

PRESS RELEASE (2025/01/09)

AI 規制と倫理的ロボットデザイン

～技術開発スピードと、AI 対応技術の規制に対する立法措置の遅さとのミスマッチ解決を目指して～

ポイント

- ① 翁岳暄准教授は、IEEE 7000 人工知能倫理標準化の枠組みを選択し、IEEE P7017™ワーキンググループを設立し、社会的ロボットの社会的影響（ELSI）に関するデザイン中心のアプローチを追求しています。
- ② 社会的ロボットの社会的影響に対する様々なアプローチをグローバルな視点から比較するため、翁准教授は Woodrow Barfield、Ugo Pagallo と共同で「The Cambridge Handbook on the Law, Policy and Regulation for Human-Robot Interaction」を編集しました。
- ③ 翁准教授は、この本の担当章で、AI のペーシング問題(※1)によって引き起こされる社会的ロボット規制のギャップを克服するための革新的な倫理的設計アプローチとして、コンプライアンス・バイ・デザイン(※2)に基づく新しい概念「リーガル・ヒューマン・ロボット・インターラクション (L-HRI)」(※3)」を提案しています。

概要

九州大学高等研究院 稲盛フロンティアプログラムの翁岳暄准教授（東北大学学際科学フロンティア研究所（クロスアポイントメント））は、2024年11月にケンブリッジ大学出版局の新書 The Cambridge Handbook on the Law, Policy and Regulation for Human-Robot Interaction にて、「Ethical Design and Standardization for Robot Governance」の章を執筆しました。

近年、AI ペーシング問題を解決するために、各国や国際機関が標準化アプローチを用いて以下の 3 つの異なるアプローチを行っています（図参照）。

- ・タイプ A（整合規格）：AI 規制に関する法律をサポートする補足的な技術規範として使用されるアプローチ
- ・タイプ B（ソフト・ロー）：政策指針や道徳原則を拘束力のない柔軟なルールとして適用されるアプローチ。
- ・タイプ C：タイプ A とタイプ B のハイブリッドアプローチ。AI 開発者が責任ある研究とイノベーション（RRI）を実施できるように、重要な倫理的配慮を社会技術的基準に変換することを目的とする。

本章では、技術進歩の急速なスピードに対応できていない立法措置の遅れに起因する、AI ペーシング問題について考察しています。重要な AI 技術の規制を検討することに加え、人間とロボットの日常的な相互作用に内在する倫理的、法的、社会的影響リスクを利害関係者が確実に管理できるよう、拘束力のない柔軟な AI 倫理基準に依拠した規制の枠組みについても論じています。AI 倫理基準をヒューマノイドロボットや表現ロボットの開発プロセスに組み込むことで、ロボット開発者は、ロボット規制のために制定された法律に抵触することなく、責任あるイノベーションと研究の原則を盛り込むことができます。本章では、2 つのケーススタディを通して、倫理的ロボットデザインのアプローチを探求し、その可能性と限界を検証し、リーガル・ヒューマン・ロボット・インターラクション (L-HRI) の実施におけるコンプライアンス・バイ・デザインの有用性を実証します。

研究者からひとこと：

産業界では、メーカーが製品を製造する際の指針となる技術仕様や手順を提供するため際に標準規格が度々用いられます。標準化プロセスは通常 2 つの方法で行われます。(1)専門機関による一連のルールの押しつけ (USB3.1、IEEE802.11 など)、または(2)特定の製品における市場支配の実現 (Windows OS など) です。しかし、AI ガバナンスにおける標準化について語る場合、議論は産業標準のみを限定とした標準ではなく、政策指針、道徳原則、社会技術標準を考慮したより広範なアプローチを取るべきです。本章では、倫理標準化のプロセスが AI 規制のベース問題にどのように対処できるかを検討しています。本章では、社会的ロボットに関する倫理的、法的、社会的影響をどのように標準化し、ロボット開発者や製造者が従うことができる具体的な設計要件を開発すべきかについて主に議論しています。

