エグゼクティブ・サマリー

人工知能と自律システム（AI／AS）の潜在能力を十分に活用するためには、現状認識を越えるとともに、より高い計算能力や問題解決能力の追求以上のことをする必要がある。

また、これらの技術が我々の道徳的価値観や倫理原則の面で人間と調和するよう、確実を期さねばならない。そしてAI／ASは、機能的な目標を達成して技術的な課題に取り組むだけでなく、人々にとって有益となるようにふるまう必要がある。これにより、日常生活におけるAI／ASの有意義な普及に必要な、人間とテクノロジーの間の高いレベルの信頼を構築することが可能になる。

アリストテレスが明らかにした「エウダイモニア」とは、人間の幸福を社会にとって最高の美徳と定義する姿勢である。おおまかに「繁栄」と翻訳されているエウダイモニアの恩恵は、我々が望む生き方を決定するのに倫理的配慮が役立つ、意識的な熟考から生まれるのである。

AI／ASの創造をその利用者と社会の価値観に沿ったものにすることにより、人間の幸福の増幅を、アルゴリズム時代における進歩の指標として優先させることが可能になる。

This document does not represent a position or the views of IEEE but the informed opinions of Committee members providing insights designed to provide expert directional guidance regarding A/IS. This translation is provided for convenience. The English language version of this document is the original and official version of record. In the event of any conflict between the English and translated version (words, terms, phrases, concepts, etc.) the original version of this document (created in English and available here) governs.

エグゼクティブサマリー

IEEEグローバル・イニシアティブについて

AIおよび自律システムにおける倫理規定に関するIEEEの国際指針(IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems)(以下、「IEEEグローバル・イニシアティブ」という)は、世界160カ国以上で40万人以上の会員を持ち、人類に資する技術発展に尽力している世界最大の技術専門家組織である米国電気電子学会(IEEE)のプログラムである。

IEEEグローバル・イニシアティブは、タイムリーな問題を特定して統一的見解を見つけるために、人工知能および自律システム・コミュニティにおける多数の声をまとめる機会を提供する。

IEEEはCreative Commons Attribution-Non-Commercial 3.0 United States Licenseに基づき、倫理的設計(*Ethically Aligned Design*: EAD)を提供する。

企業または個人はライセンス条件に従い、適宜この内容を採用することができる。また、標準化開発を含む正式なIEEEプロセスへの提出に、EADの内容と主題が選択されることが期待される。

IEEEグローバル・イニシアティブおよびEADは、IEEE TechEthics™ プログラムとして知られる、テクノロジーにおける倫理に関するオープンで広範かつ包括的な対話を促進するために、IEEEで開始された広範な取り組みに貢献している。

エグゼクティブサマリー

IEEEグローバル・イニシアティブのミッション

すべての技術者が、自律的でインテリジェントなシステムの設計と開発において倫理的配慮を優先するように教育、訓練を受け、その権限を与えられるよう確実を期する。

ここで「技術者」とは、AI／ASの技術を社会のために現実のものとする大学、組織そして企業など、研究、設計、製造またはメッセージの発信に携わる者を意味する。

本書は、人工知能、法と倫理、哲学、学術領域の政策、科学、政府および企業部門の分野における、全世界で100人以上のオピニオンリーダーの意見をまとめたものである。我々の目標は、AI／AS技術者の仕事に対して重要な情報源を提供するこのようなオピニオンリーダーたちからの見識および提案を、倫理的設計が今後数年間にわたって提供することである。この目標を達成するために、現在のバージョンの倫理的設計（EAD v1）では、人工知能および自律システムの分野における課題と候補提案を明確化する。

IEEEグローバル・イニシアティブの第2の目標は、倫理的設計に基づいたIEEE標準に対する提案を提供することである。IEEE P7000™（システム設計中に生じる倫理的問題への対処に関するモデルプロセス：*Model Process for Addressing Ethical Concerns During System Design*）は、本イニシアティブから始まった最初のIEEE標準化プロジェクト（承認済み・開発中）である。また、AI／ASの倫理問題に対する本イニシアティブの実際の影響を説明し、さらに2つの標準化プロジェクト、IEEE P7001™（自律システムの透明性）およびIEEE P7002™（データプライバシー・プロセス）が承認された。

