AI倫理への取り組み：富士通グループAIコミットメントの制定

Approach to AI Ethics: Formulation of Fujitsu Group AI Commitment

中田　恒夫　　荒木　達樹　　土屋　哲　　中尾　悠里　　Naseer, Aisha
荒堀　淳一　　山本　隆彦

あらまし

人工知能（AI）技術が大きな便益をもたらす一方で，偏ったデータで学習したAIが不公平な判断を下すマイナス面も報告されている。さらなる技術の進歩が深刻な副作用を生まないようにAI倫理が議論され，近年多くの組織，企業からガイドラインが発表された。富士通は，2015年にFUJITSU Human Centric AI Zinraiを発表した時点で「人と協調する，人のためのAI」を主張し，AIの倫理的な側面に注意を払ってきた。これまでの30年以上にわたるAIの研究開発・社会実装における実績を踏まえ，AI倫理への取り組みを具体的でわかりやすい形で表した「富士通グループAIコミットメント」を2019年3月に発表した。欧州有識者会議AI4Peopleと連携し，ここが提唱するAI倫理原則に則った上で，富士通の主要ステークホルダーである，お客様，人，社会，株主，社員に向けたメッセージとした。

本稿は，AI倫理に関する世の中の動向とともに，富士通のAIコミットメントの基本的な考え方について述べる。

Abstract

While AI technology provides significant benefits, the downsides of AI have also been reported on, as in AI that used biased data for learning and made unfair decisions. To prevent further advancements in the technology from causing serious side effects, AI ethics are being discussed and many organizations and companies have recently announced new guidelines regarding AI. At the time of release of FUJITSU Human Centric AI Zinrai in 2015, Fujitsu proposed “collaborative, human centric AI” and has paid attention to the ethical aspects of AI. Based on the results of over 30 years of R&D and social implementation of AI, we announced the Fujitsu Group AI Commitment in March 2019, which expresses our approach to AI ethics in a concrete and easy-to-understand manner. We cooperated with AI4People, Europe’s expert forum, and follow the AI ethics principles that it advocates to send a message through the Commitment to customers, people, society, shareholders and employees, who are Fujitsu’s major stakeholders. This paper presents the social trends related to AI ethics and describes the basic concept of Fujitsu’s AI commitment.

1. まえがき

人工知能（AI）技術の進歩が，大量データを用いた新しい予測・最適化手法をもたらし，人々に大きな便益を与えている。その一方で，フェイクニュースや，犯罪予測AIが特定の人種に対して不公平な判断を下すといった負の側面も報告されており，今後これらの問題が身近で深刻になることが懸念される。そのため，AIがもたらす利便性を保ちつつ，有害な副作用を防ぐ施策が喫緊の課題である。

この施策の制度面からのアプローチとしてAI倫理が重要視されており，関連する議論が各方面で行われている。近年多くの組織や企業から，AIを研究・開発・提供・運用する際の倫理要件をまとめたAI倫理ガイドラインが発表されている。

このような状況の中，富士通でも，2015年のFUJITSU Human Centric AI Zinrai発表時に「人と協調する，人のためのAI」のコンセプトを掲げ，それを基にAIの倫理的な側面に注意を払ってきた。また，2019年3月にはAI倫理方針「富士通グループAIコミットメント」（以下，AIコミットメント）[(1)]を発表した。これは，これまでの30年以上にわたる富士通のAIの研究開発・社会実装における実績を踏まえ，AI倫理への取り組みを具体的にわかりやすい形で表している。

本稿では，AI倫理が重要視される背景，世の中の取り組み，および富士通のAIコミットメントの根底にある考え方を述べる。

2. AI倫理

本章では，AI倫理の定義およびAI倫理に関する世の中の動向について述べる。

2.1 AI倫理とは？

科学技術分野における倫理の定義に関しては医療分野が先行しており，40年にわたる議論と実践が行われている。患者に医療サービスを提供する立場の日本看護協会では，「社会生活を送る上での一般的な決まりごと」[(2)]と倫理を定義し，倫理要綱を導いている。同様の決まりごとである法は，国家による強制力を持つ「倫理の最小限」と言われ，倫理が法を含む。その一方で，マナーや集団内の規則も広く倫理と認識されている。前者を広義の倫理，後者を狭義の倫理と呼び，図-1の関係にあるものと考える。[(3)]

