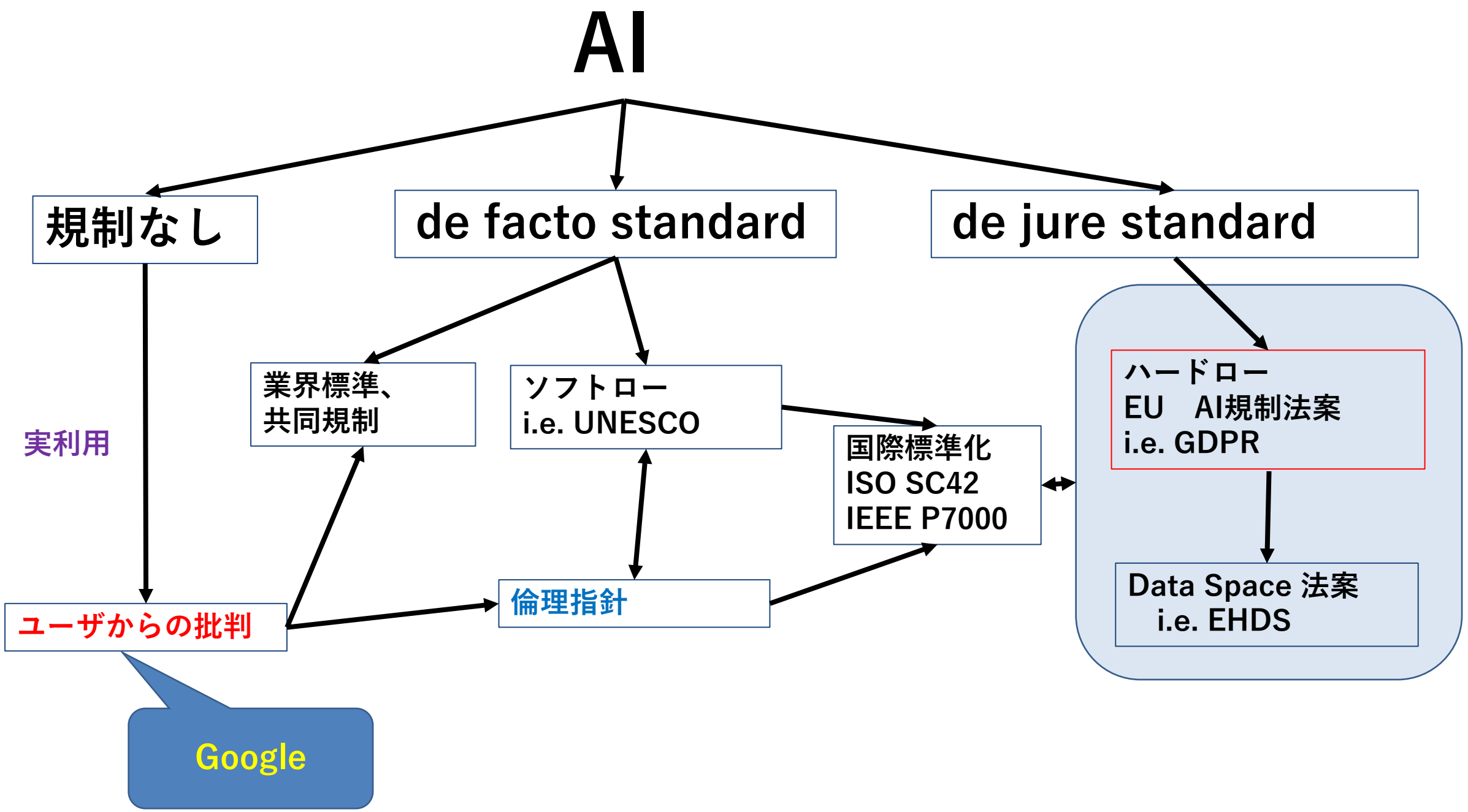
AIのトラストに係るEUの法制度

中川裕志
理化学研究所・革新知能統合研究センター


AI
規制なし
de facto standard
de jure standard
業界標準、
共同規制
ソフトロー
i.e. UNESCO
国際標準化
ISO SC42
IEEE P7000
ハードロー
EU　AI規制法案
i.e. GDPR
実利用
Data Space 法案
i.e. EHDS
ユーザからの批判
倫理指針
Google

