



一橋大学イノベーション研究センター

東京都国立市中2-1
<http://www.iir.hit-u.ac.jp>

本ケースの著作権は、筆者もしくは一橋大学イノベーション研究センターに帰属しています。本ケースに含まれる情報を、個人利用の範囲を超えて転載、もしくはコピーを行う場合には、一橋大学イノベーション研究センターによる事前の承諾が必要となりますので、以下までご連絡ください。

【連絡先】一橋大学イノベーション研究センター研究支援室
TEL: 042-580-8423 e-mail: chosa@iir.hit-u.ac.jp

社会規範の差異が人工知能（AI）の規制・イノベーションに与える影響

～欧州 AI 動向から見る知的対話システムの倫理的リスクに係る地域的差異～

2021年9月10日

一橋大学イノベーション研究センター

市川類

概要

技術・イノベーションは、一般的に、人間社会に対して大きなメリットをもたらす一方、その利用方法によっては、人間社会の有する社会規範（倫理）に抵触する可能性がある。このため、必要に応じ、当該技術・イノベーションに対して何らかの規制・制度が構築され、規律（ガバナンス）がかけられることになる。

その際、デジタル・AI技術については、特にプライバシー、公平性などの人権に係る社会規範（倫理）への影響が指摘されており、このため、近年、いわゆるAI倫理として、世界各国・機関においてAI原則が制定されてきている。その際、このような社会規範は、国・地域によって異なるため、AI原則はもちろんのこと、各国・地域によって、その規制・制度が異なるものとなる可能性がある。

このような認識のもと、本ワーキングペーパーでは、特に「人間-機械」関係に係る社会規範に係る欧州と日本とでの差異が、両地域における規制・制度動向にどのような影響を与えているかについて考察することにより、AI技術の一部であるとともに「人間-機械（AIシステム）」のインターフェースとして今後重要な役割を担う「知的対話システム」について、今後その開発・利用において生じ得る倫理面でのリスクの可能性について明確化する。

具体的には、まずは、知的対話システムに係る今後の技術の発展方向を見据えた上で、次に、「人間-機械」関係に係る欧州と日本の文化上の差異によって、欧州、日本それぞれのAI原則に異なった特徴が示されていることを分析する。その上で、これらを踏まえた上で、最近発表された欧州のAI法案を分析することにより、将来的な知的対話システムの開発・利用においては、公平性、プライバシーなど世界的に比較的共有された人権に係る視点に加え、「人-機械」関係に係る文化の差異によって生じるリスクにも留意する必要があることを示す。また、今後のAI規制及び社会規範に係る国際的な調和も含めた方向について考察する。

目次

概要	1
目次	2
1. 問題意識	3
2. 知的対話システムとその発展の方向	5
(1) これまでの対話型システムの位置づけ	5
(2) 知的対話システムの発展の経緯と現状	6
(3) 今後の知的対話システムの方向	8
3. AI原則に見る「人-機械」関係と文化の影響	9
(1) AI原則の流れとその体系・構造	9
(2) AI原則に見る「人間-機械」関係：欧州と日本の考え方の違い	11
(3) 「人間-機械」関係に係る文化の影響（ヒューマノイドロボットの事例）	12
4. 欧州のAI法案の構造と知的対話システムへの影響	15
(1) 欧州のAI（規制）法案の経緯と概要	15
(2) 知的対話システムの位置づけとその背景	20
(3) 知的対話システムの開発に係る倫理的・規制リスク	22
5. 今後の課題／規制とイノベーションの共進化	23
(別添) 欧州AI法案（付属書Ⅲ）	26

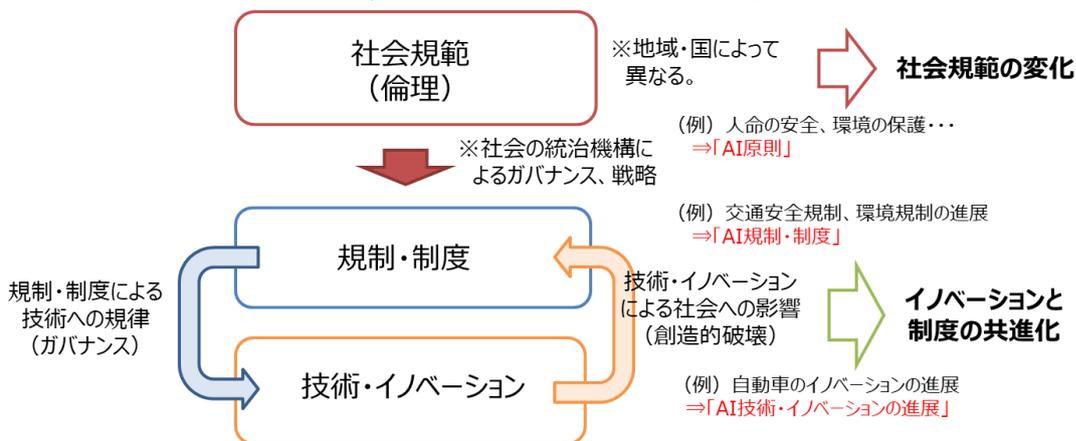
1. 問題意識

＜検討の枠組み：技術・イノベーションと制度・規制の共進化＞

技術・イノベーションは、一般的に、経済社会の発展や国民生活の向上など人類に対して、多くのメリットをもたらす一方で、その利用方法によっては、これまで人類がコミュニティとして培ってきた、安全性、人権の保護などの社会規範（倫理）に抵触する可能性がある。

その場合、当該国・地域（コミュニティ）における政治的な統治（ガバナンス）機構に依るものの、必要に応じ、当該技術・イノベーションに対して何らかの規制・制度を構築し、規律（ガバナンス）をかけることになる。また、その規律がその後の当該技術・イノベーションの進展や方向にも影響を与えることになり、その結果、長期的に見た場合は、技術・イノベーションと制度・規制の共進化が進展することになる。

図1：社会規範と技術・イノベーションと規制・制度との関係



その際、その社会規範（倫理）としては、これまで、特に製造・エネルギー技術などを中心に、安全性（生命、身体の保護）に係るものが中心であった（広義には環境の保護なども含む）のに対し、近年急速に進展・普及しつつあるデジタル・人工知能（AI）技術においては、プライバシーの保護、公平性・非差別性の確保などを含む基本的人権に係る社会規範（倫理）も大きな議論になりつつある。このような中、近年、世界各国・機関において AI 原則を制定する動きがある。

一方、この社会規範とは、必ずしも人類で共有している唯一のものがある訳ではなく、歴史的文化的な経路依存的経緯に基づき、各コミュニティそれぞれにおいて異なるものが醸成されてきている。特に、AI 関連では、ヒューマノイドロボットを事例として、「人間・機械」関係に係る認識（社会規範）が、欧州と日本とでは異なることが指摘されている。

本ワーキングペーパーにおいては、このような社会規範の差異に伴い、国・地域によって技術イノベーションに係る異なる規制・制度が構築され得ることを示すことにより、今後、イノベーションに与える影響について考察する。

<事例：知的対話システムに係る倫理的リスクの地域的差異>

上記の枠組みのもと、本ワーキングペーパーでは、その事例として、AI技術のうち、特に、知的対話システムを取り上げる。近年のAI技術等の進展に伴い、知的対話システムの技術レベルが急速に向上してきている。それに伴い、チャットボットなどのビジネスでの導入・普及や、音声アシスタント・AIスピーカー、動物型対話システムなど製品・サービスの導入・普及などが進展しつつあり、今後、知的対話システムは、従来にも増して、より人間らしい振る舞いを行うようになることが見込まれる。

この知的対話システムは、社会のデジタル化・DX化が進む中で、単に、単体として社会に導入されるのではなく、様々なビジネスモデルの中で企業・社会のデジタルシステムの中に組み込まれ、「人間」と「機械」（AIシステム）のインターフェースとして役割を果たすことになる。その結果、人間と深く関わることでゆえに、特に、「人間」と「機械」（AIシステム）との関係を考慮した倫理問題に対応する必要があるが生じる。

しかしながら、このような「人間」と「機械」のインターフェースである知的対話システムが、今後の技術の発展に伴い、倫理問題に関してどのようなリスクを生じさせる可能性があるのかについては、ほとんど分析がなされてきていないのが現状である。その際、特に、人間と機械（AIを含む）の関係に対する考え方は、上述の通り、その地域で歴史的に醸成されてきた文化によって異なるため、日本の常識的な倫理観が必ずしも世界で通じる訳ではない。

このような問題意識の下、特にAIの倫理問題・人権問題に関して関心の高く、議論をリードしている欧州におけるAI原則に関して、日本のAI原則との比較を通じ、その特徴を分析するとともに、最近発表されたAI法案を分析することにより、今後の知的対話システムの技術発展によって生じ得る倫理面でのリスクの可能性について明確化する。

<本ワーキングペーパーの構成>

このため、まず、第2章において、これまでの対話型システムの設計に係る原則について触れた上で、最近の知的対話システムの種類について、対話者・媒体の範囲（マルチモーダル／参加者数）の視点と、対話可能領域の範囲（目標／タスク／ドメインの範囲）の視点の二軸に分類し、今後の知的対話システムの発展方向を見据える。

その上で、第3章において、これまでの世界各国のAI原則の流れを振り返るとともに、特に「人間」と「機械（AI）」との関係において、欧州のAI原則は、日本のAI原則と比較してどのような特徴を有しているか、また、その文化的背景について、「人間」と「ロボット」との関係に係るこれまでの議論を踏まえ分析する。

それらの議論を踏まえ、第4章において、最近、欧州で提案されたAI法案に関し、その経緯及び体系を整理した上で、本AI法の各条文が、今後の知的対話システムの発展に対して、どのような倫理面でのリスクの可能性があるかにつき、その背景も含めて考察する。その結果、将来的な知的対話システムの開発・実用化においては、公平性、プライバシーなどに係る視点に加え、「人-機械」関係に係る文化の差異によって生じるリスクにも留意する必要があることを示す。

また、それらを踏まえた上で、今後技術のグローバル化が進み、イノベーションと制度の共進化が見込まれる中、これらの規制に係る国際ハーモナイゼーションや国際的な社会規範の変化の方向について、議論・考察を行う。

2. 知的対話システムとその発展の方向

(1) これまでの対話型システムの位置づけ

デジタル技術の進展の歴史の中で、「知的対話システム」以前に、従来より、Human・Computer Interaction の位置づけのもとで、「対話型システム」に係る研究が進められてきている。一般的に、対話型システムとは、人間（ユーザ）が設定した目標に対して、ユーザとシステムが対話をしながら作業を進めていく仕組み¹として位置づけられ、実際に、ISO (2006)²では、「対話」とは「目標達成に向けた、人とインタラクティブシステムとの相互作用（ユーザーによる情報入力行為及びシステムによる応答の連鎖）」と定義している。

このような対話型システムの設計において、人間工学等の観点からの望ましい視点として、シュナイダーマンによる対話型システムの設計原則（8つの黄金律）³や、ISO (2006)による「人間工学-人とシステムとのインタラクション-対話の原則」（7つの原則）⁴が定められている。また、ISO (2019)⁵では、対話型システムに係る、人間中心設計の原則、計画、活動等について定められている。これらの対話型システムに係る原則は、基本的に、人間（ユーザ）が、自らの意図・目的の達成に向けてコンピュータを操作するにあたって、より人間が使い易くするため、対話形式による画面での表示等の手順などを定めたものである。

¹ <https://www.atmarkit.co.jp/fwcr/rensai/usability06/01.html>

² ISO 9241-110 (2006) 「人間工学-人とシステムとのインタラクション-対話の原則」
<https://kikakurui.com/z8/Z8520-2008-01.html>