AI倫理とは，AIが人や社会に向けてサービスを提供する際にAIが満たすべき規範である。ここでのAIは，医療における看護師の立場に近いと言える。そのため，看護倫理での定義を参考にして，本稿ではAI倫理を，「AIと人が協調する上でAIが守るべき決まりごと」と定義する。これは，将来，規制化・法制化される規則も含む，広義の倫理に相当する。

2.2 AI倫理への関心の高まり

AI倫理に関する考察は，古くは1950年に小説の中で提唱されたロボット三原則[(4)]に見られる。そして，自動運転車が直面する「トロッコ問題」[(5)]として改めて注目を集めるようになった。その後，2015年頃からAIが倫理上の問題を引き起こす事例が相次いで報告された。ボットが不適切な発言を繰り返した事例，[(6)] 顔認識ソフトウェアが人に不適切なタグを付けた事例，[(7)] 人材採用支援AIに性差別が入り込んだ事例，[(8)] 再犯率の予測に使われたアルゴリズムが結果的に人種差別的であった事例[(9)]などが挙げられる。これらは，AIが人間の尊厳や公平性を損ねる危険性があることを露わにし，倫理的なAIの構築が不可欠であることを世に知らしめた。

2.3 AI倫理原則制定の動き

AIが倫理的に振る舞うことが社会から求められているにもかかわらず，「倫理的なAI」を実現する技術が確立されていない。一方で，AI技術の進歩

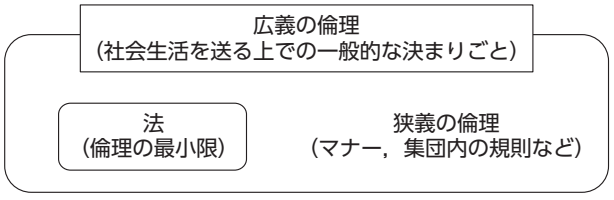

図-1　法と倫理の関係

とAIの適用が進んでおり，倫理上の深刻な問題を引き起こす危険性が高まっている。

この状況に対応するため，AIの研究・開発・提供・運用が適正に実施されるように，AI倫理原則を定める動きが活発になっている。**表-1**に，倫理に関連する主なAI原則を示す。

これらは法的な強制力を持たず，それぞれの組織・企業の立場を表明したものである。欧州ではこれを基にした規制案の策定が進んでおり，数年以内に規制・法制化がなされると見られている。

2.4 AI倫理とビジネス

AI倫理は技術や原則にとどまらず，ビジネス要件にもなりつつある。2018年9月カナダ政府は，AIサービス，ソリューション，プロダクトの調達要件[10]の必須項目として，AI倫理への対応を掲げた。そのため，供給者はフレームワーク，ガイドライン，評価ツール，テスト環境などを通じて，AI倫理への取り組みを明記しなければならない。この動きは，ほどなく世界中に広がると考えられ，AIプロバイダは速やかにAI倫理への対応を進めねばならない。

3. AI倫理に対する富士通の取り組み

富士通は，2009年にHuman Centric Intelligent Societyという言葉で，情報通信技術が切り拓く持続可能な未来の姿を表現した。また，富士通が長年培ったAI技術を体系化したFUJITSU Human Centric AI Zinrai[11]を2015年に発表した際も，Human Centric AI（人間中心のAI）を大きな特徴として掲げた。これは，AIが人間の尊厳を尊重し，AIがもたらす便益が人の幸福や自由のため，あるいは公益になることを意味する。一方，日本政府も2019年3月に内閣府が「人間中心のAI社会原則」[12]を正式決定した。また同年4月に欧州委員会が発表したAI倫理ガイドライン[13]では，AIシステムは「human-centric」でなければならないと主張している。このように，「人間中心」はAI倫理を考える上で重要な考え方となっている。

各所からAI倫理原則が発表されている中で，AI倫理への対応を2015年から謳(うた)っていた富士通も，AI倫理に関する取り組みを具体的に説明すべきと考え，AIコミットメントを策定した。[1] これはお客様や社会に対して，富士通がAIの開発者，提供者として自らを律し，AIがもたらす価値を広く社会に普及させる役割を果たす決意表明である。

