



## 社会イノベーション事業に AI を活用するための 日立的 AI 倫理への取り組み

2021年2月22日 発行  
株式会社 日立製作所  
Lumada Data Science Lab.

# 目次

巻頭言 .....	3
はじめに .....	4
1.AI 倫理を取り巻く状況.....	5
1.1 背景 .....	5
1.2 AI 倫理に関わるルール等の状況 .....	8
2.日立の AI 倫理原則について .....	10
2.1 AI 倫理原則の整備に向けた日立の取り組み .....	10
2.2 日立の AI 倫理原則に対する基本的な考え方 .....	11
2.3 AI の社会実装における着眼点 .....	12
3.日立の AI 倫理原則を遵守するための取り組み .....	18
3.1 AI 倫理原則の具体的な適用 .....	18
3.2 AI 倫理の社内教育について .....	19
3.3 AI 倫理原則の今後の改訂方針 .....	20
おわりに.....	21
出典 .....	22

## 巻頭言

日立製作所 名誉フェロー  
小泉英明

地球上で最も繁栄している現生人類は、高度な「社会」を造ることで生物種の頂点に立ちました。社会を造るためには、自分のためだけを考えるのではなくて、自分以外の人々の立場や生命を感じる心を持つことが、その根底に必須です（altruism）。それは時空を超えた広い意味での共生（ともいき）にも繋がります。自分以外の人々にも心を配ること、言い換えれば「温かな心」(warmheartedness)こそ、「倫理」(ethics)そのものだと思います。アリストテレスによって「賢慮」(phronesis)を中心とした『ニコマコス倫理学』が整理され、カント最晩年の『人倫の形而上学』には「他者の幸福を促進すること」として述べられました。

一世紀ほど前に、日立製作所が日立鉱山のなかに生まれたころ、心ある人々は鉱山が周囲の人々と共存することを第一に考えて、苦心惨憺の末に 156 m の大煙突を山の上に建てました。その周囲の山々に高層気象の観測小屋（時にゾンデを使用）を多数配置し、排ガスの状況と精錬生産量を一部連動させて、社会に開かれた経営システムを構築しました。日本の近代環境史における金字塔といわれています。

その鉱山には、発電・送電によって得られた電力や空気圧による動力システム、電動の昇降機や軌道車両、鉱脈探索や情報・通信システム、そして文化施設である劇場・病院・学校が生まれました。大きくとらえれば地域社会の「人々の安寧とよりよき生存」(Human Security and Well-Being)をめざす活動でありました。「企業は社会とともにある」という日立の考え方はそこに確立されたと考えられます。

その心は脈々と日立グループのなかに流れ続けて、「日立の社会イノベーション」(Hitachi Social Innovation)へと繋がって行きました。海外には「Industry 4.0」というドイツの工学アカデミーから生まれた思想が流れるなかで、日本はデジタル革命に「Society 5.0」という「社会と人間」の視座を加えた独自の思想を生み出しました。日立の思想は、今、世界へと発信されつつあります。

一方、国際連合の活動に日本も貢献していますが、その一つが「STI for SDGs」(Science, Technology and Innovation for Sustainable Development Goals)です。倫理と目標を明確にした日立 AI (人工知能)の寄与が期待されています。

そしてまた、グローバルに考え、現場から実行する。すなわち「Think globally, act locally」の想いで綴られたのが、この「日立の AI 倫理への取り組み」です。AI 倫理に真摯に取り組んで行こうとする、日立の若手の想いがしっかりと込められています。さらに多くの方々のお智慧を頂戴して、めざす水準を高めて参りたいと存じます。

皆様のご指導とご鞭撻を衷心よりお願い申し上げます。

## はじめに

AI (Artificial Intelligence、以降 AI) は、1956 年にダートマス会議でその名前が生み出されて以後、様々な研究や発展が繰り返され、現代では膨大なデータを背景とした機械学習やディープラーニング等によってビジネスや生活の非常に多くの領域で活用されています。

そして、今後も AI の用途が高度で多岐に亘ることを念頭に、日本や各国において AI が倫理的な面にも配慮して活用されるための考え方やルール等が検討されています。

内閣府の「人間中心の AI 社会原則」では、経済発展と社会的課題の解決をめざす “Society5.0” の実現のための重要な要素として AI を扱う上での基本理念が示され、EU でも「信頼できる AI のための倫理ガイドライン」(Ethics guidelines for trustworthy AI) として AI の活用のための倫理的な枠組みが示されています。

民間においても、関係する学会や団体など、AI の開発や利活用に関わる様々な立場から AI の取り扱い指針が整理されつつあります。

このホワイトペーパーでは、株式会社 日立製作所 (以降「日立」) における倫理面に配慮した AI の開発、利用の規範となる「社会イノベーション事業に AI を活用するための AI 倫理原則」(以降、「AI 倫理原則」)、および AI 倫理原則に則した具体的な取り組みについて紹介します。

