

IEEE:標準化によるAIのELSI (Ethical, Legal, Social Issues) 問題の解決

2018/01/18

日本電気(株) 標準化推進部 江川尚志

ELSI: AIの社会受容に向け考えるべきこと

内閣府「人工知能と人間社会に関する懇談会
(～2017/3)」の議論のごく一部

Ethics (倫理)

- AIを用いた恋愛ビジネスはどこまで許されるか
- AIに与えるデータを選択することで、最適化とそれによる人間行動の誘導が可能に
- 人工知能による与信審査で、その手続が不透明な場合、人工知能が人間の価値を測ることに
- 人間の生死や尊厳に関わる判断をインテリジェントICTに委ねてよいか

Law (法)

- AIによる事故、誤動作等の責任の所在; 人工知能等を使わないリスクも
- AIを道具として利用した創作物と、AI自身による創作物（人の関与が少ない）について、前者には権利が発生するが後者には発生しない。しかし、両者を区別するのは困難
- 医療や金融等についての自分の個人情報から近親者の健康や資産等が推測可能な場合
- ルールがないことによる萎縮効果

Society (社会) (内閣府懇談会では更に経済、社会、教育に分類)

- 労働流動性が高まったときに転職等のためのコストが必ず生じるがそのコストを誰が支払うべきか
- 従来であれば何人もの部下を使って総合力で進めていた仕事がAIと個人だけで実行可能に
- 人工知能による新サービスや人間の代替が進展すれば、例えばパターンリズムの問題が現実化
- 情報リテラシーだけでなく法リテラシーの教育を、また、科学コミュニケーションならぬ法コミュニケーションといったものと考えていくべき

これを実務者が「使える」ルールに置き換える必要性

自主規制、共同規制としての標準

現代の情報経済における直接的な「政府規制」の困難さ

- 技術的・ビジネス的イノベーションの速度
- 規制策定に必要な専門的知識
- 安全・安心やプライバシー等、法による画一的定義が困難な領域の拡大
- 利用者への実質的規制能力（アーキテクチャ・利用規約等）を有するプラットフォームの台頭
- グローバル環境での一国政府規制能力の限界
- 表現の自由への配慮

政府規制を補う自主規制、共同規制（時に規制なし）の一種として標準が使われる

規制なし	特に規制の必要なく、市場自身が問題の発生を抑止あるいは解決している
自主規制	業界団体等による自主的な規制によって当該問題が適切に解決されている（政府による一般原則の提示は存在し得る）
共同規制	自主規制と政府規制の混合措置により問題が解決されている（政府の自主規制補強措置が存在する）
政府規制	目的とプロセスが政府によって定義されており、政府機関によるエンフォースメントが担保されている

規制弱

応答的対応

規制強

IT戦略本部シェアリングエコノミー検討会議 第1回会合資料1-7（生貝直人委員資料）を参考に作成

IEEEにおけるAI倫理の標準化

2016年4月テーマ洗い出しを開始。合意したものから具体的標準化開始

論点・標準化テーマを洗い出し



具体的な標準の策定

Global Initiative on Ethics of Autonomous and Intelligent Systems

一般原則

AI, ASへの価値組み込み

倫理的研究・設計を導く方法論

汎用AIと超知性の安全性と受益者

.....

.....

計18委員会（技術討議は13委員会）

特定された論点とそれに対する勧告はEADという文書に（現在ver.2）

P7000:倫理的設計のモデルプロセス

P7001:自律システムの透明性

P7002:データプライバシーのプロセス

P7003:アルゴリズムバイアス

.....

.....

P7010まで計11プロジェクト。
当面はこれで打ち止め、必要に応じ追加

委員会名称: Methodologies to Guide Ethical Research and Design

共同議長: Raja Chatila (Co-Chair) – CNRS

Corinne J.N. Cath – (Co-Chair) PhD student, U-Oxford

下記論点について問題の明確化、勧告と参考文献を記述

学際的な教育と研究

- 倫理がAI/S関連の学位取得過程に十分に組み込まれていない
- 学際的で異文化にまたがる教育モデルの必要性
- AI設計の中に組み込まれる文化特有の価値の明確化

企業の実務とAI/S

- 産業界のための価値に基づく倫理の文化や実務の欠如
- 価値に基づくリーダーシップの欠如
- 倫理上の懸念の表明促進の欠如
- 技術コミュニティのオーナーシップや責任感の欠落 (技術関係者が自分の責任と思う事と、一般の人々や法曹や社会学の関係者が思う倫理上の責任には一般に乖離がある)
- ステークホルダーの巻き込み

AI/S技術の開発や試験のための研究の倫理

- 組織の倫理委員会のリソース不足

透明性の欠如

- 不十分な文書化が倫理的設計を妨げる
- アルゴリズムの首尾一貫性のなさや監督の欠如
- 独立の評価委員会の欠如
- ブラックボックスの使用

IEEE Ethically Aligned Design, ver.2 (2017/12より)

