

データサイエンティストのキャリアパス形成に  
向けた施策方針について【補論】

令和4年3月24日

大学共同利用機関法人情報・システム研究機構

## はじめに

この補論では、2019年3月に公開された情報・システム研究機構「データサイエンティストのキャリアパス形成に向けた施策方針について」（以下、前回レポート）で指摘された点のうち、以下について、その後の動向などをまとめた。前回レポートでは、

- ・ データサイエンティストのキャリアパスは一定でなく、利用可能スキルは所属組織のデータ活用状況に左右されている
- ・ データ・プロダクトの学術基盤（応用領域）の整備、すなわち、品質・性能の確保、安全性の確保、評価方法の統一などを進める新しい学術基盤の構築によって、プロダクトの社会実装・社会受容が促進され、ひいてはデータサイエンティストのキャリアパスが形成される

といった点などがまとめられている。

これらの点と各補論の対応関係は以下のようになっている。

品質・性能の確保、安全性の確保、評価方法の統一など → 補論 1. 品質の問題

データ・プロダクトの社会実装・社会受容 → 補論 2. データサイエンスと倫理  
補論 3. DX 推進

利用可能スキルが所属組織のデータ活用状況に左右されること → 補論 4. 実データを使った教育

前回レポートにおける統計検定合格者（社会人）に対する調査（データサイエンティストのキャリアパスなどの現況） → 補論 5. リーディング DAT プログラム受講者を対象とした調査

2022年3月

## 目次

<b>補論 1.</b>	<b>品質の問題</b> .....	<b>1</b>
1.1.	品質とは.....	1
1.2.	事後品質の評価.....	1
1.3.	品質ガイドラインなどの主な動向.....	2
<b>補論 2.</b>	<b>データサイエンスと倫理</b> .....	<b>4</b>
2.1.	「社会」の倫理と「個人」の倫理.....	4
2.2.	「機械学習と公平性に関する声明」.....	5
2.3.	国内外のデータサイエンス技術に関する主な政策・法制の動向.....	6
2.4.	国際的なデータサイエンス技術に関する主な議論の動向.....	9
<b>補論 3.</b>	<b>DX 推進</b> .....	<b>11</b>
3.1.	DX 推進の現状.....	11
3.2.	DX 人材の育成.....	12
<b>補論 4.</b>	<b>実データを使った教育</b> .....	<b>12</b>
4.1.	国内外における実践例.....	13
4.2.	ループリック例.....	15
4.3.	使用データの選択.....	16
<b>補論 5.</b>	<b>リーディング DAT プログラム受講者を対象とした調査</b> .....	<b>19</b>

## 補論 1. 品質の問題

### 1.1. 品質とは

製品やサービスの品質とは、それらの物理的・抽象的特性のことである。その価値が顧客により判断されれば、製品やサービスは市場適性をもつことになる。価値とは、良い、望ましい、または重要なものに関する基準として個人または社会によって受け入れられる道徳的、社会的、または審美的な原則<sup>1</sup>のことで、製品やサービスの品質はこの原則に一致することが求められる。

統計的機械学習に代表される技術はデータから帰納的に開発されるため、価値や市場適性を予め担保しにくい。このような技術は、CACE (Changing Anything Changes Everything) 性と呼ばれる、入力やパラメーターを少しでも変更すると全体に影響が及んでしまう性質を持っているといわれている。また、訓練データは有限であるため、事前に全ての要求をカバーしておくことは不可能である。それだけでなく、入力と出力、あるいはパラメーターと出力の間の因果関係も人間が理解可能な範囲で明らかとはいえないことがあるため、見方によっては製品・サービスの説明可能性が低いといわれることもある。

したがって、近年の統計的機械学習に代表される技術に基づいた製品・サービスについては、事後品質をどのように評価するのか、説明可能性に対するアカウントビリティの担保についての議論<sup>2</sup>が必要になってきている。

### 1.2. 事後品質の評価

独立行政法人情報処理推進機構は、品質指標としてプロセス品質評価指標とプロダクト品質評価指標を挙げている<sup>3</sup>。プロセス品質評価指標は、製品・サービス開発の過程で実施された作業を測定し評価することによって、その品質を間接的に評価するものである。例えば、作業充当率や作業実施率という指標値は、開発された製品・サービスの品質が担保できると考える。一方、プロダクト品質評価指標は、開発の過程で作成される各種の中間成果物や最終段階で作成される成果物そのものを評価する指標である。例えば、ソフトウェア開発においては、ソースコード中のコメント割合、不具合収束率や不具合修正率などの指標から

---

<sup>1</sup> American Psychological Association, “APA Dictionary of Psychology,” 2022. <<https://dictionary.apa.org>>

<sup>2</sup> 丸山宏 (2020). 機械学習応用システムにおける品質と説明可能性について. 独立行政法人情報処理推進機構 AI 白書編集委員会 (編) AI 白書 2020, 角川アスキー総合研究所.

<sup>3</sup> 独立行政法人情報処理推進機構 技術本部 ソフトウェア・エンジニアリング・センター (2012). 組込みソフトウェア開発向け 品質作り込みガイド. 独立行政法人情報処理推進機構. <<https://www.ipa.go.jp/files/000005146.pdf>>

品質が推定される。

しかしながら、品質は、製品・サービス提供時の特性だけでなく、それらが運用される環境や運用のされ方にも影響を受けてしまう。データから帰納的に開発される製品・サービスにおいては、例えば安全性という事後的な指標を考えたとき、その指標のフィードバックに対して開発者がバイアスなしで対応することの困難さは容易に想定できる<sup>4</sup>。また、こうした製品・サービスの開発は、従来のウォーターフォール型開発とは異なり、開発要件が変化するを見越して小さなサイクルで開発をまわすアジャイル開発をはじめとする手法が採られることがある。このような開発手法の性質を加味した評価指標については、いくつかの文献で検討が行われている<sup>5</sup>。

### 1.3. 品質ガイドラインなどの主な動向

統計的機械学習に代表される技術に基づく製品やサービスの市場への投入が進んだことにより、こうした技術の有効性や信頼性を定量的に評価し、性能を保証することへの要求が増大している。ここでは、いくつかの動向を概観する。

#### 1) 米国電気電子学会の活動

2016年4月に米国電気電子学会 (Institute of Electrical and Electronics Engineers, IEEE) は、産業連携プログラムとして自律・知的システムの倫理に関するグローバル・イニシアチブ (Global Initiative on Ethics of Autonomous and Intelligent Systems) を開始し、2019年6月には報告書「倫理的に調整された設計」(Ethically Aligned Design, First Edition)<sup>6</sup>を公表している。この報告書では、AIシステムや開発者が満たすべき8つの一般原則がまとめられている。

---

<sup>4</sup> 丸山, *op.cit.*(2)

<sup>5</sup> 例えば以下のような文献がある。

Borg, M. (2022, January). *Agility in Software 2.0—Notebook Interfaces and MLOps with Buttresses and Rebars*. In *International Conference on Lean and Agile Software Development* (pp. 3-16). Springer, Cham.  
Jackson, S., Yaqub, M., & Li, C. X. (2019). The agile deployment of machine learning models in healthcare. *Frontiers in big Data*, 7.

Kupiainen, E., Mäntylä, M. V., & Itkonen, J. (2015). Using metrics in Agile and Lean Software Development—A systematic literature review of industrial studies. *Information and software technology*, 62, 143-163.

Atwal, H. (2019). *Practical DataOps: Delivering Agile Data Science at Scale*. Apress.