日立の AI 倫理原則は、50 年以上の実績のある IT (Information Technology : 情報技術) をベースとした AI 技術の知識は勿論のこと、高品質な OT (Operational Technology : 制御、運用技術) や脳科学研究などを通じて培った科学技術における倫理に関する知見も含めて、規範となる考え方、行動を定めたものです。

日立は、多様化、複雑化する課題に迅速にかつ的確に対処し、社会イノベーション事業を推進するために、AI 倫理原則を遵守し、信頼できる AI の開発と利活用を図ることで、人間の尊厳が守られ快適で強靱な持続可能社会の実現と世界中の人々の QoL (Quality of Life) の向上に貢献します。

本書が広く読まれ、みなさまにご参考いただくことを期待しています。

-----  
AI (人工知能) : データ、情報、知識の学習により、自らの出力や処理方法を変化させる機能を持つ処理モジュール、および、このようなモジュールを構成要素として含むシステム

## 1.AI 倫理を取り巻く状況

### 1.1 背景

#### (1) AI は社会にどう貢献するか

これまで日立は、金融や医療、電力や鉄道、水道など様々な分野でのシステム開発を通じて社会にイノベーションを起こし、社会全体の発展に貢献してきました。これらのシステムの多くは、人の判断による高度な操作と、ルールベースによる自動制御の組み合わせで支えられています。

ルールベースの自動制御には、数値化されたルールが必要となります。人の判断による高度な操作は、経験や勘などの暗黙知に頼るところが大きく、「ルールを明文化する」ことが難しいため、自動化の壁となっていました。

AI はディープラーニングといった仕組みの活用によって、これまで熟練者にしか為しえなかった暗黙知を学習し、システムが高度な判断を自律的に行えるようにすることで、この自動化の壁を乗り越えようとしています。

熟練者のような高度な判断をもたらす AI には、移動や生産のほか、社会のあらゆる場面でのさらなる効率化や、高度化をもたらすことが期待されています。エネルギー効率の観点では、自然環境の保護にも貢献できるでしょう。また、システムの無人化、自動化の実現によって危険な場所への人員を配置せざるを得ない状況を減らし、安全にも貢献できると考えられます。

日立においても、人の手では解決できなかった課題、例えば、広大な海洋にあるごみの検出や、血液検査装置の検査精度とスループットの大幅な向上などに AI が絶大な効果を発揮してきました。

最近では、機械学習やディープラーニングの高度化により、人の行動予測や評価のほか、社会インフラのような人々の生活や安全に密接に関わる領域まで AI の利活用の範囲を広げつつあります。

我々の生活を振り返ると、e コマースサイトや SNS でのレコメンデーション機能のほか、カメラによる顔の認識、エレベーターの制御など生活のあらゆる場面で既に AI が活用されており、もはや珍しいものではありません。ビジネスにおいても、AI 自体を特殊なもの、先進的なものと捉える事は過去のものになりつつあり、ビジネスを変えていくために AI を活用することが必然になってきています。

AI が社会のあらゆる場面で当たり前存在するようになるに従い、集められるデータの種類、量は今後も飛躍的に増加し、これらのデータによってさらに様々な恩恵が生まれ、我々の生活の質がさらに向上することが期待されます。

## (2) AI を高度に活用することへの倫理的な懸念

一方、AI の活用が多様化、高度化するに従い、AI には倫理面で様々な問題が指摘されています。

例えば、開発者の意図しない動作により人の生命や健康が脅かされることや、システムが停止し、社会全体に大きな影響が及ぼされることなどの懸念が示されています。また、個人情報扱う分野を中心に AI の判断結果が差別、偏見、格差を結果的に助長する事例も出ており、大きな社会問題となっています。

これらの問題は、主に AI の特徴である学習を通じて「判断を自律的に変化させる」仕組みによって生じます。倫理面から AI を見た場合の懸念点はいくつかありますが、ここでは代表的な例として「AI による自動化、無人化」、「学習データの品質」、「環境変化への対応」、「透明性の確保」という 4 つを挙げます。

1 つめの「AI による自動化、無人化」は、本来は AI の大きなメリットです。技術の進展により、人がこれまで担ってきた作業の中で AI が代替できる割合が飛躍的に拡大しています。一方、これまで人が担っていた意思決定等を AI で自動化、無人化することで、責任の所在が曖昧になる可能性があります。また、人の尊厳や生命に関わる意思決定を AI のみに委ねてもよいか、という点も依然として議論が必要です。

そのため、AI で自動化、無人化して問題ないかを慎重に検討し、責任の所在等を明確化するとともに、必要に応じて人が介在する仕組みの整備や、自動化した処理を定期的に評価、見直すことが求められます。

2 つめの「学習データの品質」については、AI が学習する「正解」データを基に評価基準を変動させることで最適化を図る場合に生じる可能性のある問題です。これまでのルールベースシステムでは、開発時点で定められたルールの上でインプット情報が処理、出力されていたため、その後の出力内容と経緯は予見できるものでした。

一方、学習による最適化を特徴とする AI は、学習データの品質が低い場合、つまり、「バイアス」や、「不適切な内容」をデータに含む場合、開発時点では予見できない結果が後になって出力される可能性があります。