「対話 (dialogue)」：目標達成に向けた、人とインタラクティブシステムとの相互作用（ユーザーによる情報入力行為及びシステムによる応答の連鎖）。

「インタラクティブシステム (interactive system)」：ユーザーが遂行する仕事の能率を上げるように、ユーザーからの情報入力を受け取り、ユーザーに出力の伝達を行う、ハードウェア及びソフトウェアの組合せ。

³ ベン・シュナイダーマン (1995) 『ユーザーインタフェースの設計 やさしい対話型システムへの指針 (原題: Designing the User Interface: Strategies for Effective Human-Computer Interaction)』
<https://uxmilk.jp/64295>

岡田謙一、西田正吾、葛岡英明、仲谷美江、塩澤秀和「IT Text ヒューマンコンピュータインタラクション (改訂2版)」 (2016)
(参考 HP)

<https://www.atmarkit.co.jp/fwcr/rensai/usability06/01.html>

⁴ ISO 9241-110 (2006) 「人間工学-人とシステムとのインタラクション-対話の原則」
<https://kikakurui.com/z8/Z8520-2008-01.html>

「対話の原則」：① 仕事への適合性、② 自己記述性、③ ユーザーの期待への一致、④ 学習への適合性、⑤ 可制御性、⑥ 誤りに対する許容度、⑦ 個人化への適合性

⁵ ISO 9241-210:2019 「人間工学-人とシステムとのインタラクション-第210部：対話型システムの人間中心設計」

https://webdesk.jsa.or.jp/books/W11M0090/?bunsyo_id=ISO%209241-210:2019

https://webdesk.jsa.or.jp/books/W11M0090/?bunsyo_id=JIS%20Z%208530:2021

具体的な設計に係る原則としては、①「ユーザ、タスク及び環境の明確な理解に基づく設計」、②「ユーザは設計及び開発の全体を通して関与」、③「ユーザの視点からの評価に基づいて設計を方向付け、改良」、④「プロセスを繰り返す」、⑤「ユーザエクスペリエンスを考慮して設計」、⑥「様々な専門分野の技能及び視点をもつ人々を設計チームに加える」を規定している。

図 2：対話型システムの設計原則

＜シュナイダーマンの対話設計に係る 8 つの黄金律＞
『ユーザーインターフェースの設計 やさしい対話型システムへの指針』（1995）

8 つの黄金律	概要
達成感を与える対話の実現	操作をやり遂げた満足感、安心感を与えることは新たな行動への推進力になる。
短期記憶の負担の減少	短期記憶には限りがあるので、その容量に見合うように表示方法を工夫する。
フィードバックの提供	すべての操作結果に対して、状況変化を提示する必要があるが、実行頻度と実行の影響度により応答の情報量を変化させることが望ましい。
一貫性	一貫した操作手段、同一の用語の使用、コマンド形式の統一などの類似した状況に対して常に同じ対応が取れるようにする。
ショートカットの用意	上級者のために、省略形、特殊キー、隠しコマンド、マクロ機能などのショートカットを用意し、対話の回数や入力の数をおくし、メッセージをスキップすることにより応答時間の短縮や表示速度を向上させる。
主体的な制御権の提供	ユーザを応答者としてではなく主体的な操作者として取り扱う。ユーザを不安や不機嫌にするような応答や要求をしてはならない。
逆操作	可能な限り操作を可逆にすることにより、エラー回復が容易になると同時に安心感が提供され、ユーザの試行錯誤が容易となる。
簡単なエラーの処理	システムによる早期のエラー検出を行い、単純で分かりやすいエラー回復方法を提供する。そして回復が不可能となるような致命的なエラーは起きないようにする。

＜ISOにおける対話の7原則＞
ISO 9241-110 (2006)「人間工学—人とシステムとのインタラクション—対話の原則」

7 つの原則	概要
仕事への適合性	ユーザーが仕事を完了するうえでの助けとなり完了を促進する。
自己記述性	ユーザーがシステムとの対話において、自分が何についての対話をしているか、対話のどのステップにいるのか、どのような操作が許されてどのように操作を実行すればよいかを常に明らかである。
ユーザの期待への一致	対話が状況に応じて予想されるユーザの必要性及び広く受け入れられている習慣と調和している
学習への適合性	対話において、ユーザーがシステムの使い方を学習することを支援しその案内を与える
個人化への適合性	個人の能力及び必要性に応じてインタラクション及び情報の提供を変更できる
可制御性	ユーザーが目標を達成するまで、やり取りの方向及びペースを主導し制御できる
誤りに対する許容度	入力で明らかな間違いがあったにもかかわらず、ユーザーによる最小限の修正で意図する結果が得られる

(注) いずれも、順序を並び替えている。

これに対し、今回対象とする知的対話システムも、人間とコンピュータ間のインターフェースという意味では、対話型システムと同じであり、また、特に知的対話システムの初期段階のものは、いわゆる現在の対話型システムとほぼ同じであると考えられる。しかしながら、知的対話システムでは、その対話形式による「手順」を、より「知的」にすることにより、ユーザの入力媒体手法などを拡大するとともに、単なるコンピュータ・端末の操作という目的を超えて、対話可能領域の更なる拡大されたシステムであると位置づけられる。

(2) 知的対話システムの発展の経緯と現状

もともと、人工知能研究の多くは、「人間が知能を使って行う活動」をシステム（機械）にさせようとする研究⁶である。その際、人間の「対話」という活動は、人間の知能を使って行う本質的な活動の一つであることから、「知的対話システム」は、当初から、人工知能研究の主要な対象として位置づけられてきた。

具体的に、AIによる知的対話システムの研究は、MITのジョセフ・ワイゼンバウム氏が1964～1966年に書いたイライザ（Eliza）に始まるとされる⁷。このイライザは、音声入力ではなく、また、スクリプト型（パターンマッチング型）の動作をするという非常にシンプルな設計であり、今からみれば、機能は非常に限定されている。しかしながら、このイライ

⁶ 人工知能学会「人工知能って何？」

「人工知能の研究には二つの立場があるからです。一つは、人間の知能そのものをもつ機械を作ろうとする立場、もう一つは、人間が知能を使ってすることを機械にさせようとする立場です」「そして、実際の研究のほとんどは後者の立場にたっています」

<https://www.ai-gakkai.or.jp/whatsai/AIwhats.html>

⁷ SPJ (2020/1/4)「対話システムを構成する 2 つの仕組みと、フレームワークとは？」

<https://spjai.com/interactive-system/>

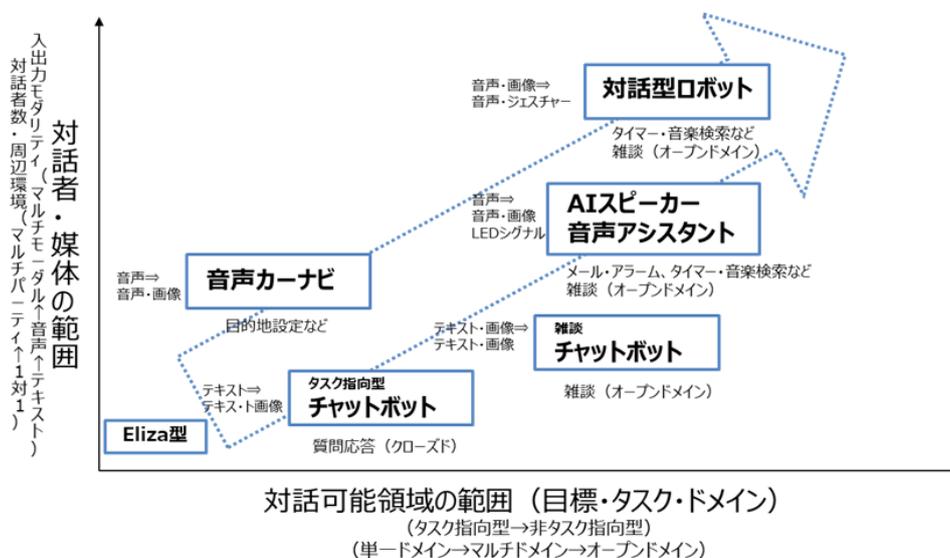
Wikipedia “ELIZA” <https://ja.wikipedia.org/wiki/ELIZA>

ずは、後の全ての対話型システムの土台となったとされ、その後、「人工無能」と言われつつも、イライザ型の対話型システムの普及は進展し、また、その後のエキスパートシステムの開発・普及に伴い、提供する情報の対象範囲やその量の拡大により、機能が拡充してきた。

一方、近年、インターネット接続と検索技術の進展による提供する情報の量・範囲が圧倒的に増大してきたことに加え、自然言語理解・文章生成技術や機械学習技術などの AI 技術の進展、特に音声処理技術の進展により、知的対話システムの機能は急速に向上しつつある。そのような中、実際に、チャットボット⁸や音声アシスタントなどに代表される、知的対話システムが急速かつ多く実用化されつつあり、今後の市場規模の大幅な拡大も見込まれている⁹。

このような知的対話システムの機能の発展を、対話者・媒体の範囲（入出力のモダリティ、参加者の数等）と、対話可能領域の範囲（目標・タスクの対象範囲、ドメインの範囲等）で分類すると、図3の通り。

図3：知的対話システムの発展の方向¹⁰



⁸ チャットボットの最近の実用化の動きについては、以下を参照。
日経 XTech（佐藤 雅哉）「何でも答える「自動対話 AI」人手不足、解消の切り札」（出典：日経コンピュータ 2017年5月11日号 pp.34-41）
<https://xtech.nikkei.com/it/atcl/ncd/17/050100019/>

⁹ 山本俊樹（野村総合研究所）「対話型 AI システムの業務適用の進め方」『IT Solution Frontier』2018.10

https://www.nri.com/-/media/Corporate/jp/Files/PDF/knowledge/publication/it_solution/2018/10/ITSF181003.pdf

三竹 保宏（NTT コミュニケーションズ）「対話型 AI の現状と未来」『三竹レポート：AI 活用の現場から』（2019年2月号掲載）

https://www.bcm.co.jp/solution-now/cat-solution-now/2019-02_1674/

¹⁰ 出典：筆者作成（中野幹生（ホンダ・リサーチ・インスティテュート・ジャパン）「対話システム構築入門 2018年度人工知能学会全国大会（第32回）チュートリアル」（2018年6月5日）p11 - p12 を参照）

https://www.ai-gakkai.or.jp/jsai2018/wp-content/uploads/2018/06/tutorial_nakano_jsai2018.pdf

また、形態としても、単体コンピュータの接続された端末から、インターネットと接続された端末（スマホ、AIスピーカーなど）へ、さらに、一部には、ロボット機能も含めた専用端末（接客用ロボット、動物系ロボットなど）に発展していく方向にある。

（３） 今後の知的対話システムの方向

現時点の知的対話システムについては、まだ「人間にははるかに遠い」という状況にあると考えられるが¹¹、更なるAI技術の進展により、今後とも、その能力は確実に向上し、また、利用範囲も広がることが見込まれる。その際、将来の知的対話システムの方向を完全に予想することは困難であるが、上記図3で示した2軸に従うと、以下のような議論を行うことができる。

<(3)-1. 対話可能領域の拡大：対話から会話へ>

従来の対話型システムは、上述のとおり、原則、目標／タスク指向型であり、また、音声カーナビ、タスク指向型チャットボットなど、現在実用化されている多くの知的対話システムも、基本的には、プロバイダー（企業）が設定した範囲（ドメイン）内において、ユーザの考える「目標」（タスク）を実現すべく対話を行うという目標・タスク指向型である。

