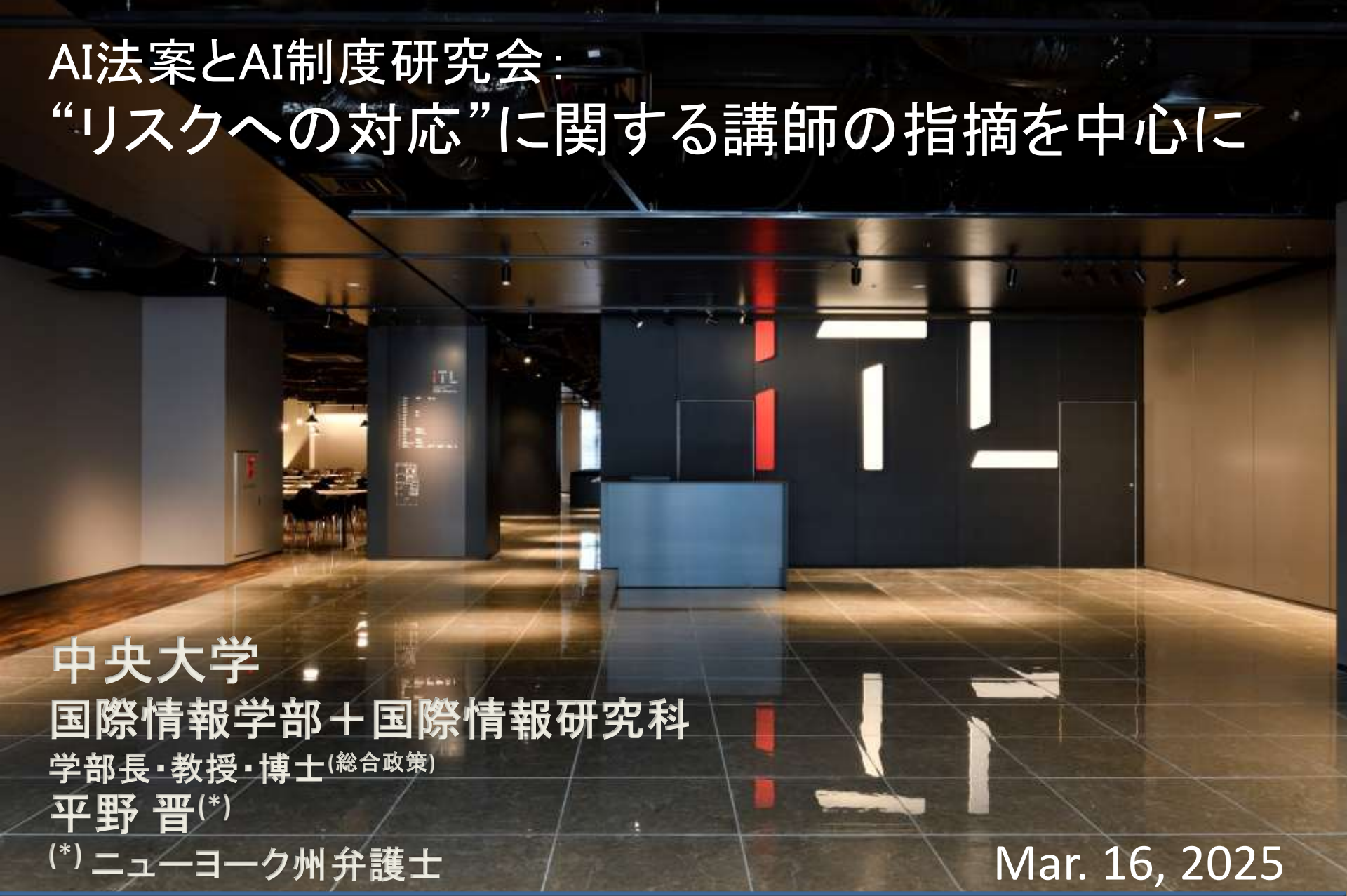
AI法案とAI制度研究会：
“リスクへの対応”に関する講師の指摘を中心に



中央大学
国際情報学部＋国際情報研究科
学部長・教授・博士(総合政策)
平野 晋(*)
(*) ニューヨーク州弁護士

Mar. 16, 2025

ELSI大学サミット

自己紹介

https://c-research.chuo-u.ac.jp/html/100003450_ja.html (last visited Mar. 15, 2025).

