

AI社会論研究会講演資料

標準化を用いたAI倫理の方向付け IEEEのイニシアチブを中心に

2017/03/01

NEC 標準化推進部 江川尚志
(t-egawa@ct.jp.nec.com)

元々は通信技術の研究者

- インターネット、電話技術の中でも特にfuturisticな技術を研究
(技術英語はともかく法律や哲学の英語は駄目です。本日の講演資料に誤訳があったら是非ご教授を)
- アーキテクチャが専門。だから「それは私とは関係ない」だけは絶対に言いません
巨大な通信ネットワークの設計には、技術は当然として、法律、経済、政治、社会、文化、宗教、すべてが関わる
「知らないです、不勉強です」は言います (事実ですし)

先日までITUでSDNという通信技術標準化の分科会議長

- ITU: 国連傘下の、通信に関する専門機関
- ITU-T SG13 JCA-SDN議長、Q.14ラポータ
- 総務省やNICTのご指導・ご協力を得つつ、公的な標準化活動に10年強たずさわる
近年は約100日/年海外出張
→ これだけ出張すると国内や現場の議論から遊離した「標準化屋」になりがち。
実際、その道を順調(?) に歩んでいた
- 2016年末で辞任；AIのELSIやオープンソースを通じた標準化に活動をシフト

イントロダクション：そもそも標準とは

カードのサイズは国際標準で決まっている

日本工業規格

JIS

X 6301 : 2005

(ISO/IEC 7810 : 2003)



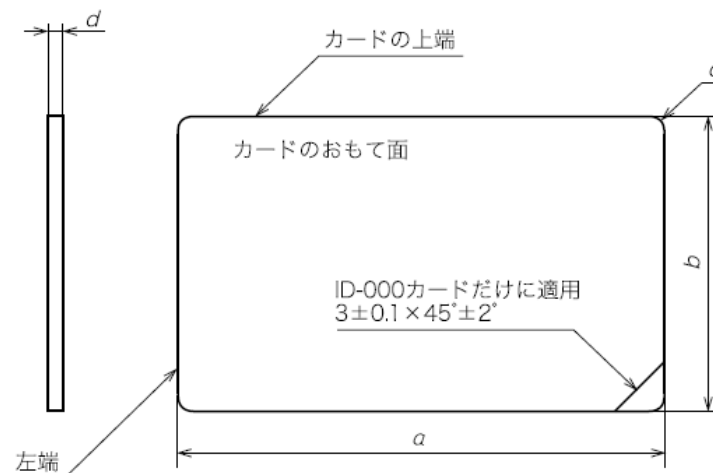
識別カードー物理的特性

Identification cards—Physical characteristics

単位 mm

序文 この規格は、2003年に第3版として発
characteristics を翻訳し、技術的内容及び規格票の
なお、この規格で点線の下線を施してある“参

1. 適用範囲 この規格は、カードの材質、構成
にカードという。)の物理的特性について規定する
用途で定義されるIDカードの特性について規定し
この規格で規定するパラメタに対する供試カー
この規格は、IDカードが満たすべき物理的特性
及び機器的要素の双方を勘案したうえでの最低限
が、この一連の規格の目的である。使用したこと
度使われた後であるかについて、この規格では考
と協議することが望ましい。

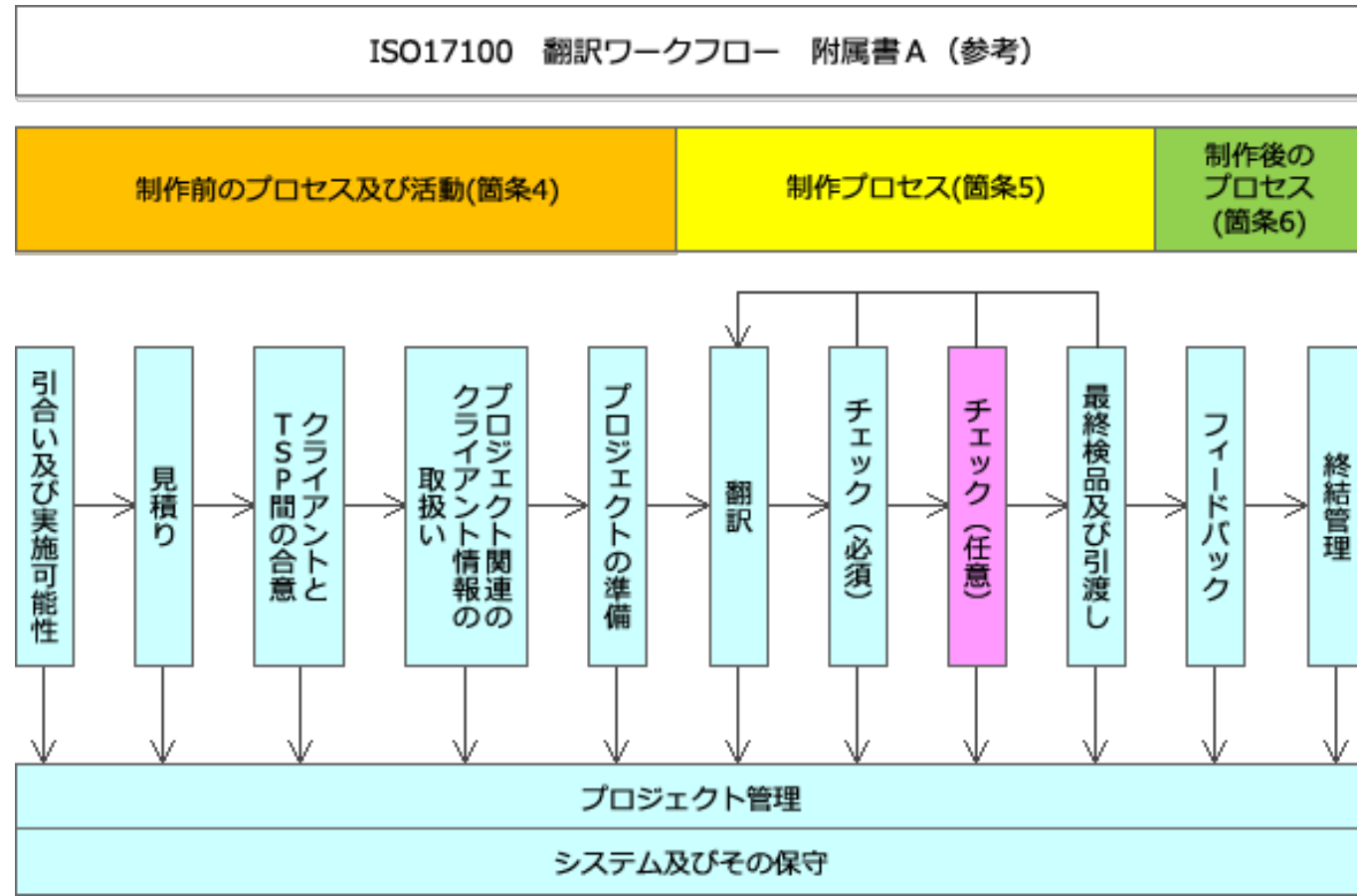


	a (幅)		b (高さ)		c (角の丸み)		d (厚さ)	
	最大	最小	最大	最小	最大	最小	最大	最小
ID-000 未使用カード	25.10	24.90	15.10	14.90	1.1	0.9	0.84	0.68
ID-1 未使用カード	85.72	85.47	54.03	53.92	3.48	2.88	0.84	0.68
ID-1 返却カード	85.90	85.47	54.18	53.92	3.48	2.88	0.84	0.68

翻訳サービスの品質保証も国際標準で

ISO 17100:2015 Translation services - Requirements for translation services

適正な翻訳サービスの提供に必要な所定のプロセス要件を翻訳サービス提供者が満たしていることを評価 ⇒ 翻訳サービス提供者の認証制度



メンバーのスタイルや売り出し方を標準化した アイドルビジネスも

- 同時に複数の場所で公演可能。メンバーが交代してもサービス継続
- ウィーン少年合唱団が本分野の先駆者

© 江藤学先生

標準の9個の機能: ISO設立25周年記念出版物より

1. 単純化
2. 互換性の確保
3. 伝達手段としての標準
4. 記号とコードの統一
5. 全体的な経済の効果;
6. 安全・生命・健康の確保
7. 消費者の利益の保護
8. 消費社会の利益の保護
9. 貿易の壁の除去