一方、近年の知的対話システムにおいては、この対応可能領域（ドメイン）の範囲が拡大するとともに、このようなタスク指向型のシステムに加えて、いわゆる雑談型のチャットボットや音声アシスタントなどの「非タスク型」のシステムが開発・導入されつつある。今後、動物型ロボットなどの形態への発展の動きとも融合し、「娯楽・癒し」「教育・コーチング」などの様々な分野も含めて、様々な非タスク型のシステムの開発が進展することも想定される。また、このようなシステムにおいては、従来よりも長い時間の「会話」が可能に続くようになるとともに、単にユーザの質問に回答するだけでなく、状況に応じて、システム自らがユーザに対して話しかけるようになることも考えられる。

これらの動きに伴い、導入されるビジネスモデルも変化しつつある。これまでの企業における知的対話システムの導入は、企業の既存の特定のサービスの提供に係るユーザに対するインターフェースとして利用するものが大半であった。これに対し、近年は、非タスク型を中心に、対話自体そのものを一種の主要サービスとして位置づけて、提供する事例も多くみられつつあり、場合によっては、当該知的対話システムが、企業の他のサービスなどに係る購買等のプロモーションのツールとして、利用されることも可能となりつつある¹²。

<(3)-2. 対話者・媒体の拡大：パーソナル化>

従来の対話型システムにおいては、ユーザが入力したテキストによる特定の指示に対して、特定の応答をするものが中心であった。その後、近年の音声処理技術の進展により、音声による入力に対して応答する知的対話システムが急速に導入・普及しつつある¹³。

¹¹ 狩野 芳伸（静岡大学）「コンピューターに話が通じるか 対話システムの現在」『情報管理』2017 vol.59 No.10 pp658-665

https://www.jstage.jst.go.jp/article/johokanri/59/10/59_658/pdf-char/ja

¹² 実際に、例えば、アマゾン社の Alexa は、アマゾン社の購買サービスのインターフェースとしても機能している。

¹³ 駒谷和範（大阪大学 産業科学研究所）「音声対話システムの構成と今後」『パテント 2019』Vol. 72 No.8 (p92-p101)

今後は、ジェスチャー、仕草・表情、周辺状況などの画像認識による入力に対応した応答や、ロボットなど物理的動作・仕草による出力などのマルチモーダル化が更に進展すると考えられる。特に、テキストや音声でもある程度は可能ではあるが、特に画像認識の進展により、人間の属性や感情を把握した上で応答する対話システム¹⁴も登場することが想定される。

また、既に多くの対話型システムでは、別途照合される ID 等を通じて、個人を見分けることが可能であるが、今後の知的対話システムでは、音声認識に加え、顔認識を初めとする画像認識等を通じて、ユーザが ID を入力しなくとも個人を特定することが可能となり、その結果、パーソナル化された対話が行われることになる。その際、システムを通じてやりとりされた多量の個人データは、上述の属性や感情に係る情報も含めて、ビッグデータ解析を通じて最適化され、当該個人に対して最適な対話を行うようなシステムが構築されていくことになる。さらに、これらの個人データは、知的対話システム自体だけではなく、プロバイダーの他のサービスにも利用されることが可能となる可能性がある。

それでは、このような知的対話システム技術の発展を想定した場合、近年動きつつある AI 原則や AI 規制との関係では、どのような問題・リスクが生じ得るのであろうか。次章以下においては、AI 原則の流れと特に積極的に取り組みつつある欧州の動きについて、考察・分析する。

3. AI 原則に見る「人-機械」関係と文化の影響

(1) AI 原則の流れとその体系・構造

<AI 原則のこれまでの流れ>

近年の急速な AI 技術の発展の中、AI 技術は人間社会に大きな変革をもたらす一方、その利用方法によっては、人間社会の有する価値観にも抵触しかねない。このような認識の下、概ね 2017 年以降、世界各国の多くの政府、非政府機関、国際機関が、AI の開発等にあって倫理面で留意すべき視点・観点等をまとめた、いわゆる AI 原則に係る提案を行っている¹⁵。

特に、2017 年 6 月の総務省の「国際的な議論のための AI 開発ガイドライン案」¹⁶における開発原則や、2019 年 3 月の内閣府の「人間中心の AI 社会原則」¹⁷（以下、「AI 社会原

<https://system.jpaa.or.jp/patent/viewPdf/3307>

¹⁴ 東中 竜一郎, 岡田 将吾, 藤江 真也, 森 大毅、「対話システムと感情」『人工知能』2016 年 31 巻 5 号 p. 664-670I

https://www.jstage.jst.go.jp/article/jjsai/31/5/31_664/article/-char/ja/

¹⁵ 例えば、日本・人工知能学会（2017 年 2 月）、日本・総務省（2017 年 6 月）、日本・内閣府（2019 年 3 月）、欧州・AI-HLEG（2019 年 4 月）、FLI（2017 年 1 月）、IEEE（2019 年 6 月）、モントリオール大学（2018 年 12 月）、中国（2019 年 6 月）、OECD（2019 年 5 月）、G20（2019 年 6 月）、UNESCO（2021 年末決定予定）

¹⁶ 総務省情報通信政策研究所「国際的な議論のための AI 開発ガイドライン案」「AI ネットワーク社会推進会議 報告書 2017」2017 年 7 月 28 日

https://www.soumu.go.jp/menu_news/s-news/01iicp01_02000067.htm

¹⁷ なお、原則の検討は内閣府で行われたが、最終決定は、内閣官房で行われている。

則」という)、2019年4月の欧州の「信頼できる AI に向けた倫理ガイドライン (以下、「AI 倫理ガイドライン」という)¹⁸などを踏まえ、2019年5月に、OECDにおいて、「信頼できる AI のための責任あるスチュワードシップに係る原則」(以下、「OECD AI 原則」という)を含む「AI に関する理事会勧告」¹⁹が決定された(42か国が採択)。この OECD の AI 原則は、翌月の G20 の AI 原則²⁰としても採択されている。

<AI 原則の全体の体系・構造>

この OECD AI 原則は、①包摂的成長、持続的成長、幸福、②人間中心の価値と公平性、③透明性と説明可能性、④頑健性、セキュリティ、安全性、⑤説明責任の5つの大項目からなる。なお、この OECD AI 原則以外を含め、世界各国・機関の AI 原則は、それぞれ独自の括り方によりまとめているため、必ずしも単純比較はできないものの、キーワードのみで記述の有無を評価すると、時期によって内容には変遷はあるものの、概ね類似する項目が記載されているとされる²¹。

一方、これらの内容に関し、「(人間として守るべき)社会規範」と、「(それを達成するための人類としての)ガバナンス手法」に分けて整理・体系化すると、完全に切り分けることはできないものの、全般関係、安全関係(生命、身体、財産の保護)、人権・公平性関係(自由、公平性、プライバシーなど)に分けられるとされる。すなわち、全体としては、人間社会の有する社会規範に対し、説明責任を中心とするガバナンス体制が求められる一方、主として安全性に対しては、AI の信頼性・制御可能性の取組みが、また、主に人権に対しては、AI の透明性・説明可能性などの取組みが求められるという構図である(図4参照)²²。

内閣官房統合イノベーション戦略推進会議「人間中心の AI 社会原則」(2019年3月29日)

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

¹⁸ High-Level Expert Group on Artificial Intelligence, “Ethics Guidelines for Trustworthy Artificial Intelligence”, 2019.4.8

<https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

¹⁹ OECD “Recommendation of the Council on Artificial Intelligence”, 2019.5.22

<https://www.oecd.org/tokyo/newsroom/forty-two-countries-adopt-new-oecd-principles-on-artificial-intelligence-japanese-version.htm>

²⁰ G20, “G20 AI Principles” 2019.6.11

<https://www.mofa.go.jp/mofaj/gaiko/g20/osaka19/jp/documents/>

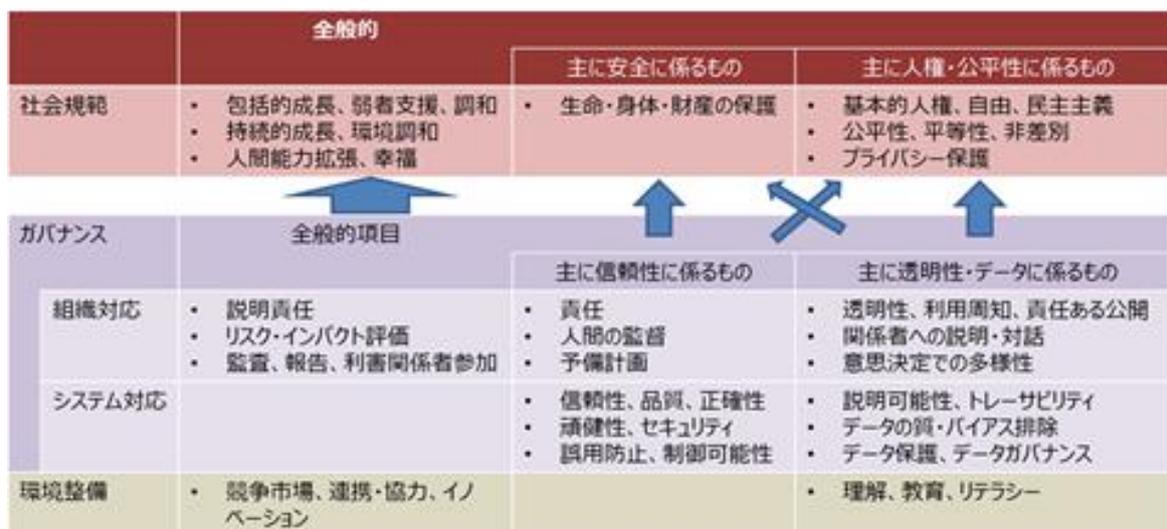
²¹ 中川 裕志(理化学研究所)「AI 倫理指針の動向とパーソナル AI エージェント」総務省 学術雑誌『情報通信政策研究』第3巻第2号 Journal of Information and Communications Policy Vol.3 No.2, I - 1~23 (2020年3月)

https://www.soumu.go.jp/iicp/journal/journal_03-02.html

²² 市川類(一橋大学)、「AI 原則の体系化と今後のガバナンスの方向 ~デジタル・AI におけるイノベーションと社会制度の共進化~」IIR ワーキングペーパー、WP#20-15、2020/10/02

<http://pubs.iir.hit-u.ac.jp/admin/ja/pdfs/show/2432>

図4：AI原則の構造化・体系化（社会規範とガバナンス課題の関係）²³



一般的に、製造・エネルギー技術などこれまでの多くの技術では、安全性に係る対応が課題であったが、AI・デジタル技術では、人権・公平性に関わる対応が多く求められており、これらが、いわゆるAIの「倫理問題」の中心的課題として特徴づけられる。

（2） AI原則に見る「人間-機械」関係：欧州と日本の考え方の違い

これらのAI原則のうち、OECD AI原則は、世界の主要国（OECD：42か国、G20）で合意されたものであり、特に、基本的人権・自由・民主主義を旨と国々を中心に、広く受け入れられたものとして位置づけられる。

しかしながら、このOECD原則の策定にあたり、主要な参考になったとされる欧州のAI倫理ガイドラインと日本のAI社会原則の記述内容を比較すると、双方とも、筆頭の項目として「人間-AI関係」について記述しているものの、その内容は大きく異なる。

具体的には、欧州のAI倫理ガイドラインでは、その筆頭の原則として「人間の自律性の向上」を掲げ、「AIシステムと交流する人間は、それらに対して完全かつ効果的に自己決定を維持しなければならない」などと、人間によるAIシステムに対する自己決定権を明示するとともに、7要件のうち筆頭要件である「人間の代理・監督」において、基本的人権に対するインパクト評価の実施を規定するとともに、「人間の代理（Human Agency：人間主体）」、「人間の監督（Human Oversight）」という2つの概念を全面的に打ち出している。前者は、人間に対するAIシステムの決定であることの通知、また、後者は、人間のAIシステムに対する監督関係を規定するものであり、いずれも人間とAIシステムとを切り分け、AIシステムを人間の監督下に置くと言う発想であると言える。