日本の状況

- ソフトロー
  - 総務省　AI開発ガイドライン(2017)、AI利活用ガイドライン(2019~)
  - 内閣府　人間中心AI社会原則(2019)
  - 各企業のAI倫理指針

➢各企業は、EUのハードローの状況を伺い、それに対応する（後追い？）方策
  ➢EUは米国は技術及び市場支配力でハードローで対抗
  ➢米国内でもハードローを好む傾向あり。

➢日本でもハードローを作るべきか　→ No

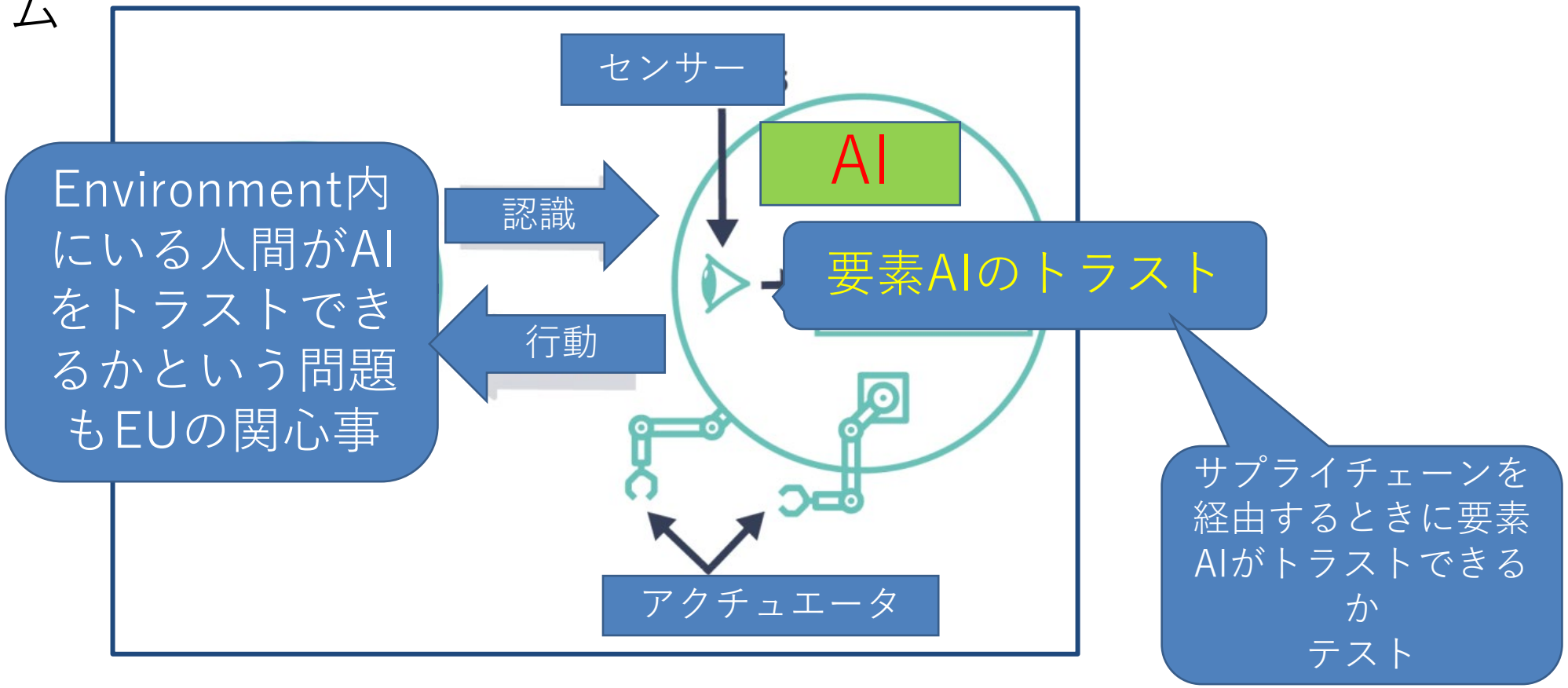
EUにおけるAIトラストの系譜

1. European Commission　HLEG : High-Level Expert Group on Artificial Intelligence
   Ethics Guidelines for Trustworthy AI (2018)