そのため、従来のシステムとは異なる発想でのシステム検討やデータ設計、テストが求められます。

3 つめの「環境変化への対応」は、社会環境の変化によって「正解」の定義が変化することによる問題です。こちらは AI の学習データによる変化とは反対に、それを活用する側の変化を示しています。

2 つめに述べた学習データの品質が作成当時の倫理観で正しいものであっても、運用中に状況が変化した場合、AI が新しい状況に最適化されるまでの間、不適切な判断を下す可能性があります。

社会環境は変化を伴うものであり、学習データにおける「正解」のモニタリングも、入出力データと同様に重要となり、運用管理の方法にも工夫が求められます。

4 つめの「透明性の確保」については、AI で問題が発生した場合の原因の特定が難しい、という問題です。AI は、学習データを基に評価ルールを最適化させるため、出力結果に疑義を指摘された場合でも、その原因や経緯を分析し、改善点や責任の所在を明確にすることが非常に難しいという特徴があります。

現在はこの問題を解決する手法として XAI（Explainable AI: AI を説明するための技術）の検討が進んでいますが、高度な検討が必要で、一般化にはいくつかのハードルが存在します。

### **(3) AI の適用領域の広がり**

先に述べた AI を利活用する上での懸念は、これまでは主に IT 分野で議論されてきたものでしたが、AI の利活用は人の安全や社会インフラに関わる OT の分野でも進みつつあります。

鉄道分野での運行管理やエネルギー分野での送配電計画の最適化、医療や製造分野での機器メンテナンスの高度化など、OT の代表的な領域では、社会生活や人の安全に関わる重要な意思決定は人によるものが中心でした。これからも人の判断の重要性は変わりませんが、今後は AI も意思決定の重要な役割を担うと想定する場合、先述の 4 つの問題には特に配慮する必要があります。

そして、OT では大型機械の制御を伴うものなど、判断次第で重大な死傷事故に繋がりがねないシステムが多くあり、学習によって自律的にルールが変化する AI を短絡的に適用することには懸念があります。そのため、AI の適用にあたっては運用上での信頼が得られる AI 利活用の仕組みが必要となります。

## 1.2 AI 倫理に関わるルール等の状況

AI の活用の広がり懸念等を踏まえ、各国政府や国際機関、民間団体等で、AI 倫理も含めた適切な利活用を推進するためのルール等の検討が進められています。

AI を使う上での基本的な考え方を示す、いわゆる「原則」を整理する動きは以前からあり、最初は 2016 年 9 月の Partnership on AI による信条 (Tenets) だと言われています。2017 年 1 月に The Future of Life Institute (FLI) がアシロマ AI 原則 (Asilomar AI Principles) を発表し、その後、日本や欧州各国等、国レベルでも検討が進みました。我が国では、内閣府において 2019 年 3 月に、Society5.0 を実現するための AI の理念として「人間中心の AI 社会原則」が策定されています。

また、2019 年 5 月には、初めて複数国で合意された OECD による AI 原則 (OECD Principles on AI) が公表され、同年 6 月の G20 首脳会合では、首脳宣言の附属文書として、G20 AI 原則 (G20 AI Principles) が合意されています。

また、このような原則をより詳細に解説する動きもあり、欧州では 2019 年 4 月に欧州委員会にて「信頼できる AI のための倫理ガイドライン」(Ethics guidelines for trustworthy AI) が示されています。我が国でも、総務省の AI ネットワーク社会推進会議において「AI 利活用ガイドライン」が公表されています。

現時点において AI を規制する包括的な法制度はありませんが、欧州委員会の AI 白書 (WHITE PAPER On Artificial Intelligence - A European approach to excellence and trust) に示されたように雇用や遠隔生体認証等、特定の分野に絞った規制を提案する動きがあり、実際に、サンフランシスコ市のように公共機関による顔認証技術の使用を規制する条例を設けている事例も存在します。