4. AI倫理原則の客観性・網羅性に向けた施策

富士通はAIコミットメントを策定する際に，AI

表-1　倫理に関連する主なAI原則

発表した組織・企業	名称	年	原則数
Partnership on AI	Tenets[14]	2016	8
Future of Life Institute	Asilomar AI Principles[15]	2017	23
IEEE（Institute of Electrical and Electronics Engineers）	Ethically Aligned Design[16]	2017	5
An Initiative of Université de Montréal	Montréal Declaration Responsible AI_[17]	2018	10
マイクロソフト	Future Computed	2018	6
European Group on Ethics in Science and New Technologies	Statement on Artificial Intelligence, Robotics and ‘Autonomous’ Systems[18]	2018	9
Select Committee on Artificial Intelligence（英国）	AI in the UK: ready, willing and able?[19]	2018	5
IBM	Everyday Ethics for Planning	2018	5
AIネットワーク社会推進会議（総務省）	AI利活用原則案	2018	10
ソニー	ソニーグループAI倫理ガイドライン	2018	7
AI4People[20]	An Ethical Framework for a Good AI Society	2018	5
内閣府	人間中心のAI社会原則	2019	7
High-Level Expert Group on AI（EU）	Ethics Guidelines for Trustworthy AI	2019	4

※網掛け部分の6種は，AI4Peopleの5原則策定の際に参考にされたもの。

倫理の各原則が必要である根拠（客観性）と，考慮されるべき内容全体がカバーされること（網羅性）が不可欠と考えた。この二つを担保するために，外部のAI倫理有識者会議AI4Peopleと連携した。

AI4Peopleは2017年11月に設立された，AIの社会的インパクトを議論する欧州初の国際的なフォーラムであり，富士通も創立メンバーとして参画している。初年度である2018年度の活動目標は，AI倫理原則を制定し，Good AI Society構築に向けた提言[21]をまとめることであった。

AI倫理原則作成において後発のAI4Peopleが採用した方法は，有力なAI倫理原則6種（表-1の網掛け部分）を比較検討して，普遍的な原則にまとめあげるものであった。その結果AI4peopleでは，生命医学倫理[22]の4原則全てと，AI倫理特有の原則一つに集約された。以下にそれら5原則を示す。このうち原則1～4が生命医学倫理と共通である。

原則1：与益(注)（Beneficence）

AIシステムは，人と共同体の幸福・福利のために設計・開発されなければならない。価値や富を生み出すだけでなく，公平な社会，すべての人が参加できる社会，持続可能な社会への貢献を含む。

原則2：無危害（Non-maleficence）

AIシステムは，人に危害を加えてはならない。危害の対象は，身体だけでなく，尊厳，自由，プライバシー，安心・安全も含む。

原則3：人の自律（Autonomy）

AIシステムよりも人の判断や決定が上位に位置する。AIが判断・決定・指示・指令を出したとしても，人は決定権，根拠を知る権利，従わない権利を有する。AIに決定を委ねる権利は当人だけが有し，いつでもそれを撤回できる。

原則4：正義（Justice）

AIがもたらす便益がすべての人に広く行き渡り，負の側面が特定の人や集団に偏らないこと，および差別がないことを意味する。損害に対して，補償や救済の手立てが用意されていることも含む。

原則5：説明可能（Explicability）

生命医学倫理では，説明可能性の要件が「自律」原則に含まれ，患者に十分な情報提供がなされた上で，患者自身が納得して治療を受けられることを求めている。医療の世界と異なり，AIサービスではAIと人間が価値観を共有することを前提にできない。AIが「考えている」内容や判断の基準を，サービスの受け手である人間が理解できるように説明されることで信頼が生まれる。AIが下した判断の根拠がわかりやすく説明されること，判断に問題があると人間が感じた際に，そこに至るプロセスを再現し，原因を特定できること，規制上・法律上の問題が生じた際に，第三者が監査できることを保証するのが，説明可能原則である。

この5原則は，AI倫理を網羅的にカバーすると同時に，生命医学倫理で長年受け入れられてきた原則に則っており，客観性が担保されている。これは，AIの技術が進歩し，社会への適用が広がっても，有効であると考えられる。

5. 富士通グループAIコミットメント

先述したように，AIコミットメントは富士通がお客様，社会を含むステークホルダーに向けた決意表明である。その目的は，単純に「AIが満たすべき決まりごと」を定めた倫理原則を満足するAIを送り出すことにとどまらない。富士通グループの理念に基づき，AIの研究・開発・提供・運用に携わる者の責務としてお客様・社会との接点を重視し，「お客様起点のAI」「信頼できるAI」「責任あるAI」を提供する。このために「富士通グループでAIに関わるすべての役員・従業員」が従う行動指針を規定する。