2. **EU AI white paper (2020)**

3. **EU AI regulation acts (2021)**

- AIの定義
- AIとは、特定の目標を達成するために、環境を分析し、ある程度の自律性を持って行動を起こすことにより、知的な行動を行うシステム



Trustworthy AI

- Training Dataに恣意的なバイアスが入り込んでいない

- 倫理性
  - 基本権：尊厳、自由、平等と連帯、市民の権利と公正
  - 公益を最大化しながら、個人の権利と自由を保護する
  - 法令遵守

- 人間中心　＝　常に人間が上位の決定権者

- Trustworthy AI
  ＝　倫理性　＋　人間中心
  ＋　技術的なトラスト（工学的ツールとしての信頼性が高いという意味）

Trustworthy AI の条件

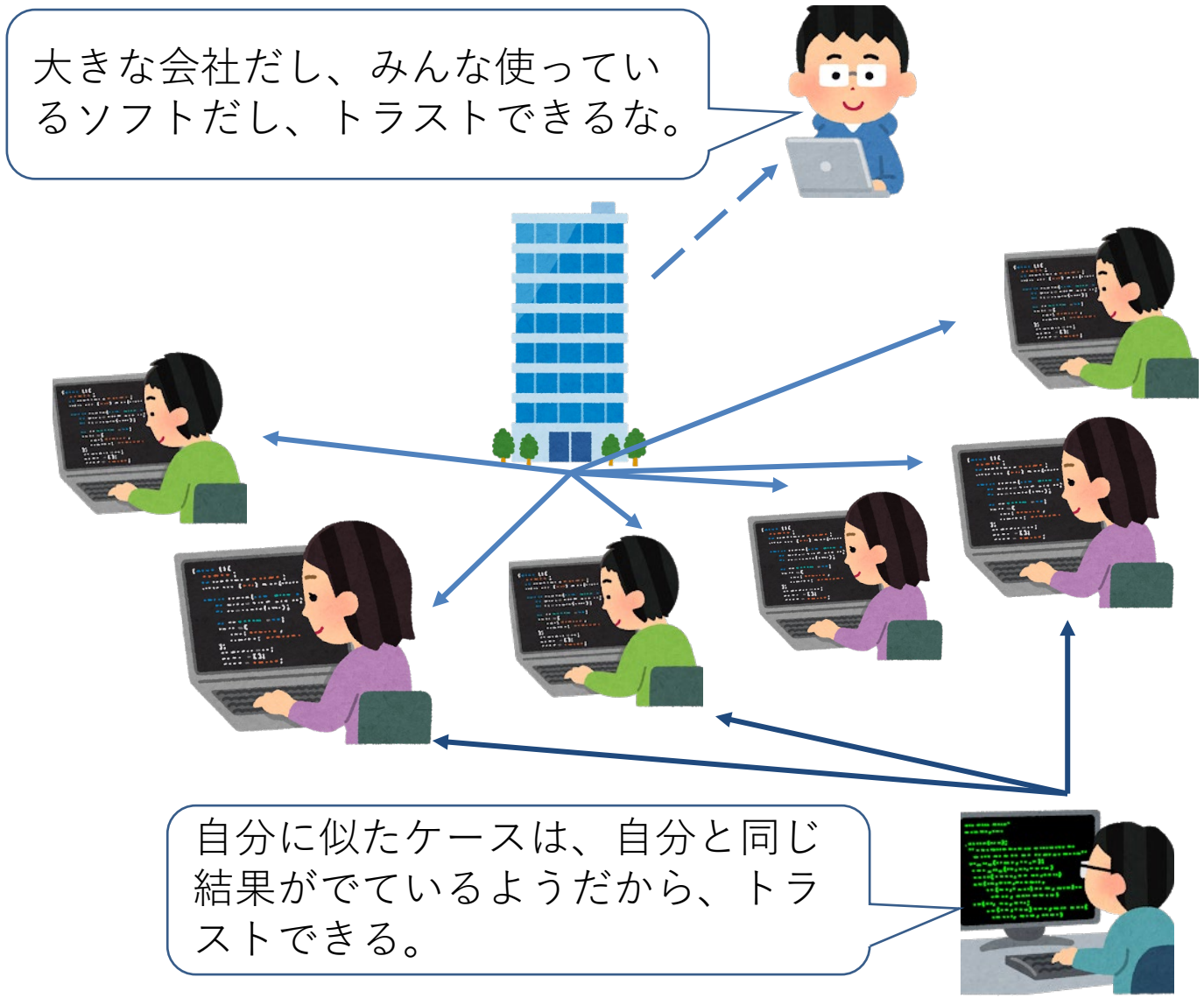
この１．～１０．がEUが考えるAIのトラスト．

ようするにEUは自律的なAIを信じていない

EUは人間が制御できるように限定したAIにしたいという意図

- 1. アカウンタビリティ
- 2. データガバナンス　（バイアス除去）
- 3. 誰でも使えるAI
- 4. 自律的に動くAIガバナンスする
  - ➢ 人間が監視，早期介入できる
- 5. 差別しない
- 6. 人間の自律性を尊重し促進