T.R.B. Sanders, *The aims and principles of standardization*, 1972

標準とは何か、標準化屋は延々と議論し標準化。
だが、本質は「人々が尊重する」であろう

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

- 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

AIの社会実装に向け標準が担える機能と進行中の標準を概観、
有意義なAI関連標準作成のために取るべき行動を整理

■ イントロダクション：そもそも標準とは

■ 様々な標準：AIの規制に適した標準に向けて

■ AIを巡る標準化

- 米国科学技術政策局 'The National Artificial Intelligence Research and Development Strategic Plan'
- IEEE SA 'Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems' と P7000 シリーズ 標準化 プロジェクト
- BS8611:2016 'Guide to the ethical design and application of robots and robotic systems'
- (ご参考) ITU-T Y.3013 'Socio-economic assessment of future networks by tussle analysis'

■ 有意義な標準化のために

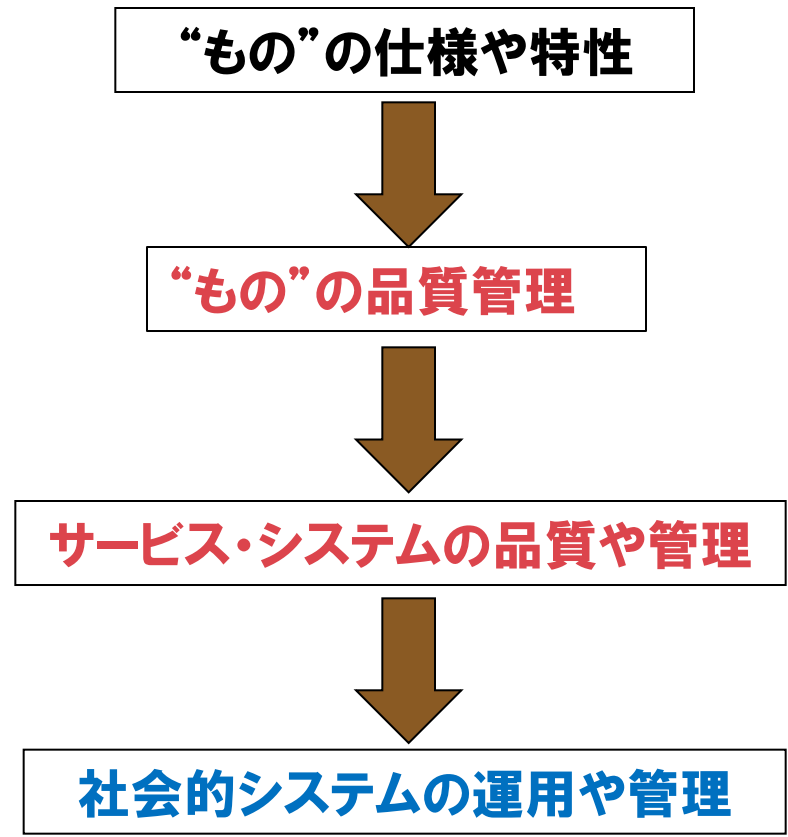
様々な標準

AIの規制に適した標準に向けて

*標準、基準、規範、規格と、日本語では多様な表現があるが、これらを包含して「標準=スタンダード」と表記する

標準化の対象領域は拡大中

標準化の対象はモノやサービスから社会システムまで



◆ISOの技術委員会(TC)名称の例

- TC1 ねじ
- TC8 船舶および海洋技術
- TC42 写真
- TC61 プラスチック
- TC176 品質管理及び品質保証
- TC210 医療用具の品質管理
- TC224 上下水道サービスの品質基準
- TC228 観光及び関連サービス
- TC232 人材育成と非公式教育サービス
- TC242 エネルギーマネジメント
- TC268 コミュニティにおける持続可能な開発 (スマートシティ)
- TC309 組織のガバナンス

■ 認証制度がある標準は、**認証取得が受注の必須要件**となることがある

- コモンクライテリア (CC:ISO/IEC 15408)

情報技術セキュリティの観点から、情報技術に関連した製品及びシステムが適切に設計され、その設計が正しく実装されていることを評価

⇒ITセキュリティ評価及び認証制度 (JISEC)

- 情報セキュリティマネジメント (ISMS:ISO/IEC 27001)

事業の活動全般及び直面するリスクを考慮して、文書化された管理体制を確立、導入、運用、監視、見直し、維持し、かつこれを継続的に改善すること

⇒ISMS適合性評価制度



- サイバーセキュリティマネジメント (CSMS:IEC 62443-2-1)

産業用オートメーション及び制御システムを対象として、その構築から運用・保守に渡ってサイバー攻撃から守るためのセキュリティ対策を実施し、システムを運用するもの

⇒CSMS適合性評価制度



強制法規に取り込まれる標準

- 日本工業規格(JIS)は任意規格であるが、法令からJISが引用されると、**強制力が発生する**
- 発効中の法令約8千件の5%でJISを引用
- 保健医療分野では、情報化の促進のために**厚生労働省標準規格**が定められ、**推奨標準**となっている

種類		制定機関	発効件数	JIS引用件数
法令	憲法・法律	国会	1,961	3
	政令・勅令	内閣	2,225	20
	府令・省令	各府省	4,098	364
	(合計)		(8,284)	(387)
法令の 補足	告示	各府省	件数不明	~400
	通達	各府省	件数不明	件数不明

2016年11月現在



標準の作成の仕方による分類

①デジュール標準 (de jure standard)

- “de jure”はフランス語の「法律上で正式の」の意味
- 公的な機関で明文化された手続きによって作成
- ≒「共同規制」

②フォーラム・コンソーシアム標準

⇒デファクト標準に分類されることもある

- 複数の参加者の合議によって運営される標準化団体で作成
- 同一分野で競合する標準化団体も存在しうる
- ≒「自主規制」

③デファクト標準 (de facto standard)

- “de facto”はフランス語の「事実上の」の意味
- 市場競争を勝ち抜き、実質的に市場を浸透している標準
- ≒市場の力のみに基づく「規制なし」

(例) ICカード、パスポート



(例) DVD、Blu-ray



(例) iOS、Windows



主要なデジュール国際標準化機関

()内は設立年次

機関 (創立年)	ITU (1865) ※国連の機関		ISO (1947)	IEC (1906)	
	ITU-T	ITU-R	※民間の公益法人	ISO/IEC JTC1(1987)	※民間の公益法人
標準化分野	通信一般 (無線を除く)	無線通信 (放送を含む)	工業標準全般 (電気一般を除く)	情報技術	電気一般
日本の 代表機関	総務省 (情報通信審議会)		日本工業標準調査会：JISC (経済産業省の審議会組織)		
国内の関係 標準化機関	TTC	ARIB	日本規格協会 各種工業会 学会 等	情報処理学会 JBMIA JEITA	日本規格協会 電気学会 各種工業会 等
海外の関係 標準化機関	ETSI(欧州) ATIS(米国)		CEN(欧州)	INCITS(米国)	CENELEC (欧州) EIA(米国)

＊他には国連CEFACT等の特定分野での国際標準化組織が存在する

フォーラム標準化団体の例

名称	標準化分野	参加単位	参加数
IETF (Internet Engineering Task Force)	通信プロトコル等インターネット関連技術	個人	1000～2000人
W3C (World Wide Web Consortium)	HTML, HTTP 等Webにおける各種技術	企業・組織	～350団体
USB-IF (USB Implementers Forum)	USB インターフェイスの仕様	企業	～180社
OMA (Open Mobile Alliance)	携帯電話のアプリに関する規格	企業・組織	～110団体
Unicode Consortium	文字符号化方式の規格	企業・組織 ・個人	～40団体 ～100人
AUTOSAR (Automotive Open System Architecture)	自動車搭載ソフトウェアの仕様	企業	～100社
CEA (Consumer Electronics Association；全米家電協会)	テレビや家庭内ネットなどの規格	企業	～2000社
IEEE-SA (IEEE Standards Association)	電力、電子、通信など幅広い分野における規格	個人 (企業・組織)	～15000人 (～200団体)

◆標準化を主に行う団体と、標準化部門を持つ業界団体や学会とがある

デファクト標準の例

多くのデファクト標準は、特定企業の規格として始まり、後にデジュールやフォーラム標準となる。例えば現在、MS OfficeのファイルフォーマットはISO/IEC標準。HTMLはW3C規格。コンテナはISO標準。

ICT分野

MS Windows, MS Office

HTML

携帯電話: Android OS

CPUアーキテクチャ Intel x86

PDF (Adobe Acrobat)

QUERYキーボード



物流



コミュニケーション



新しいデファクト標準化手法：オープンソースソフトウェア

議論する暇があったらさっさと作って市場を占有してしまえ



コンピュータのプログラムを作り、**無料で公開**、**誰でも貢献可能**

- 週末プログラマのお勉強レベルから、東京証券取引所の取引システムの基盤まで
- 最先端の巨大システム向けソフトが無料で公開されている
- 膨大なソフトが絶えずつくられ、ごく一部が生き残る多産多死モデル