全5項目から成るAIコミットメントは，富士通が各ステークホルダー（お客様，人，社会，株主，社員）に対して，倫理原則を満たすAIをどのような形で提供するか，あるいは提供することを目指すかを宣言する形とした。AI4peopleのAI倫理5原則とステークホルダーのマトリックスを作ることで，ステークホルダー視点の倫理原則を導いた。AI4peopleのAI倫理5原則とAIコミットメントの各項との関係を，表-2に示す。

以下に，AIコミットメントの5項目について詳しく述べる。

（注）　しばしば「善行」とも訳される。

表-2　AIコミットメントとAI4PeopleのAI倫理5原則との関係

		AI4PeopleのAI倫理5原則				
		与益	無危害	自律	正義	説明可能
ステークホルダー	お客様	✔	✔	✔	✔	✔ (1)
	人	✔	✔ (2)	✔ (4)	✔ (2)	✔ (5)
	社会	✔	✔	(3)	✔	✔ (5)
	株主	適切なAI倫理指針を発行することで貢献				
	社員	AI倫理規則・細則でカバー（2019年度中）				

※表中の(1)～(5)は，当該領域をカバーするAIコミットメントを示す。

(1) AIによってお客様と社会に価値を提供します

富士通がAIを提供する際の大きな指針である。AIシステムを構築し，お客様にそれを提供して終わりではなく，初期段階からお客様とCo-creation（共創）することでイノベーションを生み出し，「継続的に発展するAI」によってお客様に価値を提供し続ける。社会の情報通信インフラを担う富士通のビジネス特性から，これが社会への貢献にもつながる。

(2) 人を中心に考えたAIを目指します

人に対する与益，無危害，正義原則に対応する。人のためになるAIを構築し，かつ多様な価値観，多様な能力を持つ人々がそれぞれの想い，状況に合わせて能力を発揮できるように，AIが支援できるようにしていく。同時に，AIがもたらし得る弊害を抑えるために，差別軽減・除去，セキュリティの確保，プライバシーの保護を含めた品質の確立に注力し，人が安心して使えるAIを目指す。

(3) AIで持続可能な社会を目指します

社会に対する与益を中心に，無危害，正義原則にも対応し，地球環境を持続させるためにAIを活用していくことを示す。富士通グループがCo-creationを通じて，持続的に社会に貢献していくのがHuman Centric Intelligent Societyの考え方である。この活動と，国際社会がSDGs（Sustainable Development Goals）の達成に向けて取り組む方向性は一致している。

(4) 人の意思決定を尊重し支援するAIを目指します

人に対する自律原則に対応し，AIの提案や判断よりも人間が上位に来ること，すなわち人が最終判断を行う権利を常に有することを意味する。

(5) 企業の社会的責任としてAIの透明性と説明責任を重視します

説明可能原則を基にしつつ，富士通が社会を支える情報インフラを担うという立場から，より強い表現にしたものである。ステークホルダーに富士通のAIを信頼してもらうために，AIの透明性を確保し，AIが下した判断に関する十分な情報を提供することを掲げる。AI技術が万能ではないことを認識し，不具合を減らすための努力を続けるとともに，不具合が発生したとしてもその影響を扱える範囲にとどめ，再発防止に向けた手立てをすぐに打てるようなシステム構築を目指す。

ここで述べた五つの「約束」で目指している内容は，現時点ですぐに実現できるものではない。AI技術が進歩，社会への適用が急速に進んでいる現在，「信頼できるAI」を構築するための技術的・制度的検討が不可欠である。このAIコミットメントが制度整備の第一歩であると同時に，技術開発戦略を構築する上での基盤となる。これは固定されたものではなく，技術動向，社会におけるAIの受容状況，AI倫理の議論の深まりに応じて，随時見直されるべきものである。

6. むすび

本稿では，富士通が2019年3月に発表した「富士通グループAIコミットメント」の内容と，そこに至る考え方を述べた。

この文書自体は，AIの研究・開発・提供・運用に関する現時点の指針を述べたものである。これ

は，日常的な活動において発生する課題に対する具体的な解決策ではない。引き続き，この内容に基づいて，現場の判断に用いることができる規則・細則などの制定を行う。またこれに並行して，富士通での適用経験を活かし，世界各地で進められているAI倫理の規制・法制化が適切に実施されるように働きかけていく。

参 考 文 献

(1) 富士通：富士通グループAIコミットメント.
https://pr.fujitsu.com/jp/news/2019/03/13-1.html

(2) 日本看護協会：看護実践情報.
https://www.nurse.or.jp/nursing/practice/rinri/text/basic/what_is/index.html

(3) 小野原雅夫：まさおさまの　何でも倫理学.
https://blog.goo.ne.jp/masaoonohara/e/8cf1e287a4c420ec976acdc38e893f87

(4) アイザック・アシモフ：わたしはロボット．東京創元社（1976）.