- 7. プライバシーの尊重
- 8. ロバスト
  - 信頼性と再生成性，正確さ，攻撃への耐久性と立ち直り，止めるときの計画を予め立てておく．
    - ➢ 普通のソフトだと思い込みたがっている
- 9. 安全性
- 10. 透明性

トラストの実態


大きな会社だし、みんな使っているソフトだし、トラストできるな。
自分に似たケースは、自分と同じ結果がでているようだから、トラストできる。

EU的トラストを実現する技術

- Technical methods
  - アーキテクチャ
  - X-by-design --- Privacy-by-design , Security-by-design, etc.
  - テストと検証（実用の供する前）
  - トレーサビリティと監査可能性（実用に供した後）
  - XAI

- Non Technical methods
  - 法的規則
  - 標準化
  - アカウンタビリティを確保するガバナンス
  - codes of conduct（行動規範）
  - 倫理的な考え方を育むための意識化と教育
  - ステークホルダー間の対話, 社会との対話
  - 多様性確保と「だれかを取り残さない」ような設計を行うチーム

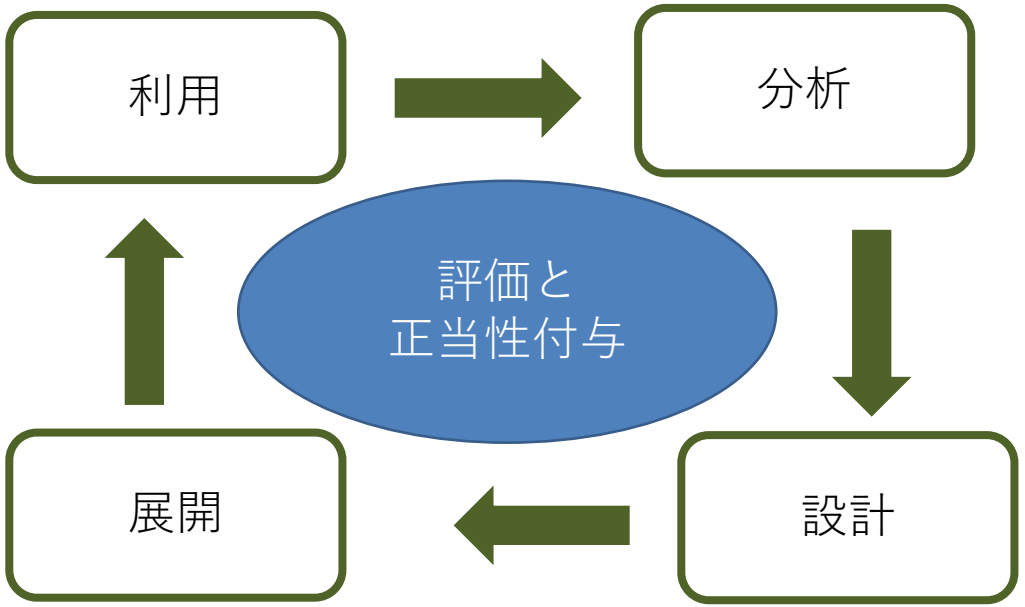
EUはこの設計から実装の流れでAIのトラストを実現するつもりだが，
アジャイルな考え方が不足気味