これに対し、日本のAI社会原則においては、筆頭の基本理念として「人間の尊厳が尊重される社会」を、また、筆頭の原則として「人間中心の原則」を掲げているものの、その内容は、過度に依存しないこと、悪用されコントロールされないことといった一般的な原則を述べているに留まっている（図5参照）。なお、これに関連して、このAI社会原則の下に

²³ 出典：市川類（一橋大学）、「AI原則の体系化と今後のガバナンスの方向～デジタル・AIにおけるイノベーションと社会制度の共進化～」IIRワーキングペーパー、WP#20-15、2020/10/02
<http://pubs.iir.hit-u.ac.jp/admin/ja/pdfs/show/2432>

位置づけられる総務省の開発原則案では、この「人間-機械」関係に係る原則の一つとして「制御可能性」が挙げられおり、したがって、日本ではむしろ悪用防止に係る技術的な要件として捉えていると位置づけられる。

図5：日欧のAI原則における「人間-機械（AIシステム）」関係の記述

<倫理ガイドライン（欧州）>	<人間中心AI社会原則（日本）>								
<p align="center">【4つの倫理原則】</p> <table border="0"> <tr> <td data-bbox="215 504 566 705"> <p>人間の自律性の尊重 (Respect for Human Autonomy)</p> <p>(略) AIシステムと交流する人間は、それらに対して完全かつ効果的に自己決定を維持しなければならない。AIシステムは、人間を、不正に従属させ、強制し、欺き、強制し、操作し、条件付け、群れさせてはならない。代わりに、AIは人間の認知的・社会的・文化的技能を増強し、補完し、強化するものでなければならぬ。(略)</p> </td> <td data-bbox="582 504 758 705"> <p>被害の防止 (Prevention of Harm)</p> <p>公平性 (Fairness)</p> <p>説明可能性 (Explicability)</p> </td> </tr> </table> <p align="center">【7つの鍵となる要件】</p> <table border="0"> <tr> <td data-bbox="215 772 502 1064"> <p>人間の代理と監督 (Human Agency and Oversight)</p> <p>基本的人権：基本的人権を損なうリスクがある場合には、基本的人権に係るインパクト評価を行わなければならない。 人間の代理 (Human Agency)：ユーザは、AIシステムに関連する自律的な意思決定であることを知らされなければならない。 人間の監督 (Human Oversight)：人間の監督は、AIシステムが人間の自律性を損なったり他の悪影響を及ぼしたりしないようにするのに役立つ。</p> </td> <td data-bbox="518 772 774 1086"> <p>技術的頑強性と安全性 (Technical robustness and safety)</p> <p>プライバシーとデータガバナンス (Privacy and data governance)</p> <p>透明性(Transparency)</p> <p>多様性、非差別、公平性 (Diversity, non-discrimination and fairness)</p> <p>社会的・環境的厚生 (Societal and environmental wellbeing)</p> <p>説明責任(Accountability)</p> </td> </tr> </table>	<p>人間の自律性の尊重 (Respect for Human Autonomy)</p> <p>(略) AIシステムと交流する人間は、それらに対して完全かつ効果的に自己決定を維持しなければならない。AIシステムは、人間を、不正に従属させ、強制し、欺き、強制し、操作し、条件付け、群れさせてはならない。代わりに、AIは人間の認知的・社会的・文化的技能を増強し、補完し、強化するものでなければならぬ。(略)</p>	<p>被害の防止 (Prevention of Harm)</p> <p>公平性 (Fairness)</p> <p>説明可能性 (Explicability)</p>	<p>人間の代理と監督 (Human Agency and Oversight)</p> <p>基本的人権：基本的人権を損なうリスクがある場合には、基本的人権に係るインパクト評価を行わなければならない。 人間の代理 (Human Agency)：ユーザは、AIシステムに関連する自律的な意思決定であることを知らされなければならない。 人間の監督 (Human Oversight)：人間の監督は、AIシステムが人間の自律性を損なったり他の悪影響を及ぼしたりしないようにするのに役立つ。</p>	<p>技術的頑強性と安全性 (Technical robustness and safety)</p> <p>プライバシーとデータガバナンス (Privacy and data governance)</p> <p>透明性(Transparency)</p> <p>多様性、非差別、公平性 (Diversity, non-discrimination and fairness)</p> <p>社会的・環境的厚生 (Societal and environmental wellbeing)</p> <p>説明責任(Accountability)</p>	<p align="center">【基本理念（3）】</p> <table border="0"> <tr> <td data-bbox="837 504 1141 705"> <p>人間の尊厳が尊重される社会 (Dignity)</p> <p>我々は、AIを活用して効率性や利便性を追求するあまり、人間がAIに過度に依存したり、人間の行動をコントロールすることにAIが利用される社会を構築するのではなく、人間がAIを道具として使いこなすことによって、人間の様々な能力をさらに発揮したり、(以下略)</p> </td> <td data-bbox="1157 504 1364 705"> <p>多様な背景を持つ人々が多様な幸せを追求できる社会 (Diversity & Inclusion)</p> <p>持続ある社会 (Sustainability)</p> </td> </tr> </table> <p align="center">【AI社会原則（7）】</p> <table border="0"> <tr> <td data-bbox="837 795 1125 1041"> <p>人間中心の原則</p> <p>AIの利用は、憲法及び国際的な規範を保障する基本的人権を侵すものであってはならない。 (略) AIが活用される社会において、人々がAIに過度に依存したり、AIを悪用して人の意思決定を操作したりすることのないよう、我々は、リテラシー教育や適正な利用の促進などのための適切な仕組みを導入することが望ましい。</p> </td> <td data-bbox="1141 795 1364 1086"> <p>教育・リテラシーの原則</p> <p>プライバシー確保の原則</p> <p>セキュリティ確保の原則</p> <p>公正競争確保の原則</p> <p>公平性、説明責任及び透明性の原則</p> <p>イノベーションの原則</p> </td> </tr> </table>	<p>人間の尊厳が尊重される社会 (Dignity)</p> <p>我々は、AIを活用して効率性や利便性を追求するあまり、人間がAIに過度に依存したり、人間の行動をコントロールすることにAIが利用される社会を構築するのではなく、人間がAIを道具として使いこなすことによって、人間の様々な能力をさらに発揮したり、(以下略)</p>	<p>多様な背景を持つ人々が多様な幸せを追求できる社会 (Diversity & Inclusion)</p> <p>持続ある社会 (Sustainability)</p>	<p>人間中心の原則</p> <p>AIの利用は、憲法及び国際的な規範を保障する基本的人権を侵すものであってはならない。 (略) AIが活用される社会において、人々がAIに過度に依存したり、AIを悪用して人の意思決定を操作したりすることのないよう、我々は、リテラシー教育や適正な利用の促進などのための適切な仕組みを導入することが望ましい。</p>	<p>教育・リテラシーの原則</p> <p>プライバシー確保の原則</p> <p>セキュリティ確保の原則</p> <p>公正競争確保の原則</p> <p>公平性、説明責任及び透明性の原則</p> <p>イノベーションの原則</p>
<p>人間の自律性の尊重 (Respect for Human Autonomy)</p> <p>(略) AIシステムと交流する人間は、それらに対して完全かつ効果的に自己決定を維持しなければならない。AIシステムは、人間を、不正に従属させ、強制し、欺き、強制し、操作し、条件付け、群れさせてはならない。代わりに、AIは人間の認知的・社会的・文化的技能を増強し、補完し、強化するものでなければならぬ。(略)</p>	<p>被害の防止 (Prevention of Harm)</p> <p>公平性 (Fairness)</p> <p>説明可能性 (Explicability)</p>								
<p>人間の代理と監督 (Human Agency and Oversight)</p> <p>基本的人権：基本的人権を損なうリスクがある場合には、基本的人権に係るインパクト評価を行わなければならない。 人間の代理 (Human Agency)：ユーザは、AIシステムに関連する自律的な意思決定であることを知らされなければならない。 人間の監督 (Human Oversight)：人間の監督は、AIシステムが人間の自律性を損なったり他の悪影響を及ぼしたりしないようにするのに役立つ。</p>	<p>技術的頑強性と安全性 (Technical robustness and safety)</p> <p>プライバシーとデータガバナンス (Privacy and data governance)</p> <p>透明性(Transparency)</p> <p>多様性、非差別、公平性 (Diversity, non-discrimination and fairness)</p> <p>社会的・環境的厚生 (Societal and environmental wellbeing)</p> <p>説明責任(Accountability)</p>								
<p>人間の尊厳が尊重される社会 (Dignity)</p> <p>我々は、AIを活用して効率性や利便性を追求するあまり、人間がAIに過度に依存したり、人間の行動をコントロールすることにAIが利用される社会を構築するのではなく、人間がAIを道具として使いこなすことによって、人間の様々な能力をさらに発揮したり、(以下略)</p>	<p>多様な背景を持つ人々が多様な幸せを追求できる社会 (Diversity & Inclusion)</p> <p>持続ある社会 (Sustainability)</p>								
<p>人間中心の原則</p> <p>AIの利用は、憲法及び国際的な規範を保障する基本的人権を侵すものであってはならない。 (略) AIが活用される社会において、人々がAIに過度に依存したり、AIを悪用して人の意思決定を操作したりすることのないよう、我々は、リテラシー教育や適正な利用の促進などのための適切な仕組みを導入することが望ましい。</p>	<p>教育・リテラシーの原則</p> <p>プライバシー確保の原則</p> <p>セキュリティ確保の原則</p> <p>公正競争確保の原則</p> <p>公平性、説明責任及び透明性の原則</p> <p>イノベーションの原則</p>								

すなわち、欧州と日本は、基本的人権・民主主義といった価値観を共有する国家として、「人間中心のAI」の開発を目指すということでは共通認識を有するものの、その中での具体的な「人間-機械（AIシステム）」関係に関しては、欧州では、「人間の自己決定権」「人間の代理」「人間の監督」などの概念の下、「人間が機械（システム）を管理する」ということを枠組みとして規定しようとする考えが強いのに対し、日本では、単に悪用防止のための技術要件としてしか見ていないという違いがあると評価される²⁴。

(3) 「人間-機械」関係に係る文化の影響（ヒューマノイドロボットの事例）

このような人間-機械（AIシステム）関係に係る、欧州（西洋）と日本（東洋）での考え方の差異については、以前より、知的対話システムの一つであるロボットとの関係で議論されてきている。特に、日本におけるロボット開発においては、ヒューマノイドロボットの開発に多くの重心が置かれているが、その理由・背景として、松原、辻井、國吉など多くの日本のAI・ロボット研究者は、東洋と西洋の宗教観の違いを指摘している^{25,26}。

²⁴ なお、OECD AI原則では、「人間の監督」を「人間中心の価値」の一部として、また、「人間の代理」を「透明性」の一部として位置づけている。

²⁵ IT Media News, 「第1章-2 「アトムを実現する方法は1つしかない」 松原仁教授が語る未来」2009年05月19日

<https://www.itmedia.co.jp/news/articles/0905/19/news089.html>

²⁶ また、同様に、グロス/ニッツバーグ（2019）には、日本の人工知能研究者に対するインタビューの結果として、以下の内容を記載している。

すなわち、西洋では、キリスト教に見られるような“創造主である神”という絶対的な概念を持つ宗教が社会の根底にあるため、「神とその創造物の筆頭である人間」と「それ以外のもの」とを区別する二元論的な論理が働きやすいとされる。このため、人間が人間の形を模倣したロボットを作ることは、神への冒とくであるとして社会に拒否感があり、また、人工物（機械）が人を凌駕することへの恐怖感が社会に根強いとしている。