AIGO, OECD, Paris, France, in Sept. 2018.

CHUO UNIVERSITY
FACULTY OF
GLOBAL INFORMATICS

平野　晋

（中央大学 国際情報学部 教授 ・ 学部長
ニューヨーク州弁護士）

略歴

中央大学法学部法律学科卒 ／ コーネル大学(法科)大学院修了（LL.M）・『CORNELL INT'L L.J.』編集委員 ／ 博士(総合政策)（中央大学） ／ 富士重工業㈱法規部主任、㈱NTTドコモ法務室長などを経て、中央大学総合政策学部教授、同大学国際情報学部開設準備室長など。

国際機関/政府の委員等

- 経済協力開発機構(OECD)「AI専門家会合」日本共同代表(2019年）
- 内閣府「AI制度研究会」構成員
  - 同　「人間中心のAI社会原則会議」構成員
- 総務省「AIネットワーク社会推進会議」構成員(副議長）
  - 同　「AIガバナンス検討会」構成員(座長)　など

著書・論文

- 『ロボット法：ヒトとAIの共生にむけて（増補第2版）』（2024年, 弘文堂）
- 「AIに不適合なアルゴリズム回避論：機械的な人事採用選別と自動化バイアス」『情報通信政策研究』第7巻2号1頁（総務省, 2024年3月）
- 『アメリカ不法行為法』（2006年, 中央大学出版部）
- 『電子商取引とサイバー法』（1999年, NTT出版）　など

講演の背景

Paris, France in Oct. 2017.

内閣府「AI制度研究会」

「イノベーションの促進とリスクへの対応の両立」

出典：@首相官邸 ２０２４年８月２日

首相官邸「AI戦略会議・AI制度研究会」令和6年8月2日 https://www.kantei.go.jp/jp/101_kishida/actions/202408/02ai.html (last visited Aug. 15, 2024).

「AI法案(*)」

「イノベーションの促進とリスクへの対応の両立」

(*)「人工知能関連技術の研究開発及び活用の推進に関する法律案」

一方におきまして、今も御指摘ありましたが、ＡＩには様々なリスクがございます。これらへの対応が必要であります。座長から御説明いただきました中間とりまとめ案に沿いまして、城内大臣を中心に、平大臣ほか関係閣僚が協力し、ＡＩのイノベーション加速とリスク対応を両立させる新たな法案を早期に国会に提出できますよう、対応を進めていただきたいと存じます。

政府におきますＡＩ政策の司令塔機能を強化するため、全閣僚からなります『ＡＩ戦略本部』を設置いたします。

首 相 官 邸
Prime Minister's Office of Japan

首相官邸「総理の一日」「AI戦略会議・AI制度研究会合同会議」令和6年12月26日 https://www.kantei.go.jp/jp/103/actions/202412/26ai.html (last visited Mar. 16, 2025).