権利，
原則，
価値

Trustworthy AI のための要求

左の要求を実現うるための技術的，非技術的な方法

実質アジャイルなループなのに
アジャイルと明言していない



Trustworthy Assessment List
別紙

日本での類似の文章としては
QA4AI
http://www.qa4ai.jp/

EU　AI白書2020

- AIサービスは事前に徹底的な**リスク予測**を行うべき
- リスク発生の予測はAI技術に複雑さや発展の早さから技術的に抑えることは困難であることも意識

- AIシステム製造の**サプライチェーンの各段階**で倫理指針ないしは法制度に基づくリスク管理や公平性，非差別性を徹底することを求めている

  - 米国も中国産の基本ツールを念頭におき、サプライチェーンチェックをする方向。５G機器など。

◆基本権（人権），人間の尊厳，多様性，非差別，プライバシーと個人データの保護，　に対するリスペクトを推進する

- このような価値観を世界に広めたい

◆この辺は普通な話ですが、

- AIは既存のEU法およびEU加盟国の国内法が適用される：

  ➢だんだんAI敵視っぽくなってきて

- **EUは、AIアプリケーションに関連する潜在的なリスクに基づく証拠を明確化するために、既存の法的ないし技術的手段を最大限に活用する**

  ➢そうはいってもAIの進歩は認めざるを得ず。。。

  ◆効果的な適用と施行を確実にするために、例えば責任に関する特定の分野の既存の法律を調整または明確にする必要があるかもしれない

- サプライチェーン内の異なる経済事業者間の責任の割り当てに関する不確実性をなくすこと

➢AIシステム，AI製品のサプライチェーンの各段階の当事者が責任を分担すること.

➢EUの強い監督意識が見られます

- AIシステムがすべての**ライフサイクルフェーズ**でエラーや不整合に適切に対処できることを保証する

ちょっと現実感が？？

- AIシステムの出力は、人間によって事前にレビューおよび検証されていない限り、有効にならない

- 動作中のAIシステムの監視、およびリアルタイムで介入して非アクティブ化する機能（人間による）
  - そのために，設計段階で、AIシステムに運を課す

ちょっと現実感が？？

AI規制規則に引き継がれている

- AIを調整する必要がある場合は、たとえばEUでシステムを再トレーニングすることにより、修正する必要があり

- AIシステムが、通常適用される法的制約をを超えている場合，標準化されたEU全体のベンチマークに準拠していなければならないし，そのことが製品に明記されているべき.

  - 保護主義的な傾向が感じられる.
  - EUでは大企業中心の高リスク分野の他に中小規模開発者の保護や支援を強く打ち出していることにも特徴がある.