また、ソフトウェア開発の新しい方向性としては、最初の文献でも触れられている Software 2.0 がある。

Andrej Karpathy, “Software 2.0,” Nov 12, 2017. <<https://karpathy.medium.com/software-2-0-a64152b37c35>>

<sup>6</sup> IEEE Standards Association, “Ethically Aligned Design, First Edition (EAD1e) Infographic.” <<https://standards.ieee.org/industry-connections/ec/ead1e-infographic/>>

## 2) 国際標準化機構と国際電気標準会議による規格開発

国際標準化機構 (International Organization for Standardization, ISO) と国際電気標準会議 (International Electrotechnical Commission, IEC) が合同設置した技術委員会で、AI システムの信頼性 (透明性、検証性、説明性など) やリスクマネジメントについて規格開発が行われている。AI システムの信頼性に関する概観文書は 2020 年 5 月に出版済みである<sup>7</sup>。その他、いくつかの関連標準が出版されている。

## 3) 新エネルギー・産業技術総合開発機構、産業技術総合研究所、国立情報学研究所によるガイドライン

新エネルギー・産業技術総合開発機構は産業技術総合研究所や国立情報学研究所と共同で、2021 年 7 月に機械学習品質マネジメントガイドライン 第 2 版を公開している<sup>8</sup>。このガイドラインでは、対象となる AI システムのライフサイクル全体にわたる品質マネジメントが扱われている。機械学習を利用した AI システムにおける品質は、AI システム利用時に必要な品質を「利用時品質」、AI システム中で機械学習要素に要求される品質を「外部品質」、機械学習要素が固有に持つ特性を「内部品質」という 3 つに分類されている。内部品質の向上により外部品質が必要なレベルで達成され、さらに最終的な製品における利用時品質が実現されると考えられている。

## 4) 経済産業省、厚生労働省、消防庁によるガイドライン

経済産業省は厚生労働省・消防庁と共同で、2021 年 3 月に「プラント保安分野 AI 信頼性評価ガイドライン」を改訂、公開している<sup>9</sup>。このガイドラインでは信頼性評価に、新エネルギー・産業技術総合開発機構は産業技術総合研究所による機械学習品質マネジメントガイドライン 第 1 版の構造 (利用時品質、外部品質、内部品質) を用いて、石油・化学プラントでの AI 技術活用に求められる高度な安全性と説明責任に応えようとしている。

## 5) AI プロダクト品質保証コンソーシアムによるガイドライン

インダストリ寄りのガイドラインとしては、2018 年 4 月に設立された AI プロダクト品

---

<sup>7</sup> ISO/IEC JTC 1/SC 42 Artificial intelligence , “ISO/IEC TR 24028:2020,” May 2020. <<https://www.iso.org/standard/77608.html?browse=tc>>

江川尚志, “AI の trustworthiness の標準化状況 - バイアスを中心に -,” June 18, 2021. <<http://www.qa4ai.jp/Conf2021/20210618.SC42.pdf>>

<sup>8</sup> AIST デジタルアーキテクチャ研究センター, “成果公開,” Jul 5, 2021. <<https://www.digiarc.aist.go.jp/publication/aiqm/guideline-rev2.html>>

<sup>9</sup> 経済産業省, “プラント保安分野 AI 信頼性評価ガイドラインを改訂しました,” Mar 30, 2021. <<https://www.meti.go.jp/press/2020/03/20210330002/20210330002.html>>

質保証コンソーシアム（英文名：Consortium of Quality Assurance for Artificial-Intelligence-based products and services, 略称：QA4AI コンソーシアム）<sup>10</sup>から、2021年9月にAIプロダクト品質保証ガイドライン 2021.09版が公開されている<sup>11</sup>。このガイドラインでは、5つの評価軸（Data Integrity、Model Robustness、System Quality、Process Agility、Customer Expectation）を設定、それぞれの評価項目を示している。

## 6) 機械学習工学研究会の活動

日本ソフトウェア科学会の機械学習工学研究会<sup>12</sup>は、機械学習応用システムの開発・テスト・運用の方法論の確立と体系化を目指して活動している。2019年以降、国際シンポジウムやワークショップを開催し、海外に向けても情報発信が行われている。

## 補論 2. データサイエンスと倫理

今日のデータサイエンスにとって未知の領域であり、最大の問題は、社会の安全や利益と個人や少数派の安全や利益との間のバランスをいかにとるかという、古くからある議論に対する現代的な答えを導き出すことである。この議論は、公共政策、犯罪抑止、不正検知、自動運転、保険の契約査定、顧客離脱防止、ターゲティング広告やレコメンデーション、動画生成などで使用される技術の適切な使用方法はいかにあるべきか、と言い換えることができる。

### 2.1. 「社会」の倫理と「個人」の倫理

製品やサービスを提供、あるいは使用する上で、プロフェッショナルの倫理、個人の倫理と（文化や社会の）共通倫理の妥協点を時代にに応じて探っていくという対処方法が提案されている<sup>13</sup>。個人の倫理とは、その人自身の一連の約束事で、個人的な経験に起源があるといえる。共通倫理とは、ある文化、または社会の大部分のメンバーにより共有されている理想の集合である。この文脈では、プロフェッショナルの倫理と共通倫理が個人の倫理に対峙す

---

<sup>10</sup> AI プロダクト品質保証コンソーシアム, “QA4AI,” 2022. <<http://www.qa4ai.jp>>

<sup>11</sup> AI プロダクト品質保証コンソーシアム, “AI プロダクト品質保証ガイドライン 2021.09 版.” Sep 15, 2021. <<http://www.qa4ai.jp/QA4AI.Guideline.202109.pdf>>

<sup>12</sup> 機械学習工学研究会, “日本ソフトウェア科学会.” 2018. <<https://sites.google.com/view/sig-mlse/>>

<sup>13</sup> Harris, C. E., Pritchard, M. S., James, R. W., Englehardt, E. E., & Rabins, M. J. (2018). *Engineering ethics: Concepts and cases*. Wadsworth. なお、同書では、moral と ethic の明確な使い分けがされていないので、訳語としては「倫理」で統一する。

るものと位置付けられ、プロフェッショナルの倫理の多くは共通倫理の条項を言い換えたものであると考えられている。

プロフェッショナルの倫理と共通倫理はどちらも何らかの批判に応じて変わりうるが、同じように変わるわけではない。例えば、共通倫理に含まれる不正直の禁止は変化しそうにないと考えられる。一方で、環境に対する義務についての信念は変化の途上にあるため、プロフェッショナル側がその倫理を率先して変更すれば共通倫理の変化に影響することもあり得る。

さて、データサイエンス領域におけるプロフェッショナルの倫理は、関連学協会ごとに提示されており、例えば以下のようなものがある。

- ・ 国際統計協会 (International Statistical Institute; ISI) の Professional ethics<sup>14</sup>
- ・ 米国統計学会 (American Statistical Association; ASA) の Ethical Guidelines for Statistical Practice<sup>15</sup>
- ・ 統計関連学会連合による統計家の行動基準<sup>16</sup>

## 2.2. 「機械学習と公平性に関する声明」

2019年12月、機械学習の技術及び応用を研究している研究者コミュニティ(人工知能学会 倫理委員会、日本ソフトウェア科学会 機械学習工学研究会、電子情報通信学会 情報論的学習理論と機械学習研究会)から、機械学習の利用と公平性に関する声明が公開された<sup>17</sup>。この声明の骨子は、次の2点に集約される。

- (1) 機械学習は道具にすぎず人間の意思決定を補助するものである
  - (2) 公平性に寄与できる機械学習を研究し、社会に貢献できるよう取り組む
- そして、2020年1月には機械学習と公平性に関するシンポジウムが開催されている<sup>18</sup>。

データ・プライバシーの問題に焦点を当てたシンクタンク The Future of Privacy Forum<sup>19</sup>

---

<sup>14</sup> International Statistical Institute, “Professional ethics,” Oct 13, 2021. <<https://www.isi-web.org/about/policies/professional-ethics>>

<sup>15</sup> American Statistical Association, “Ethical Guidelines for Statistical Practice,” Apr 2018. <<https://www.amstat.org/your-career/ethical-guidelines-for-statistical-practice>>

<sup>16</sup> 統計関連学会連合, “統計家の行動基準,” Apr 22, 2017. <[http://www.jfssa.jp/wp-content/uploads/2017/09/statistician\\_20170422.pdf](http://www.jfssa.jp/wp-content/uploads/2017/09/statistician_20170422.pdf)>

<sup>17</sup> 人工知能学会 倫理委員会・日本ソフトウェア科学会 機械学習工学研究会・電子情報通信学会 情報論的学習理論と機械学習研究会, “機械学習と公平性に関する声明,” Dec 10, 2019. <<http://ai-elsi.org/wp-content/uploads/2019/12/20191210MLFairness.pdf>>

<sup>18</sup> 国立情報学研究所, “「機械学習と公平性 議論の経過と声明の内容」丸山 宏 | 機械学習と公平性に関するシンポジウム,” Jan 27, 2020. <<https://www.youtube.com/watch?v=hhG-5fQfRyY>>