(5) Jean-Francios Bonnefon：Autonomous Vehicles Need Experimental Ethics: Are We Ready for Utilitarian Cars?：arXiv:1510.03346v1，2015.
https://pdfs.semanticscholar.org/13d4/56d4c53d7b03b90ba59845a8f61b23b9f6e8.pdf

(6) Wikipedia：Tay.
https://ja.wikipedia.org/wiki/Tay_(%E4%BA%BA%E5%B7%A5%E7%9F%A5%E8%83%BD)

(7) Forbes Japan：グーグル「人間の顔をゴリラと間違えるミス」を謝罪.
https://forbesjapan.com/articles/detail/6399

(8) Reuters：Amazon scraps secret AI recruiting tool that showed bias against women.
https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G

(9) A. Chouldechova：Fair prediction with disparate impact: A study of bias in recidivism prediction instruments：Big Data，Vol.5，p.153-163，2017.
https://arxiv.org/pdf/1703.00056.pdf

(10) Public Works and Government Services Canada：Invitation to qualify for opportunities in Artificial Intelligence.
https://buyandsell.gc.ca/invitation-to-qualify-for-opportunities-in-artificial-intelligence

(11) 富士通：FUJITSU Human Centric AI Zinrai（ジンライ）-富士通のAI（人工知能）.
https://www.fujitsu.com/jp/solutions/business-technology/ai/ai-zinrai/

(12) 内閣府：人間中心のAI社会原則検討会議.
https://www8.cao.go.jp/cstp/tyousakai/humanai/index.html

(13) European Commission：Draft Ethics guidelines for trustworthy AI.
https://ec.europa.eu/digital-single-market/en/news/draft-ethics-guidelines-trustworthy-ai

(14) Partnership on AI：Tenets.
https://www.partnershiponai.org/tenets

(15) Future of Life Institute：Asilomar AI Principles.
https://futureoflife.org/ai-principles/

(16) IEEE：Ethically Aligned Design.
https://ethicsinaction.ieee.org/

(17) An Initiative of Université de Montréal：Montréal Declaration Responsible AI_.
https://www.montrealdeclaration-responsibleai.com/

(18) European Group on Ethics in Science and New Technologies：Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems.
https://ec.europa.eu/info/news/ethics-artificial-intelligence-statement-ege-released-2018-apr-24_en

(19) House of Lords–Select Committee on Artificial Intelligence：AI in the UK: ready，willing and able?
https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf

(20) AI4People：AI4People.
http://www.eismd.eu/ai4people/

(21) L. Floridi et al.：AI4People–An Ethical Framework for a Good AI Society: Opportunities, Risks，Principles，and Recommendations：Minds and Machines，Vol.28，issue 4，p.689-707，2018.
https://link.springer.com/content/pdf/

10.1007%2Fs11023-018-9482-5.pdf
(22) ビーチャム，チルドレス：生命医学倫理 第五版：麗澤大学出版会，2009.

著者紹介

中田　恒夫（なかた　つねお）
（株）富士通研究所
人工知能研究所
AI品質の研究戦略，およびAI倫理の研究戦略，ルールづくりに従事。

荒木　達樹（あらき　たつき）
Fujitsu Intelligence Technology Business Development
Data・XAI領域における新規事業創出や新たな契約モデルの開発に従事。

土屋　哲（つちや　さとし）
富士通（株）
ソフトウェア事業本部
AIサービスプラットフォームの開発に従事。

中尾　悠里（なかお　ゆうり）
（株）富士通研究所
人工知能研究所
推薦システム，公平性配慮型機械学習，アクセシビリティのユーザー観察研究に従事。

Naseer, Aisha
Fujitsu Laboratories of Europe
Trusted Technologies Research Group
AI倫理の研究，特に欧州における標準化関連団体との連携に従事。

荒堀　淳一（あらほり　じゅんいち）
富士通（株）
法務・コンプライアンス・知的財産本部
AI，デジタルアニーラ，データ利活用などの先端技術分野における政策立案に従事。

山本　隆彦（やまもと　たかひこ）
富士通（株）
法務・コンプライアンス・知的財産本部
AI倫理など新技術の社会実装で生じる課題に対するスタンダード推進の戦略策定に従事。