- もはやEUの保護主義

- AIシステムが実利用時に再学習によって動きが変化する場合，開発者側はその都度リスク管理，公平性などを確認することを求めている．

- つまり，AIシステムの仕様を倫理指針，法制度で管理しようとする指針を打ち出している．開発者にとっては非常な重荷になると予想される．

Trustworthy　AI　から
EU AI白書へ.
そしてその内容をいよいよ法律化したように見える

ARTIFICIAL INTELLIGENCE ACT
AI　規制規則

Brussels, 21.4.2021

“Proposal for a Regulation laying down harmonised rules on artificial intelligence”

URLは以下：

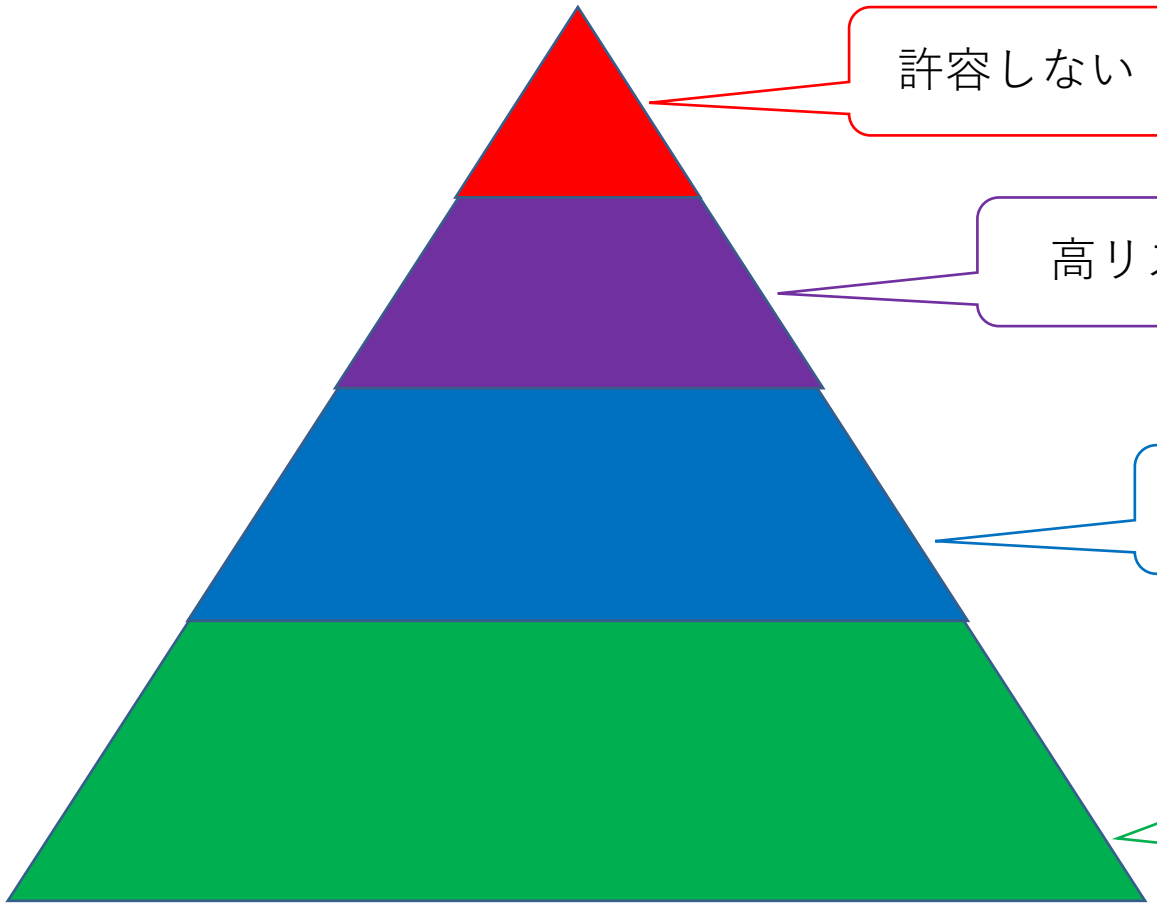
<https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence>

前文に書かれた目的

- 米国，中国などEU域外で進むAI技術に対する不信感
  - EU加盟国は、人工知能が安全であり、基本的権利の義務に従って開発および使用されることを保証するために、国内規則の採用をすでに検討しているが,
  - 国の規則が異なると、国内市場が細分化され、AIシステムを開発または使用する事業者が個別国の法律に振り回される

➢EU全体で一貫した高レベルの保護規則の確保

- オペレーターに統一された義務を課し、 公益と国内市場全体の人の権利を最優先する一様な保護規則が必要.