表 1: 主な AI 倫理に関する原則、ガイドライン等

名称	発行元	概要
信条 “Tenets” (2016 年 9 月)	Partnership on AI	メンバーが尊重すべき AI への考え方として、「人の利益追求」「関係者の関与」「オープンな調査と対話」など 8 項目を示している。
アシロマ AI 原則 “Asilomar AI Principles” (2017 年 1 月)	FLI	AI が人々に豊かな暮らしをもたらすための研究課題、倫理と価値、長期的な課題として 23 項目で原則を示している。
人間中心の AI 社会 原則 (2019 年 3 月)	内閣府 (日本)	Society5.0 の実現に AI が貢献することへの期待から、理念として尊重する 3 つの価値として「人間の尊厳が尊重される社会」「多様な背景を持つ人々が多様な幸せを追求できる社会」「持続性ある社会」を示している。
信頼できる AI のための 倫理ガイドライン “Ethics guidelines for trustworthy AI” (2019 年 4 月)	欧州委員会	信頼できる AI を実現するための枠組みとして、システムのライフサイクル全体を通じて満たすべき 3 要素として「合法的であること」「倫理的であること」「堅牢であること」を示し、これらの調和は社会の努力によって醸成されることを示している。
人工知能に関する OECD 原則 “OECD Principles on AI” (2019 年 5 月)	OECD	信頼できる AI の責任ある運用のために、「包括成長、持続可能な開発、幸福の追求」、「法、人権、民主主義、多様性の尊重」など 5 項目の原則を示し AI 提供者に推進と実施を求めている。
G20 AI 原則 “G20 AI Principles” (2019 年 6 月)	G20	G20 における AI の原則として、「信頼できる AI のための責任あるスチュワードシップに関する原則」「信頼できる AI のための国内政策と国際協力」を示している。
AI 利活用ガイドライン (2019 年 8 月)	総務省 (日本)	AI の利用者が AI を利活用するにあたり留意すべき事項をまとめ、実現のために講ずることが期待される具体的な措置について、「適正利用」「安全」「セキュリティ」「公平性」など 10 項目の原則で整理されている。
AI 白書 “WHITE PAPER on Artificial Intelligence” (2020 年 2 月)	欧州委員会	AI 活用におけるイノベーションの姿や、信頼のあるエコシステムの構築の姿として、「人による監督」「高信頼性と安全の確保」「プライバシーの保護」などに触れながらリスクと活用のあり方を示している。

## 2.日立の AI 倫理原則について

### 2.1 AI 倫理原則の整備に向けた日立の取り組み

#### (1) これまでの日立における AI 倫理の醸成

これまで、日立は IT、OT 分野での長年に亘る AI 技術の研究や製品開発の蓄積から、社会や企業の様々な課題を解決することに伴い、AI 技術に関わる倫理観の醸成に努めてきました。

日立のデジタルイノベーションを加速する取り組みである“Lumada”においても、世界の変化をデータでとらえ、柔軟に対応する能力と現場の経験を組み合わせる新たな価値を生み出すために AI の利活用における高い倫理観は欠かせない存在となっています。

この高い倫理観に支えられた Lumada の適用分野は多岐にわたり、これまで金融、産業、流通、電力、モビリティなどの分野で多くのソリューションを提供しており、今後、さらなる AI 技術の開発、流通、利活用の加速に努めています。

#### (2) 政府や国際団体など、制度検討への貢献

日立はこれまでの AI の社会実装や研究開発の知見を背景として、公的機関や主要な団体などでの制度検討に貢献してきました。

国内では 2016 年より総務省の AI ネットワーク社会推進会議に参画し、先述の「AI 利活用ガイドライン」を含む、今後の AI のあるべき姿の検討に貢献しています。

国際的には、ISO (International Organization for Standardization:国際標準化機構) のメンバーとして ISO/IEC JTC 1/SC 42 に参画し、国際標準としての AI や機械学習の活用の在り方についての議論を行なっています。また、欧州委員会への意見書提出や、WEF (World Economic Forum:世界経済フォーラム)への参加など、国内外の AI 倫理検討における先端の知見を持続けるよう、日々取り組んでいます。

## 2.2 日立の AI 倫理原則に対する基本的な考え方

ここでは、日立の AI 倫理原則の前提となる、3 つの基本的な考え方を紹介いたします。

- (1) AI が持続可能社会の実現に貢献すること
- (2) AI が人間中心の考え方の下に社会に浸透すること
- (3) AI が社会や環境の変化に長期間に亘り柔軟に対応し続けること

### (1) AI が持続可能社会の実現に貢献すること

「誰一人取り残さない」持続可能で多様性と包摂性のある社会“を実現するため、2015 年 9 月の国連サミットで「持続可能な開発目標」として“SDGs”(Sustainable Development Goals) が採択されました。

日立は、AI の開発や適用においてもこの考え方に倣い、ソリューションや製品の提供を通じて新たな社会、環境、経済価値を創出し持続可能な社会の実現に資する必要があると考えています。

日立は AI が社会全体の多様な課題を解決し、人々の QoL の向上に貢献できるよう、AI を利活用する目的や効果を明確に定め、設計および開発、検証します。

### (2) AI が人間中心の考え方の下に社会に浸透すること

AI を活用することで、プライバシーが侵害されたり、自由、公平、公正等の、個人の権利、尊厳が損なわれることがあってはなりません。日立が推進する社会イノベーション事業は、社会を構成する個人の QoL 向上が目的であるため、AI 活用におけるプライバシーの保護、公平性の実現等は非常に重要となります。

日立は、プライバシー保護に継続的に取り組んでいるほか、日立グループ行動規範において、人権を尊重し、あらゆる企業活動において性別、性的指向、年齢、国籍、人種、民族、思想、信条、宗教、社会的身分、門地、疾病、および障がいなどによる差別をしないことを宣言しています。AI の社会実装においても、この理念にのっとり、人間中心の考え方を根幹に据え、個人の権利や尊厳を侵害しないよう評価、検証を行うことで、AI が社会に受け入れられるよう努めます。