これに対し、東洋的な宗教観では、草木を含めて万物に心が宿ると考えのもと、「人間とそれ以外のもの」の捉え方が連続的であるとされる。このため、人が「心」を有するような人工物を作ることに對し大きな違和感はないとしている。Geraci(2013)²⁷も、日本では（西洋と異なり）宗教との関係でロボットが話題になることはないため、多くの人が、人間とその機械的な子孫と位置づけられるロボットとの親交を期待しており、ロボットを人間社会のパートナーとして受け入れているとしている。また、その背景として、日本では、伝統的に、非生物についても「神（神聖なるもの）」として受け入れており、したがって、ロボットも、精神を有する「神」として受け入れられ、さらに、仏教においてもそのような包摂を受け入れているとしている。

その上で、それぞれの国・地域におけるこのような宗教観の考え方の違いをベースに大衆文化も育まれ、その結果、ヒューマノイドロボットに係る異なった社会的・文化的受容が形成されているもの考えられる。すなわち、大衆文化において、西洋では、フランケンシュタインから始まり、映画「ターミネーター」などを含め、「人間に危害を及ぼすロボット」として描かれているものが多いのに対し、日本では、鉄腕アトムに代表されるように、子供と友達になるロボットが中心であり、人間に危害を及ぼすことはない（Geraci(2013), 上野(2016)²⁸参照）存在として多く描かれていることが指摘される。このため、実際に、ヒューマノイドロボットに関する社会の受容度は、日本と欧米において異なることが明らかにされている²⁹。

「日本の産業技術総合研究所（AIST）で人工知能研究センター長を務める辻井潤一は言う。「海外ではモンスターのようイメージを持たれがちですが、日本では、ロボットは人間の護り手か友達のようなものなんです」と。（略）「日本の社会が先進的テクノロジーを受け入れるのには、人口動態やポップカルチャーよりもさらに根深い理由がある、と。それは東洋では昔から当たり前の基本的な哲学によるものだ。西洋思想と違って、東洋では人間が特別でない。」（略）「「私たちには、一神教の神のような「創造主」という概念がありません」と辻井は言う。「西洋文明は常に『人間は神のコピーであり、人間には特権が与えられている』と考えています。アジアの文化には、そうした考えがありません。動物から人間まで、緩やかにつながっているのです。」と。（略）

「東京大学・知能情報システム研究室の國吉康夫教授は、「人間そっくりのヒューマノイドは、人工知能の今後の性向のために重要だ」と強調している。」（略）「「私たちは、人間のようなものを作ろうとしています」と教授は言う。「それが悪いことだとか、恐ろしいことだとは思いません。そこがおそらく欧米人と日本人の違いなのでしょう。欧米人の多くは、人間に匹敵するような別個の存在を要因できないのです」と。」

オラフ・グロス、マーク・ニッツバーグの「新たなAI大国 その中心に「人」はいるのか？」講談社(2019/12/6) 位置 No.1528

²⁷ Robert M. Geraci, “Robotics and Religion”, Encyclopedia of Sciences and Religions 2013 Edition | Editors: Anne L. C. Runehov, Lluís Oviedo, pp.2067-2072

https://link.springer.com/referenceworkentry/10.1007%2F978-1-4020-8265-8_1229

²⁸ 上野裕子（三菱UFJリサーチ&コンサルティング）、「海外から見た日本のロボット産業・技術」GLOBAL Angle 2016.7 No.124

https://www.murc.jp/wp-content/uploads/2016/08/global_1607_1.pdf

²⁹ Tatsuya Nomura, Tomohiro Suzuki, Takayuki Kanda, Jeonghye Han, Namin Shin, Jennifer Burke and Kensuke Kato, “What people assume about humanoid and animal-type robots: Cross-cultural

以上の議論の多くは、ヒューマノイドロボットを対象にしたものであるが、同様の議論は、AI、特に知的対話システムにも適用される。すなわち、このヒューマノイドロボットの開発の本質は、形態・動作が人間に類似していること以上に、「人間が知能を使って行う活動」を行うロボット（機械）の開発を意図したものであり、したがって、世間一般的には、人工知能、特に知的対話システムの開発とほぼ同義であるとみなされる。

また、特にAIに関しては、ハンス・モラベック（1998）³⁰やレイ・カーツワイル（2005）³¹など、欧米を中心に、汎用人工知能（AGI）の登場やシンギュラリティ（技術特異点）の可能性に係る議論が盛んに行われた。このシンギュラリティの議論は、一部には、トランスヒューマニズム（超人間主義）と呼ばれるような超知性体が「神（の指示を受けた救世主としての人間）の創造物」として登場し、その中で、映画「マトリックス」ごとく、人間の脳の情報をアップローディングして、不死のバーチャル世界の構築・逃避に期待するような、ある意味異端的な宗教的な動き³²も生まれる一方で、本質的には、キリスト教などにおける黙示録・終末史観と重ね併せられ、欧米の多くの識者がAGIの開発に対する批判と規制の必要性を主張したという経緯がある³³。このようなAGI・シンギュラリティに係る議論には、キリスト教などの一神教の考えが背景にあるとされ³⁴、欧米では大きな議論を巻き起こしたのに対し、日本ではこれらの欧米での議論が輸入され話題になったのみで、独自の社会的な動きは特段起きていない。

その後、AGIに対する過剰な反応は収まっているものの、上述のヒューマノイドロボットに見られるような日本と欧米（特に欧州）における文化的背景による社会の受容度の違いに加え、欧州の宗教観に基づく、AIそのものに対する社会における不安感は、次章に示す欧州のAI規制法案のベースになっているものと考えられる。

なお、文化的側面以外にも、近年の人口動態の差異も、ロボットに対する受容度に影響を及ぼしているとの指摘もある。すなわち、欧米（特に欧州）では失業率が高く、AIやロボ

analysis between Japan, Korea, and the United States”, *International Journal of Humanoid Robotics* Vol. 05, No. 01, pp. 25-46 (2008)

<https://www.worldscientific.com/doi/abs/10.1142/S0219843608001297>

Tatsuya Nomura; Keisuke Sugimoto; Dag Sverre Syrdal; Kerstin Dautenhahn, “Social acceptance of humanoid robots in Japan: A survey for development of the frankenstein syndrome questionnaire”, 2012 12th IEEE-RAS International Conference on Humanoid Robots (Humanoids 2012)

<https://ieeexplore.ieee.org/abstract/document/6651527>

Nomura, Tatsuya T., Syrdal, D.S., Dautenhahn, K., “Differences on social acceptance of humanoid robots between Japan and the UK”, *Procs 4th Int Symposium on New Frontiers in Human-Robot Interaction* (2015)

<https://uhra.herts.ac.uk/handle/2299/16345>

³⁰ Hans Moravec, “Robot: Mere Machine to Transcendent Mind（日本語版『シェーキーの子どもたち—人間の知性を超えるロボット誕生はあるのか』）

³¹ Ray Kurzweil, “The Singularity Is Near: When Humans Transcend Biology”,

³² 例えば、Robert M. Geraci, “Apocalyptic AI: Religion and the Promise of Artificial Intelligence”, *Journal of the American Academy of Religion*, Volume 76, Issue 1, March 2008, Pages 138–166

<https://academic.oup.com/jaar/article-abstract/76/1/138/682743>

また、池田光穂・井上大介「シンギュラリティ時代における宗教」

https://www.cscd.osaka-u.ac.jp/user/rosaldo/religion_at_technological_singularity.html

³³ 例えば、2017年1月のFLI（Future of Life Institute）のアシロマ原則では、「人間による制御」に加えて、「再帰的に自己改善する人工知能」の安全管理の厳格化を提唱している。

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

³⁴ 西垣通、島藺進「人工知能と宗教—『AI原論』から見えてくるもの—」（現代宗教2019）

<http://www.iisr.jp/journal/journal2019/P045-P061.pdf>

ットが雇用を奪うのではないかとの警戒感が強い一方、日本では人口減少の中にあり、AIやロボットによる業務の効率化、労働の代替が強く求められている状況にあり、社会として受容しやすい状況にあるとされる³⁵。

図6：人工知能に係る社会規範と制度・イノベーションの関係（欧州・日本比較）

		欧州	日本
社会規範	背景としての宗教	一神教（キリスト教） 「神・人」と「それ以外（機械）」の二元論 ※バーチャル世界への「意識」移転への期待	多神教（神道）・仏教 「人」と「人工物（機械）」の連続性 ※全てのモノに、神・生命が宿る
	大衆文化	人間に危害をもたらすロボット （フランケンシュタイン、ターミネーターなど） ※シンギュラリティ論に係る恐怖感	人間の友達であるロボット （鉄腕アトム、ドラえもんなど）
	人口動態	高い失業率（ロボット・AIへの懸念）	低い失業率（ロボット・AIへの期待）
社会制度・ガバナンス	AI原則	人間の自律性尊重：人間と機械の分離、人間による機械の管理 （Human AgencyとHuman Oversight）	人間中心の原則：人間による機械の悪用の防止、制御可能性
	AI規制	AIに対する強い規制を指向（予防原則）？	AIに対する弱い規制を指向？
技術・イノベーション		AI技術・知的対話システムの普及、発展	
		ヒューマノイドロボットの開発への重点	

4. 欧州のAI法案の構造と知的対話システムへの影響

（1）欧州のAI（規制）法案の経緯と概要

<経緯>

欧州委員会（EC）のAI戦略である「欧州のためのAI」³⁶が発表されたのは、2018年4月であり、これは、日本、米国、カナダ、アジア諸国などのAI戦略の発表と比較して、比較的遅い。しかしながら、その後、毎年のように、次のステップに向けた取組方針を策定してきているという意味では、欧州は、他の先進国と比較して、着実かつ積極的にAI戦略に取り組んでいると言える。一般的に、AI戦略には、研究開発、ベンチャー、データ流通、インフラ整備などのイノベーション促進の側面と、倫理・原則・法制などのガバナンス体制整備の側面があるが、欧州は、特に後者に関し、法制化に向けて、世界の先陣を切って取り組んでいることが特徴である。

³⁵ BBC, “What the world can learn from Japan’s robots”, Japan 2020

人口減少社会におけるロボットへの受容性の高さ。

Yet in Japan’s favour is its very long history of embracing robots, not fearing them. In the West, pop culture and media often frame robots as job-stealing Terminators itching to start a revolution. In Japan, they’re often cute and cuddly; anime and manga have depicted robots as things to love. Others point to a respect for inanimate objects that’s rooted in Shintosim.

<https://www.bbc.com/worklife/article/20200205-what-the-world-can-learn-from-japans-robots>

³⁶ European Commission, “Artificial Intelligence for Europe”, Brussels, 25.4.2018 COM(2018) 237 final

具体的には、2018年4月の「欧州のためのAI」において1年後までにAI倫理ガイドラインを策定することを記載した上で、6月には独立の高級専門家グループを設置し、その後、同ガイドラインのドラフトの発表、パブリックコメントの実施を経て、2019年4月に同最終版を発表し、OECD AI原則の策定にも寄与している。また、その後、当該ガイドラインを活用したパイロット事業³⁷を行い、2020年7月には、評価リストの確定版³⁸を発表している。

一方、AI倫理ガイドラインと同時に発表したECの政策文書³⁹では、今後ガイドラインをアップデートする方針を記載していた。しかしながら、2020年2月に発表した「AIに関する白書」⁴⁰では、AI規制のための新法の必要性とその枠組みを提案し、その後、パブリックコメントを行い、2020年11月にその結果をとりまとめる⁴¹とともに、2020年9月には法制化に係る初期影響評価⁴²を発表し、これについてもパブリックコメントを行っている。

これらを踏まえ、2021年4月に、「AIに関する調和ルール（AI法）の起草と特定のEU改正」⁴³案（以下、AI法案という）を発表した。今後、欧州議会等での承認が必要であり、2021年秋にも承認がなされるのではないかと見込まれている。なお、欧州議会においては、事前に議会の意向を欧州委員会（EC）作成の同法案に反映させるべく、2020年10月に「AI、ロボット、関連技術の倫理的側面に関する枠組み」⁴⁵の提言書を発表している。特に、この議会の提言書においては、AI規制枠組みの中核に、環境問題などでよく提示される「予防原則」を位置づけるべきとしていることが特徴的である⁴⁶。

³⁷ Pilot the Assessment List of the Ethics Guidelines for Trustworthy AI

<https://ec.europa.eu/futurium/en/ethics-guidelines-trustworthy-ai/register-piloting-process-0.html>

³⁸ High-Level Expert Group on Artificial Intelligence (set up by European Commission), “The Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self assessment”, European Commission B-1049 Brussels, 17th of July 2020.