IEEE P70xx標準化プロジェクト一覧

	タイトル	役職者	開始/完成目標
P7000	倫理的設計の モデルプロセス	議長: John Havens (IEEEコンサルタント、社会運動家) 副議長: Sarah Spiekermann (ウィーン経済・経営大学)	2016年9月/ 2018年末
P7001	自律システム の透明性	議長: Alan Winfield (西イングランド大) 幹事 江川尚志 (NEC)	2016年12月/ 2018年1月
P7002	データプライバシー のプロセス	議長: Michelle Dennedy (cisco) 副議長: Aurelie Pols (Mind Your Privacy(コンサルタント会社)) 幹事: Matthew Silveira (Objective Business Solutions)	2016年12月/ 2018年1月
P7003	アルゴリズムック バイアス (差別)	議長: Ansgar Koene (ノッティンガム大) 副議長: Paula Boddington (オックスフォード大)	2017年2月/ 2018年7月
P7004	子供と学生データ のガバナンス	議長: Marsali Hancock (DQ Institute, シンクタンク, NPO)	2017年3月/ 2019年2月
P7005	従業員データ のガバナンス	議長: Ulf Bengtsson (Sveriges Ingenjorer, スウェーデン大学卒エンジニア協会) 副議長: Christina Colclough (UNI Global Union, 労働組合の国際組織)	2017年3月/ 2017年12月
P7006	パーソナルデータ AIエージェント	議長: Katryna Dow (Meeco) 副議長: Gry Hasselbalch (DataEthics (シンクタンク). コペンハーゲン大教授でもある)	2017年3月/ 2017年12月
P7007	用語	議長: Edson Prestes (リオグランデ・ド・スル連邦大学) 副議長: Sandro Rama Fiorini (パリ第12大学)	2017年3月/ 2018年3月
P7008	人を倫理的に つき動かすAI	議長: Laurence Devillers (LIMSI, CNRS付属研究所) 副議長: John Sullins (ソノマ州立大学)	2017年7月/ 2018年12月
P7009	AIのフェール セーフ設計	議長: Danit Gal(北京大学, IEEEアウトリーチ委員会議長) 副議長: Alan Winfield (西イングランド大)	2017年7月/ 2018年12月
P7010	AI時代の 幸福の指標	議長: John Havens (IEEEコンサルタント、社会活動家) 副議長: Laura Musikanski (Happiness Alliance, NPO)	2017年7月/ 2018年12月

AIならでは: P7001, 03, 08; 個人情報保護系: P7002, 04, 05, 06; ソフトウェア系: P7000, 09; その他: P7007, 10
当面、これで打ち止め。必要があれば追加

開始: 上位委員会 (NESCO)でプロジェクトが設立承認された日時。完成目標: 親ソサエティでの第1回スポンサー投票目標日時。投票成功=技術的に完成とみなされる。その後議論の公平性等のチェックを経て半年後に正式標準に

P7000:倫理的設計のモデルプロセス

- 委員会名称: Model Process for Addressing Ethical Concerns During System Design (システム設計時の倫理上の懸念を扱うためのモデルプロセス)
- スポンサー(*): Computer Society, Software & Systems Engineering Standards Committee (C/S2ESC)
- 議長: John Havens (IEEEコンサルタント、社会運動家)
副議長: Sarah Spiekermann (ウィーン経済・経営大学)
- 開始: 2016/09, 第1回スポンサー投票(**) 目標: 2018末, 投票: 個人
- 技術に関わる関係者がシステム開発を開始し、分析し、設計する時に倫理上の懸念を適切に扱う事が出来るよう手順のモデルを定める
- 関連する標準: 存在せず。関連文書: IEEE, 及びIEEE Computer Society/ACMソフトウェアエンジニアの倫理綱領 (code of ethics)
- 利用者: 技術者に加え研究者、企業、大学、各種の組織など技術の具体化に関わる組織を広く想定

* スポンサー: 技術的内容に責任を持つ組織

** スポンサー投票: 承認されると、技術的には完成と見なされる。その後、役員会で手続きの公平性他がチェックされ、問題がなければ約半年後に正式標準に

リスク管理の枠組みを使い、設計プロセスへ倫理を組み込むための方法論と要求条件を規定

- 1: 新製品・新サービスが提供する価値と害を特定, 2: 組み込むべき価値(e.g., プライバシ)を特定, 3: リスク管理の観点から取るべき行動 (e.g., 設計で解決できるもの、組織論で対応するもの) を決定
- 各段階での関係者、基本的な作業フロー、考慮すべきこと (e.g., 汎用的な価値を最初に考慮せよ、大切な人のために作ると思え、他) を記述
- 具体的な価値特定手段、取るべき行動は要求条件とはしない
応用分野 (e.g., telecom) に特化した記述は例えば7000.xで記述される可能性

現在75ページ、基本的な枠組みは記述済。テーマ毎のチームが詳細記述の完成度向上に努める

(1) Ethical Business or Mission Analysis Process

Harms & benefits are collected with the help of

- Utilitarian Analysis
- Virtue Ethics
- Deontological Analysis

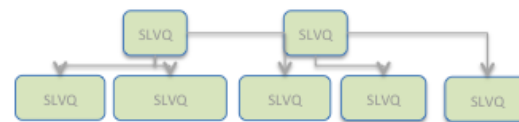
... followed by the identification & prioritization of underlying values.

(2) Value Needs Refinement Process

Refining the value spectrum with regulation & P7000 Baseline Values



& conceptually investigating all values with a view to their "System Level Value Quality Requirements" (SLVQR)



(3) Value based Risk Management Process

Setting SLVQRs as Value Requirements at Risk (SLVQR@R) & choosing system controls for each SLVQR@R
Controls are called "Value Based System Design Specifications" (VBSDS)

Ongoing participatory stakeholder involvement

Ongoing participatory stakeholder involvement

倫理的設計の中核プロセス

 **Orchestrating** a brighter world

NEC