ユーザへのアピール力は抜群。「**動かして試して下さい**」「**無料です**」

エンジニア的には、紙の標準よりも説得力があり逆らい難い

- 紙の標準化：英語が酷い、そこカンマが抜けてる、その言い回しは曖昧だ、、、
- オープンソース：昔、基本ソフトの中核 (Linux kernelのスケジューラ) での論争
「xxの方がエレガントだ」「その方式を実装する難しさ、分かっているのか。**作って持
って来い**、そうしたらまた議論しよう」

オープンソースソフトウェアが広まる技術的理由

ソフトウェアのプログラムが持つ特殊な性質

- 純粋な情報のため低コストかつ容易にコピー可能
- 腐らない。一度作れば永久に使える (理論上は)
- 天才プログラマは凡庸プログラマの100倍の生産性
そうした天才プログラマは自由時間があるし、プログラミングが大好き
一個人でもまとまった仕事の実現可能に
また、これがあるため、OSSコミュニティは徹底した実力主義

現代の製品・システムの大半は「汎用ハード+個別ソフトの組み合わせ」

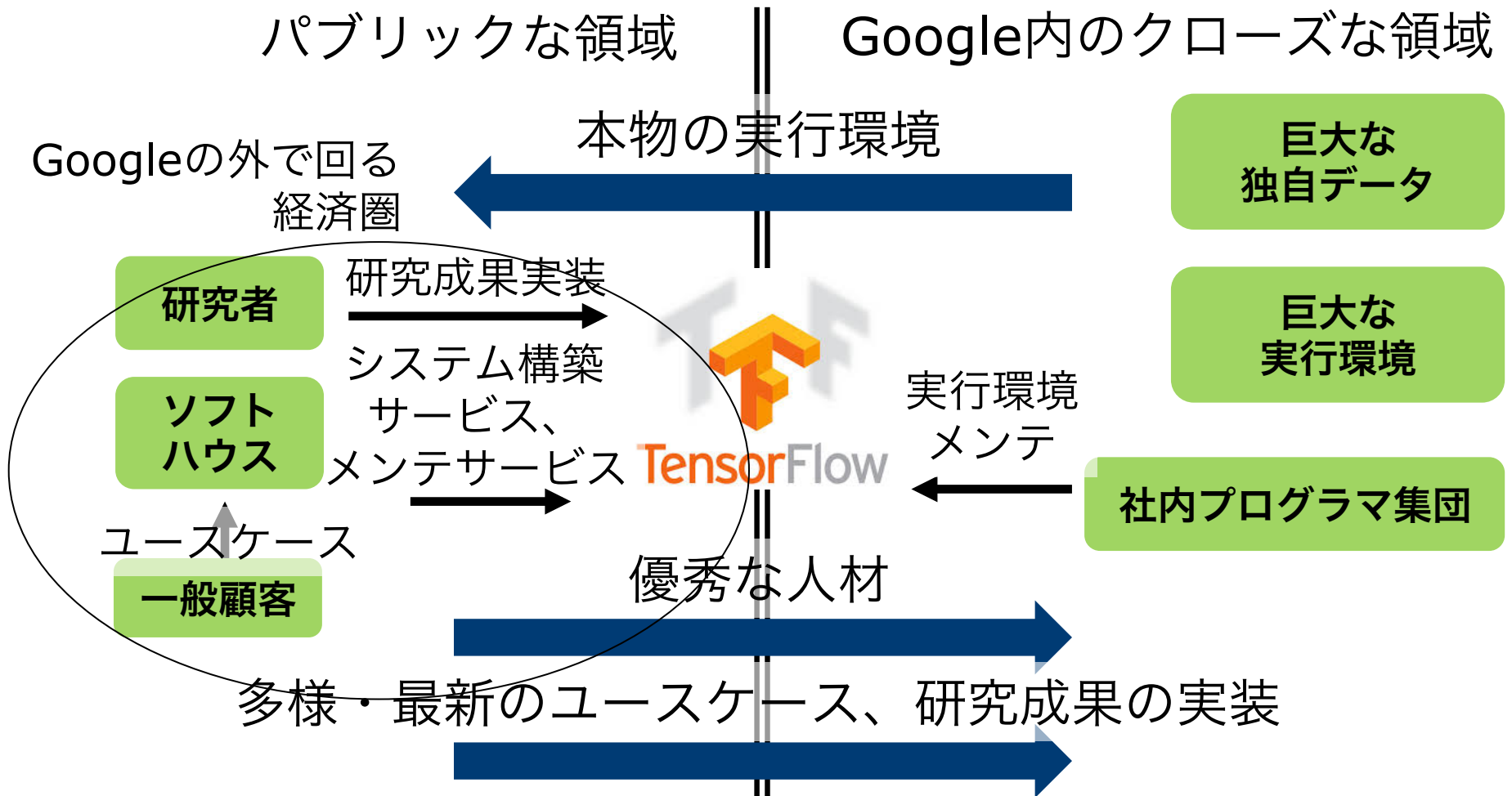
- 差異化の大半はソフトウェアで実現される
- その結果、コンピュータ産業のソフトウェアの文化が他産業に拡散中

インターネットにより世界的な分散協力体制の構築が容易に

ハードウェア向けにオープンソース・ハードウェア、オープンデザイン等の概念も提案されているが、ソフトウェア向けほど広まっていない

技術的には、近い将来、OSSの領域拡大を妨げる理由はなさそう

TensorFlow公開によるGoogleの損得勘定



- 自社OSSの周囲に研究者・開発者コミュニティが出来ればリターンは巨大
- 自社外で回る経済圏も。だが、その人々からも無形の感謝は受ける

進行中・完成済AI関連標準化(1)
米国科学技術政策局「AI研究戦略計画」

■ 米国ホワイトハウスはAIの政策を検討するため、本年5-7月に公開ワークショップを4回開催、意見募集を経て報告書を発行。これを皮切りに様々な部署が10-12月に計3本の報告書を発行

1. 「Preparing for the future of AI (人工知能の未来に備えて)」大統領府・科学技術政策局 (Office of Science and Technology Policy, OSTP)、10/12発行
AIと規制; 研究及び労働力; 公正、安全、ガバナンス; 世界におけるAI利用とセキュリティ、の主要4課題について勧告
2. 「The National Artificial Intelligence Research and Development Strategic Plan (米国AI研究開発戦略)」科学技術会議 ICT技術研究開発小委員会 (National Science and Technology Council (NSTC), Networking and Information Technology Research and Development Subcommittee)、10/13発行
AI研究開発で中国に「引用された論文数」で抜かれたことを述べ、7個の戦略を勧告
3. 「Artificial Intelligence, Automation, and the Economy (AI, 自動化、経済)」Executive Office of the President (大統領行政府)、12/21発行
AIへの投資や開発の推進 (特にサイバー防衛)、大人にも子供にもAIを見据えた将来の職業向け訓練の導入、転職を支援しAIの成果が幅広くいきわたらせること、の3政策を提言。
第2, 第3の政策は教育制度、最低賃金、労働組合、医療保険など永年の課題を詰め込んだ観あり

米国科学技術会議「AI研究戦略計画」によるAI標準化戦略 (2/2)

米国科学技術会議 ICT技術研究開発小委員会は2016年 10月「米国人工知能研究開発戦略」として連邦政府の予算によるAI研究の方針を策定

- 連邦政府の予算によるAI研究の究極の目標として、社会に便益をもたらす新たなAIに関する知識及び技術を生み出しつつ、ネガティブな影響を最小化することを提示
- 同戦略は、この目標を実現するために優先的に取り組むべき事項を下記と設定

1. AI研究に長期的投資
2. 人間とAIの協働に向けて効果的な方法を開発
3. AIの倫理的、法的及び社会的含意を理解し、それらに対処
4. AIシステムの安全性及びセキュリティを確保
5. 共有される公共的なデータセット及びAIの訓練・試験のための環境を開発
- 6. 基準及びベンチマークを通じてAI技術を計測し評価**
7. 国家のAI研究開発人材のニーズをより良く理解

総務省AIネットワーク社会推進会議第1回資料3-2 (事務局資料) を纏め直し

戦略6: 標準やベンチマークを通じたAI技術の計測と評価 (1/3)

戦略6: 標準やベンチマーク、テストベッド、AIコミュニティがそれらを受け入れる事が重要。特に下記は重要 (総務省訳では標準ではなく基準)