³⁹ European Commission, “Building Trust in Human-Centric Artificial Intelligence”, Brussels, 8.4.2019 COM(2019) 168 final

⁴⁰ European Commission, “WHITE PAPER On Artificial Intelligence - A European approach to excellence and trust”, Brussels, 19.2.2020 COM(2020) 65 final

⁴¹ European Commission, “Public consultation on the AI White Paper Final report”, November 2020

⁴² European Commission, “Inception Impact Assessment :Proposal for a legal act of the European Parliament and the Council laying down requirements for Artificial Intelligence”, Ref. Ares(2020)3896535 - 23/07/2020

⁴³ European Commission, “Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL: LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS”, Brussels, 21.4.2021 COM(2021) 206 final

⁴⁴ European Commission Brussels, 21.4.2021, COM(2021) 206 final 2021/0106 (COD) “Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS”

⁴⁵ European Parliament, “Framework of ethical aspects of artificial intelligence, robotics and related technologies”, European Parliament resolution of 20 October 2020 with recommendations to the Commission on a framework of ethical aspects of artificial intelligence, robotics and related technologies (2020/2012(INL))

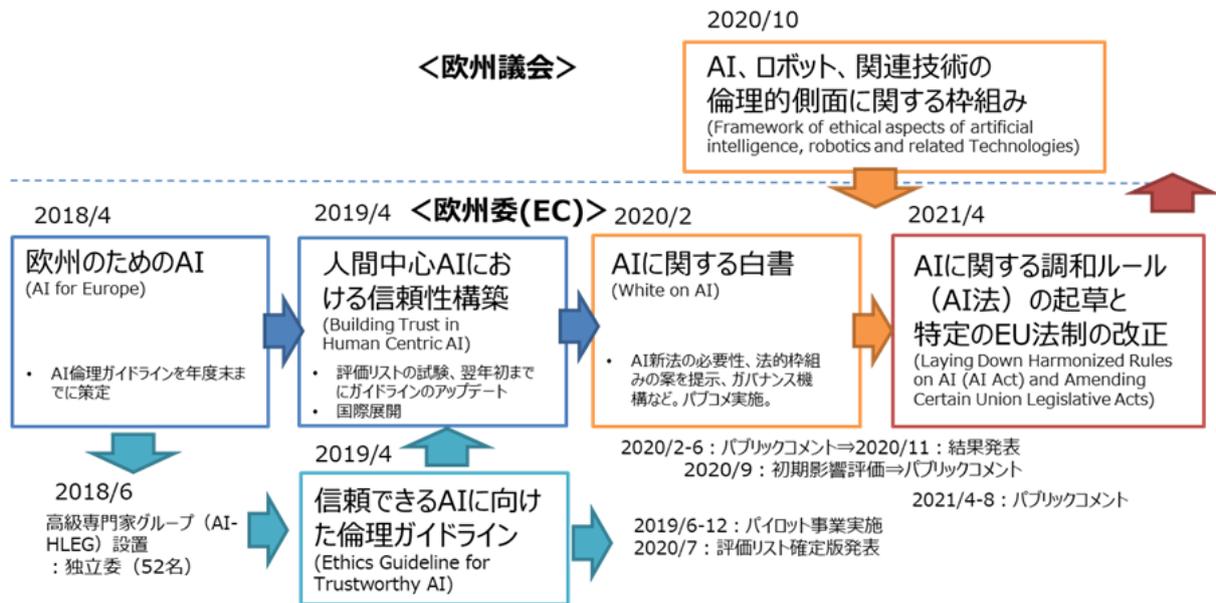
<https://www.europarl.europa.eu/news/en/press-room/20201016IPR89544/parliament-leads-the-way-on-first-set-of-eu-rules-for-artificial-intelligence>

https://www.europarl.europa.eu/doceo/document/TA-9-2020-0275_EN.html

⁴⁶ 同報告書（pp7）：“Human-centric and human-made artificial intelligence”

3. (略), stresses that citizens' trust in AI can only be built on an ethics-by-default and ethics-by-design regulatory framework (略); considers that building on such an approach should be in line with the precautionary principle that guides Union legislation and should be at the heart of any regulatory

図 7 : 欧州の AI 政策の経緯 (倫理・法制関連) 47



＜欧州の AI 法案の概要＞

上述の通り、2021年4月に欧州委員会（EC）が発表したAI法案は、本論文執筆時点においてはまだ成立しておらず、また、仮に成立したとしても、実際に適用されるのは、その1～2年後であり、それまでに、その詳細が固まっていくものと考えられる。

その前提のもとで、以下においては、同AI法案の特徴及び、特に規制対象となるAIシステムの類型、構成につき、その経緯・背景も含めて整理・分析を行う。

(i) リスクベースの規制手法の導入

まず、今回の欧州のAI法案は、リスクベースアプローチを基本としつつも、世界で初めての横断的なAI規制であり、規制色が強いのが特徴である。このリスクベースアプローチ、すなわちハイリスクのAIシステムに対して強い規制を課す一方、そうでないAIシステムについては弱い規制を課すという方針は、AI白書において提言され、本AI法案においても、採用されている。一般的に、規制を行うとした場合において、リスクベースアプローチを採用することは、合理的であると考えられる。

なお、AI白書後に発表されたECによる初期影響評価では、①産業界主導のソフトローアプローチ、②任意のラベリング制度、③法律で義務化（特定、ハイリスク、全て、の3つのオプション）、④それらの組合せのオプションを提示し、これに対し、産業界からは、①、②を望む声が寄せられていた。しかしながら、AI法案の前文（P9）においては、いく

framework for AI; calls, in this regard, for a clear and coherent governance model that allows companies and innovators to further develop artificial intelligence, robotics and related technologies;
 47 出典：EC 発表各種資料、長宗豊和（JBCE 事務局長）「欧州 AI 政策について」（経済産業法 AI 社会実装アーキテクチャー検討会資料 2021年3月5日）などより、筆者作成
https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/005_02_01.pdf

つかオプション⁴⁸がある中で、リスクベースアプローチによる横断的な規制措置と行動指針に基づく自主措置な措置の組合せを採用するとしている。

このような規制色が強いアプローチが採用されている背景として、欧州では、産業界だけでなく、市民団体が、AIに対して積極的にコメントしていることが挙げられる。実際に、AI白書のパブコメ結果においては、産業界だけでなく、市民団体が多くコメントを寄せており、その際、総じてAI白書で示されたハイリスクの範囲が狭いとして更なる規制強化を要望している⁴⁹。このように、欧州において市民団体がAIに対して積極的に発言を行っている背景には、もちろん欧州においては人権意識の高い文化・歴史的背景があることに加え、上述した欧州におけるAIに対する社会受容性の低さ・不信感が、その背景の一因であると考えられる⁵⁰。

(ii) 4段階アプローチと個別列挙された内容

AI法案は、上記リスクベースアプローチを採用するにあたって、A. 禁止するもの、B. 利用にあたって制限を課すもの（ハイリスク）、C. 透明性の要件を課すもの、D. その他（規制対象外、自主取組対象）の4段階に分けている。

その際、同法案の目的としては、上記第3章（1）での分析結果と同様、健康・安全と基本的人権の保護の二つを掲げ⁵¹、このうち、健康・安全に係るものについては、主に既存規制法の改正で対応する一方、特に基本的人権の保護に係るものを中心に、新たに規制対象とするAIシステムを段階ごとに個別に列挙するという構成になっている。

このように、対象を可能な範囲で明示化することは、法執行の安定性上重要であると考えられる。ただし、その際、Bについては、法案本体ではなく付属書Ⅲに主として分野特定で記載することとし、欧州委員会に改正権限が付与（法案第7条）されているのに対して、A、Cについては、法案本体において対象AIシステムを特定している。このことは、Bについては、今後の技術の発展を踏まえ柔軟に対応する一方、A、Cについては規制対象とする強い意図をもって法制化したものと考えられる。

また、それまでのAI白書、議会提言書などでは、ハイリスク（B）とそれ以外（D）の2段階のリスクベースアプローチが提示されていたのに対し、今回の法案では、これまでほ

⁴⁸ AI規制法案（2021）の前文（pp.9）では、以下の5つのオプションを示し、そのうち、オプション3+を採用したとしている。

Option 1: EU legislative instrument setting up a voluntary labelling scheme;

Option 2: a sectoral, “ad-hoc” approach;

Option 3: Horizontal EU legislative instrument following a proportionate risk-based approach;

Option 3+: Horizontal EU legislative instrument following a proportionate risk-based approach + codes of conduct for non-high-risk AI systems;

Option 4: Horizontal EU legislative instrument establishing mandatory requirements for all AI systems, irrespective of the risk they pose.

⁴⁹ 日本語でのまとめとしては、以下を参照

長宗豊和（JBCE 事務局長）「欧州のAI政策について」（経済産業省第5回AI社会実装アーキテクチャー検討会 資料2-1）2021年3月5日

https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/005_02_01.pdf

⁵⁰ また、欧州として、そのような状況を活かしつつ、グローバル経済の中で、戦略的に優位に立ちたいという意向も垣間見られる。

⁵¹ 前文（5）”A Union legal framework laying down harmonised rules on artificial intelligence is therefore needed to foster the development, use and uptake of artificial intelligence in the internal market that at the same time meets a high level of protection of public interests, such as health and safety and the protection of fundamental rights, as recognised and protected by Union law.”

とんど議論がなされていないまま、AとCが追加され、4段階となっている。その理由は必ずしも明らかにされていないが、その規制対象となるAIシステムの内容を見ると、市民団体などから明示的に社会的な懸念が示されていたAIシステムを規制対象にすべく新たな類型として、AとCの類型を設定したものと推測される。具体的には、Aのうち、社会的スコアの利用、公的空間での法執行目的で遠隔生体認証については、中国での利用等を踏まえて、欧州域内において大きな議論になっている⁵²。また、Cのうち、ディープフェイクに関しては、既に規制すべきとの動きがある⁵³。

なお、ハイリスク型(B)については、「自然人の生体認証と分類」、「重要インフラの管理と運用」を除き、いずれも、個人の情報に基づきAIを利用して分析・評価を行い、その結果に基づき当該自然人に対する特定の分野における公的・私的な権利に係る決定を行う場合において、差別的にならないよう公平性の観点から規制を行うものである(具体的には、別添(付属書Ⅲ)を参照)。

図6：欧州AI法案での対象AIシステム⁵⁴

分類・利用制限	対象AIシステム	
A. 受容できないAI (EUの価値観と矛盾するAI) ⇒禁止	<ul style="list-style-type: none"> 潜在意識への操作(Art.5(1)(a)) 子供や精神障害者を相手とする搾取行為(Art.5(1)(b)) 社会的スコアの一般的な利用(Art.5(1)(C)) 公的空間での法執行目的の遠隔生体認証(Art.5(1)(f)) 	○ ○
B. ハイリスクAI ⇒要件と事前適合性評価を条件	<p>規制対象製品の安全要素(Annex II)</p> <ul style="list-style-type: none"> 産業機械、医療機器等、法によって第三者認証の対象となるもの <p>特定分野のAIシステム(Annex III)</p> <ul style="list-style-type: none"> 自然人の生体認証と分類 重要インフラの管理と運用 教育と職業訓練 雇用、労働者管理、自営業の機会 必須の民間サービス、公共サービス・利益へのアクセスや享受 法のエンフォースメント 移住、亡命及び国境管理 司法運営と民主的プロセス 	○ △
C. 透明性義務を伴うAI ⇒情報/透明性の義務を条件	<ul style="list-style-type: none"> 自然人と相互作用するシステム (Art.52(1)) 感情推定や生体情報に基づくカテゴリー形成を行うシステム (Art.52(2)) ディープフェイク (Art.52(3)) 	○ ○
D. 極小リスク/リスクなしAI ⇒規制なし(自主取組)	(上記以外のAIシステム)	