<sup>19</sup> このシンクタンクはワシントン DC を拠点とし、AT & T、Facebook、Google などの企業、ビル&メリ



は、2017年12月のWeb記事の中で、機械学習アルゴリズムが機会の損失、経済的損失、社会的損失、自由の損失といった、「意図しない結果」をもたらしていることを述べている<sup>20</sup>。

- ・ 機会の損失の例：バイアスのかかった採用アルゴリズムによって、一部のグループに属する人の雇用機会が失われた。
- ・ 経済的損失の例：人種に関するバイアスのかかったアルゴリズムによって、アフリカ系米国人男性の与信限度額が10,800ドルから3,800ドルへ大幅に減らされた。
- ・ 社会的損失の例：SNSのニュースフィードはユーザーのオンラインアクティビティに合わせて調整されているので、ユーザーは世の中に対する自らの考えを裏付ける意見を見たり読んだりする可能性が高くなっている。
- ・ 自由の喪失の例：米国の再犯罪率予測に使われることが多いシステムが、アフリカ系米国人の再犯罪率を実際よりも高く評価していた。

上述の機械学習と公平性に関するシンポジウムの中でも指摘されているが、例えば機会の損失を防ぐために、センシティブ変数を予測に使用しないという方法がある（「手続きの公平性」）。しかしながら、手続きの公平性を満たしたアルゴリズムであっても、場合によっては「結果の公平性」（センシティブ変数の影響がなければ、結果の構成比は同じになる、あるいは財が平等に割り当てられる）が満たされないことが指摘されている<sup>21</sup>。なお、このような対立関係にある条件として他に、「機会の平等」（不平等な結果を受容）と「結果の平等」（逆差別を受容）などがある<sup>22</sup>。

### 2.3. 国内外のデータサイエンス技術に関する主な政策・法制的動向

#### 1) 国内の動向

2021年6月に統合イノベーション戦略推進会議は、AI戦略2019とそのフォローアップ版に続く「AI戦略2021～人・産業・地域・政府全てにAI～」を公開した<sup>23</sup>。AI戦略2019

---

ンダ・ゲイツ財団、米国国立科学財団などの財団からの共同支援を受けている。

<sup>20</sup> The Future of Privacy Forum, “Unfairness by algorithm: Distilling the harms of automated decision-making,” Dec 11, 2017. <<https://fpf.org/blog/unfairness-by-algorithm-distilling-the-harms-of-automated-decision-making/>>

<sup>21</sup> Calders, T., & Verwer, S. (2010) Three naive Bayes approaches for discrimination-free classification. *Data Mining and Knowledge Discovery*, 21, 277–292.

<sup>22</sup> Barocas, S., & Hardt, M. (2017). Fairness in machine learning. The 31st Annual Conference on Neural Information Processing Systems, Tutorial. (<https://mrtz.org/nips17/>).

Feldman, M., Friedler, S. A., Moeller, J., Scheidegger, C., & Venkatasubramanian, S. (2015). Certifying and removing disparate impact. *Proc. of the 21st ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining*, 259–268.

<sup>23</sup> 内閣府 統合イノベーション戦略推進会議, “AI戦略2021～人・産業・地域・政府全てにAI～,” Jun 11, 2021. <[https://www8.cao.go.jp/cstp/ai/aistrategy2021\\_honbun.pdf](https://www8.cao.go.jp/cstp/ai/aistrategy2021_honbun.pdf)>

では、「多様性を内包した持続可能な社会」を実現しつつ、世界規模の課題解決への貢献、大きな付加価値の創造と生産性の向上、ならびに産業競争力の強化を目指すため、教育改革、研究開発、社会実装、データ関連基盤、AI時代のデジタル・ガバメント、中小企業・ベンチャー企業への支援、そして倫理に関する統合的な政策パッケージが策定されている。これに続く AI 戦略 2021 では、これまで前提としてきた社会・経済システムが大きく変革していること、諸外国における AI 関連の動きが加速していることなどを踏まえ、社会・経済活動に真に役立つ AI の社会実装の促進に重点を置き、これまでの取り組みを加速させるとともに、新たな AI 戦略を策定している。

また、関係府省庁の AI 関連のガイドラインなどが相次いで公開されている。

- ・ AI・データの利用に関する契約ガイドライン 1.1 版（経済産業省、2019 年 12 月）<sup>24</sup>
- ・ AI 利活用ガイドライン～AI 利活用のためのプラクティカルリファレンス～（総務省、2019 年 8 月）<sup>25</sup>
- ・ AI 利活用ハンドブック～AI をかしこく使いこなすために～（消費者庁、2020 年 7 月）<sup>26</sup>
- ・ 国際的な議論のための AI 開発ガイドライン案（総務省、2017 年 7 月）<sup>27</sup>
- ・ 機械学習品質マネジメントガイドライン 第 2 版（新エネルギー・産業技術総合開発機構・産業技術総合研究所・国立情報学研究所、2021 年 7 月）<sup>28</sup>

## 2) 国外の動向

### 米国

2021 年 1 月、米国で「Executive Order on Advancing Racial Equity and Support for Underserved Communities Through the Federal Government」が発せられた<sup>29</sup>。この大統領令には、公平なデータ作業部会 (Equitable Data Working Group) の設立が盛り込まれている。連邦政府の多くのデータセットは、人種、民族、性別、障害、収入、退役軍人の地位、

---

<sup>24</sup> 経済産業省, “「AI・データの利用に関する契約ガイドライン 1.1 版」を策定しました,” Dec 9, 2019. <<https://www.meti.go.jp/press/2019/12/20191209001/20191209001.html>>

<sup>25</sup> 総務省, “AI 利活用ガイドライン～AI 利活用のためのプラクティカルリファレンス～,” Aug 9, 2019. <[https://www.soumu.go.jp/main\\_content/000637097.pdf](https://www.soumu.go.jp/main_content/000637097.pdf)>

<sup>26</sup> 消費者庁, “AI 利活用ハンドブック～AI をかしこく使いこなすために～(2020 年 7 月発行),” Jul 2020. <[https://www.caa.go.jp/policies/policy/consumer\\_policy/meeting\\_materials/review\\_meeting\\_004/ai\\_handbook.html](https://www.caa.go.jp/policies/policy/consumer_policy/meeting_materials/review_meeting_004/ai_handbook.html)>

<sup>27</sup> 総務省, “国際的な議論のための AI 開発ガイドライン案,” Jul 28, 2017. <[https://www.soumu.go.jp/main\\_content/000499625.pdf](https://www.soumu.go.jp/main_content/000499625.pdf)>

<sup>28</sup> AIST デジタルアーキテクチャ研究センター, *op.cit.*(7)

<sup>29</sup> The White House, “Executive Order on Advancing Racial Equity and Support for Underserved Communities Through the Federal Government,” Jan 20, 2021. <<https://www.whitehouse.gov/briefing-room/presidential-actions/2021/01/20/executive-order-advancing-racial-equity-and-support-for-underserved-communities-through-the-federal-government/>>

その他の主要な人口統計学的変数によって集計されていないとされ、このデータ不足が連鎖的に、公平性を促進する努力を阻害しているという考え方が背景にあることが示されている。一方、米国国立標準技術研究所 (National Institute of Standards and Technology, NIST) は、説明可能な人工知能に関する 4 原則 (Four Principles of Explainable Artificial Intelligence) を公開している<sup>30</sup>。

- ・ 説明 (Explanation) : システムは、アウトプットおよび／またはプロセスに付随する証拠または理由を提供するか、またはシステムに含めること。
- ・ 有意味性 (Meaningful) : 利用者に理解可能な説明を提供すること。
- ・ 説明の正確性 (Explanation Accuracy) : 説明は、アウトプットを生成する理由を正しく反映し、かつ／またはシステムのプロセスを正確に反映していること。
- ・ 知識の限界 (Knowledge Limits) : システムが設計された条件下で、かつ、その出力に十分な信頼性が得られたときのみ、システムが動作すること。

## 中国

中国は、2015 年以降、AI を含む先進的製造技術の導入を目指して「中国製造 2025」を中心に、2030 年頃には世界最先端の AI 先進国となるという目標の下で人材育成、基礎研究の強化、研究成果の産業化が進められている。2019 年には公平・公正、包摂、プライバシー、安全性、ガバナンスなどの原則が示されている<sup>31</sup>。