許容しない

◆公共の場での公的機関による生体（顔）認証,
◆サブリミナル
◆公的機関による社会的スコアリング,
◆性別，年齢，民族などでの差別

高リスク

◆このカテゴリーがもっとも重要で影響大.
後で述べる

限定的リスク

◆AIシステムの透明性が要求される

最小のリスク

◆企業ごとの行動規範（code of conduct） が要求される

第5条で明記された禁止すべきAI

- 法執行の目的で行う公的にアクセス可能なスペースでの「リアルタイム」**リモート生体認証（顔認証）システム**
- 人間に身体的または心理的危害を引き起こす可能性のある方法で人の行動を実質的に歪めるために**サブリミナル**技術用いるAI
- **公的機関**による個人のトラストの評価または分類（**スコアリング**）を行うAI
- 年齢、身体的または精神的障害によって特定のグループの人々の脆弱性につけこむAI

高リスクAIのカテゴリーの内容
ANNEX III: High-Risk AI Systems

- １生体認証とそれによる分類（リアルタイムと事後）　差別があってはならない
  - リアルタイムおよび「ポスト」リモート生体認証に使用することを目的としたAI
- 2.重要な生活インフラストラクチャの管理と運
  - **道路の管理と運用**における安全設備として使用することを目的
  - **交通**と水、ガス、暖房、電気の供給のためのAI
- 3.教育と職業訓練へのアクセスの可否決定
  - 教育および職業訓練機関へのアクセス可否の決定または自然な割り当ての目的で使用することを目的としたAIシステム
  - 職業訓練機関および一般の教育機関への入学のためのテストの者を評価するためAI
- 4.採用における利用、人事評価，労働者管理，雇用と解雇
  - 採用または選択、特に求人広告、アプリケーションのスクリーニングまたはフィルタリング
  - 候補者の評価面接またはテストのためのAI
  - 昇進と解雇の決定を行うために使用されるAI
  - 契約関係、タスクの割り当て、パフォーマンスの監視と評価のために使われるAI

自動運転車に必須の道路情報環境

高リスクAIのカテゴリーの内容
ANNEX III: High-Risk AI Systems

- 5.不可欠な民間サービスおよび公共サービスへのアクセス適格性評価順位付け
  - 公的機関によって、または公的機関に代わって使用されることを目的としたAI
  - 公的支援の給付とサービスに対する適格性を評価するAI
  - そのような利益およびサービスを付与、削減、取り消し、または再請求するAI
  - クレジットスコアを計算するAI
  - 消防士や医療援助などの緊急性ある処理に対する優先順位を計算するために使用されるAI

- 6.法執行機関が個人の状況に立ち入る
  - 個人のリスクを高めるために法執行機関が使用することを目的としたAI
  - 再犯または刑事犯罪の潜在的な犠牲者のリスクを評価するAI
  - 法執行機関がポリグラフなど感情状態を検出するために使用するAI
- 法執行機関がディープフェイクを検出するために使用するAI
  - 第52条（3）で透明性に関与するとして言及されている

高リスクAIのカテゴリーの内容
ANNEX III: High-Risk AI Systems

- 7.移住、庇護および国境管理での利用
  - 公的機関がポリグラフや感情状態を検出するためとして使用することを目的としたAI
  - 公的機関がセキュリティリスク、不法移民のリスク、または健康リスクを含み、加盟国の領土に入ろうとする、または入国した人を検査，検証するAI
  - 渡航文書の信憑性と裏付けとなる文書セキュリティ機能をチェックすることにより、本物ではないドキュメントを検出するAI
  - 公的機関による審査を支援することを目的としたAI
  - 庇護、ビザ、居住許可の申請および関連する苦情ステータスを申請する人の適格性を審査するAI

- 8.司法当局が事実を調査および解釈するのを支援するAI

高リスクAIをEU市場に売り出す，あるいは実利用する前にやならけばならないこと

- ここが本AI規制規則におけるEUの本音.