「AI法案(*)」

「イノベーションの促進とリスクへの対応の両立」

(*)「人工知能関連技術の研究開発及び活用の推進に関する法律案」

AI制度の在り方につきましては、昨年8月から、AI戦略会議の下に設置したAI制度研究会におきまして、検討を開始し、AI関係者へのヒアリングや議論、パブリックコメントを重ね、本年2月4日に「中間取りまとめ」を決定したところでございます。この「中間取りまとめ」を踏まえ、「AI法案」の詳細についての検討を進めてまいりました。

本法案は、AIは、生産性の向上や労働力不足の解消などのメリットをもたらす一方で、偽情報・誤情報の拡散、犯罪の巧妙化などのリスクも存在することから、国民生活の向上や経済社会の発展の実現には、AIによるイノベーション促進とリスク対応を両立させることが重要との考えのもと、AI政策の司令塔機能を強化する、内閣総理大臣を本部長とする「AI戦略本部」の設置、政府が推進すべきAI政策の基本的な方針等を示す「AI基本計画」の策定、AIの適正性確保のための国際規範に即した「指針」の整備、AIの動向に関する情報収集や国民の権利利益を侵害する事案の調査などについて、規定しているものであります。


内閣府
Cabinet Office

城内内閣府特命担当大臣記者会見要旨 令和7年2月28日

1.発言要旨

内閣府
Cabinet Office

城内内閣府特命担当大臣記者会見要旨 令和7年2月28日
https://www.cao.go.jp/minister/2411_m_kiuchi/kaiken/20250228kaiken.html (last visited Mar. 16, 2025).

内閣府「AI制度研究会」
「イノベーションの促進とリスクへの対応の両立」

内閣府ホームページ「中間とりまとめ（案）」
https://www8.cao.go.jp/cstp/ai/ai_kenkyu/5kai/shiryou1.pdf
(last visited Mar. 14, 2025).

講師が着目するリスク

- 「不適切な」AI利用
  - ✓ 内閣府「AI制度研究会」→「AI法案(*)」
  - ✓ “ADM:” Automated Decision-Making、又は Algorithmic Decision-Making
- AIへの過度な依存
  - ✓ 内閣府「人間中心のAI社会原則」
  - ✓ 平野『AIとヒトの共生にむけて』

(*)「人工知能関連技術の研究開発及び活用の推進に関する法律案」

「イノベーションの促進と
リスクへの対応の**両立**」

ELSIは、ハンドルとブレーキ

【高橋構成員】

２点確認したい。１点目は、事務局への確認である。資料１の冒頭で「ＩＣＴインテリジェント化の影響とリスク」について議論があった。技術開発はアクセルにあたる。人文社会科学にはハンドルという側面もあるが、ブレーキでもある。ブレーキを踏む人を確認することやそのためのルールが、正しいものになっているからこそアクセルを踏めるだろう。例えば、遺伝子工学では倫理委員会が大学毎にあるが、そういう組織があるからこそ技術開発が進められてきた。そういった理解でよいか。

ICTインテリジェント化影響評価検討会議（AIネットワーク化検討会議「第１回 議事概要」５～６頁, 2016年2月2日（理研生化学シミュレーション研究チーム チームリーダー高橋恒一構成員発言））
https://www.soumu.go.jp/main_sosiki/kenkyu/iict/index.html (last visited Aug. 15, 2024)..

（１） 構成員

須藤 修（座長）	東京大学大学院情報学環教授
平野 晋（座長代理）	中央大学大学院総合政策研究科教授
石井 夏生利	筑波大学図書館情報メディア系准教授

エンジンとハンドル/ブレーキ

S T E M

Science,
Technology,
Engineering, and
Mathematics

理数工学 系

E L S I

Ethical,
Legal, and
Social
Implications

人文/社会科学 系

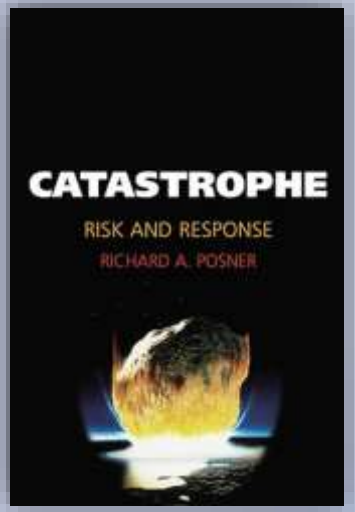
“信頼に値する AI”

“人間中心の AI”

【参考】リチャード A. ポズナー判事の指摘

"[s]cientists want to advance scientific knowledge rather than to protect society from science; the policy maker's ordering of values is the reverse. Not that scientists are indifferent to public safety; but it is not their business and sometimes it is in competition with their business

RICHARD A. POSNER, CATASTROPHE: RISK AND RESPONSE 99 (2004)(emphasis added).



