算がうまくいっているのかが分からない。

ポイント ビジネス現場でAIのカラクリと結果の妥当性を説明できないといけない

- ✓ 投資運用などでのリスクの説明
- ✓ 執行猶予期間や刑期の短縮の判断

49

AI（生成モデル）の影 フェイク〇〇

研究不正ガイドライン文科省

1. ねつ造
2. 改ざん
3. 剽窃



すべて**ノイズ**のインプットから発生した地球上にない(?)顔画像
<https://www.whichfaceisreal.com/>



- 捏造とは、存在しないデータ、研究成果等を作成すること。
- 改ざんとは、研究資料、機器・過程を変更する操作を行い、データ、研究活動によって得られた結果等を真正でないものに加工すること。

ポイント エキスパートの知識や経験を越えるAI技術の登場にも留意する

作業の代替

50

Fairnessが鍵 バイアスは技術的課題

- ✓ Interpretability 解釈可能性
- ✓ Accountability 説明責任
- ✓ Reproducibility 再現性
- ✓ Transparency 透明性

AI技術開発の面で

- 白人男性の顔認識率は、女性マイノリティのものより相当高い。
- 自動運転での事故は誰が責任をとるのか？

ポイント 社会へのAI技術の埋め込みには注意しないと、ビジネスとして成立しない

ELSI
 倫理的・法的・社会的課題 (Ethical, Legal and Social Issues)

新技術の社会実装の観点から

2018年5月25日に施行
 「一般データ保護規則(GDPR)」

日本での「人間中心のAI社会原則」
 の立案(2019年3月)

51

ビッグテックへの強い風当たり

アンドロイドの機器数は30億台超え 世界の人口78億
 アップルの機器数は16億5千万台
 フェイスブックの利用者は35億超え(LINEの利用者は1億9千万弱)

(2020/08/08 朝日新聞の記事を一部参考)

米国

- 2019/09/05 米、グーグルに制裁金180億円 ユーチューブで同意なく子供情報収集(産経Web)
- 2019/09/10 50州・地域がグーグル調査 独禁法違反、包囲網狭まる(産経Web)
- 2021年1月 「(一部の会社による利用者のプライバシー侵害を許せば)我々は人としての自由そのものを失う」(byティムクック@アップルCEO) ターゲティング広告を非難
- 2021年2月 「SNS市場が帰る市場 レディット問題」(日経Web)
- 2021/03/14 「研究者2人を解雇したGoogle AI倫理対立の深刻度」(日経Web)
- 2021/04/27 追跡型ネット広告、転換期に Appleが新OSで自主規制(日経Web)
- 2021/05/27 「各国がAI規制を」 グーグルCEO、国際ルールを求める(朝日Web)
- 2021年 「グーグルはコンテンツを作るクリエイターから酸素を奪っている」(byブラッド・スミス@マイクロソフト社長) ニュース記事を使う際の対価支払い要求
- 2021/06/12 「米下院 GAFA規制法案提出 巨大ITの事業分割視野」(朝日Web)
- 2021/10/06 フェイスブックから内部告発「子供への精神的苦痛を看過」

52

ヨーロッパでの動き

■ 2021年1月「巨大ITの発信力制御を EU、米に呼びかけ」ダボス・アジェンダにて(日経Web)

2021年4月 EUにおける包括的AIの規制提案

- 公法・行政規制的なもの
- プロバイダーがEU域内におらずとも、AIシステムによる結果がEU域内に影響を及ぼす場合には適用対象
- リスクの大きさ・重要度に応じてAIを ①禁止されるAI、②ハイリスクAIシステム、③一定のリスクを持つAIシステム、④行動規範を自主的に採用するAI の4つに分類し、分類に応じた措置で対応する形式
- AI製品は様々なコンポーネントが組み合わされており、サプライチェーンが非常に長くなる傾向にあるが、トラブル時の原因や責任所在の確認に関しては、官からの要求に応じてディストリビューター、インポーター等が協力することが義務づけ

2021年11月に向けて準備中 ユネスコのAI倫理勧告案

- 政策措置の中にAIに対して法人格を与えるべきではないとの記載がある。法技術の問題として検討の余地があり得る。
- アメリカの論文では完全自動運転車に法人格を認め、そこに一定のお金をプールさせないと運用できない形にし、事故が起きた際、被害者に自動的にお金を支払う仕組みの提案がなされている。また、刑事的な考え方においては、罰金刑を科して、所有者である株主や経営者等に打撃を与えるといった考え方も
- ユネスコの勧告は非拘束であり、強制力はない。また、アメリカはユネスコを離脱している。

(内閣府 人間中心原則会議 会議録から抜粋し要約)

53

アウトライン

- a. データサイエンス教育強化の背景と経緯
- b. 社会の動き：DX、IoT、インダストリー4.0
- c. ビッグデータと機械学習
- d. 帰納法の弱点と社会実装での課題
- e. 将来の社会像、教育の姿

54

5年後、10年後の未来社会

■ データガバナンス

- 情報銀行、データ銀行：生み出すデータは自分のもの
- データの利用を認めるかわりに見返りも！
- ビッグデータの有効利用で社会全体でメリットを享受
- 個人情報の取り扱いに関しても社会全体で折り合いをつける機会が増大
- どこにどこまで許諾するかは自分でわからないので、代理人(AIエージェント)が存在感を示す。

■ AI・DSテクノロジー

- SNSに自主規制がすすむ。もっとすすんで、SNSの機能について法規制も。
- ヨーロッパのAI規制の考え方は、米国とかなり隔たりがある。日本は中庸。
- あらゆる業種でデータ分析にもとづく意思決定が普及
- Withコロナでの働き方と、AI・DSを基盤とする業種は親和性が高い。

✓ **トラスト技術、トラストシステムの成熟が鍵。** (注：言葉は会社名を指しているものではありません)



55

人生の分水嶺

人間の生き方(職業)を変えてきた観点での革命

紀元前 農耕革命



260年前 動力革命



70年前 情報革命



10年前 知能革命



- 『仕組み』をつくるものが勝利を得る世界
- AI・データサイエンスに関わる知識とデザインセンスが大切



- 2012年
- ✓ 深層学習ImageNetが、物体認識コンペで圧倒的な性能を示す
 - ✓ オバマ大統領政権のビッグデータイニシアチブの開始

(日本語) 単行本 - 2021/6/17
樋口が1, 2章執筆

56

文理融合はもはや「ナンセンス」、当たり前

- ✓ 理系の基本的素養（ツールとしての数学とプログラミング力）を身につけ、文系的センスを生かした社会応用力をもたないと、これからの荒波は生き残れない。
- ✓ これからのビジネスはマトリックス戦略が肝になる。ビジネスで活躍できる人材になれる。
- ✓ ビッグデータをもとにした情報化社会においては、「自分の望むライフスタイル」を実現するのにストレートに有効