- 多岐にわたるAI標準の開発 (詳細: 次ページ)
- AI技術ベンチマークの確立 (詳細: 次次ページ)
- AIテストベッドをより広く利用可能とすること
AIの全分野においてテストベッドは重要である。政府には大量のユニークなデータがある。これを研究者が利用可能とするテストベッドは、他では得難い機会を与える。
- AIコミュニティの標準やベンチマークへの関与
政府がリーダーシップと調整を行い、ユーザ、インダストリ、アカデミアを巻き込んでコミュニティとして標準やベンチマークを作成し、広く利用されるようにすべき。過去の例としてはNISTの情報検索の方法論を評価するText Retrieval Conference (TREC)、顔認識のFace Recognition Technology (FERET) (後にFRVTに進化) がある。

テストベッドを作りコンテストを行い、その評価基準を標準とするシナリオか

「多岐にわたるAI標準の開発」

急進展するAIに合わせたAI標準の開発が必要。標準によりカギとなるobjectiveの機能や相互接続性、信頼性や安全性が確保できる。例えばIEEE P1872-2015 (Standard Ontologies for Robotics and Automation) は概念と用語を整理。

その上で、必要な標準として下記を列挙。

■ ソフトウェアエンジニアリング：システムの複雑性や持続性、セキュリティの管理、創発的な振る舞いのモニタと制御

■ パフォーマンス：正確性、信頼性、頑健性、アクセシビリティとスケーラビリティの保証

■ 尺度：パフォーマンスやコンプライアンスに影響する因子の定量化

■ 安全性：システム、人間とコンピュータの相互作用、制御系、規制へのコンプライアンスについてのリスク管理とハザード解析を評価

■ 使い勝手：インターフェースや制御が効果的で、効率よく、直感的であることの保証

■ 相互接続性：交換可能なコンポーネントやデータ、トランザクションモデルを標準や互換インターフェースを通じて規定

■ セキュリティ：情報のCIA (秘匿性、完全性、可用性) とサイバーセキュリティを記述

■ プライバシー：情報が処理、転送、保存される際の保護を制御

■ 追跡可能性：イベント (の実行、試験、完了) の記録、データのキュレーション

■ ドメイン：ドメイン特有の標準の用語や対応するフレームワークの定義

「AI技術ベンチマークの確立」 試験計画や手順を規定した試験方法を開発し標準化すべき

ベンチマークが持つ機能

- 試験と評価から成る
- 標準の開発、および標準への準拠において、定量的評価を与える
- 目標を示し、戦略的に選ばれたシナリオへとイノベーションを導く
- AIの科学や工学の進展の指標にもなる

ベンチマークでの定量的評価項目 (例)

- 正確性、腹圧性、信頼と権限、リスクと不確実性、説明性、意図しないバイアス、人間の能力との比較、経済的インパクト

過去、ベンチマークにより技術進歩を導いた例

- NISTによる緊急対応ロボット向けの試験手順と評価方法
顧客には購入時の目安を、開発者には利用時に必要な機能を示した
ASTM internationalでE54.08.01として標準化
ロボカップレスキューロボット他で用いられる
- IEEEとNISTによるARIAC (Agile Robotics for Industrial Automation Competition)
ロボットシステムの機敏さを、フロアを動くロボットが製造ラインのエラーの検出、新製品が投入された時のプログラマの手間、等を条件として競う

AIには、過去の例以上にドメイン横断、汎用的な標準、ベンチマーク、テストベッドが必要

進行中・完成済AI関連標準化(2) IEEEのAI倫理イニシアチブとP7000シリーズ標準化

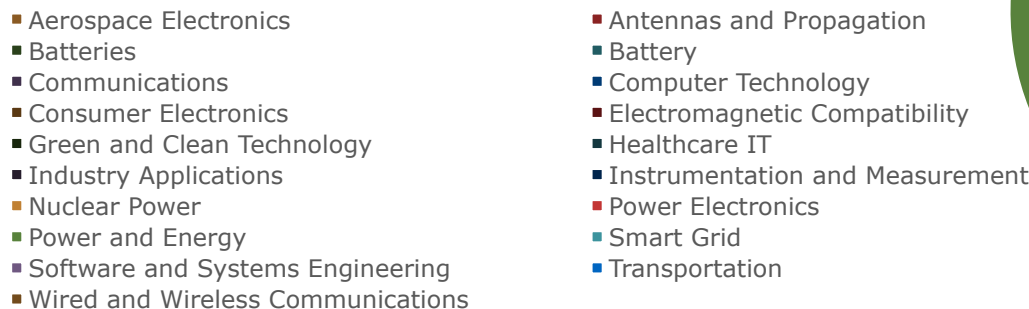
IEEE (Institute of Electrical and Electronics Engineers) 情報通信や電力を中心とした米国に本部を持つエンジニアの団体

- エンジニアの団体として学会、教育、資格認定等とならび標準化も行う
仲間内の規約としての標準
- メンバーの分布 (2013) : 49.8% 米国; 23.4% インド、中国、日本、他環太平洋地域;
18.3% ヨーロッパ、中東、アフリカ; 4.1% カナダ; 4.4% ラテンアメリカ
- メンバーの業界分布 (2013): 39.7% インダストリー; 34.3% アカデミア; 7.8% 政府
機関; 5.9% 自営業; 4.3% 退職者; 2.1% 無職; 5.9% その他

IEEE Standards Association (IEEE SA)

IEEEの分野別標準化プロジェクト数

- IEEEの標準化部門。1973年に設立されたStandards Boardを
発展的に解消、1998年に現在のStandards Associationに
- IEEE 802.3 (イーサネット), 802.11 (WiFi) の標準が有名
- プロジェクト数では電力系が最多



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。合意したもののから具体的標準化開始

論点・標準化テーマを洗い出し

Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous 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17委員会（技術討議は12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



具体的な標準の策定

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

(P7004:子供と学生のデータのガバナンス)

(P7005:従業員データのガバナンス)

(P7006: パーソナルデータAIエージェント)

P7004-06は上位委員会に提案中/承認待ち

IEEE AI倫理を考えるグローバルな率先垂範 (イニシアチブ)

名称: Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems (AIとAS(*)において倫理を深く考えるグローバルな率先垂範 (イニシアチブ))

論点を洗い出し標準化テーマを特定するための活動 (標準化は行わず)

設立: 2016年4月; 活動期間は2017年末まで (IEEE SAの活動は時限が基本)

- 設立前から非公式に議論が行われ、2016年3月時点で65名、10月時点で120名が参加
- 活動期間終了後は、アライアンス等へと発展的に解消することを検討

議長: Raja Chatila 仏国立科学研究センター (CNRS) インテリジェントシステム&ロボット研究所 (ISIR) 所長

傘下に17の委員会 (うち技術討議は12) を設置して**AIの倫理を中心に法律や雇用にも配慮して論点整理**

- 委員会の例: 一般原則、AI/ASへの価値組み込み、倫理的研究と設計を導く方法論 他

成果目標

1. Ethically Aligned Design (EAD, 倫理的整合のとれた設計) 文書の作成
委員会のテーマごとに論点とそれに対する勧告をまとめた文書
2. 成果をP7000シリーズ標準化プロジェクトに反映

* 純粋なソフトはAI、物理的実体を持つロボットはASと呼んでいる模様。あまり区別されておらず、委員会毎にも少し表現が揺らぐ。AIS (Autonomous Intelligent Systems) と呼ぶ場合も

IEEE イニシアチブ：執行委員会

- Raja Chatila, Executive Committee Chair, The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems (仏CNRS ISIR所長)
- Kay Firth-Butterfield, Executive Committee Vice-Chair, The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems (英国判事)
- John C. Havens, Executive Director of The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems (米、IEEEコントラクタ、社会運動活動家)
- Greg Adamson, President, IEEE Society on Social Implications of Technology; オーストラリア、メルボルン大、Electrical and Electronic Engineering
- Ronald C. Arkin, Regents' Professor, Associate Dean for Research in the College of Computing at Georgia Tech (米)
- Stephen L. Diamond, Global Standards Officer and General Manager, Industry Standards Office, EMC Corporation (米)
- Virginia Dignum, Associate Professor, Faculty of Technology Policy and Management, TU Delft (独)
- Philip Hall, Member, IEEE-USA Government Relations Council (オーストラリア/米、メルボルン大 Electrical & Electronic Engineering, RelmaTech CEO (英), ミシガン在住)
- Eva Schulz-Kamm, Political Affairs & Public Co-Creation, NXP (米)
- Raj Madhavan, Founder & CEO, HumRobTech, LLC, USA & Distinguished Visiting Professor of Robotics, Amrita University, India (印)
- Richard Mallah, Director of Artificial Intelligence Projects Future of Life Institute (米)
- AJung Moon, Co-founder Open Roboethics initiative (カナダ、ブリティッシュコロンビア大Ph.D candidate)
- Monique Morrow, CTO New Frontiers Engineering at Cisco (米)
- Francesca Rossi, Full Professor of Computer Science University of Padova (伊)
- Yu Yuan, IEEE Digital Senses Initiative議長 (中/米)