- 以下の各者ごとに，細かい指示がたくさんだされている.

  - 製造業者,
  - 配布業者,
  - サービス事業開発者（provider),
  - 事業者代表,
  - 公的ユーザ　　（個人的に利用するだけのユーザは除く）

  ◆EU域外国の業者にも適用される

高リスクAIの開発，運用で守ること

- AIシステムの全利用期間におけるリスク管理システム：9条
- データガバナンス：バイアスなし，エラーなし，十分な代表性：10条
- 技術文書の作成：11条
- 利用状況レコードの保存する機能：12条
- ユーザへの内容説明（精度など）と適切な指示，人間が監視する規則の整備：13条
- 人間が監視する出力の限界値，精度．緊急停止ボタンの設置，監視は2名以上で確認するなど：14条
- 継続的に学習するAIの精度，技術的ロバストさ，サイバーセキュリティの確保，敵対的標本データの不使用：15条

CE marking

- 高リスクAIの守るべき事を宣言し，CE markingとして貼付する：49条
  - 配布者はCE markingを確認する：27条

- CE markingは、製造業者または輸入業者が、欧州経済領域（EEA）内で販売される製品の欧州の健康、安全、および環境保護基準への準拠を確認するための管理上のマーキング

- 品質指標や認証マークではない
- CE markingは、EEA規格に準拠して製造されたEEA外で販売された製品にも付ける
- 米国で特定の電子機器を販売するために使用されるFCC適合宣言のようなもの

- CE markingは、製品が健康、安全、および環境保護に関するEU基準を満たしているというメーカーの宣言
- 製品が原産国に関係なく、欧州経済領域のどの部分でも自由に販売できることを示す

高リスクAIが実用に供された後

- マズいことが起きたとき，即時に必要な訂正を行動をとる：21条
- 高リスク**AI**に関するEUのデータベースへの登録：60条
- 市販後の動作のモニタリングと稼働状況レコードの保持：61条
- インシデント報告義務：62条

- （利用規制）違反及びデータガバナンス要件に対する違反の場合は、三千万ユーロ又は全世界の年間総売上の6%が上限の制裁金：71条
- その他の違反の場合は、二千万ユーロ又は全世界の年間総売上の4%が上限の制裁金：71条

命令系統

1. European Commission：　欧州委員会
2. The European Parliament and of the Council：欧州議会及び理事会
3. European Artificial Intelligence Board (“Board”)：新たに設置：56条～59条
4. National Supervisory Authorities：EU内各国の監督権限を持つ組織

ご清聴
ありがとうございました

付録：EU:2面市場を念頭おくプラットフォーム規制法

- デジタル市場法（ＤＭＡ）　2020 年12 月法案提出　2022年７月ＥＵ閣僚理事会採択（成立）
- 種々のプラットフォームと主要なIT企業（コア・プラットフォーム・サービス）が対象
- コア・プラットフォーム・サービスのうち，影響力の強い企業をゲートキーパーとして標的化
- 独占禁止およびアプリストアへのアクセス自由化の観点からゲートキーパーへ以下のような制約を行う法案
  - 他のオンライン仲介業者における別の価格や条件をつけた販売を許容すべし
  - コアプラットフォーム経由のエンドユーザーとビジネスユーザーとの契約を自由化すべし
  - ビジネスユーザーのコア・プラットフォーム利用上の非公開データをゲートキーパーが競争に利用することの禁止
  - ビジネスユーザーのアプリストアへのアクセスについて差別禁止

- その他，かなり厳しい制約があり，業界からは反対の意見が多かった.