エグゼクティブサマリー

本書の構成と内容

「倫理的設計」は8つのセクションから成り、各セクションはAI／ASに関連する具体的なトピックに対応している。これらのトピックは、IEEEグローバル・イニシアティブの特定の委員会で詳細に議論されている。また各委員会のセクションには、これらのトピックに関する課題と候補提案がリストされている。以下に、各委員会の概要および各委員会のセクションで取り上げられている問題を記載する。

1 | 一般原則

一般原則に関する委員会（The General Principles Committee）は、全てのタイプのAI／ASに当てはまる、倫理的懸案事項の概要を説明している。内容は以下の通り。

1. 人権の最高の理念を具現化する。
2. 人類と自然環境に対する最大の利益を優先させる。
3. AI／ASが社会技術的なシステムとして進化するのに伴い、リスクおよび悪影響を軽減する。

委員会は、AI／AS設計の新しい倫理ガバナンスの枠組みの中で、委員会が特定した原則、課題および候補提案が最終的に将来の規範と基準を支え、土台を形成することを目指している。

課題：

- AI／ASが人権を侵害しないことをどのように保証できるか（人権原則の確立）
- AI／ASに責任があることをどのように保証することができるか（責任原則の確立）
- AI／ASの透明性はどのようにして保証できるか（透明性の確立）
- AI／AS技術から得られる恩恵を拡大し、悪用されるリスクを最小限に抑えるにはどうすればよいか（教育と意識に関する原則の確立）

2 |自律知能システムへの価値観の組み込み

社会に有益な自律知能システム（AIS）を開発するためには、技術コミュニティが関連のある人間の規範や価値観を理解し、そのシステムへの組み込みが可能であることが極めて重要である。

自律知能システムへの価値観の組み込みに関する委員会（The Embedding Values into Autonomous Intelligence Systems Committee）は、以下の点において設計者を支援することにより、3つのアプローチとしてAISに価値観を組み込むというより大きな目的に取り組んでいる。

1. AISの影響を受ける特定のコミュニティの規範と価値観を特定する。

2. そのコミュニティの規範と価値観をAIS内に実装する。
3. そのコミュニティにおける人間とAISの間の、これらの規範と価値観の整合性および適合性を評価する。

課題：

- AISに組み込む価値観は普遍的ではなく、むしろ主にユーザーのコミュニティやタスクに特有のものである。
- モラルの過負荷：AISは通常、互いに矛盾する場合のある規範および価値観の多様性による影響を受ける。
- AISは、特定のグループのメンバーに不利益をもたらす、組み込み型のデータまたはアルゴリズム的傾向を持つ場合がある。
- 関連する（特定のコミュニティにおけるAISの特定の役割の）一連の規範が特定された際、そのような規範をどのようにして計算アーキテクチャに組み込むべきかが不明である。
- AISに実装される規範は、関連するコミュニティの規範と互換性を持つ必要がある。
- 人間とAISとの間に適切な信頼関係を築く。
- AISの価値観整合性に関する第三者評価。

3 |倫理的研究と設計を導く方法論

現代のAI／AS組織は、人間の幸福、エンパワーメントおよび自由が、AI／AS開発の中核となるよう確実を期するべきである。このような意欲的な目標の達成が可能な機械を作成するために、倫理研究・設計手法に関する委員会（the Methodologies to Guide Ethical Research and Design Committee）は課題と候補提案を作成し、世界人権宣言で定義されている人権のような人間の価値観が、当委員会のシステム設計方法論によって生み出されるよう確実を期する。価値観志向の設計手法は、倫理指針に基づく人間の進歩に向けて、AI／AS組織にとって不可欠な重点事項となる必要がある。機械は人間の役に立つべきであり、その逆はではない。この倫理的に健全なアプローチは、ビジネスと社会の両方に対し、AIの経済的アフォーダンスと社会的アフォーダンスの維持の間に均等なバランスをもたらすことを保証する。

課題：

- 倫理が学位プログラムに含まれていない。
- AI／ASの異なる問題を説明するために、学際的・異文化的な教育のモデルが必要である。
- AI設計に組み込まれた文化的に特有の価値観を差別化する必要がある。
- 価値観に基づいた倫理文化と慣習が産業界に欠如している。
- 価値観を意識したリーダーシップが欠如している。

エグゼクティブサマリー

- 倫理的懸案を提起するエンパワーメントが欠如している。
- 技術コミュニティのオーナーシップや責任が欠如している。
- AI／ASの最良の状況のためにステークホルダーを参画させる必要がある。
- 資料が不十分なことにより倫理的な設計が阻害されている。
- アルゴリズムに対する監視に一貫性がない、または不十分である。
- 独立したレビュー組織が不足している。
- ブラックボックス・コンポーネントを使用している。