### (3) AI が社会や環境の変化に長期間に亘り柔軟に対応し続けること

人々の価値観や考え方は社会や環境に伴い移り変わります。AI が適用されるシステムにおいては、普遍的な価値観を維持しつつ、人々の価値観の変化に適応した高い価値を提供し続けること、旧弊によって人に不利益をもたらさないことが求められます。

そのため、日立は、長期的に AI の品質が確保できるよう、社会や環境の変化を的確に捉えて AI のライフサイクルを維持管理することが重要と考えています。

## 2.3 AI の社会実装における着眼点

日立は、情報通信を中心とした IT システムと、電力や水道、鉄道などの社会インフラや産業流通などの OT システムを通して社会イノベーションに貢献しています。これまで AI の適用は IT システムが中心でしたが、今後は OT システムへも領域を広げていくことに鑑み、ここでは IT と OT、両者の違いに触れながら具体的な実装を検討する上で日立が取り組んでいる 7 つの着眼点について紹介します。

### ① 安全重視

IT システムと OT システムの区別に関わらず、全てのシステムにおいて安全の確保は日立の普遍的な価値観であり、AI においても、IT、OT それぞれの特性を踏まえた評価と対策が必要です。

特に、発電機や鉄道車両、工作機械などの大型機器の制御に関わる OT システムにおいては、多くの場合で機器の制御が人の安全と密接に結びついています。また、IT 分野においても、医療診断支援システムや電子商取引システムなどは、人の生命や財産に関わる重要な意思決定に影響を及ぼします。

日立では、これまで数多くのシステムを手掛けてきたノウハウを活かし、AI の適用にあたり必要な安心安全が担保されるよう、継続的な対策を講じています。これには、効率より安全を優先する考え方を AI へ組み込むことや、AI の判断に異常が発生した場合に安全のための制御を行なうフェイルセーフ、適切な縮退を行なうフェイルソフト、機能維持を行なうフォルトトレラントなどの考え方の組合せを含みます。

### ② プライバシー保護

個人情報の適切な取扱いは、特に IT システムにおいて重要な要素となります。日立は、個人のプライバシー、利益を侵害しない AI の実現に向けて、長年実施してきたプライバシー保護対策のノウハウを活用し、効果的な対応を継続的に行なっています。様々な場面での AI の適用に伴い、プライバシーの保護は、後述するセキュリティへの対応と同様に今後一層重要となります。そのため、学習データに個人情報を用いる場合はその使用を最小限にすると共に、個人を識別できないようにする加工、データの暗号化やアクセス可能な人員の制限など様々な対策が必要となります。

OT システムへの AI 適用においては、AI が取扱う情報は制御対象機器の温度、電圧など機器の動作に関わるものが中心であるものの、例えばカメラ映像を AI が解析して人の作業を支援するソリューションなどプライバシーへの配慮が必要な場合があり、無視はできません。

### ③ 公平性実現

AI の判断結果が人種や性別などに基づく差別や偏見を助長しないよう、配慮が必要です。前述したように、日立は、日立グループ行動規範において、人権を尊重し、あらゆる企業活動において差別をしないことを宣言しています。

これには学習データの入力と出力双方での配慮を必要とします。日立は、入力時点で学習データの種類や要素別の分量、正解と定義する内容にバイアスを内包させない仕組みを講じています。出力時点においても、AI が学習したデータや正解の定義が、意図しない差別や偏見を生む可能性を考慮し、それぞれのシステムの継続的な特性評価と対策を講じることで、公平性の確保に努めています。これには、AI の設計開発だけでなく、運用も含めた総合的な仕組みの構築を含みます。また、グローバルに求められる水準の確保を前提とすうえ、システムが適用される地域の習慣や特性にあわせ、AI の公平性が保たれるよう、柔軟な対応に努めています。

### ④ 適正開発・利活用

設計された用途や動作条件から逸脱して利用され、本来の目的から異なる使い方をされないよう、日立は、システムの具体的かつ長期的な利活用シーン（ユースケース）を示し、AI の悪用や誤った使い方の可能性の排除に努めています。

AI の適正開発と利活用においては、運用当初だけでなく、社会状況やサービス内容の変化に合わせた機能の見直しが継続的に行なわれることが重要です。

例えば、IT システムにおいては運用のサイクルにおいてサービスそのものを見直す機会が比較的多く、適用されるAIも当該システムの機能拡張やその時点での価値観の変化にあわせた素早い柔軟な対応が可能です。

一方、OT システムでは社会インフラを中心に 24 時間 365 日連続稼働することを前提としたものもあり、運用開始後に AI を見直す機会が限られる場合があります。また、運用のサイクルは一般的な IT システムよりも長い、10～20 年あるいはそれ以上の長期に亘ることも多くあります。そのため、当初の開発時点で長期的に変化しにくい要素に学習データを絞り、学習によって判断結果が変動する範囲を限定するなどの工夫が求められます。