Conflict of the First Priorities

Policy Makers
E·L·S·I
1st Priority of Value: **Public Safety**

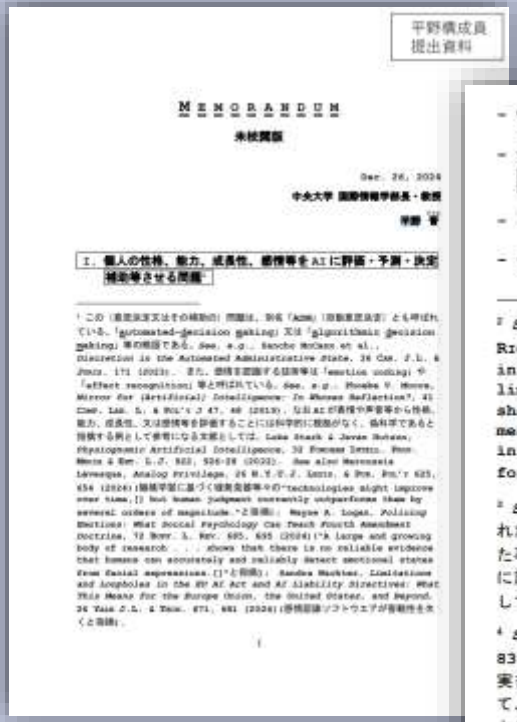
versus

Scientists
S·T·E·M
1st Priority of Value: **Advancement of Scientific Knowledge**

AI制度研究会:「リスクへの対応」

- 「不適切な」AI利用
  - ✓ 内閣府「AI制度研究会」→「AI法案」
  - ✓ “ADM:” Automated Decision-Making
    Algorithmic Decision-Making