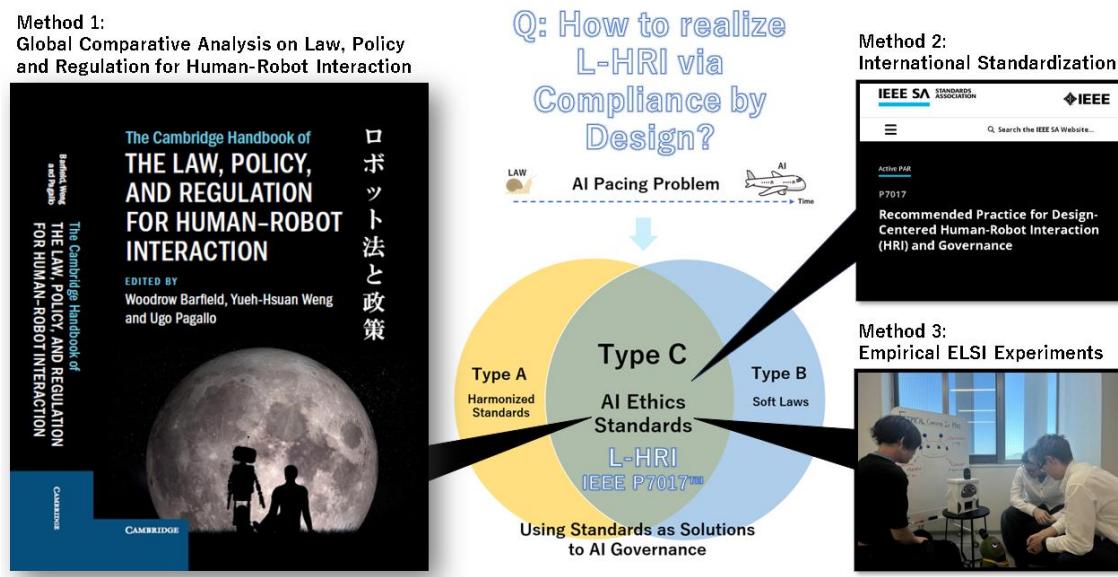


図. ロボットの倫理設計と標準化のための 3 つの方法論

【研究の背景と経緯】

この章では、ロボットガバナンスにおける AI 倫理の標準化の使用について議論します。具体的には、急速な技術進歩のスピードに追いつけていない立法措置の遅れにより、AI 搭載技術の規制に対する課題を考察します。重要な AI 技術の規制を検討することに加えて、この章では、日常的な人間とロボットの相互作用に内在する倫理的、法的、社会的影響 (ELSI) リスクを利害関係者が管理できるよう、拘束力のない柔軟な AI 倫理基準に依存する規制の枠組みを提唱します。ヒューマノイドや表現力のあるロボットの開発プロセスに AI 倫理基準を組み込むことで、ロボット開発者は、ロボット規制のために制定された「ハード・ロー」と矛盾することなく、責任ある革新と研究の原則を取り入れることができるでしょう。この章では、2 つのケーススタディを通じて、倫理的なロボット設計のアプローチを探求し、その可能性と限界を検討し、リーガル・ヒューマン・ロボット・インターフェース (L-HRI) を実装する上での「倫理的に整合性のある設計」と「社会システム設計」フレームワークの有用性を実証します。

【研究の主な内容と成果】

この章では、AI技術の急速な進歩と、それに対応する法整備の遅れという「AIペース問題」に対処するため、倫理的デザインと標準化の重要性を議論しています。特に、ロボットガバナンスにおいて、従来の法律による規制だけでなく、AI倫理基準を活用することが提案されています。

(1) AIペース問題の解決策：法律による規制は硬直的で変化に対応しにくいため、倫理基準や技術標準を柔軟に適用するアプローチが有効であると主張されています。

(2) 3つの標準化アプローチ：

タイプA：AI規制に関する法律をサポートする補足的な技術規範として使用されるアプローチ。

ISO 13482などの安全基準が例として挙げられています。

タイプB：政策指針や道徳原則を拘束力のない柔軟なルールとして適用されるアプローチ。

各国のAI倫理ガイドライン（例：IEEEの「Ethically Aligned Design」や日本の「人間中心のAI社会原則」）が紹介されています。

タイプC：タイプAとタイプBのハイブリッドで倫理的配慮を技術標準に落とし込むアプローチ。

これにより、AI開発者が責任ある研究とイノベーションを促進できるとされています。

(3) 倫理的デザインの重要性：倫理的デザインは、ロボットの設計段階から倫理的、法的、社会的な影響（ELSI）を考慮することで、リスクを予防的に管理する役割を果たすとしています。