## ⑤ 透明性・説明責任重視

上述の「安全重視」「プライバシー保護」「公平性実現」「適正開発・利活用」や、後段に述べる項目の配慮には細心の注意を以って取り組んでいますが、技術の進歩や社会情勢の絶え間ない変化の中で、完全かつ継続的にそれらを実現することには困難を伴うことを、日立は認識しています。そのため、日立は、AI の判断結果の根拠の検証と、説明などの透明性確保に継続的に努めるとともに、検証結果に基づき将来に向けた改善に取り組めます。これは AI の利活用に対する長期的な安心を提供し、持続可能な社会の実現に AI が貢献するための前提であると考えています。

日立は、日立グループの行動規範において、多様なステークホルダーとの信頼関係を維持、発展させるため公正で透明性の高い情報開示を行なうことを宣言しています。AI は、その特性として出力結果の過程を示すことが非常に難しいという問題を内包していますが、XAI の研究などを通じてこの問題の克服に努め、AI においても説明責任を果たすよう努めます。

## ⑥ セキュリティ重視

AI の適切な運用にあたり、「データが漏えいしないこと」「データが改ざんされないこと」「データが破壊されないこと」といったセキュリティ面の配慮は IT システムと OT システムの双方に重要な要素となります。

日立は、日立グループ行動規範においてサイバー攻撃などの脅威に組織的に取り組み、適切な対策を講じることを宣言しています。

サイバー攻撃に対しては、一般的には IT システムでは機密性の確保が、OT システムでは可用性の確保が特に重要視されるなど、システムの特性に応じた様々な対策が必要です。セキュリティ対策には NIST (National Institute of Standards and Technology: 米国国立標準技術研究所) や NISC (National center of Incident readiness and Strategy for Cybersecurity: 内閣サイバーセキュリティセンター) をはじめ様々な機関からガイドラインが示されており、日立はこれらに加えて長年のセキュリティ対策の経験から、システムの特性に応じた様々な分析と対策を講じてシステム全体の継続的な保護を図っています。

## ⑦ 法令遵守

AI の利活用においては、先述のとおり様々なガイドラインや、関連するルールが存在します。

日立は、AI が適用される国や地域に適用される条約や法令等を遵守した上で AI を実現し、運用します。

### (3) 日立の AI 倫理原則について

ここまで述べた日立の国内外での様々な AI 倫理に関わる団体での活動による知見や、IT および OT システムで求められる着眼点を考慮し、日立はこれらを網羅する形での AI 活用の指針として「社会イノベーション事業に AI を活用するための AI 倫理原則」（以降、「AI 倫理原則」）を策定、運用しています。

#### ■日立の AI 倫理に対する考え方

日立の AI 倫理原則を述べるにあたり、前文として日立の AI 倫理の考え方を述べています。日立が AI の活用によって社会課題の解決や QoL 向上への貢献をめざしていること、日立の長年の IT、OT システムの実績と AI の知識、脳科学研究などを通じて培った科学技術における倫理に関する知見を統合したものととして AI 倫理原則を定めていることを示しています。

表 2: 日立の AI 倫理原則 前文

日立製作所（以下、「日立」）は、社会イノベーション事業に AI を適切に活用することにより、社会価値・環境価値・経済価値を一層向上するとともに、人間の尊厳に立脚した快適で持続可能な社会の実現と、世界の人々の Quality of Life (QoL) の向上に貢献することをめざしています。そのために、日立は、100 年以上にわたり取り組んできた高品質な OT（Operational Technology、制御・運用技術）に、50 年以上の実績のある IT（Information Technology、情報技術）をベースとした AI 技術の知識と、脳科学研究などを通じて培った科学技術における倫理に関する知見を統合し、AI の利活用を推進します。日立は、AI 倫理原則として、AI の計画・社会実装・維持管理の 3 つのフェーズにおける行動規準と、全てのフェーズに共通する 7 つの実践項目を定め、この原則に基づき AI を利活用します。

### ■ AI 活用における 3 つの行動規準

日立の AI 倫理原則では、2.2 項で述べた考え方を基に AI の開発と実装、長期的な利活用の規準として以下の 3 つの行動規準を掲げています。AI の活用は社会イノベーション事業の推進にあたり必要不可欠であり、潜在する様々な AI のリスクを抑制して高いレベルでの安心、安全を継続的に実現するための方針として定めたものです。

表 3: 日立の AI 倫理原則における 3 つの行動規準

#### **I 持続可能社会実現のために、AI の開発、利活用を計画します**

AI の潜在的な倫理的リスクを抑制し、新たな価値の創出を図るには、計画の時点において、サービス、ソリューションやプロダクトに AI を利活用しようとする目的が適切であることが重要です。日立は、社会の多様な課題を解決し、快適で強靱な持続可能社会の実現や、世界中の人々の QoL 向上のために AI の開発と利活用を計画します。

#### **II 人間中心の視点で、AI を社会実装します**

AI の判断結果が、個人を尊重し、社会の利益に資するよう、責任をもって AI を社会実装し、AI と人間との共生を可能とすることが重要です。日立は自由、公平、公正の理念にのっとり、人間中心の視点で AI の社会実装を行うとともに、AI が想定通りに機能するよう検証に努めます。