## 欧州

欧州では、2018 年 6 月に欧州委員会の Khalil Rouhana が「欧州の AI」(Artificial Intelligence for Europe) という AI の活用に関する提案を行った<sup>32</sup>。この提案では、官民による AI に対する投資の拡大、AI がもたらす社会的・経済的な変化への準備、倫理的・法的枠組みの確保という 3 つのアプローチが示されている。倫理的・法的枠組みの確保という点について、欧州委員会は 2018 年 6 月に AI に関するハイレベル専門家グループ (High-Level Expert Group on Artificial Intelligence, AI HLEG) を設置した。AI HLEG は 2019 年 4 月以降、4 つの成果物を公表しているが、倫理に直接関連するのは以下の 2 つである。

- ・ 信頼できる AI のための倫理ガイドライン<sup>33</sup> (2019 年 4 月公表) : AI に関する人間中心

---

<sup>30</sup> Phillips, P., Hahn, C., Fontana, P., Yates, A., Greene, K., Broniatowski, D., & Przybocki, M. (2021). *Four Principles of Explainable Artificial Intelligence, NIST Interagency/Internal Report (NISTIR)*, National Institute of Standards and Technology, Gaithersburg, MD, [online], <https://doi.org/10.6028/NIST.IR.8312>, [https://tsapps.nist.gov/publication/get\\_pdf.cfm?pub\\_id=933399](https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=933399) (Accessed February 23, 2022)

<sup>31</sup> 情報処理推進機構 AI 白書編集委員会 (2020). AI 白書 2020. 角川アスキー総合研究所.

<sup>32</sup> Khalil Rouhana, "Artificial Intelligence for Europe," Jun 27, 2018.

<[https://ec.europa.eu/futurium/en/system/files/ged/ai\\_for\\_europe\\_k\\_rouhana\\_27.06.pdf](https://ec.europa.eu/futurium/en/system/files/ged/ai_for_europe_k_rouhana_27.06.pdf)>

<sup>33</sup> High-Level Expert Group on Artificial Intelligence, "Ethics guidelines for trustworthy AI," Apr 8, 2019. <<https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>>

のアプローチ（人間の自主性の尊重、危害の防止、公平性、説明可能性）を提唱し、AI システムが信頼できるために満たすべき 7 つの主要な要件（人間の活動と監視、堅固性・安全性、プライバシーとデータガバナンス、透明性、多様性・被差別・公平性、社会・環境の幸福、アカウントビリティ）を示している。

- ・ 信頼できる AI のための最終評価リスト<sup>34</sup>（2020 年 7 月公表）：倫理指針をアクセス可能で動的な自己評価チェックリストに変換した実用的なツール。

2020 年には「AI 白書」(White Paper on Artificial Intelligence – a European approach to excellence and trust)<sup>35</sup>で普及・促進と信頼できる安全な開発を可能にするための政策オプションが提示され、2021 年 4 月 21 日に「AI 規則案」(Proposal for a laying down harmonized rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts)<sup>36</sup>が公表された。この規則案では、多くの具体的な使用方法を 4 段階のリスク基準（受容できない AI システム、ハイリスク AI システム、透明性義務を伴う AI システム、極小リスク／リスクなし AI システム）に類型化した上で、それぞれについてルールや使用方法を設定している。

#### 2.4. 国際的なデータサイエンス技術に関する主な議論の動向

政策や法制は数あるアプローチの一つであり、いくつかのアプローチを組み合わせることによって、データサイエンス技術を適切に使用していくことができると考えられる。国際標準化機構などが定める技術仕様の標準は安全基準づくりに有効と考えられるし、技術の使用にあたっては倫理教育も重要と考えられる。データサイエンス技術の使用に関する議論のいくつかの動向についてまとめる。

##### 1) 人工知能に関する OECD 原則

2019 年 5 月、経済協力開発機構（Organization for Economic Cooperation and Development, OECD）は人工知能に関する OECD 原則（Recommendation of the Council on Artificial Intelligence）<sup>37</sup>を公開し、この原則は 42 カ国に採択された。この原則は、AI シ

---

<sup>34</sup> High-Level Expert Group on Artificial Intelligence, “Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment,” Jul 17, 2020. <<https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment>>

<sup>35</sup> European Commission, “White Paper on Artificial Intelligence – a European approach to excellence and trust,” Feb 19, 2020. <<https://digital-strategy.ec.europa.eu/en/consultations/white-paper-artificial-intelligence-european-approach-excellence-and-trust>>

<sup>36</sup> European Commission, “Proposal for a laying down harmonized rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts,” Apr 21, 2021. <<https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1623335154975&uri=CELEX%3A52021PC0206>>

<sup>37</sup> OECD, “Recommendation of the Council on Artificial Intelligence,” May 22, 2019.

システムが健全、安全、公正かつ信頼に足るように構築されることを目指す国際標準で、概要は以下の通りである。

- ・ AI は包摂的成長と持続可能な発展、暮らし良さを促進することによって、人々と地球環境に利益をもたらすものでなければならない。
- ・ AI システムは、法の支配、人権、民主主義の価値、多様性を尊重するように設計され、また公平公正な社会を確保するために適切な対策が取れるようにすべきである。
- ・ AI システムの透明性を確保し責任ある情報開示を行うべきである。
- ・ AI システムはその運用期間中は健全で安定した安全な方法で機能させるべきで、起こりうるリスクを常に評価、管理すべきである。
- ・ AI システムの開発、普及、運用に携わる組織及び個人は、上記の原則に従いその正常化の責任を負うべきである。

## 2) UNESCO による「人工知能の倫理に関する勧告」

2021 年 11 月、国連教育科学文化機関 (United Nations Educational, Scientific and Cultural Organization, UNESCO) では、AI 倫理の国際規範である「人工知能の倫理に関する勧告」(Recommendation on the Ethics of Artificial Intelligence)<sup>38</sup>を策定し、総会で採択された。ただし米国は 2017 年に UNESCO を脱退しており、この採択には関わっていない。勧告では尊重すべき価値として、①人権と基本的自由および人間の尊厳の尊重、保護および促進、②環境と生態系の繁栄、③多様性と包括性の確保、④平和で公正な相互接続された社会での生活を挙げ、安全性、公正性、プライバシーとデータの保護、透明性と説明可能性など 10 の原則を示している。

## 3) AI に関するグローバルパートナーシップ

2020 年 6 月、「人間中心」の考えに基づいた、責任ある AI の開発と使用に取り組む国際的なイニシアティブである「AI に関するグローバルパートナーシップ (Global Partnership on Artificial Intelligence, GPAI) <sup>39</sup>が設立された。これは、政府・国際機関・産業界・有識者などからなる官民多国間組織で、G7 を中心とする 15 カ国・地域が設立メンバーとなり、現在では 25 の国と地域が参加している。GPAI には「責任ある AI」、「データ・ガバナンス」、「仕事の未来」、「イノベーションと商業化」、「AI とパンデミック対応」の 5 テーマの作業

---

<<https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>>

そして、2019 年 6 月の G20 首脳会合では、この OECD 原則を附属文書とし、G20 AI 原則が合意されている。

<sup>38</sup> UNESCO, “Recommendation on the Ethics of Artificial Intelligence,” Nov 24, 2021. <<https://unesdoc.unesco.org/ark:/48223/pf0000380455>>

<sup>39</sup> Global Partnership on Artificial Intelligence, “GPAI: Global Partnership on Artificial Intelligence.” <<https://gpai.ai>>

部会が設置されており、AI 原則と実践の橋渡しに寄与することを目的としたプロジェクトが進められている。2022 年 1 月には、AI 原則の実践に関する国際的な議論への理解を深めることを目的に、国内向けのシンポジウムが経済産業省と総務省によって共催されている<sup>40</sup>。

### 補論 3. DX 推進

2018 年に「未来投資戦略 2018—「Society 5.0」「データ駆動型社会」への変革—」が閣議決定されている。「Society 5.0」「データ駆動型社会」のような社会における取り組みの例として、デジタルトランスフォーメーション (digital transformation, DX)がある。DX とは、広義には IT の浸透によって人々の生活がより良い方向に変化することであるが、ビジネス用語としては企業がデータとデジタル技術によって競争優位性を確立することである。