以上15名。米国が多いが所長はフランスで、それ以外も英語圏を中心に多数

IEEE イニシアチブ：委員会

Executive Committee (執行委員会), 事務局的な4委員会: Drafting, Standards, Ecosystem Mapping, 'Lexonomy' に加え12委員会が活動

General Principles (一般原則)

Embedding Values into Autonomous Intelligent Systems (AI/ASへの価値組み込み)

Methodologies to Guide Ethical Research and Design (倫理的研究と設計を導く方法論)

Safety and Beneficence of Artificial General Intelligence (AGI) and Artificial Superintelligence (ASI) (汎用AIと超知性の安全性と受益者)

Personal Data and Individual Access Control (個人情報とアクセス制御)

Reframing Autonomous Weapons Systems (自律兵器の再定義)

Economics/Humanitarian Issues (経済および人道的事項)

Law (法)

Classical Ethics in Information and Communication Technologies (ICTにおける古典的倫理)

Mixed Reality (複合現実)

Affective Computing (感情を読み表す計算機)

Effective Policymaking for Innovative Communities Involving Artificial Intelligence and Autonomous Systems (EpicAI) (創造的な社会のための効果的政策立案)

IEEE イニシアチブの成果 (1): EAD

■ 名称 : Ethically Aligned Design: A Vision for Prioritizing Human Wellbeing with Artificial Intelligence and Autonomous Systems (EAD v1, 倫理的整合のとれた設計: AIとASにおいて人の幸福を優先するためのビジョン第1版)

■ 第1版 (for public discussion版, EADv1)
2016年12月発行; 全136ページ

■ 第2版: 2017年秋

このために下記の活動を予定

- コメント募集: ~3月6日
- コメント募集結果発表: 4月10日
- イニシアチブ全体会合での討議: 6月5, 6日

■ 内容 : 各委員会 (の分野) 毎に論点を数点挙げ、その論点の背景、勧告案、参考文献を記述

- (例): 自律システムへの価値組み込み委員会

論点例 : 組み込むべき価値は往々にして特定のユーザーコミュニティや課題に依存

勧告案 : 万人向けの規範や価値のセット構築は非現実的と認め、特定のコミュニティや課題向けの規範セットを特定すべき

- 立上げが遅かった4委員会は第1版では論点整理まで。勧告案は6月全体会合で発表

IEEE イニシアチブの成果 (1-1):一般原則

委員会名称: General Principles (一般原則)

共同議長: Kay Firth-Butterfield, U-Texas
Alan Winfield, U-West of England

下記の論点を議論し、勧告案を作成

人類の利益: AI/ASが人権を侵害しないことを保証する方法

勧告案: 標準化や規制当局を含むガバナンスフレームワークの設立、既存の
・将来の法的枠組みを政策や技術の討議へと翻訳する方法論の確立

責任: AI/ASに結果責任を負わせる方法

勧告案: 立法と司法による製造過程や利用現場でのAIの責任の明確化、システムの設計者や開発者が利用者の文化的規範の多様性を認識する必要性、規範を創造するマルチステークホルダーのエコシステム設立、登録システムの整備 (利用目的、訓練データ (可能ならば)、データ源、アルゴリズム等々)

透明性: AIやASの透明性を確保する方法

勧告案: システムの透明性を測定し、試験し、適合性を客観的に評価できる標準の開発。

教育と注意喚起: AIやASの便益を最大化し、誤用のリスクを最小化する方法

勧告案: 誤用について人々に注意喚起。そのため人々に倫理や安全の教育をSNS等新しい方法で行い、また法の執行機関を教育し市民と協力可能に

IEEE イニシアチブの成果 (1-2):AI/ASへの価値組み込み

■ 委員会名称: Embedding Values into Autonomous Intelligent Systems (AI/ASへの価値組み込み)

■ 共同議長: AJung Moon: Open Roboethics initiative, and Ph.D candidate, U-British Columbia

Francesca Rossi: U-Padova, currently at the IBM Research

下記の論点を議論し、勧告案を作成

■ AIシステムのための規範や価値の特定

- 組み込むべき価値は往々にして特定のユーザーコミュニティや課題に依存
- 要求過多。様々な相矛盾しうる規範や価値に従うことが求められる
- 特定グループに不利なアルゴリズム的バイアスを持つ可能性

■ AIシステムへの規範や価値の組み込み

- 規範セットが特定されても、実装方法は決して明らかではない

■ 人とAIシステムとの間で規範や価値の整合が取れていることを評価し確認

- 組み込まれた規範は人間コミュニティの規範と両立する必要がある
- AIに対する適切なレベルの信頼の確立
- 整合性の第三者評価

AI/ASへの価値組み込み委員会の勧告案（1）

AIシステムのための規範や価値の特定

- 論点：組み込むべき価値は往々にして特定のユーザーコミュニティや課題に依存
勧告案：**万人向けの規範や価値のセット構築は非現実的と認め**、特定のコミュニティや課題向けの規範セットを特定すべき
- 論点：要求過多。様々な相矛盾しうる規範や価値に従うことが求められる
勧告案：より大きなコミュニティの価値や規範を反映するよう、優先順位付けを行う。このためにCommon Good Principle(*)は有用である。設計段階での優先順位は明確化が必要。利用時には優先順位を変更する必要がある場合も
- 論点：特定グループに不利なアルゴリズム的バイアスを持つ可能性
勧告案：バイアスは学習データの片寄り等により簡単に意図せずして起きると認識するのが第一歩。良い設計指針はその意味で有用。また本分野では様々な研究が進行中であり、そうした研究を行い教育に反映させる必要あり。特定の弱者の集団をユーザーとするAIは多く、そうしたAIの開発時には関連する利害関係者、関係する専門家、規制当局、政策立案者を開発の初期から巻き込むことが望ましい

AIシステムへの規範や価値の組み込み

- 論点：規範セットが特定されても、実装方法は決して明らかではない
勧告案：研究を加速すべき。こうした研究は、AIシステムの安全性や信頼性向上にもつながる

* 初出1992年の倫理学の論文

AI/ASへの価値組み込み委員会の勧告案 (2)

■ 人とAIシステムとの間で規範や価値の整合が取れていることを評価し確認

- 論点：組み込まれた規範は人間コミュニティの規範と両立する必要がある
勧告案：システムが完成したら厳しくチェックすべき
- 論点：AIに対する適切なレベルの信頼の確立
勧告案：**透明性と検証可能性**がカギ。システムは透明性や検証可能性を示す機能を持つことが望ましいが、適切なレベルを達成する方法は研究途上であるし、**人は自分の能力を超えるシステムを過度に信頼する傾向**がある。従って**単なる注意書きでは不十分**であり、想定外の使い方をされないようにすること、「ユーザはシステム運用時に自分の役割をどう考えるだろうか」の認識を設計時に含める事、が重要。対話的に「出来ません」「自信がありません」とシステムが表明することも有用
- 論点：整合性の第三者評価
勧告案：規制当局などの評価者が満足する透明性を達成するため、変更は記録を取り、場合によっては航空機のブラックボックス的なシステムを導入すべき。これは完全ではないが診断に有用となる。また2018年施行のEU一般データ保護規則 (GDPR) は個人情報に基づき自動的に決定する時には、個人は後で理由説明を求め異議を申し立てることができる¹と定めている。理由説明は未解決の研究テーマであるが、機械学習より論理ベースAIの方が透明性が高い等の特徴はあり、シミュレータ等も有用。また規制当局は必要な透明性を定義すべき。信頼できるAI/ASシステムの評価基準の定義もすべき

委員会名称: Methodologies to Guide Ethical Research and Design (倫理にかなう研究や設計を行う方法論)

共同議長: Raja Chatila (Co-Chair) – CNRS
Sarah Spiekermann (Co-Chair) – Vienna-U
Corinne J.N. Cath – (Co-Chair) PhD student, U-Oxford

下記の論点を議論し、勧告案を作成

学際的な教育と研究

- 倫理が学位取得過程に含まれていない
- 学際的で異文化にまたがる教育モデルの必要性

個々の文化に特有の価値に従った差異化

- ビジネスの実務とAI
- 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
- 倫理上の懸念を表明する力の欠如 (従業員等が問題に気付いても声を上げられない)
- 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