課題：

- より多様な領域にわたる高い自律性を伴う複雑な目的関数を最適化する能力によって測られるようなAIシステムの性能の向上につれて、予期せぬ動作や意図しない動作が一層危険になる。
- 安全性を、未来のより汎用的な機能を持つAIシステムに組み込むことは難しい。
- 自律性と機能が強化されたAIシステムの開発と展開において、研究者および開発者は次第に複雑さを増す倫理的・技術的な安全性に関する問題に直面する可能性が高い。
- 将来のAIシステムは、農業革命または産業革命規模で世界に影響を及ぼすような能力を備える可能性がある。

4 |汎用人工知能（AGI）と人工超知能（ASI）の安全性と恩恵

未来の高度なAIシステム（汎用人工知能またはAGIとも呼ばれる）は、農業革命または産業革命規模で世界に変革をもたらす可能性があり、これは前例のないレベルで世界に繁栄をもたらし得る。
汎用人工知能（AGI）と人工超知能（ASI）の安全性と恩恵に関する委員会（Safety and Beneficence of Artificial General Intelligence（AGI）and Artificial Superintelligence（ASI）は、複数の課題と候補提案を提供し、変革の良い形での実現が、同じ目的を共有するAIコミュニティの共同努力を通じて成されるよう支援している。

5 |個人情報と個別アクセス制御

個人情報に関する重要な倫理的ジレンマは、データの非対称性である。個人情報および個別アクセス制御に関する委員会（Personal Data and Individual Access Control Committee）はこの非対称性に対処するために、個人情報の定義、アクセスや管理に対する各自のアイデンティティの管理者としての根本的な必要性の証拠となる問題および候補提案を明確化した。委員会は、完全な解決策は存在しないこと、またあらゆるデジタル・ツールはハッキングされる可能性があることを認識している。しかしながら、委員会は明るい未来のために、人々が自意識をコントロールする

エグゼクティブサマリー

データ環境の有効化を推奨し、データの非対称性を根絶できるツールの例や改善された取り組みを提供する。

課題：

- 個人は、アルゴリズム時代に自分の個人情報をどのように定義し整理することができるか。
- 個人識別情報の定義と範囲とは何か。
- 個人情報に関する管理の定義はどのようなものか。
- 個人を尊重するために、データアクセスをどのように再定義できるか。
- どのようにして、個人を尊重するように個人情報に関する同意を再定義することができるか。
- 共有することが重要でないように見えるデータは、個人が共有したくない情報の推測に使用され得る。
- 正しいインフォームド・コンセントのために、データの取扱者はデータのアクセスと収集による結果（良い結果および悪い結果）が個人に開示されていることをどのように保証するか。
- パーソナライズされたAIやアルゴリズム監視者を個人で所有できるか。

6 |自律兵器システムの再構築

身体的危害を加えるように設計された自律システムは、従来の兵器や危害を及ぼすようには設計されていない自律システムと比較して、更なる倫理的影響をもたらす。これらに関する職業倫理は、広範囲に及ぶ懸念事項に対応する高い基準を持つことが可能であり、持つべきである。大まかに、自律兵器システムの再構築に関する委員会（the Reframing Autonomous Weapons Systems Committee）は以下の点を提案している。

- 技術組織が、人間による兵器システムの有意義な管理が社会にとって有益であることを受け入れる。
- アカウンタビリティを保証する監査証跡により、そのような管理を確実にする。
- これらの技術を生み出している者が自分たちの仕事の意味を理解している。
- 職業倫理規定が危害を加えることを意図した創作物に適切に対処する。

課題：

- プロフェッショナルな組織の行動規範には、製作物に可能な範囲内で、メンバーの製作物、作成した人工物や化学物質に、メンバーの守るべき価値観や基準の順守を同様に強制できない重大な抜け穴がしばしば存在する。
- 人工知能、自律システム、自律兵器システム（AWS）の重要な概念に関する定義についての混乱は、重大な問題に関するより本質的な議論の妨げとなる。
- AWSは、その性質上、秘密裏に帰属不明の状態で運用されやすい。