倫理的デザインのための方法論：プライバシー・バイ・デザイン（PbD）は、プライバシー保護に焦点を当てた設計手法として紹介されていますが、ヒューマン・ロボット・インタラクション（HRI）における倫理的な課題を包括的に扱うことは難しいと指摘されています。価値に配慮したデザイン（VSD）は、人間の価値観を技術設計に取り入れる手法として紹介されています。また、思索的デザインやユニバーサルデザインも紹介されています。

(4) リーガル・ヒューマン・ロボット・インタラクション（L-HRI）：ロボットと人間の相互作用が法的要件を満たすように設計する必要性が強調されています。ロボットが法律を遵守するためのアプローチとして、コンプライアンス・バイ・デザインの概念が提案されています。倫理的デザインの適用事例として、データ保護における同意の取得プロセスが詳細に検討されています。IEEE Std 7007などの基準を参考に、倫理的な概念を形式化し、ロボットが倫理規範を理解し行動するための基盤を構築する方法が示唆されています。

(5) 社会システムデザイン：ロボットの社会性を設計する上で、ロボットが置かれる環境や社会システムを考慮する必要性が指摘されています。ロボットの具現化が、人間のプライバシー意識や行動に影響を与えることが実験によって示されています。

【今後の展開】

倫理的デザインと標準化は、責任あるロボット開発のために不可欠であり、今後の研究では、ロボットの信頼性、倫理的な影響、法的責任などを考慮する必要があると結論付けています。この章では、AI技術の急速な進歩に対応するための、倫理的デザインと標準化という新しいアプローチの重要性を示しており、法的な規制だけでなく、倫理的な観点からもロボット開発を推進していく必要性を強調しています。

【用語解説】

(※1) AI のペーシング問題

技術開発のスピードと、AI 対応技術の規制に対する立法措置の遅さとのミスマッチが明らかとなっている問題のことを言う。

(※2) コンプライアンス・バイ・デザイン

法的・倫理的基準をロボットの設計、機能、行動に直接組み込むアプローチ。その目的は、機械が人間のようにハイレベルな法的・倫理的基準を直接に認識処理しなくとも、ロボットが最初から法的要件や倫理基準に準拠するように設計されていることを保証することである。これは、ロボットの行動を誘導するセーフガード、行動制約、技術的手段を組み込むことによって可能になり、許容される法的・倫理的境界の範囲内でロボットが動作することを保証する。

(※3) リーガル・ヒューマン・ロボット・インターラクション (L-HRI)

ロボットの設計、開発、使用のライフサイクルに、法的・倫理的基準の遵守を直接組み込む設計中心のアプローチ。

【謝辞】

本研究は、JST 創発的研究支援事業【JPMJFR222C】の助成を受けています。

【書籍情報】

掲載書籍：The Cambridge Handbook on the Law, Policy and Regulation for Human-Robot Interaction

Woodrow Barfield, Yueh-Hsuan Weng and Ugo Pagallo (Eds), Cambridge University Press

タイトル：24 - Ethical Design and Standardization for Robot Governance

著者名：Yueh-Hsuan Weng

ISBN：9781009386708

DOI：10.1017/9781009386708.029

【お問合せ先】

<研究に関すること>

九州大学高等研究院

稻盛フロンティアプログラム 深教授 翁 岳煊（ウェン ユエ シュアン）

電話：+81 92-802-5432

電子メール：weng.yueh.hsuan.647@m.kyushu-u.ac.jp

<報道に関すること>

九州大学 広報課

TEL：092-802-2130 FAX：092-802-2139

Mail：koho@jimu.kyushu-u.ac.jp