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

■ 委員会名称: Safety and Beneficence of Artificial General Intelligence (AGI) and Artificial Superintelligence (ASI) (汎用AIと超知性の安全性と受益者)

■ 共同議長: Malo Bourgon, Machine Intelligence Research Institute
Richard Mallah, Cambridge Semantics / Director of AI Projects,
Future of Life Institute

下記の論点を議論し、勧告案を作成

■ 技術

- AIシステムの能力が高まり、より複雑なシステムの最適化が可能となり、自律性が高まった時、予期せぬ振る舞いの危険性も高まる
- 昔ながらの安全性の概念を、将来の、全くアーキテクチャの異なる汎用AIに適用するのは困難かもしれない

■ 一般原則

- AIの能力や自律性が高まるにつれ、研究者や技術者はより複雑な倫理的、技術的な安全上の課題に直面する
- 将来のAIは農業革命や産業革命に匹敵する衝撃を世界に与えるかもしれない

■ 委員会名称: Personal Data and Individual Access Control (個人情報とアクセス制御)

■ 共同議長: Michelle Dennedy, Vice President, Cisco
John C. Havens, Executive Director, 本The IEEE Global Initiative

下記の論点を議論し、勧告案を作成

■ 個人データの保護

- アルゴリズムの時代において個人は「個人データ」をどのように定義し組織化すべきか
- 個人情報 (PII, Personal Identifiable Information) の定義と範囲
- 個人情報の「制御」の定義

■ 個人情報へのアクセスと同意

- 個人を尊重した「データへのアクセス」の再定義
- 個人を尊重した「同意」の再定義方法
- 些末な情報から共有したくない推論が行われる
- データ取扱い者が、消費者から本当に同意を取り付けたと確信する方法

■ 個人情報管理

- 人が個人AIやアルゴリズム的守護者を持つことは可能か

IEEE イニシアチブの成果 (1-6):自律兵器の再定義

委員会名称: Reframing Autonomous Weapons Systems (自律兵器の再定義)

議長: Richard Mallah, Future of Life Institute

下記の論点を議論し、勧告案を作成

プロフェッショナル達の倫理規定には往々にして重要な抜け穴 (例えば、人道問題になり得る兵器を開発し兵士に与えたりする)

AI/AS, 自律兵器 (AWS, Autonomous Weapon Systems) の概念定義には混乱があり、議論をはざまに落とし込んでいく

自律兵器はその特性上、密やかかつ匿名の利用を招く

自律兵器の結果責任をうやむやにする方法が色々ある

自律兵器のふるまいは予測できない可能性がある。学習はこの問題を大きくする

自律兵器開発の合法化は中期的に地政学的危険性を招く先例となる (自律兵器同士で発砲し合い意図しない紛争が起きたり、自律兵器に頼って構築した戦略バランスは、ソフトの進化により一夜にして崩れたりする恐れがある)

人が監督しなければ、余りにも簡単にはずみで人権侵害が起き、緊張が高まる

自律兵器の直接的、間接的顧客は多様性に富んでおり、兵器拡散や誤用の問題を複雑で問題含みにする

自律兵器は紛争を急速に拡大する (人より反応が速いため、自律兵器同士で対峙すると人よりも急速に紛争が拡大)

自律兵器の設計保証を検証する標準の欠如

自律兵器と、準自律兵器の倫理上の境界の理解は混乱しがち

委員会名称: Economics/Humanitarian Issues (経済および人道的事項)

議長: Raj Madhavan, Humanitarian Robotics Technologies

下記の論点を議論し、勧告案を作成

自動化と雇用

- AIやASはメディアで間違った理解をされており、人々を混乱させている
- 自動化は市場に閉じた問題に過ぎないと思われる (公衆衛生を始め市場外も色々ある)
- 雇用問題の複雑さが無視されている
- 技術の進歩が速すぎ、伝統的な再教育方法では追いつかない
- AIの政策がイノベーションを遅める可能性がある

説明責任と平等な利用

- AIやASの技術が利用できる場所は世界的にみて偏っている
- 個人情報へのアクセスや理解の欠如

AIの利益を途上国に

- 本IEEEイニシアチブへの途上国の参加増
- AI/ASが途上国と先進国の間、先進国の中、途上国の中での経済や権力構造の格差を拡大する可能性

IEEE イニシアチブの成果 (1-8):法

■ 委員会名称: Law (法)

■ 共同議長: Kay Firth-Butterfield, AI Austin / U-Texas
Derek Jinks, U-Texas Law School

下記の論点を議論し、勧告案を作成

■ AI/ASの説明責任と検証可能性を改善する方法

■ AIに透明性があり人権を尊重することを保証する方法（法が政府に求める透明性、関与、正確性の原則を国際機関から地方自治体までで実現する方法）

■ AIシステムで死傷者が出た時、法的説明責任を保証できる設計

■ 個人情報の完全性を尊重するAI/ASシステムの設計と展開

IEEE イニシアチブの成果 (1-9):ICTにおける古典的倫理

■ 委員会名称: Classical Ethics in Information and Communication Technologies (ICTにおける古典的倫理)

■ 議長: ---

下記テーマについて議論。結果は6月公表予定

■ 機能、目的、アイデンティティ、代理

■ スタートアップと大企業のための倫理上の課題設定

■ 議論から行動へ

■ 古典的倫理学の方法論を通じた責任と結果責任

■ 自律システム設計における文化バイアスの洗い出し

■ IT, ICT, AIに対する擬人化されたアプローチ

■ 倫理用語、それを通じた価値の深掘り

IEEE イニシアチブの成果 (1-10): 複合現実

■ 委員会名称: Mixed Reality (複合現実)

■ 共同議長: Monique Morrow, Cisco
Jay Chair, IEEE SA

下記の論点を議論。勧告案は6月公表予定

■ (AIで強化された)複合現実の来たる世界でセレンディピティを掴み取るには

■ 複合現実における肉体と精神の結合の在り方、治験への応用とリスク

■ ソフトウェアへ感覚を明け渡すことの意味。感覚を複合現実明け渡すとは精神を明け渡すことなのか?

■ 幻想の世界の中で地理的制約や部族的関係が薄れ、アイデンティティが突然現れ消えゆくとき文化的共同体の未来は

■ 適切な複合現実の提供には物理状態を継続的に観測する必要があるが、これは悪魔とのダンスではないのか

IEEE イニシアチブの成果 (1-11):感情を読み表す計算機

■ 委員会名称: Affective Computing (感情を読み表す計算機)

■ 共同議長: Ronald Arkin, Georgia Tech
Joanna Bryson, U-Bath

下記の論点を議論。勧告案は6月公表予定

■ 人間の可能性を支えるシステム; 人間の自立が損なわれる他の問題への対処

■ システムが嘘をついたら

■ いちゃいちゃするシステムが出来たら

■ システムが文化的タブーを踏み越えたら

■ システムが自分の感情を持ったら

■ システムが感情をかきたて人を操ったら

IEEE イニシアチブの成果 (1-12):効果的政策立案

■ 委員会名称: Effective Policymaking for Innovative Communities Involving Artificial Intelligence and Autonomous Systems (EpicAI)
(創造的な社会のための効果的政策立案)

■ 共同議長: David Bray, FCC
Michael Kringsman, industry analyst

下記の論点を議論。勧告案は6月公表予定

■ 公的セクタ (政府、NPO, PPP、公的セクタの人々) が公共サービスを向上させるため、より速やかにAIや自律システムを受け入れる方法

■ 公的セクタ、民間セクタが協力し過度な制約なしにAIや自律システムの革新的な利用法を探求するため、最善な勧告を作る方法

IEEE イニシアチブの成果 (2): P7000シリーズ標準

EADの議論の中で特定された標準化項目は順次標準化プロジェクトに。現在4プロジェクトが開始済、3個が提案中。全て**標準 (not ガイドライン)**

■ P7000: 倫理的設計のモデルプロセス

システム開発時に倫理上の懸念を扱う手順のモデルを規定

■ P7001: 自律システムの透明性

透明性のレベルを測定・試験可能に。「本製品の透明性はレベル3」

■ P7002: データプライバシーのプロセス

個人情報を用いる製品やサービスでのシステム/ソフトウェアエンジニアリングプロセスの要求条件を規定

■ P7003: アルゴリズムミックバイアス

アルゴリズム開発での「負のバイアス」（人種差別や性差別など）を特定し除去する方法論を規定

以下は現在、上位委員会に提案し承認待ちのプロジェクト

■ (P7004: 子供と学生のデータのガバナンス)

■ (P7005: 透明性ある従業員データのガバナンス)