講師が特に指摘したリスク



平野構成員
提出資料

MEMORANDUM

未校閲版

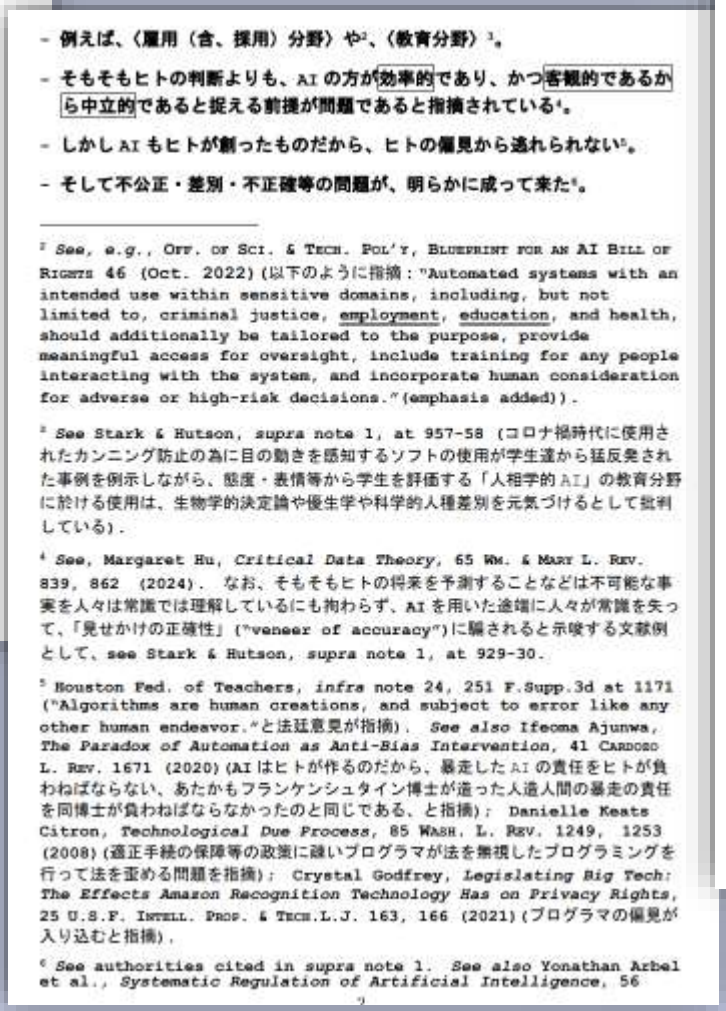
Dec. 26, 2024

中央大学 国際情報学部長・教授

平野 晋

I. 個人の性格、能力、成長性、感情等をAIに評価・予測・決定・補助等させる問題

[illegible]



- 例えば、〈雇用（含、採用）分野〉や[2]、〈教育分野〉[3]。
- そもそもヒトの判断よりも、AIの方が効率的であり、かつ客観的であるから中立的であると捉える前提が問題であると指摘されている[4]。
- しかしAIもヒトが創ったものだから、ヒトの偏見から逃れられない[5]。
- そして不公正・差別・不正確等の問題が、明らかに成って来た[6]。

---

[2] *See, e.g.*, Off. of Sci. & Tech. Pol'y, Blueprint for an AI Bill of Rights 46 (Oct. 2022)（以下のように指摘："Automated systems with an intended use within sensitive domains, including, but not limited to, criminal justice, employment, education, and health, should additionally be tailored to the purpose, provide meaningful access for oversight, include training for any people interacting with the system, and incorporate human consideration for adverse or high-risk decisions."（emphasis added））.

[3] *See* Stark & Hutson, *supra* note 1, at 957-58（コロナ禍時代に使用されたカンニング防止の為に目の動きを感知するソフトの使用が学生達から猛反発された事例を例示しながら、態度・表情等から学生を評価する「人相学的AI」の教育分野に於ける使用は、生物学的決定論や優生学や科学的人種差別を元気づけるとして批判している）.

[4] *See*, Margaret Hu, *Critical Data Theory*, 65 Wm. & Mary L. Rev. 839, 862 (2024). なお、そもそもヒトの将来を予測することなどは不可能な事実を人々は常識では理解しているにも拘わらず、AIを用いた途端に人々が常識を失って、「見せかけの正確性」（"veneer of accuracy"）に騙されると示唆する文献例として、see Stark & Hutson, *supra* note 1, at 929-30.

[5] Houston Fed. of Teachers, *infra* note 24, 251 F.Supp.3d at 1171（"Algorithms are human creations, and subject to error like any other human endeavor."と法廷意見が指摘）. *See also* Ifeoma Ajunwa, *The Paradox of Automation as Anti-Bias Intervention*, 41 Cardozo L. Rev. 1671 (2020)（AIはヒトが作るのだから、暴走したAIの責任をヒトが負わねばならない、あたかもフランケンシュタイン博士が造った人造人間の暴走の責任を同博士が負わねばならなかったのと同じである、と指摘）; Danielle Keats Citron, *Technological Due Process*, 85 Wash. L. Rev. 1249, 1253 (2008)（適正手続の保障等の政策に疎いプログラマが法を無視したプログラミングを行って法を歪める問題を指摘）; Crystal Godfrey, *Legislating Big Tech: The Effects Amazon Recognition Technology Has on Privacy Rights*, 25 U.S.F. Intell. Prop. & Tech.L.J. 163, 166 (2021)（プログラマの偏見が入り込むと指摘）.