⁵² 「EU、AI めぐる初の規則案を公表--監視社会を防げないとの批判も」

<https://japan.zdnet.com/article/35169750/>

「EU商業会議所、中国の「社会信用システム」に警鐘」

<https://jp.reuters.com/article/china-eu-business-socialcredit-idJPKCN1VI0CT>

⁵³ 「世界各国ではディープフェイクをどのように規制しているのか？」

<https://note.com/hidekiikedan/n/n652f7b763ac0>

⁵⁴ 出典：EC 発表資料及び経済産業省「EUのAIに関するフレームワーク」(経済産業省第1回AI原則の実践の在り方に関する検討会 資料5)より筆者作成
経済産業省「EUのAIに関するフレームワーク」(経済産業省第1回AI原則の実践の在り方に関する検討会 資料5) 2021年5月11日

https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/2021_001.html

(注) 青字は主として「健康・安全」に関わるもの、黒字は主として「基本的人権」に関わるもの。また、右側の枠の丸印は、知的対話システムに大きく関連するもの。

(2) 知的対話システムの位置づけとその背景

それでは、この欧州の AI 法案においては、どのような知的対話システムがどのような観点から規制対象とされているのであろうか。以下において、それぞれ A～C の 3 つの規制類型ごとに考察する。

<A. 受容できない AI (サブリミナルなど) >

AI 法案第 5 条第 1 項(a)と(b) では、人間の行動を物理的に歪曲し、物理的又は心理的に人(自ら又は他人)に危害を与えるような AI システムのサービスの提供や利用を禁止している。具体的には、(a)では、人の意識を超えたサブリミナル手法の使った AI システムを、また、(b)では、年齢や物理的・精神的障害により脆弱性を有する特定のグループを搾取する AI システムを規制(禁止)対象としている。

第 2 章(3)-1 で記載したとおり、今後、知的対話システムは、対話を通じて、利用する人間(ユーザ)の行動に対して影響を及ぼすものに発展していくことを考えると、理論上は、同規制の対象になることが考えられる。要は、ユーザに対し悪影響を与える知的対話システムに係る規制であり、法律上は、特に悪意のある 2 種類の AI システムを明示的に禁止対象にしたものと位置づけられる。

なお、同じ禁止措置の対象である「社会的スコアの利用(第 1 項(c))」、「公的空間での法執行目的で遠隔生体認証(第 1 項(d))」については、上述のとおり、社会的に大きな懸念が示されているのに対し、この 2 種類の AI システムが、どのような理由で禁止措置に明示されることになったかは不明である⁵⁵。ただし、欧州の AI ガイドラインの筆頭要件に「人間の代替(Human Agency)」が記載され、AI システムによる人間への働きかけに対する不信感が社会の根底にある⁵⁶中で、もともと欧州では広告規制でサブリミナルの禁止条項が存在すること⁵⁷、また、子供などの脆弱性を有するグループの保護の必要性が謳われていたことなどを背景に、特に悪意のあるものを限定的に禁止措置として規定したのではないかと推測される。

<C. 透明性要件対象の AI システム>

AI 法案第 52 条第 1 項において「「自然人と交流(Interact)する AI システム」については、自然人に対し、交流している相手が AI システムであることを通知しなければなら

⁵⁵ 経産省資料によると、具体的な事例としては、以下を挙げている。

(a)トラック運転手に長時間運転させるために、非可聴域の音を聞かせるにあたり、この効果を最大化する周波数を AI により見つける。

(b)音声アシスタントが組み込まれた人形が、未成年者に対してゲームを装って、次第に危険な行為へと駆り立てる。

⁵⁶ 具体的には、AI ガイドラインでは、以下のように記載されている。

"AI systems can sometimes be deployed to shape and influence human behaviour through mechanisms that may be difficult to detect, since they may harness sub-conscious processes, including various forms of unfair manipulation, deception, herding and conditioning, all of which may threaten individual autonomy."

⁵⁷ 角田美穂子(横浜国立大)「EU の広告規制」(「JARO 設立 30 周年記念」社団法人日本広告審査機構(JARO)) 2004 年

<https://www.jaro.or.jp/jaro30/pdf/1-9.pdf>

<https://www.jaro.or.jp/jaro30/>

い」としている。これは全ての知的対話システムを対象にした規制であり、このように透明性の要件を課す発想は、上述の「Human Agency」すなわち、「（人間ではなく）AIの決定であることを知らせるべき」という、欧州ならではの思想に基づくものである。実際に、AI倫理ガイドライン以前に発表された欧州委員会の報告書⁵⁸には、チャットボットに対する不信感が記載されている⁵⁹。

ただし、「利用の状況から判断して明らかな場合は除く」としていることから、AIスピーカーやロボット型の知的対話システムなどについては、明らかに除外されると考えられる。また、チャットボットなどインターネットを通じて提供するサービスについては、本透明化義務の対象となりうるものの、一般的には、人間になりすまして対話する必要性は乏しいことから、本義務付けはさほど問題はないと考えられる。

また、AI法案第52条第2項においては、「感情認識システム」⁶⁰や「生体分類システム」⁶¹の利用者は、その対象となる自然人に対して、その旨通知することが求められる。第2章(3)-2.で記載した通り、今後の知的対話システムの発展では、画像認識技術等により、対話の対象となるユーザの属性（性別、年齢等）や感情などの判定を行って応答するシステムが多く導入されることが見込まれ、それらは、この透明化規制の対象になることが考えられる。

本規制も、これまで議論なく、今回の法案で初めて規制の対象とされた類型であり、その理由は必ずしも明らかではない。ただし、欧州ならではの Human Agency の議論の延長で、人間と機械（AIシステム）は異質なものであり、人間であれば当然のように認識できる感情や性別・年齢であっても、機械（AIシステム）が認識するには、人間の了承が必要との考え方にに基づき規制対象としたものと推測される。

一方、日本では、例えば既に性別・年齢などを判別してお勧めの商品を提示する自販機が10年以上前に導入されている⁶²。今後、欧州の規制が仮に日本でも導入されるとすると、どのようにして通知をするのかを検討する必要性が生じるとともに、これらの技術に係るイノベーションを阻害する可能性もあると考えられる。

⁵⁸ European Commission, European Group on Ethics in Science and New Technologies, “Statement on Artificial Intelligence, Robotics and ‘Autonomous’ Systems”, Brussels, 9 March 2018

⁵⁹ 具体的には、以下の通り記載されている（p11）。

”Without human intervention and control from outside, smart systems today conduct dialogues with customers in online call-centres; speech recognition interfaces and recommender systems of online platforms, e.g. Siri, Alexa and Cortana, make suggestions to users. Beyond the straightforward questions of data protection and privacy, we may ask whether people have a right to know whether they are dealing with a human being or with an AI artefact. Moreover, the question arises whether there should be limits to what AI systems can suggest to a person, based on a construction of the person's own conception of their identity.”

⁶⁰ 定義は、以下の通り。

(34) ‘emotion recognition system’ means an AI system for the purpose of identifying or inferring emotions or intentions of natural persons on the basis of their biometric data;

⁶¹ 定義は、以下の通り。

(35) ‘biometric categorisation system’ means an AI system for the purpose of assigning natural persons to specific categories, such as sex, age, hair colour, eye colour, tattoos, ethnic origin or sexual or political orientation, on the basis of their biometric data;

⁶² 日経 XTech 「年齢と性別を判別して商品をオススメ、JR 品川駅にデジタルサイネージ搭載の自販機登場」（2010年8月10日）

<https://xtech.nikkei.com/it/article/NEWS/20100810/351132/>

<B. ハイリスク AI>

AI 法案第 6 条に基づき付属書Ⅲに記載された項目のうち、1 項目目の「自然人の生体認証と分類」では、上述の生体分類に加え、個人認証を行うシステムについて規制対象としている。ただし、「リアルタイム又は事後の遠隔認証」に対象を限定しており、ある意味で、プライバシーの保護の延長と言える。今後、知的対話システムの対話可能者数が拡大し、副特定多数の中から「リアルタイム又は事後の遠隔」にて個人を特定するような場合においては規制対象となりうる。

また、付属書Ⅲの 3～8 項目は、それぞれ「教育と職業訓練」「雇用、労働者管理、自営業の機会」「必須の民間サービス、公共サービス・利益へのアクセスや享受」「法のエンフォースメント」「移住、亡命及び国境管理」「司法運営と民主的プロセス」に係る利用分野限定の規制であり、また、その内容は、上述のとおり、いずれも、個人の情報に基づき AI を利用して分析・評価を行い、その結果に基づき当該自然人に対する特定の分野における公的・私的な権利に係るような重要な意思決定を行う場合において、差別的にならないよう公平性の観点（基本的人権の保護の観点）から規制を行うものである（別添参照）。

したがって、知的対話システムにおいても、対話を通じて個人情報収集し、それを利用して特定分野での重要な意思決定・判断を行うシステムとして利用する場合には、本規制の対象になるものと考えられる。例えば、知的対話システムを通じて、受験生に対し試験を行い、その結果を踏まえて可否を判定するようなシステムが開発されれば、本規制の対象となりうる。

（3）知的対話システムの開発に係る倫理的・規制リスク

今後、知的対話システムの開発においては、その倫理リスクを考慮して開発する必要がある。その際、「倫理」とは、その国・地域における歴史的な経緯によって醸成された文化によって異なるものであり、一方、近年のグローバル化を踏まえると、本国だけではなく、世界各国それぞれの倫理感を意識した上で開発することが望ましい。その際、特に、人権意識が一般的に高く、また、「人間・機械」関係に係る認識で、日本と異なった文化を有する欧州の動向は、今後の開発における倫理上のリスクとしては留意すべき点になる。

そのような認識の下、上記（2）においては、欧州の AI 法案では、どのような知的対話システムが規制対象になりうるかについて、その背景も含めて考察してきた。これらの考察結果を、今後の知的対話システムの今後の倫理的リスクとして捉え、知的対話システムの発展方向に係る 2 軸（第 2 章参照）を踏まえて整理すると、図 7 の通りになる。

- まずは、知的対話システムの技術が向上し、対話システムの応答が、より「人間」に近づいていく中で、当該知的対話システムが、人間であると誤認されることがないように、システムであることの明示・通知が求められるリスクがある。
- 次に、今後、対話者・媒体の範囲が拡大し、入出力モーダルがマルチ化し、感情や性別・年齢などを認識し、また、顔認証など個人認証を行うようなシステムを開発・実用化する場合には、プライバシー保護などの観点に加え、「人間・機械」関係の視点も含めた倫理上の対応が求められるリスクがある。
- また、今後、対話可能領域の範囲が拡大し、知的対話システム自体がユーザに対して重要な意思決定を担うような場合には、公平性の観点から倫理上の対応が求められる