■ (P7006: パーソナルデータAIエージェント)

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)
- 開始: 2016/09, 第1回スポンサー投票(**)目標: 2018/07, 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、上位委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

P7001:自律システムの透明性

委員会名称: Transparency of Autonomous Systems (自律システムの透明性)

スポンサー(*): Vehicular Technology SocietyのIntelligent Transportation Systems (VT/ITS) とRobotics and Automation SocietyのStanding Committee for Standards (RAS/SC) が共同

議長: Alan Winfield (西イングランド大)

開始: 2016/12, 第1回スポンサー投票(**)目標: 2018/01, 投票: 個人

透明性のレベルを測定し、試験可能に

- 設計者が標準への適合レベルを評価可能に。「本製品の透明性はレベル3」等
- 透明性向上のための示唆提供。例: センサや内部状態は安全なストレージに保存

背景: 社会にAI/ASが受容されるために透明性が必要に

- ロボットの動作の理由を理解可能とし、予期せぬ動作は問いただし可能にする必要
- 事故発生時の原因究明や証拠提出の必要性

関連する標準: 存在せず

* 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、上位委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

P7002:データプライバシーのプロセス

委員会名称: Data Privacy Process (データプライバシーのプロセス)

スポンサー(*): Computer Society, Software & Systems Engineering Standards Committee (C/S2ESC)

議長: Michelle Dennedy (cisco)

開始: 2016/12, 第1回スポンサー投票(**)目標: 2018/01, 投票: 個人

従業員や顧客、その他の外部の関係者の個人情報を用いる製品やサービスでの、プライバシー上の懸念に配慮したシステム/ソフトウェアエンジニアリングプロセスの要求条件を規定

●関係者がプライバシー上のチェック、適切に行われている事の確認が可能に

対象ライフサイクル: 開発から品質保証、価値実現まで

関連する標準: ISO/IEC JTC 1/SC40/WG1の38500シリーズ、特に38505-1, The application of ISO/IEC 38500 to the governance of data, ISO/IEC 29100, Privacy framework

利用者: 個人情報を含むシステムの関係者

* 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、上位委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

P7003:アルゴリズムミックバイアス

- 委員会名称: Algorithmic Bias (アルゴリズムミックバイアス)
- スポンサー(*): Computer Society, Software & Systems Engineering Standards Committee (C/S2ESC)
- 議長: Ansgar Koene (ノッティンガム大)
- 開始: 2017/02, 第1回スポンサー投票(**)目標: 2018/07, 投票: 個人
- アルゴリズムの開発において「負のバイアス」(人種や性別など法的に禁じられている差別、法的ではない差別を共に想定)を特定し除去する方法論を規定
- バイアスを制御するための検証用データセットの選択手順と基準、アルゴリズムの利用限界を規定し、伝え、対象外の利用を防ぐガイドライン、他を記述することを想定
- 将来的に本標準に基づく認証を想定「本サービスは人種や性別に対する差別がないことをP7003に基づき認証します」
- 関連する標準:存在せず

* 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、上位委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

提案中・承認待ちの標準化プロジェクト

NesCom (新規標準化プロジェクトを審議する委員会) 3月会合に提案中のプロジェクト; ボード承認見込: 4月

P7004: Child and Student Data Governance (子供と学生のデータのガバナンス)

- スポンサー (*): Computer Society/Learning Technology (C/LT)
- 第1回スポンサー投票(**)目標: 2019/02, 投票: 個人

P7005: Transparent Employer Data Governance (透明性ある従業員データのガバナンス)

- スポンサー (*): Computer Society/Software & Systems Engineering Standards Committee (C/S2ESC)
- 第1回スポンサー投票(**)目標: 2017/12, 投票: 個人

P7006: Personal Data AI Agent (パーソナルデータAIエージェント)

- スポンサー (*): Computer Society/Software & Systems Engineering Standards Committee (C/S2ESC)
- 第1回スポンサー投票(**)目標: 2017/12, 投票: 個人

* 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、上位委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

進行中・完成済AI関連標準化(3)
英国国家標準 BS8611:2016

■ 名称: Guide to the ethical design and application of robots and robotic systems' (ロボットおよびロボットのデバイス(*)の倫理的設計と応用のガイド)

■ British Standard。ロボット≒AIと見なせば、世界初のAI倫理に関する**国家標準**。ただし**ガイドライン**であり強制力はない

■ アカデミーが中心に作成と聞く

■ 安全設計の標準の考え方をベースに、下記を記述

- 倫理上の危機とは何であり、それを特定する方法
- 倫理上の指針と対応策
- 守るべき設計指針
- 検証と確認
- 参考情報

* ロボット: 二つ以上の軸についてプログラムによって動作し、ある程度の自律性をもち、環境内で動作して所期の作業を実行する運動機構。ロボティックデバイス: 産業用ロボット又はサービスロボットの特徴を満たすが、プログラムできる軸数又は自律性の程度に不足のある運動機構 (ISO 8373/JIS B 0134:2015)

BS8611:倫理上の危機と特定方法

倫理上の危機: 社会的/分野別/商業・金融上/環境上に分けられる

起き得る危機: 信頼の欠如、プライバシー、過度の依存等々

結果として起きるリスク: 使われなくなり、データが流出し、人々の能力や意欲が低下等々

リスクの緩和方法: 高信頼な設計、データの制御、依存の明確化等々

検証・確認方法 (ユーザ/専門家 (の助言) /ソフトウェア的的手法等々)

上記をリスクマネジメントの手法を使い起き得る危機を特定

- 不正な利用や操作ミス、状況の不確実性、心理的な影響等々に注意してリスク評価すべき
- 参考: ISO31000: リスクマネジメントの原理とガイドライン, ISO 12100: 機器の安全のための設計の一般的原理

学習するロボット (製造時には必ずしも動作が確定しない)

- 学習の程度を「環境 (に合わせて適切に動作)」「パフォーマンス向上」「戦略的 (作業の必要性を判断し必要なら変更等々)」の3段階に分け、ロボットが変化したら容易に分かるようにすべき

全般的な指針

- 責任を負うのは人であってロボットではない
- プライバシー・バイ・デザイン
- ロボット関係者は社会への技術の周知に努める
- 特に下記を規定
 - プライバシーと秘密
 - 人間の尊厳と人権（見られ・聞かれたくない場合のためスイッチ・オフを可能に等）
 - 文化的多様性の尊重
 - 人間性の損失への配慮（製造ラインをあまりに高速で動かすべきではない等）
 - 法律、利益とリスクのバランス
 - 個人と組織の責任
 - 社会的責任
 - インフォームド・コンセント（医療分野のロボットは無論、それ以外でもリスクや便益、限界などを納得した上での同意が必要）
 - インフォームド・コマンド（人間の命令が機器内部の処理手順に適切に翻訳される必要があり、もし出来ないなら停止すべき）
 - ロボット中毒（つついロボットを使ってしまうようになる事を悪用してはならない）
 - ロボットへの依存（過度に依存するようになってはならない）
 - ロボットの擬人化（子供など脆弱なユーザが使うロボットを擬人化させる場合、ロボットの使い方や振る舞いについての説明文書を添付する）

分野毎の設計指針：リハビリ、医療、軍事について規定

軍事：人の責任はロボットを使おうとも免除されるべきではない

守るべき設計指針

- 安全設計の概念を援用して倫理的な設計を行うべき
本質倫理設計 (立体交差にすれば交通事故は起きない) が望ましいが、不可能だったり実
際的でない時は信号機相当の安全対策ほかを設けるべき

検証と確認

- 安全設計と異なり「倫理上のスペック」は必ずしも得られると限らない
- 考えられる対策
他分野で使われる手法：冗長性を持たせる、独立した安全のためのシステムを設ける、ソ
フトウェアならば形式手法を使う、確率的な検証を行う
これらを使いユーザに依る検証、ソフトウェアによる検証、専門家のガイダンス、経済的
・社会的アセスメント、法的アセスメント、コンプライアンステスト等を行うべき
- 自律システムは予期せぬ振る舞いをする場合あり
対策：システムを分割し検証を容易に、倫理上の監視を行う独立したシステムを組込

進行中・完成済AI関連標準化 (4)
(ご参考：ITU-T Y.3013)