[6] *See* authorities cited in *supra* note 1. *See also* Yonathan Arbel et al., *Systematic Regulation of Artificial Intelligence*, 56

- 日本では（も）人事採用AIのベンダ[7]が活発に売り込み、かつそれを利用する企業も少なくないように見受けるけれども、「やり過ぎ」に注意が必要ではないか[8]。

Ariz. St. L.J. 545, 557 (2024); Godfrey, *supra* note 5, at 168（顔認識技術――FRT――が感情を読み取れるという主張が批判されていると指摘）, 雇用にAIを利用すると差別的結果が生じるリスクを指摘する文献としては、see, *e.g.*, Sonderling et al., *infra* note 12.

[7] 人事採用にAIを使うことが不適切であることは、日本でも、例えば平野が座長を務める「AIネットワーク社会推進会議・AIガバナンス検討会」の第2回会合（平成30年[2018年]12月10日）に於ける講者のご発表に於いても既に指摘されていた。「資料1 早稲田大学 大湾先生 御発表資料：人事データ活用への関心とガイドライン作成に向けての議論」
https://www.soumu.go.jp/main_content/000589116.pdf (last visited Sept. 12, 2024). 平野自身の文献については、平野晋「AIに不適合なアルゴリズム回避論：機械的な人事採用選別と自動化バイアス」『情報通信政策研究』第7巻2号1頁（総務省,2024年3月）
https://www.soumu.go.jp/iicp/journal/journal_07-02.html (last visited Sept. 24, 2024);「（資料1）AIの判断に対するヒトの最終決定権の限界：Human-in-the Loopの問題」*in* 総務省「情報通信法学研究会 令和5年度」2023年9月6日
https://www.soumu.go.jp/main_sosiki/kenkyu/hougakuken/R05_siryou.html (last visited Sept. 24, 2024).

[8] 「やり過ぎ」についての批判としては、例えば顔の表情から感情を読み取ることは困難であると指摘されているにも関わらず、大企業やスタートアップ企業等がこれを売り込んでヒトの一生を左右している問題が、連邦取引委員会（FTC）委員等による共著論文に於いて次のように指摘されている。この指摘が日本にも当てはまることがないことを、筆者は祈るばかりである。

> A review that analyzed more than a thousand studies on emotional expression concluded that "[e]fforts to simply 'read out' people's internal states from an analysis of their facial movements alone, without considering various aspects of context, are at best incomplete and at worst entirely lack validity, no matter how sophisticated the computational algorithms."[] Nevertheless, large companies [] --plus a host of well-funded start-ups-- continue to sell questionable affect recognition technology, and it is sometimes deployed to grant or deny formative life opportunities.
>
> A striking example of the use of affect recognition is in hiring.

Rebecca Kelly Slaughter et al., *Algorithms and Economic Justice: A Taxonomy of Harms and a Path Forward for the*

「AI制度研究会 構成員提出資料」33/50～46/50頁
https://www8.cao.go.jp/cstp/ai/ai_kenkyu/5kai/shiryou2.pdf (last visited Mar. 16, 2025).

「不適切な」AI利用
欧米に於けるADM(*)批判と規制

(*) ADM: Automated Decision-Making又は Algorithmic Decision-Making

アメリカの学説や政策

- ✖ **就活者の選別をAIに委ねたり依拠することは、不正確（エビデンス欠如）及び／又は差別的（不公正）なので望ましくない。**

See 平野晋「AIに不適合なアルゴリズム回避論」『情報通信政策研究』7巻2号1頁（2024年3月）https://www.jstage.jst.go.jp/article/jicp/7/2/7_1/_html/-char/ja; 平野晋「Human in the Loopの問題」*in* 総務省_情報通信法学研究会, 2023年9月6日https://www.soumu.go.jp/main_content/000899843.pdf ；「資料１ 早稲田大学 大湾先生 御発表資料：人事データ活用への関心とガイドラ イン作成に向けての議論」https://www.soumu.go.jp/main_content/000589116.pdf .