リスクが生じる。また、知的対話システムの応答・やりとりがユーザに対して悪影響を与えていないか、事前に確認することが求められる可能性がある。

図 7：欧州 AI 法から見る知的対話システムに係る規制リスク

	欧州AI法案での規制内容	備考（主な視点）
対話システム自体	・ AIシステムであることのお知らせ (Art.52(1))	人-機械関係に係る視点
対話者・媒体の範囲		
感情認識・生体分類	・ 感情認識・生体分類を行うことのお知らせ (Art.52(2))	人-機械関係に係る視点
生体認証・分類	・ 遠隔生体認証・分類に係る規制 (Annex III 1(a))	プライバシーに係る視点
対話可能領域の範囲		
目的・意思決定	・ 特定の利用分野での意思決定への利用に係る規制 (Annex III 3.~8.)	公平性に係る視点
ユーザ影響	・ ユーザに悪影響を与える特定のシステムの禁止 (サブミナル、子供対応など) (Art.5(1)(a)(b))	人-機械関係に係る視点

これらの内容を踏まえると、知的対話システムに係る倫理的リスクには、従来のプライバシーの延長に係るもの、AIによりリスクとして新たに注目されている公平性・非差別性に係るものに加え、「人間-機械」関係に係る文化的認識な差異によりリスク認識が異なるものがあると理解することができる。

なお、この人間-機械 (AI システム) 関係に係る認識の差異の問題は、機械の自律性の問題はあるものの、結局のところ、人間の機械に対する社会としての信頼性の問題である。また、機械 (システム) は人間 (開発者) が作るものであることを踏まえると、結局は、人間の人間 (開発者) に対する社会として信頼性の問題であるとも解釈することができる。このような中、開発者においては、社会の認識する倫理的なリスクを理解した上での知的対話システムの開発が求められる。

5. 今後の課題／規制とイノベーションの共進化

<知的対話システムの倫理的リスクに係る課題>

本論文では、AI 原則・規制に先導的に取り組んでいる欧州の動向を踏まえて、今後の知的対話システムの開発に係る倫理上のリスクについて議論してきた。上述の通り、欧州は、一般的に、人権意識が高く、また、「人間-機械」関係に関し日本と異なる文化・倫理観を有していることから、国内では気づかないような視点も少なくない。このような流れの中、一部海外の民間企業などでは、知的対話システムに係る倫理上のガイドラインを作る動き⁶³

⁶³ 例えば、マイクロソフト社は、対話型 AI に係るガイドラインを作成している。また、HP 社は、対話型 AI におけるプライバシーとの人間性との関係に指摘している。

マイクロソフト社「責任ある対話型 AI」

<https://www.microsoft.com/ja-jp/ai/ai-lab-conversational-ai>

マイクロソフト社「Guidelines for Responsible Conversational AI」

<https://aidemos.microsoft.com/responsible-conversational-ai>

HP 社「対話型 AI とチャットボットの増加」

もあり、日本企業においても、今後、本論文で記載したリスクを踏まえて、リスクマネジメントに取り組むことが期待される。

その際、本論文で取り上げた知的対話システムに係る倫理上のリスクは必ずしも網羅的なものではないことに留意する必要がある。上述の通り、今回議論の対象とした欧州の AI 法案は、本論文執筆時点では、まだ決定された訳ではなく、また、今後とも、技術の進展に伴い、規制対象となる AI システムが見直しされ、また、追加される可能性もある。また、世界的に見て、欧州が AI に対する規制を先導しているのは確かではあるが、他の地域の動向も把握する必要がある。

<社会規範と人工知能に係る規制とイノベーションの共進化>

また、今回の欧州の AI 法案が、必ずしも、今後の世界の AI 法制に係る標準になる訳ではないことに留意する必要がある。

もちろん、欧州は、国際標準化の推進も含めて、本 AI 法案の内容を、世界的に普及拡大を図るべく取り組む方針である。しかしながら、東洋的な思想・宗教的な背景を有する日本においては、現時点で、AI 規制に係る法制化に向けた動きは、必ずしも力を得ていない。実際に、日本では、経済産業省が、2021年7月に報告書を発表⁶⁴しているが、現段階では、AI システムに対する横断的な義務規定は不要とし、当面は、リスク評価やマネジメントのあり方も含めたガイドラインを提供が有効としている⁶⁵。

このような中、技術・イノベーションのグローバル化の進展する一方で、世界各国の規制制度が国際的にハーモナイズされていくのか、またさらには、加えて世界各国の多様な（狭義での）社会規範についても変化が生じ、ハーモナイゼーション（調和化）が長期的に進展していくのかは、今後の課題であろう。過去においては、プライバシーやそれに基づく個人情報・データ保護法制は、1980年代以降概念化され、その後各国において多少の差はあるものの世界的に法制化されてきたという経緯がある一方、欧州から提案されたデータベース法制は世界で受け入れられなかった経緯がある。そのような中、今回議論しているような「人間・機械」関係に係るガバナンス体系や社会規範が、世界においてどのように受け入れられていくかは、今後の関心課題である。

その上で、技術・イノベーションと制度・規制の共進化がどのように進展するかも検討課題となる。一般的に、技術・イノベーションが進化するにつれ、制度・規制も必要に応じ強化されるなどの変化が生じるであろう。また、制度・規制の導入に技術・イノベーションの速度・方向も変化するであろう。さらに、技術・イノベーションの普及が進むと、社会規範自体も変化することが想定される。例えば、知的対話システムや AI の普及が進むと、人間と機械との関係においても新たな共存関係が構築され、その結果、人間の機械に対する倫理観＝社会規範も変化し、その結果、規制などの制度も見直されていくと言うダイナミクスが生じうる。

<https://www.hpe.com/jp/ja/japan/insights/reports/conversational-ai-and-the-rise-of-the-chatbots-1805.html>

⁶⁴ 経済産業省 AI 原則の実践の在り方に関する検討会「我が国の AI ガバナンスの在り方 ver. 1.1 AI 原則の実践の在り方に関する検討会 報告書」令和3年7月9日

https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/index.html

⁶⁵ 同認識のもと、「AI 原則実践のための ガバナンス・ガイドライン ver. 1.0」を発表している。経済産業省 AI 原則の実践の在り方に関する検討会 AI ガバナンス・ガイドライン WG「AI 原則実践のための ガバナンス・ガイドライン ver. 1.0」令和3年7月9日

このような観点からは、欧州議会が提言しているような予防原則を、特に AI 技術などの急速に変化するような技術に対して適用するとした場合⁶⁶、その後のイノベーションを阻害するとともに、社会規範のダイナミックな変化をも抑制する可能性が高いことから、むしろ状況に応じて臨機応変に規制・制度を見直す体制が望ましいと考えられる。

なお、このような技術に係る規制法制は、狭義での社会規範のみによって決定されるものではないことに留意することが必要である。上述で主に議論の対象とした欧州と日本は、先進国の中では米国などと比較して、デジタル・AIに係る産業の競争力が相対的に弱い国・地域である。一方、デジタル・AIに係る研究面、産業面での競争力を有する米国においては、規制の創設の検討にあたっては、デジタル・AIに係る産業への影響・インパクトがより強く求められることが戦略的にもまた政治的にも求められる構造にあると言える。このような要因もあってか、米国の AI 政策においては、デジタル・AI 導入にあたってのガバナンスに関しては、政府に限って検討する方向にあるものの、民間企業に対しては、現時点で AI 規制を導入する動きは全くなく、むしろ、対中国との関係でデジタル・AI 関連産業の競争力を強化する方向に大きく舵を切りつつある⁶⁷。

このようなことを踏まえると、今後の制度（規制）と技術（イノベーション）の共進化については、世界各国の政治・経済情勢を含めたさらに複雑な関係に大きく影響されるものであることも考慮する必要がある。

(以上)

⁶⁶ 例えば、欧州議会の報告書（2020, pp.7）では、予防原則のアプローチを、AI 規制枠組みの中心に据えるべきと記載している。

⁶⁷ 例えば、以下を参照。

National Security Commission on Artificial Intelligence, “Final Report”, 2021-03-01

<https://www.nscai.gov/>

<https://www.nscai.gov/wp-content/uploads/2021/03/Full-Report-Digital-1.pdf>

(別添) 欧州 AI 法案 (付属書Ⅲ)

分類	対象となる AI システム
1. 自然人の生体認証と分類	(a)自然人に係る「リアルタイム」及び「事後」の遠隔生体認証のために利用されることを意図した AI システム
2. 重要インフラの運営と操業	(a)道路交通や水・ガス・熱・電力の供給のマネジメントと操業において安全部分として利用されることを意図した AI システム
3. 教育と職業訓練	(a)自然人に対し、教育や職業訓練機関へのアクセスや割当を決定する目的で利用されることを意図した AI システム (b)教育や職業訓練機関において学生を評価する目的や、教育機関への入学に共通に必要なとされる試験への参加者の評価のために利用されることを目的とした AI システム
4. 雇用、労働者管理、自営業へのアクセス	(a)自然人の採用又は選抜のために利用されることを意図した AI システム。特に空きポストの広告、応募者のスクリーニング・フィルタリング、インタビューや試験の過程における候補者の評価。 (b)仕事関連の契約関係における昇進や終結に係る意思決定や、当該関係における業務配分、人のパフォーマンスや行動の監視と評価のために利用されることを意図した AI システム。
5. 必須の私的・公的サービス・便益へのアクセスと享受	(a)公的機関が、あるいは公的機関のために、自然人に係る公的支援に係る便益あるいはサービスの適格性の評価のために利用されることを意図した AI システム。それらの便益・サービスの付与、削減、無効、再要求に係るものも含む。 (b)自然人の信用価値を評価、又は、その信用スコアを構築するために使用される AI システム。ただし、小規模のプロバイダーが自ら利用するためのサービスに含まれる AI システムは除く。 (c)消防員あるいは救急隊を含め、緊急事態対応サービスの派遣あるいは派遣の重点決定のために利用されることを意図した AI システム。
6. 法執行	(a)法執行当局が、自然人が罪を犯す又は再犯するリスク、あるいは犯罪の潜在的被害者となるリスクを評価するために、自然人に係る個人のリスク評価を行うために利用することを意図した AI システム。 (b)法執行当局が、生体監視装置と同様のツールとして、あるいは自然人の感情の状況を検知するために利用されることを意図した AI システム (c)法執行当局が、52条(3)に規定するディープフェイクを検知するために使用することを目的とした AI システム (d)法執行当局が、刑事犯の捜査、訴追の過程において、証拠の信頼性を評価するために使用されることを意図した AI システム (e)法執行当局が、自然人のプロファイリングや、個人の特性・特徴や自然人・グループの過去の犯罪行動の評価に基づき、実際あるいは潜在的な刑事犯の発生・再発生を評価するために利用することを意図した AI システム (f)法執行当局が、刑事犯の発見、捜査、起訴の過程において、自然人をプロファイリングするために利用されることを意図した AI システム (g)データにおける知られていないパターンを明確化し、隠れた関係を発見するため、異なるデータ源や異なるデータフォーマットで利用可能な、複雑な関連・被関連の多量のデータセットを検索することにより、法執行当局が使えるような、自然人に係る犯罪分析のために利用されることを目的とした AI システム
7. 移民、亡命、国境管理マネジメント	(a)所管官庁が、生体監視装置と同様のツールとして、あるいは自然人の感情の状況を検知するために利用されることを意図した AI システム (b)所管官庁が、加盟国に入学しようとするあるいは入学した自然人によって生じる、セキュリティリスク、非定常的な移民リスク、健康リスクなど各種のリスクを評価するために利用されることを意図した AI システム (c)所管官庁が、自然人の旅行証明書や関連書類や真正性を確認し、そのセキュリティの特徴をチェックすることにより、非真正の書類を検知するために利用されることを意図した AI システム (d)所管省庁が、地位を求めて申請する自然人の適格性に関して、亡命、査証、居住許可や関連する要請の審査を支援することを意図した AI システム
8. 司法の管理と民主プロセス	(a)司法当局が、事実と法を調査・解釈し、事実の完全なセットを方に適用することを支援することを意図した AI システム