#### **III 提供価値が長期間にわたり持続するよう、AI を維持管理します**

AI の社会実装後も、AI が長期間かつ安定的に価値を提供し続けることが重要です。日立は、AI の提供価値が社会や環境の変化に応じ継続的に受け入れられるよう維持管理に努めます。



## ■ AI 倫理原則を遵守するための 7 つの実践項目

先述の 3 つの行動規準の遵守にあたり、2.3 項で述べた着眼点を基に 7 つの実践項目を定め、実際の IT および OT システムの研究および開発などに適用しています。

表 4: 日立の AI 倫理原則を遵守するための 7 つの実践項目

<p><b>1. 安全重視</b></p> <p>AI および AI を活用したシステムやソリューションが想定通りの動作をするよう品質を検証し、利用者やそれに関係する者の生命や健康をはじめとする人権、財産、名誉、信頼、信用を守ることに努めます。また、同時に、地球環境の破壊あるいは悪化を防止し、人々に安全な暮らしを提供する AI の実現、運用に努めます。</p>
<p><b>2. プライバシー保護</b></p> <p>AI の学習、評価、運用に利用する入力データおよび AI が出力するデータに関して、個人情報適切に扱い、プライバシーを含む権利を保護するよう AI の実現、運用に努めます。</p>
<p><b>3. 公平性実現</b></p> <p>AI の判断結果が、多様なステークホルダーの利益に資するとともに、人種、性別、国籍などによる差別や偏見を発生させたり助長させたりすることがないよう AI の実現、運用に努めます。</p>
<p><b>4. 適正開発・利活用</b></p> <p>AI が設計された用途や動作条件から逸脱した使い方をされないよう、AI の具体的な利活用シーン(ユースケース)におけるリスクポテンシャルを踏まえた開発を行い、AI の適正な利活用に努めます。また、利用者や運用者等に対し利用方針、利用条件等を示すことで AI の適正な運用に努めます。さらに、AI が利用される動作環境の変化、AI やその判断結果に対する人々の意識や社会状況の変化などを随時確認し、AI の適切な維持管理に努めます。</p>
<p><b>5. 透明性・説明責任重視</b></p> <p>AI の判断結果の根拠などを検証し説明できるよう AI の透明性確保に努めるとともに、AI やその判断結果に関して用途や状況に応じ、説明責任を果たすよう努めます。</p>
<p><b>6. セキュリティ重視</b></p> <p>情報漏洩、改ざん、システムの破壊、サービスの妨害などを防止するよう、システムや運用レベルでの対策も含め、セキュリティを重視した AI の実現、運用に努めます。</p>
<p><b>7. 法令遵守</b></p> <p>利用される国や地域の法令を遵守した、AI の実現、運用を行います。</p>

### 3. 日立の AI 倫理原則を遵守するための取り組み

#### 3.1 AI 倫理原則の具体的な適用

AI 倫理原則を運用するために、日立は社会イノベーション事業の先進の協創プロジェクトが多く集まる Lumada Data Science Lab (以下“LDSL”) の専門チームにおいて、実際のシステムへの適用に向けた支援や、社内教育に取り組んでいます。

具体的には、AI 倫理原則に沿った AI の導入等が行われるように事前評価を行うとともに、そのライフサイクルを通じた適正な運用に努めています。評価を効率的、効果的に進めるためにチェックリストや関連ドキュメントの整備を図り、後述するように社内の普及啓発を推進しています。

また、専門チームの運営においては、日立内のプライバシー保護諮問委員会<sup>※1</sup>が協力することでプライバシー保護に関わる専門知識を効率的に活用するとともに、AI 倫理コンサルティング<sup>※2</sup>を展開している株式会社日立コンサルティングのメンバーも参画し、日立グループ全体の AI 倫理の知見を集約しています。

さらに、AI 倫理に関する取り組みを客観的に評価し、より良い運用を図るため、外部有識者から構成される AI 倫理アドバイザリーボードを設置しています。AI 倫理アドバイザリーボードは、案件への示唆だけでなく、AI 倫理原則に係る取り組みの改善等についても助言を行なっております。<sup>※3</sup>。

(参考) AI 倫理アドバイザリーボード メンバー (2021 年 2 月現在、五十音順)

新保 史生	(慶應義塾大学 総合政策学部 教授)
鈴木 晶子	(京都大学大学院 教育学研究科 教授)
中川 裕志	(理化学研究所 革新知能統合研究センター チームリーダー)
長田 三紀	(情報通信消費者ネットワーク 代表)
森 亮二	(英知法律事務所 弁護士)

※1 パーソナルデータの利活用における日立のプライバシー保護の取り組み (日立製作所)

[https://www.hitachi.co.jp/products/it/bigdata/bigdata\\_ai/personaldata\\_privacy/](https://www.hitachi.co.jp/products/it/bigdata/bigdata_ai/personaldata_privacy/)

※2 AI 倫理コンサルティング (日立コンサルティング)

[https://www.hitachiconsulting.co.jp/business/public/ai\\_logic/](https://www.hitachiconsulting.co.jp/business/public/ai_logic/)