EUのAI Act

- ✖ **雇用**や**教育等**の分野に於けるAI利用は、厳しい規制対象。
- ✖ **職場**と**教育機関**に於いて感情を推認するAI利用は、原則として禁止。

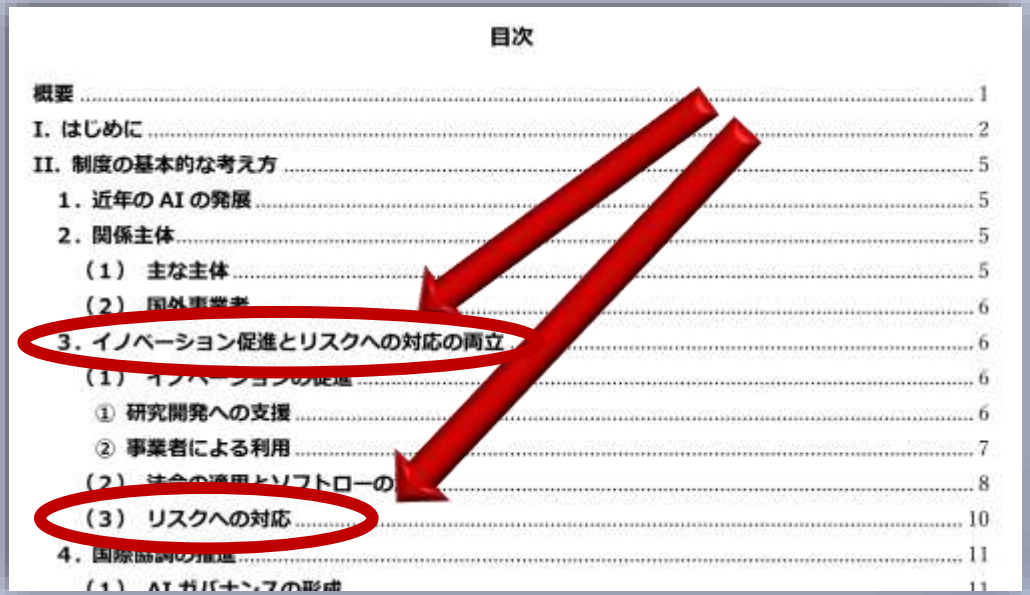
☑不正確　☑不透明　☑説明無責任

- 就活生・求職者に対するAI利用（濫用）例：
  - 「Jaredジャレッド」という名前で高校時代に「ラクロス」をやっていたら、仕事で成功する。
  - ゲームをやらせて仕事上の成功の資質が予測できる。
- 実証的/科学的根拠（エヴィデンス）を欠いている。
- データマイニングで相関関係を示せても、因果関係は証明されていない。
- peer reviewによる再現性を通じた科学的正しさが証明されない。
- 営業秘密の壁。

See e. g., 平野晋「Human in the Loop の問題」総務省_情報通信法学研究会, 2023 年9月6日 https://www.soumu.go.jp/main_content/000899843.pdf (last visited Mar. 16, 2025).

内閣府「AI制度研究会 中間とりまとめ~~（案）~~」

や用途が様々存在する中で、開発、提供、利用といった AI のライフサイクルの各場面において顕在化する可能性のあるリスクとは、いかなる種類の AI モデルのどのような性質に起因するリスクであって、誰にどのような影響を与えるものかといった要素を分析する必要がある。その前提として、まずは AI の開発、利用等に関する実態を調査・分析し、社会全体で認識を共有した上で必要な対応を適時適切に行うことが重要である。例えば、不適切な AI による求職者の選別や AI による消費者の混乱に対する懸念があり、政府は実態の把握に努めるとともに必要な対策を検討することも重要である。また、政府による実態の把握のため、各主体に協力の要請を行う



目次

内閣府「AI制度度研究会 中間とりまとめ（案）」11頁
https://www8.cao.go.jp/cstp/ai/ai_kenkyu/5kai/shiryou1.pdf
(last visited Mar. 15, 2025).