エグゼクティブサマリー

- AWSの行動の説明責任が無効にされる可能性のある方法が複数存在する。
- AWSは予測可能ではない（その設計および作戦での使用に依存する）。学習システムにより、予測可能な利用の問題を難しくする。
- AWSの開発を合法化することは、中期的に見て地政学的に危険な前例を作る。
- 戦場から人間の監視を排除することは、不慮の人権侵害や意図せぬ緊張の高まりにあまりにも容易につながりかねない。
- AWSの直接的および間接的な顧客の多様性は、兵器の拡散および乱用の複雑で問題のある状況につながる。
- 何もしなければ、AWSにおける自動化は紛争の拡大を促進する。
- AWSの設計保証検証の基準がない。
- AWSと半自律型兵器システムの仕事における倫理的境界の理解は、わかりにくい場合がある。

7 | 経済的／人道的問題

日常生活における人間の介入を減らすことを目指す技術、方法論およびシステムは急速な進化を遂げており、個人の生活をさまざまな形で変えようとしている。経済的／人道的問題に関する委員会（the Economics/Humanitarian Issues Committee）の目的は、ヒューマンテクノロジーの世界的なエコシステムを形成する主な要因の特定、経済的および人道的な影響への対応、重要な難所を解消することによって実現できる解決に向けた重要な機会の提案である。委員会による提案の目標は、人間、その機関および情報主導型先進テクノロジーの間の関係における主要な問題に関連する現実的な方向性の提案を行い、これらの問題に関する専門的、指導的、および助言的思考により豊富な情報を与えられ得る学際的・産業横断的な対話を促進することである。

課題：

- メディアにおけるAI／ASの誤った解釈は、一般の人々の混乱を招いている。
- 自動化は、通常、市場の状況のみにおいて考察されることはない。
- ロボティクス／AIに関して、利用の複雑さが軽視されている。
- 労働力（再）訓練の既存の方法に対する技術的な変化が速すぎる。
- あらゆるAI政策は、革新を遅らせる可能性がある。

エグゼクティブ サマリー

- AIと自律技術は世界中で同じように利用できるわけではない。
- 個人情報に関し、利用する機会や理解が不足している。
- IEEEグローバル・イニシアティブでは、新興国の積極的な参加が必要である。
- AIや自律システムの出現は、先進国と新興国の間および各国内の経済格差や権力構造の相違を悪化させる可能性がある。

8 | 法律

AI／ASの初期の開発は、多くの複雑な倫理的問題を引き起こしてきた。これらの倫理的な問題は、多くの場合、直接的には具体的な法律的課題となるか、副次的な法的問題につながる。法律に関する委員会（the Law Committee）は、この分野の弁護士にとって、差し迫った必要性のある分野であるにもかかわらずこれまでのところ弁護士や学者をほとんど惹きつけていない多くの仕事があると考えている。弁護士は、これらの分野における規制、ガバナンス、国内法および国際法に関する議論に加わり、AI／ASによってもたらされる人類および地球にとっての大きなメリットを、未来のために慎重に管理する必要がある。

課題：

- 自律的でインテリジェントなシステムのアカウンタビリティと検証可能性をどのように改善できるか。
- AIの透明性およびAIが個人の権利を尊重することをどのようにして保証することができるか。例えば、国民は政府やそのAIを信頼して自らの権利を守ることができるはずなのに、国際統治機関、政府および地方自治体が国民の権利を侵害するようなAIを使用している場合などである。
- AIシステムは、これらのシステムに起因する損害に対する法的責任を保証するようにどのように設計できるか。
- 自律的でインテリジェントなシステムは、個人情報の完全性を保ちながら、どのように設計され展開され得るのか。

我々の新しい委員会とその現在の業務は、「倫理的設計」の最後に記載している。

エグゼクティブ サマリー

本書の作成方法

本書は、IEEE Standards Associationのプログラムであるindustry Connectionsプログラムのプロセスに従い、オープンで協調的な合意形成手法を用いて作成している。Industry Connectionsは、組織や個人が先進テクノロジーの問題に関する考え方を洗練させ、改善する際に、潜在的な新しい標準化活動や、標準化に関連する製品およびサービスの創出を支援し、組織や個人の間のコラボレーションを促進する。

本書の引用方法

「倫理的設計」の第1版を引用する際には、以下の通りに出典を記載すること。

人工知能および自律システムにおける倫理的考察のためのIEEEグローバル・イニシアティブ、倫理的設計：*人工知能と自律システムによる幸福を優先するためのビジョン*、第1版、2016年。
http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html

The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems. *Ethically Aligned Design: A Vision For Prioritizing Wellbeing With Artificial Intelligence And Autonomous Systems*, Version 1. IEEE, 2016.
http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html