#### 3.1. DX 推進の現状

日本では、何らかの形で DX に取り組んではいるものの、米国に比べ十分な成果が出ている企業は少ない<sup>41</sup>。情報処理推進機構が行った調査では、十分な成果が出ていると回答した日本企業は 17.0%であるのに対し、米国では 56.7%であった。経済産業省は日本における DX 推進の現状を一連のレポートにまとめているが、このような状況はこれまでに指摘されてきた課題が十分に達成できていない可能性も考えられる。2018 年の「IT システム『2025 年の崖』の克服と DX の本格的な展開」という副題が付けられた DX レポート<sup>42</sup>では、多くの企業の経営層では DX の必要性を理解しているが、2つの大きな課題があると指摘されている。①既存システムが事業部門ごとに構築され複雑化・ブラックボックス化している。②業務自体の見直しを含めた経営改革の実行方法が不透明である。2020 年の DX レポート<sup>43</sup>では、デジタル変革に対する現状への危機感を持つ国内企業は増加しているものの、「DX の取り組みを始めている企業」と「まだ何も取り組めていない企業」に二極化しつつあるとし、

---

<sup>40</sup> 経済産業省, “GPAI (Global Partnership on AI) シンポジウムを開催します,” Jan 11, 2022. <<https://www.meti.go.jp/press/2021/01/20220111003/20220111003.html>>

<sup>41</sup> 独立行政法人情報処理推進機構 (2021). DX 白書 2021: 日米比較調査にみる DX の戦略、人材、技術。独立行政法人情報処理推進機構。

<sup>42</sup> 経済産業省, “DX レポート～IT システム「2025 年の崖」の克服と DX の本格的な展開～,” Sep 7, 2018. <[https://www.meti.go.jp/shingikai/mono\\_info\\_service/digital\\_transformation/pdf/20180907\\_03.pdf](https://www.meti.go.jp/shingikai/mono_info_service/digital_transformation/pdf/20180907_03.pdf)>

<sup>43</sup> 経済産業省, “DX レポート 2 (中間取りまとめ),” Dec 28, 2020. <<https://www.meti.go.jp/press/2020/12/20201228004/20201228004-2.pdf>>

これから取り組むべきいくつかのアクションが提案されている。全体の 9 割を占める DX 未着手企業や DX 途上企業（DX 実施が散発的な企業）においては、①DX の認知・理解、②製品・サービス活用による事業継続（DX 利用のファーストステップ）が挙げられている。変革プロセスにある、残りの企業に対しては、短期的対応として④DX 推進体制の整備、⑤DX 戦略の策定、⑥DX 推進状況の把握が、中長期的対応として⑦デジタルプラットフォームの形成、⑧ユーザー企業とベンダー企業の共創<sup>44</sup>、研究開発などによる変革の加速、⑨DX 人材の確保がそれぞれ挙げられている。

### 3.2. DX 人材の育成

DX 推進には、それを担う人材の育成や確保（社外との連携を含む）が重要である<sup>45</sup>。しかしながら、日本では量、質ともに不足していることが指摘されており<sup>46</sup>、社員の学び直し（リスキル）へ取り組むことの重要度が増している。米国に比べ、日本では AI、IoT、データサイエンスなどに対するリスキルの方針を示している企業が少ないとされる。情報処理推進機構の調査では、米国は 72.1%の企業がその方針を持っているが、日本では「実施していないし検討もしていない」企業が 46.9%にのぼっている。

## 補論 4. 実データを使った教育

これまで、実データをデータサイエンス教育で使用するものの重要性は何度か指摘されてきた。例えば、統計学カリキュラムにおいては 3 つの重要な要素があることが指摘されている<sup>47</sup>。

- ① 統計計算の幅を広げる
- ② 計算リテラシーと論理的思考を深める
- ③ 実践的なデータを使った計算

---

<sup>44</sup> ユーザー企業とベンダー企業の共創が推進されれば、ユーザー企業とベンダー企業の垣根がなくなり究極的な産業の姿が実現されるとされ、このようなデジタル変革後の産業や企業の姿が「追補版」としてまとめられている。

経済産業省，“DX レポート 2.1（DX レポート 2 追補版）”，Aug 31, 2021. <<https://www.meti.go.jp/press/2021/08/20210831005/20210831005-2.pdf>>

<sup>45</sup> 経済産業省，“デジタルトランスフォーメーションを推進するためのガイドライン（DX 推進ガイドライン）Ver. 1.0,” Dec 12, 2018. <<https://www.meti.go.jp/press/2018/12/20181212004/20181212004-1.pdf>>

<sup>46</sup> 独立行政法人情報処理推進機構, *op.cit.*(40)

<sup>47</sup> Nolan, D. & Temple Lang, D. (2010). Computing in the statistics curricula. *The American Statistician*, 64, 97-107.

また、米国統計学会による「統計教育の評価と指導のためのガイドライン:大学報告書(2016年改訂版)」(Guidelines for Assessment and Instruction in Statistics Education (GAISE) College Report 2016)<sup>48</sup>では、統計教育に対して6つの提言がされており、3番目の提言が直接該当する。

- ① 統計的思考を教える
- ② 概念的な理解に重点を置く
- ③ 実データを文脈と目的に合わせて統合させる
- ④ アクティブ・ラーニングを促進する
- ⑤ コンセプトの探求とデータ分析にテクノロジーを活用する
- ⑥ 学生の学習を改善し評価するアセスメントを活用する

#### 4.1. 国内外における実践例

##### 1) 数理・データサイエンス教育強化拠点コンソーシアムによるモデルカリキュラム

数理・データサイエンス教育強化拠点コンソーシアムは、2020年4月に、分野を問わず全ての大学・高専生(約50万人卒/年)を対象にしたリテラシーレベルの教育のモデルカリキュラム<sup>49</sup>を公開している。このモデルカリキュラムでは、「数理・データサイエンス・AI」は、かつての「読み・書き・そろばん」的な素養とされており、

- ・ なぜ、数理・データサイエンス・AIを学ぶのか理解すること
- ・ 社会でどのように活用され新たな価値を生んでいるのか理解すること
- ・ AIの得意なところ、苦手なところを理解し、人間中心の適切な判断ができるようになること
- ・ 社会の実データ、実課題を適切に読み解き、判断できるようになること

などの点が重要とされている。そして、実データ、実課題を用いた演習など、社会での実例を題材に数理・データサイエンス・AIを活用することを通じて、現実の課題と適切な活用法に対する学びがカリキュラムに取り入れられることが推奨されている。

また、2021年3月には、主に学部3、4年生を対象としたモデルカリキュラム(応用基礎レベル)<sup>50</sup>を公開している。このモデルカリキュラムでは、応用基礎レベルの教育をリテ

---

<sup>48</sup> American Statistical Association, “Guidelines for Assessment and Instruction in Statistics Education (GAISE) College Report 2016,” Jul 2016. <[https://www.amstat.org/docs/default-source/amstat-documents/gaisecollege\\_full.pdf](https://www.amstat.org/docs/default-source/amstat-documents/gaisecollege_full.pdf)>

<sup>49</sup> 数理・データサイエンス教育強化拠点コンソーシアム, “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～データ思考の涵養～,” Apr 2020. <[http://www.mi.u-tokyo.ac.jp/consortium/pdf/model\\_literacy.pdf](http://www.mi.u-tokyo.ac.jp/consortium/pdf/model_literacy.pdf)>

<sup>50</sup> 数理・データサイエンス教育強化拠点コンソーシアム, “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～,” Mar 2021. <[http://www.mi.u-tokyo.ac.jp/consortium/pdf/model\\_ouyoukiso.pdf](http://www.mi.u-tokyo.ac.jp/consortium/pdf/model_ouyoukiso.pdf)>



ラシーレベル教育と専門教育とを繋ぐ「橋渡し教育」として位置づけている。特に AI が引き起こしうる未来を理解した上で、データサイエンスをさまざまな専門分野へ応用・活用し、現実の課題解決、価値創造を担う人材を幅広く育成しようとしている。そして、リテラシーレベルと同様に実データ、実課題を用いた演習に加え、この応用基礎レベルでは各専門分野の特性に応じた演習や PBL などを効果的に組み入れることによって、実践的スキルの習得を目指すことが推奨されている。