※3 AI 倫理アドバイザリーボードは 2021 年 2 月に設置しており、AI 倫理原則等の作成時には関与していませんが、今後、ボードからの助言を踏まえ、原則を含む AI 倫理の取り組み全般の改善を図る予定です。

### 3.2 AI 倫理の社内教育について

AI の社会実装に適切な倫理観を導入するには、組織や制度の整備だけでなく、AI に関わる者それぞれが AI 倫理の重要性を理解し、実践することが重要です。

そのため、AI 倫理の社内普及活動として、AI 倫理原則の説明と運用のためのチェックリストやフロー、関連ドキュメントについて、社内教育を推進しています。ここには、AI の定義のほか、差別や透明性などの定量化の難しい項目の一定の指針や、システム検討についての考え方を含みます。

また、AI によって引き起こされ得る問題や社会に与える影響のほか、法制度やビジネス動向、考えられる対策を「誰が」「いつ」「どのように対策すべきか」の観点で実践的に教育することで、AI 倫理の重要性について社員の理解を深めています。

特に、OT システムにおいては公平性や透明性の観点が従前のシステム検討では入りにくいことから、想定事象も含めた問題事例のケーススタディも交えて丁寧な教育に取り組んでいます。

この取り組みは有識者による講義のほか、広く啓発するための e ラーニングも含めて取り組んでいきます。そして、国内だけでなくグローバルに展開し、各国の実情に合わせた適切な AI の実装ができるよう努めていきます。

将来的には、お客様や社外への啓発活動など、AI 倫理の考え方を日立だけのものではなく、社会全体に還元しより良い AI 社会を実現させるため活動に発展させることをめざしています。

### 3.3 AI 倫理原則の今後の改訂方針

AI 倫理の考え方は社会情勢などによって変化するため、各国のルールや発生事象にあわせた見直しを逐次図ることが重要です。

日立では、おおむね半期に一度を目安に、それまでの適用状況や日立内外で問題となった事象、国や国際機関、業界でのルール策定や改訂の状況を鑑みて LDSL を中心に AI 倫理原則の見直しを図ります。

改訂した場合には都度公開し、日立の最新の取り組みを広くご理解いただくと共に、AI が倫理の観点からより良いものへとなっていくよう努めてまいります。

## おわりに

本書は、既にも人々の生活に欠かせない存在にまで成長している AI 技術のさらなる発展にあたり、日立の AI 倫理に対する考え方と取り組みについてまとめたものです。

AI は今後も多様な用途で活用が進み、AI を取り巻く倫理観も変化し続けることが見込まれます。

日立は、社会イノベーション事業に携わるものとして、今後も高い倫理感を持って AI 技術の開発と適用に取り組んでまいります。また、AI における倫理の考え方は広く共有し、社会全体により有益な AI が浸透するよう、今後も多様な意見を取り入れながら、継続的に改善を図っていきます。

## 出典

### **EC: Ethics Guidelines for Trustworthy AI**

<https://ec.europa.eu/futurium/en/ai-alliance-consultation>

### **EC: White Paper on Artificial Intelligence: a European approach to excellence and trust**

[https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust\\_en](https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en)

### **FLI: アシロマの原則**

<https://futureoflife.org/ai-principles-japanese/>

### **G20 : G20 AI Principles**

<https://www.g20-insights.org/wp-content/uploads/2019/07/G20-Japan-AI-Principles.pdf>

### **OECD: OECD Principles on AI**

<https://www.oecd.org/going-digital/ai/principles/>

### **内閣府: 人間中心の AI 社会原則**

<https://www.cas.go.jp/jp/seisaku/jinkouchinou/pdf/aigensoku.pdf>

### **総務省: AI 利活用ガイドライン**

[https://www.soumu.go.jp/main\\_content/000624438.pdf](https://www.soumu.go.jp/main_content/000624438.pdf)

### **日立製作所 : パーソナルデータの利活用における日立のプライバシー保護の取り組み**

[https://www.hitachi.co.jp/products/it/bigdata/bigdata\\_ai/personaldata\\_privacy/](https://www.hitachi.co.jp/products/it/bigdata/bigdata_ai/personaldata_privacy/)

### **日立コンサルティング : AI 倫理コンサルティング**

[https://www.hitachiconsulting.co.jp/business/public/ai\\_logic/](https://www.hitachiconsulting.co.jp/business/public/ai_logic/)

## 本書について

本書は、2021年2月現在、社会イノベーション事業の先進の協創プロジェクトが多く集まる「Lumada Data Science Lab.」を中心に、AIを倫理的に活用するための取り組みについてまとめたものです。

本書に記載されている情報は、今後、予告なく変更されることがあります。

---

### 社会イノベーション事業にAIを活用するための、日立のAI倫理への取り組み

---

2021年2月22日 初版発行

著者 株式会社 日立製作所 Lumada Data Science Lab.  
株式会社 日立コンサルティング

編集・発行 株式会社 日立製作所 Lumada Data Science Lab.  
〒185-8601 東京都国分寺市東恋ヶ窪一丁目 280 番地

---