

海外の議論の動向

平成 31 年 4 月
平 事 務 局

IEEE「倫理的に調整された設計 第1エディション」を公表【2019年（平成31年）3月25日】

- 3月25日、米国電気電子学会（IEEE）の自律的及び知的システムの倫理（Ethics of Autonomous and Intelligent Systems(A/IS)）に関するグローバル・イニシアチブは、報告書「倫理的に調整された設計(Ethically Aligned Design(EAD) 第1エディション)」を公表。
- 同エディションは、特に倫理課題から技術への橋渡しを目的とした標準であるIEEE P7000™シリーズや用語集の作成など「原則から実行へ（From Principles to Practice）」を重視。
- EADの概念的枠組の柱は大きく①人間の普遍的な価値（Universal human values）、②政策の自己決定とデータの仲介（Political Self-Determination and Data Agency）、③技術的な信頼性（Technical Dependability）の3つに分類される。また、それらに紐付けられる一般原則として、①人権（Human Rights）、②幸福（Well-being）、③データ仲介★（Data Agency）、④効能★（Effectiveness）、⑤透明性（Transparency）、⑥アカウントビリティ（Accountability）、⑦悪用への警戒（Awareness of Misuse）、⑧技能★（Competence）が掲げられている。（★印が本エディションで加わったもの）

欧州委員会「信頼できるAI（Trustworthy AI）のための倫理ガイドライン」公表【2019年（平成31年）4月8日】

- 4月8日、欧州委員会は、選定した52名の専門家グループ（HLEG）により作成された信頼できるAIのための倫理ガイドラインを公表。
- 同ガイドラインでは、信頼できるAIのためには合法的、倫理的、及び、頑健であるべきとし、その上で基本的人権に基づき尊重すべき4つの倫理原則（人間の自律性の尊重、危害の防止、公平性、説明可能性）、および7つの要求条件（人間の営みと監視、技術的な頑健性と安全性、プライバシーとデータガバナンス、透明性、多様性・無差別・公平性、環境及び社会の幸福、アカウントビリティ）を掲げ、さらにそれらを評価するためのチェックリスト（Assessment list）を列挙している。
- 掲げられた内容は非拘束的なものとしてAIを開発・利用する全ての関連するステークホルダーを対象としており、企業・団体等の各ステークホルダは、本ガイドラインの内容を咀嚼し、自らの憲章・行動規範等に適用することにより、信頼できるAIへの関与の表明が可能。
- 今後、上記チェックリストについては、広く継続的にレビューを行い、2020年に取りまとめる予定。

豪州産業イノベーション科学省 「豪州のAI倫理フレームワーク」案公表、意見募集開始 【2019年（平成31年）4月5日】

- 4月5日、豪州産業イノベーション科学省（the Department of Industry, Innovation and Science）は、同国の科学産業研究機構（CSIRO）の研究部門Data61により作成された「豪州のAI倫理フレームワーク」に関する議論ペーパーを公表すると共に、意見募集を開始した（5月31日まで）。
- 同ペーパーでは、コアとなる8原則（（1）純粋な利益を生み出すこと、（2）危害を与えないこと、（3）規制および法令遵守、（4）プライバシー保護、（5）公平性、（6）透明性と説明可能性、（7）競争力、（8）アカウンタビリティ）を掲げた上で、データガバナンス、自動決定、人間の振る舞いの予測などの課題に対するこれらの原則の在り方及び、原則に対しどうしていくか（倫理フレームワーク）を記載。
- 倫理フレームワークでは、原則から実施へと題し、影響評価、レビュープロセス、リスク評価、ベストプラクティスガイドライン、教育、産学連携、AIのモニタリング、（問題が起きたときに）助けを求めるメカニズム、コンサルテーションの在り方に触れている。