## 2) 統計エキスパート人材育成プロジェクト

統計数理研究所を中核機関とした統計エキスパート人材育成コンソーシアムによって、統計エキスパート人材育成プロジェクトが 2021 年 7 月から開始された<sup>51</sup>。このプロジェクト創設は、いわゆる「骨太の方針 2020」(2020 年 7 月 17 日閣議決定)<sup>52</sup>において、「データサイエンス教育や統計学に関する専門教員の早期育成体制等を整備する」とされていることが背景にある。このコンソーシアムにおける育成対象は、大学統計教員(大学院修士水準の統計学の講義や統計活用研究の指導を行うことができる人材)と大学院生(大学院修士水準の統計エキスパート)である。米国などに比べて、日本の統計研究人材は少なく、高度な統計学のスキルを有する人材の育成を重点的に進めることが重要と考えられている。また、このプロジェクトでは、統計人材育成エコシステム(このコンソーシアムで育成された大学統計教員がコンソーシアム参画機関などで統計学の教育・研究の中核となるとともに、参画機関などにおいて次の統計エキスパートを育成する)の構築が計画されている。育成にあたっては、育成対象者の各専門領域におけるデータを中心として、さまざまな実データが使用されている。

## 3) ボストン大学における MSSP プログラム

ボストン大学では大学院修士で実践中心のプログラムが提供されている<sup>53</sup>。プログラムを修了した学生には、実践統計修士号(Master of Science in Statistical Practice (MSSP))が与えられる。多くのデータサイエンス教育では、実データを使用した科目や講義回は教育プログラムの終盤に配置されたり、理論、方法論やプログラミングを学習することの意義を高

---

<sup>51</sup> 統計数理研究所, “「統計エキスパート人材育成プロジェクト」を開始,” Jul 12, 2021. <<https://www.ism.ac.jp/kouhou/news/20210712.html>>

<sup>52</sup> 内閣府, “経済財政運営と改革の基本方針 2020,” Jul 17, 2020. <<https://www5.cao.go.jp/keizai-shimon/kaigi/cabinet/2020/decision0717.html>>

<sup>53</sup> Boston University College of Arts & Sciences: MS in Statistical Practice, “A Practice-Centric Program,” <<https://www.bu.edu/mssp/>>

Masanao Yajima, “Statistics Practicum: Placing 'Practice' at the Center of Data Science Education,” Feb 26, 2021. <[https://estat.sci.kagoshima-u.ac.jp/SESJSS/PDF/JCOTS21\\_D1S1P1\\_yajima\\_slides.pdf](https://estat.sci.kagoshima-u.ac.jp/SESJSS/PDF/JCOTS21_D1S1P1_yajima_slides.pdf)>

Kolaczyk, E. D., Wright, H., & Yajima, M. (2021). Statistics Practicum: Placing “Practice” at the Center of Data Science Education. *Harvard Data Science Review*. <https://doi.org/10.1162/99608f92.2d65fc70>

めるために配置される。しかしながら、このプログラムでは、教育プログラムの中核に実践を配置し、理論、方法論などの学習は、その実践が効果的に行われるように計画されている点に特徴がある。学生は理論や方法論などの学習と同時に、学内の統計相談 (statistical consulting service) やパートナーシップを結んでいる企業などとのプロジェクト (external partner projects) に参加することが計画されている。こうした実践の評価は、自己評価、ピア評価、あるいは共同評価などが組み合わされて行われている<sup>54</sup>。

## 4.2. ルーブリック例

実データを使う教育が課題解決型のプロジェクト方式になる場合には、できた／できなかったという区分では学習者のパフォーマンスを適切に評価できないことがあるため、「どの程度できたか」というように質的に評価することがある。その際には、学習到達度を示す評価基準が観点と尺度とともに表にまとめられたもの (ルーブリック) が使用される。データサイエンス教育においては、PPDAC サイクル<sup>55</sup>に基づくルーブリックと、CRISP-DM<sup>56</sup>に基づくルーブリックがある。ここでは、その例を概観する。

### 1) PPDAC サイクルに基づくルーブリック例

PPDAC サイクルに基づくルーブリックの例として、プロジェクト型統計的探究力の評価ルーブリック (Statistics Project-based Assessment Rubric Table, START)<sup>57</sup>がある。これは、PPDAC サイクルを使って行ったプロジェクト型学習におけるパフォーマンスを PPDAC のフェーズごとに評価するものである。①P (Problem)、②P (Plan)、③D (Data)、④A (Analysis)、⑤C (Conclusion)、⑦プレゼンテーション、⑧総合達成度の 7 項目 5 段階で評価される。

### 2) CRISP-DM に基づくルーブリック例

---

<sup>54</sup> Dochy らの方法にルーツを持つルーブリックに基づくとされている。

Dochy, F. J. R. C., Segers, M., & Slujsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher education*, 24(3), 331-350.

<sup>55</sup> Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, 67(3), 223-265.

<sup>56</sup> Wirth, R., & Hipp, J. (2000, April). CRISP-DM: Towards a standard process model for data mining. In *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining* (Vol. 1, pp. 29-40).

Berthold, M. R., Borgelt, C., Höppner, F., Klawonn, F., & Silipo, R. (2020). *Guide to Intelligent Data Analysis: How to Intelligently Make Sense of Real Data*. London: Springer

<sup>57</sup> 深澤弘美・櫻井尚子・和泉志津恵 (2018). 統計的探究プロセスとその評価. 統計数理, 66, 15-36.

PPDAC サイクルの改良版とみなされている CRISP-DM<sup>58</sup>に基づくルーブリックでは、CRISP-DM がウォーターフォール型開発のような一方通行のプロセスではないことを反映してか、プロセスのフェイズごとの評価は行われない。パフォーマンスは、例えば、①データの価値（現代社会での適用可能性、起業可能性、持続可能性の問題、提案されたアイデアが社会に与える全体的な影響など）、②代替案とイノベーション（同じドメイン、同じデータやアプリケーションの代替案との比較など）、③技術的なツール（目的に応じた適切な使い分けなど）、④プロジェクトエフォート（データ収集や統合の労力などに関する問題の解決策や回避策、チームワークなど）、⑤説明の質という 5 項目について評価されるとともに、⑥貢献度と⑦協調性に関してグループ内でメンバーによって相互評価される<sup>59</sup>。

#### 4.3. 使用データの選択

実際のビジネスデータを企業から教育用途で提供してもらう際には、いくつかの留意点がある。例えば、データ提供企業に対しては、学生のインターンシップ、博士課程レベルの研究協力、従業員の生涯教育への協力などへの配慮といったベネフィットが提供できなくてはならないだろう<sup>60</sup>。

経済産業省は、2021 年 3 月に「AI・データサイエンス人材育成に向けたデータ提供に関する実務ガイドブック」<sup>61</sup>を公開している。このガイドブックでは、企業にとっての利益を上げ、リスクを下げる枠組が 4 つ整理されている。4 つの枠組みとは、①ハッカソン型（データ受取側がハッカソンを開催する）、②有償コンサル型（ハッカソンの運営に加え、データ受取側がデータ提供側にコンサルティングを提供し、課題の発掘支援と解決に必要なデータの特典・加工などをデータ受取側が有償で請け負う）、③共同研究型（データ提供側とデータ受取側が共同研究し、成果を共有）、そして④データ提供型（データ提供側がその対価として一定の金銭と引き換えに、データ受取側にデータを提供）である。そして、データ提供側とデータ受取側が実務面で直面するであろう課題について、各枠組の契約雛形に基づいて論点が整理・解説されている。

一方、データ活用の基盤作りも続けられている。例えば、以下のような取り組み、あるいはリソースがあり、中には無償で使用できるデータもある。

---

<sup>58</sup> Daniel Frischemeier, “Data Science Education for secondary school students and Civic Statistics Education for preservice teachers (Special lecture at JCOTS 2020),” Feb 28, 2020. <[https://estat.sci.kagoshima-u.ac.jp/SESJSS/PDF/Presentation\\_DataScienceEducation\\_27022020.pdf](https://estat.sci.kagoshima-u.ac.jp/SESJSS/PDF/Presentation_DataScienceEducation_27022020.pdf)>

<sup>59</sup> Martínez-Plumed, F. & Hernández-Orallo, J. (2021). Project-based learning for scaffolding data scientists' skills. *16th International Conference on Computer Science & Education (ICCSE)*, 758-763.

<sup>60</sup> Priestley, J., & McGrath, R. (2021). *Closing the Analytics Talent Gap: An Executive's Guide to Working with Universities*. Auerbach Publications.