ITU-T Y.3013: 利害衝突を分析する社会経済的アセスメント

■ 網技術の設計・標準化時の関係者の利害衝突を分析・調停する手法を規定

■ 利害衝突を3ステップで特定、調停できるまで繰り返す

1. 利害関係者の特定

机上検討、インタビュー、ロールプレイ等を駆使

2. 利害衝突の特定

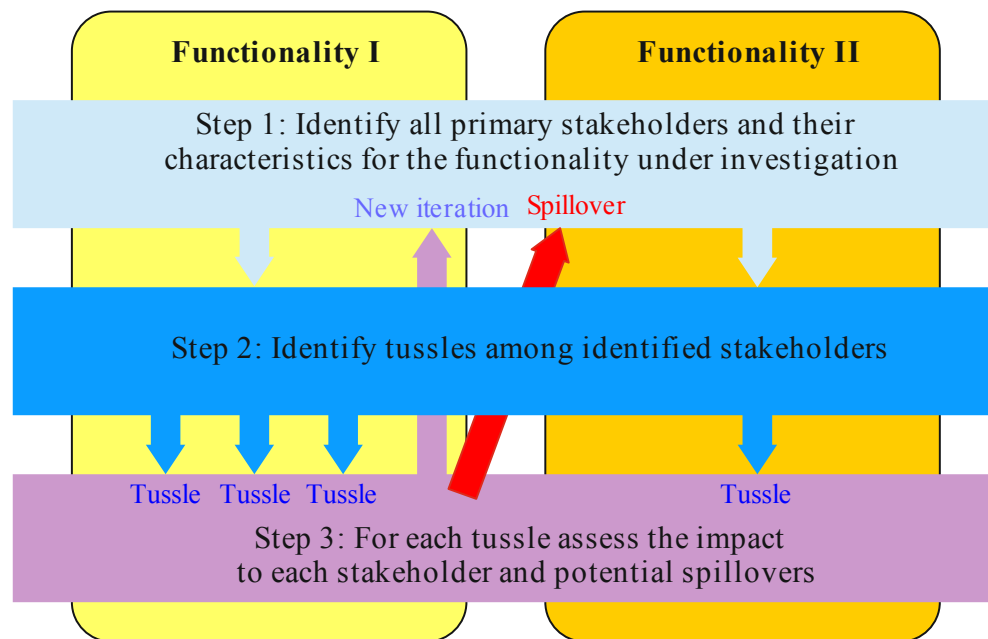
ロールプレイ、机上検討、リスク管理、MACTOR法、SWOT分析、インタビュー等を駆使

3. 利害衝突の止揚

個々の利害、個々の関係者毎にインパクトを分析し、衝突回避策を検討。ゲーム理論他を駆使

■ 欧州SESERV (Socio-economic Services) プロジェクト成果の標準化

- 勧告完成後、ITU理事会に対し今後の全ITU標準に本プロセスを義務付ける提案、却下



Y.3013(14)_F01

有意義な標準化のために

有意義な標準を作るために

目的の明確化：世界平和のためか、ビジネスのためか

- 国際単位系(メートルや1秒の定義) は今は世界平和のための標準
(かつては違った。イギリスとフランスの覇権争いの場。今もうるう秒の議論に残る)
- そもそも標準は必要か

ビジネスのための標準ならば、ビジネスモデルの明確化

- 標準化する領域、させない領域の明確化(オープン・クローズ戦略)が第一歩
- それを基に、関連する知的財産権の整理も必要
- 「日本発」「弊社がリード」だけでは金にならない

タイミング

- 標準化とはイノベーションを打ち切り「ここまでは終わった話」とする作業。十分な事例蓄積前に行うと弊害が大きい

仲間づくり

- 顧客は、標準を実際に使う人は巻き込めているか。敵は、味方は誰か

組織体制：特に戦略レベルの人材と意思決定は重要

- 組織全体レベルの意思決定者、日常レベルの戦略決定者、現場で戦うチーム、の3段階があることが理想

標準は、自分の想いを実現する手段の一つ。手段にはコストが伴う
無闇に作ればよいものではない

標準化はイノベーションを止めます

- 標準化とは「ここまでは終わった話」にするための活動。イノベーションをしたい向きは、標準を土台に、次のステップで力を発揮してください、が標準化

標準は利益率を下げます

- 参入障壁が下がり利益率低下。それを補うほどの市場の拡大がないと厳しい
DVD等では一般に標準化で市場が50-100倍になると言われる
理想は、自社に繋がる部分は標準(オープン)化してコスト削減し、コア技術は標準化せず(クローズ)に保つことだが、現実には決して容易ではない

標準化活動それ自体はコストです

- 特許ならば特許使用料が取れるが、標準規格では金はまず取れない

苦勞して作った標準でも使われないことが大半です

- 標準を作ること自体は難しくない。だが顧客が調達仕様に指定する標準、は難しい

標準化は守りの活動が大半です

- 人脈が重要なため、継続参加が重要。公的国際標準化機関の場合、提案の9割は日本以外から。攻められてあたふた、がどうしても多くなる

「日本発の国際標準」への固執は弊害が多い

■ 一時期「日本発の国際標準」が持て囃された。今も時々マスコミに登場
仕組みを作る者が最もおいしい思いをする、との発想
これ自体は正しい

■ だが、これ自体を目標とすると話がゆがむ

- 日本の国内市場は、無視がどうしても不可能なほど巨大ではない
「日本はこう決めた。者ども、従え」と言える国力がない。米国、中国にはある
→ 日本、を言いすぎると諸外国の反発を招き、使われない標準へ一直線
- 民間において「日本」の意味は必ずしも明確ではない
国内のライバルのために自社の海外部門を切り捨てる、はあり得ない
- 日本国内にも様々な利害関係者がいる
B2B2Cビジネスにおいて、最初と2番目のBの、BとCの利害は異なる
- 話が変な方向へ行きがち
考え抜かれたビジネスモデルなしでの標準化はビジネスに繋がらない
要のポストを取ることは重要だが、役職者や提案の数で成果を図ると話が歪む

大いに妥協せよ。他人を巻き込め。そして「これは俺が作った標準」と
思って貰え。それが、使われる標準への道だ（小職の昔の上司）

世界に通用する製品やサービスを作る、それが日本発標準への近道だ。
製品の競争力のなさを標準で補うことは出来ない（内海元ITU事務総局長）

標準化は「枝切り」作業

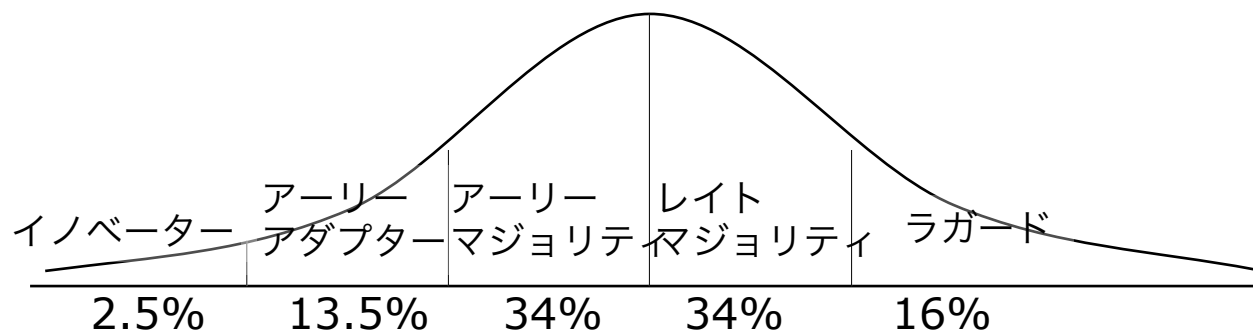
- イノベーションを打ち切り「ここまでは終わった話」「イノベーションは他で行いましょう」が標準化

アーリーアダプターに普及し終わる直前の標準化が理想

- 実際の事例がある程度蓄積
- 仕様の混乱がマスには及ばない

最近はやまる傾向

- 「競争の場の定義」を標準化で行う
用語、概念、守るべきプロセスなどを定め、安心して競争に参加できるようにする
- 標準化には早くとも数か月、普通は年単位の時間がかかる。早めに着手
- 標準化機関間の競争



E. ロジャースのイノベーター論

標準化の体制

一般的な標準化

組織全体レベルの
意思決定者

日常レベルの
戦略決定者

議長

現場で戦うチーム

標準化活動を、組織としての
戦略の中に位置づけ。

「xxと組み、xxと戦い、xx達成が
君たちのミッションである」

標準化活動の具体的内容を概要レ
ベルで把握、後方から現場に指示。
更には外部人材を使い標準化・オ
ープンイノベーションを遂行

日々の現場で他組織の人材と協力
・対立しつつ活動

オープンソース
ソフトウェア

会社幹部

External R&D
manager

Community
manager

Programmers

小職は、現場で戦うことは出来ます。でも方針と問い合わせ先が必要です。
ご支援を

 **Orchestrating** a brighter world

NEC