<sup>61</sup> 経済産業省, “AI・データサイエンス人材育成に向けた データ提供に関する実務ガイドブック,” Mar 2021. <[https://www.meti.go.jp/policy/it\\_policy/jinzai/AIdataguide.pdf](https://www.meti.go.jp/policy/it_policy/jinzai/AIdataguide.pdf)>

- ・ Awesome Public Datasets<sup>62</sup> : 公共データソースのカテゴリーごとのリスト
- ・ BuzzFeedNews<sup>63</sup> : BuzzFeed のニュースルームが提供するオープンソースのデータ、ツールなど
- ・ CKAN<sup>64</sup> : データハブやデータポータルを強化するためのオープンソースの DMS
- ・ ClinicalTrials.gov<sup>65</sup> : 米国国立医学図書館により運営されている臨床試験データベース
- ・ data.europa.eu<sup>66</sup> : EU の機関や加盟国のオープンデータポータル
- ・ DATA.GO.JP<sup>67</sup> : 日本の公共データの案内・横断的検索を目的としたオープンデータのデータカタログ
- ・ data.gov<sup>68</sup> : 米国の行政機関が作成したマシンリーダブルなデータセット
- ・ data.gov.in<sup>69</sup> : インド政府の省庁や組織が公開するデータセット、ドキュメントなど
- ・ data.gov.uk<sup>70</sup> : 英国政府のオープンデータ
- ・ Data is Plural<sup>71</sup> : Jeremy Singer-Vine の Data Is Plural weekly newsletter に掲載されているデータソースのアーカイブ
- ・ データ活用社会創成プラットフォーム<sup>72</sup> : 大学・研究機関で共創する産学官連携のためのデータプラットフォーム
- ・ データ流通推進協議会<sup>73</sup> : データカタログ作成ガイドラインの公開などをはじめとしたデータ流通環境の整備
- ・ データサイエンス共同利用基盤施設<sup>74</sup> : 大学等の多様な分野の研究者に対する大規模データ共有（生命科学、極域環境科学、人間・社会分野、人文学）、データ解析支援、および人材育成を実施

---

<sup>62</sup> “Awesome Public Datasets.” <<https://github.com/awesomedata/awesome-public-datasets>>

<sup>63</sup> “BuzzFeed News.” <<https://github.com/BuzzFeedNews>>

<sup>64</sup> CKAN, “The world’s leading open source data management system.” <<https://ckan.org>>

<sup>65</sup> National Library of Medicine, “ClinicalTrials.gov.” <<https://www.clinicaltrials.gov>>

<sup>66</sup> Publications Office of the European Union, “data.europa.eu.” <<https://data.europa.eu/en>>

<sup>67</sup> デジタル庁, “DATA.GO.JP.” <<https://www.data.go.jp>>

<sup>68</sup> The U.S. General Services Administration, “The home of the U.S. Government’s open data.” <<https://data.gov>>

<sup>69</sup> Government of India, “data.gov.in.” <<https://data.gov.in>>

<sup>70</sup> UK Government, “data.gov.uk | Find open data.” <<https://data.gov.uk>>

<sup>71</sup> Jeremy Singer-Vine, “Data Is Plural.” <<https://www.data-is-plural.com>>

<sup>72</sup> データ活用社会創成プラットフォーム共同研究基盤, “mdx 大学・研究機関で共創する産学官連携のためのデータプラットフォーム.” <<https://mdx.jp>>

<sup>73</sup> 一般社団法人データ流通推進協議会, “Data Trading Alliance.” <<https://data-trading.org>>

<sup>74</sup> データサイエンス共同利用基盤施設, “データサイエンスの推進を通じて科学の発展や社会のイノベーションへの貢献を目指す.” <<https://ds.rois.ac.jp>>

- ・ Data World<sup>75</sup> : Kaggle Datasets と同様、ユーザーから提供された様々なデータセット、ならびに企業がデータを保管・整理するためのプラットフォーム
- ・ G 空間情報センター<sup>76</sup> : 官民などが保有する G 空間情報データの提供
- ・ Google’s dataset search engine<sup>77</sup> : 特定のトピックやソースのデータ検索エンジン
- ・ 人文学・社会科学データインフラストラクチャー構築推進事業<sup>78</sup> : 人文学・社会科学研究に関するデータを分野や国を超えて共有・利活用する総合基盤
- ・ 情報銀行推進委員会<sup>79</sup> : 安心、安全な情報銀行サービスのための認定事業を実施
- ・ Kaggle Datasets<sup>80</sup> : (大抵の場合) 要約情報とプレビュー付きのデータセット、自分のデータセットを投稿して公開または非公開にすることも可能
- ・ 公的統計マイクロデータ研究コンソーシアム<sup>81</sup> : 日本における公的統計マイクロデータの研究利用（二次利用）を促進
- ・ 気象ビジネス推進コンソーシアム<sup>82</sup> : 気象データのビジネス利活用推進
- ・ 公共交通オープンデータ協議会<sup>83</sup> : 公共交通の運行データや施設データの提供
- ・ Hugging Face<sup>84</sup> : NLP データセットを含む約 2000 のデータセット、ならびにデータの説明、使用目的と制限、操作方法、バイアス、学習データと学習手順情報、多くの一般的な指標での評価結果が掲載されたモデルカード
- ・ 日本語の歴史的典籍の国際共同研究ネットワーク<sup>85</sup> : 「日本語の歴史的典籍」約 30 万点を画像データ化し、既存の書誌情報データベースと統合させた「新日本古典籍総合データベース」開発
- ・ Makeover Monday<sup>86</sup> : データ可視化に関する自律的なコミュニティ
- ・ r/datasets/<sup>87</sup> : データセットを共有し、見つけ、議論する場

---

<sup>75</sup> data.world, inc, “data.world.” <<https://data.world/datasets/open-data>>

<sup>76</sup> 一般社団法人社会基盤情報流通推進協議会, “G 空間情報センター.” <[https://www.geospatial.jp/gp\\_front/](https://www.geospatial.jp/gp_front/)>

<sup>77</sup> Google, “Dataset Search.” <<https://datasetsearch.research.google.com>>

<sup>78</sup> 日本学術振興会, “人文学・社会科学データインフラストラクチャー構築推進事業.” <<https://www.jsps.go.jp/j-di/index.html>>

<sup>79</sup> 一般社団法人日本 IT 団体連盟情報銀行推進委員会. <<https://www.tpdms.jp>>

<sup>80</sup> Kaggle, “Datasets.” <<https://www.kaggle.com/datasets>>

<sup>81</sup> 大学共同利用機関法人情報・システム研究機構, “公的統計マイクロデータの利活用推進に向けて.” <<http://jmodc.org>>

<sup>82</sup> 気象ビジネス推進コンソーシアム, “気象ビジネス推進コンソーシアム.” <<https://www.wxbc.jp>>

<sup>83</sup> 公共交通オープンデータ協議会, “公共交通オープンデータセンター.” <<https://www.odpt.org>>

<sup>84</sup> Hugging Face <<https://huggingface.co/datasets>>

<sup>85</sup> 人間文化研究機構 国文学研究資料館 古典籍共同研究事業センター, “日本語の歴史的典籍の国際共同研究ネットワーク構築計画.” <<https://www.nijl.ac.jp/pages/cijproject/>>

<sup>86</sup> Makeover Monday, “Data Sets.” <<https://www.makeovermonday.co.uk/data/>>

<sup>87</sup> r/datasets, “Datasets.” <<https://www.reddit.com/r/datasets/>>

- ・ UCI Machine Learning Repository<sup>88</sup>：機械学習コミュニティ向けのデータセット
- ・ USAFacts<sup>89</sup>：米国の人口、政府の財政、および政府の社会への影響に関するデータとレポートの提供
- ・ VisualData<sup>90</sup>：コンピュータビジョンのデータセット

## 補論 5. リーディング DAT プログラム受講者を対象とした調査

統計数理研究所で実施されている統計思考力育成事業リーディング DAT (Data Analytics Talents)<sup>91</sup>受講者を対象にアンケート調査を行った。リーディング DAT は、現代社会で必要とされる統計数理の知識とスキルを持ったデータサイエンティスト (Leading Data Analytics Talents)の育成を目的としたプログラムである。初心者の一歩先から、プロジェクトを先導しマネジメントできるレベルまで、スキルアップに必須の統計科学の知識を学ぶことができる。

### 1) 調査方法

対象者 2018、2019 年度のリーディング DAT 受講者、計 498 名であった。

手続き Web 調査を行った。アンケート依頼はメールで行ったが、一段階手前でアンケート依頼メールの受け取り可否について対象者に確認した。アンケートの質問項目は、前回レポート (情報・システム研究機構「データサイエンティストのキャリアパス形成に向けた施策方針について」) で報告された調査と同一であった。

### 2) 調査結果

回答の回収率は 20.1%で、100 名からの回答を得た。このうち 4 名の回答者は約 3 分の 1 に無回答であったため、集計・分析から除外した。

---

<sup>88</sup> Center for Machine Learning and Intelligent Systems, “UCI Machine Learning Repository.” <<https://archive.ics.uci.edu/ml/index.php>>

<sup>89</sup> USAFacts, “Our nation, in numbers: Government data to drive fact-based discussion.” <<https://usafacts.org>>

<sup>90</sup> VisualData, “VisualData Discovery.” <<https://visualdata.io/discovery>>

<sup>91</sup> 統計数理研究所 統計思考院, “統計思考力育成事業 リーディング DAT(Data Analytics Talents)” <<https://www.ism.ac.jp/shikoin/training/leadingdat/index.html>>

**回答者属性** 回答者はどの年代においても男性の比率が高く、40代の回答者が最も多かった(図1)。また、専門領域は理工系の回答者の割合が多かった(図2)。所属組織の業種は、製造業、IT関連業、次いで情報サービス業の順に多かった(図3)。

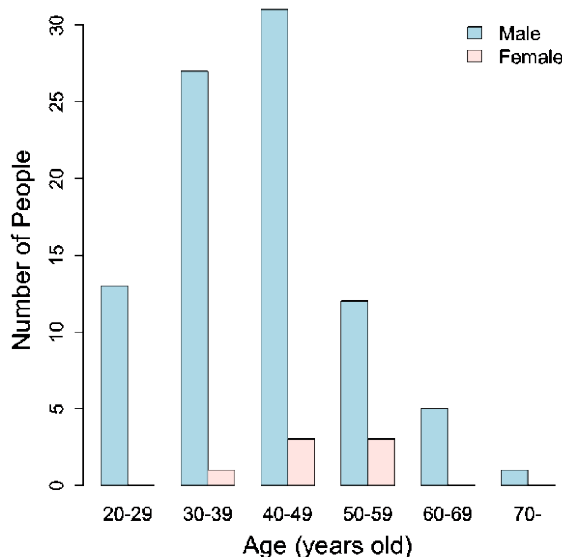


図1. 回答者の年齢と性別

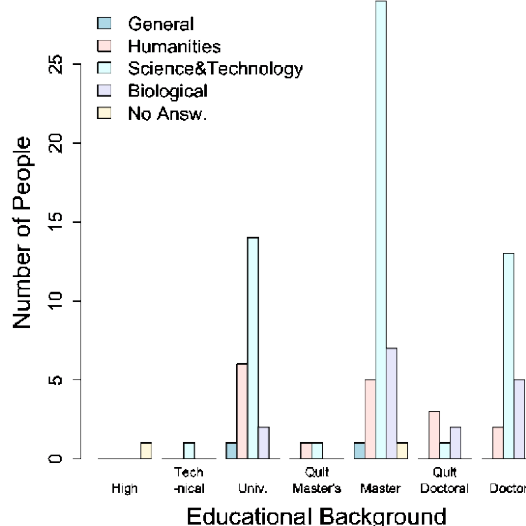


図2. 回答者の最終学歴と学生時の専門領域

**クラスター分析** 今回の分析では、以下の質問項目に対する回答を対象として k-means 法で2クラスターに分類した。クラスター数はシルエット法の結果とクラスター内安定性を参考に決定した。

- 質問 8 (データ分析業務に係る年収)
- 質問 9 (データ分析関連業務の頻度)
- 質問 10 (利用しているデータの種類)
- 質問 11 (データ活用の際の担当)
- 質問 13 (組織内のデータ活用に関するポジション)
- 質問 16 (利用可能な統計系スキル)
- 質問 20 (利用可能な情報系スキル)
- 質問 24 (利用可能なビジネス系スキル)
- 質問 28 (経験したことのある事業)
- 質問 29 (自分のスキルが発揮できているか否か)
- 質問 30 (データ活用の専門家としてキャリアを積みたいか否か)
- 質問 31 (データ活用の専門家としてキャリアを積みたい理由)

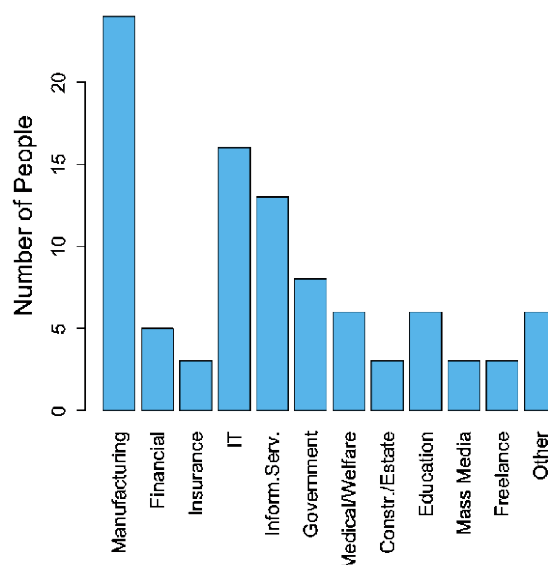


図3. 回答者が所属している組織の業種

なお、これらのデータには名義尺度と順序尺度が混在しているため、質問項目毎に固有ベク

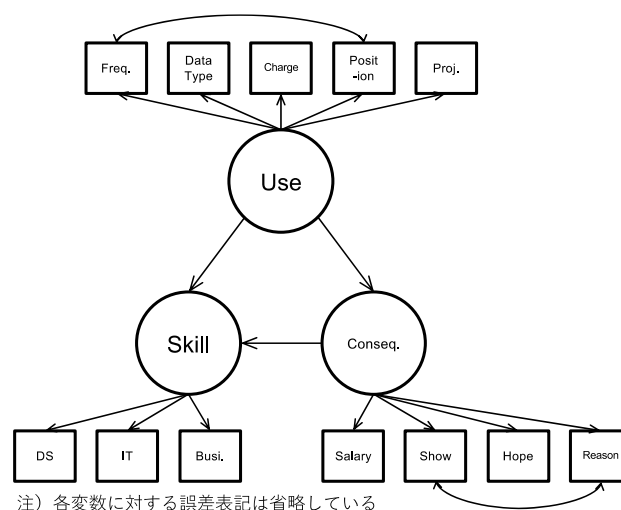
トルを求め標準化して分析に利用した。

2 クラスターを対象とした多集団同時分析  
次のようなモデルを想定して、多集団同時分析を行った。

モデル:スキル獲得 (Skill)はデータ活用状況 (Use)とスキル利用結果 (年収等やスキル利用に係る認識;Conseq.)により規定される。また、スキル利用結果はデータ活用状況により規定される。

モデルを吟味した結果、弱測定不変モデルを選択した (AIC: 2978.7, SBC: 3148.0, SRMR:

0.122, RMSEA: 0.079)。モデルの当てはまりはあまりよくなかったが、平均構造を導入してクラスター間の違いを検討した。



### 3) 調査結果に対する考察

分析結果と質問項目毎の回答傾向の比較から、各クラスターに次のようなペルソナを当てはめることができた。

クラスター1 (n=34)：自分のスキルが生かせていると感じている現場寄りの実務家

統計系スキル、情報系スキル、ビジネス系スキルはどれも多様であるが、統計的プログラミングスキルは低い。自分のスキルが仕事で生かせていると感じている。

クラスター2 (n=62)：主に製造業での業務経験豊富なデータサイエンス実務家

統計系スキル、情報系スキル、ビジネス系スキルはどれも多様で、統計的プログラミングスキルも高い。実務では、製造業での業務経験が豊富であるが、全体として、まだ十分に自分のスキルが生かしきれていないと感じている。

スキルレベルに応じたデータ利活用を質的に高めていくことが利用可能スキルの多様化と高度化につながる可能性がある。また、クラスター2の傾向から窺えるスキルレベルとデータ活用状況のギャップはキャリアパス形成に負の効果をもたらすことが、前回レポートに続いて示唆された。