「リスクへの対応」: AI法案

「人工知能関連技術の研究開発及び活用の推進に関する法律案」*in* 内閣府「第127回 通常国会」https://www.cao.go.jp/houan/217/index.html（last visited Mar. 15, 2025）.

（調査研究等）

第十六条

国は、国内外の人工知能関連技術の研究開発及び活用の動向に関する情報の収集、不正な目的又は不適切な方法による人工知能関連技術の研究開発又は活用に伴って国民の権利利益の侵害が生じた事案の分析及びそれに基づく対策の検討その他の人工知能関連技術の研究開発及び活用の推進に資する調査及び研究を行い、その結果に基づいて、研究開発機関、活用事業者[*]その他の者に対する指導、助言、情報の提供その他の必要な措置を講ずるものとする。

（強調及び注書付加）

[＊]「活用事業者」とは、「人工知能関連技術を活用した製品又はサービスの開発又は提供をしようとする者その他の人工知能関連技術を事業活動において活用しようとする者」をいう。（第7条）

（基本理念）

第三条

．．．。

４　人工知能関連技術の研究開発及び活用は、不正な目的又は不適切な方法で行われた場合には、犯罪への 利用、個人情報の漏えい、著作権の侵害その他の国民生活の平穏及び国民の権利利益が害される事態を助長するおそれがあることに鑑み、その適正な実施を図るため、人工知能関連技術の研究開発及び活用の過程の透明性の確保その他の必要な施策が講じられなければならない。

５　．．．。

（強調付加）

（国の責務）

第四条

国は、前条に定める基本理念（以下「基本理念」という。）にのっとり、人工知能関連技術の研究開発及び活用の推進に関する施策を総合的かつ計画的に策定し、及び実施する責務を有する。

（強調付加）

（活用事業者の責務）

第七条

人工知能関連技術を活用した製品又はサービスの開発又は提供をしようとする者その他の人工知能関連技術を事業活動において活用しようとする者（以下「活用事業者」という。）は、基本理念にのっとり、自ら積極的な人工知能関連技術の活用により事業活動の効率化及び高度化並びに新産業の創出に努めるとともに、第四条の規定に基づき国が実施する施策及び第五条の規定に基づき地方公共団体が実施する施策に協力しなければならない。

内閣府「AI制度研究会 中間とりまとめ~~（案）~~」

（２）政府等による利用

政府が AI を利用することにより、行政サービスや業務等の質・効率を向上させることができるほか、政府がユースケースやその有用性を示し、具体事例、留意点等を周知することにより、AI の活用を促進し、国内の AI 市場の発展等にも貢献できるため、政府が率先して AI を利用していくことは重要である。ただし、国民の権利利益に重大な影響を及ぼしかねないものについては、AI の出力結果を自動的に採用することのリスク[11]を踏まえ、慎重に取り組むべきである。

地方自治体についても、行政サービスの大きな部分を占めており、国民生活への影響も大きいため、AI の利用を推進し、行政サービスや業務等の質・効率を向上させていくことが重要である。また、地方自治体における AI の利用の推進にあたっては、各自治体の先進的な取組[12]を含む利用事例を参考にするほか、各自治体の地域課題に応じた AI の利用も重要である。

---

[11] 海外におけるリスク発現事案としては、米国における失業保険に関する事例（ミシガン州の失業保険庁が請求者の詐欺を検出するために使用した統合データ自動化システムは、2013 年から 2 年間で 93%のエラー率を記録し、2 万人の州民を詐欺で誤って告発した。）や、教師の失職につながった事例（ヒューストン独立学区は、教師が生徒の学力成長に及ぼす影響を推定する付加価値モデル（VAM）を導入し、2007 年から 2016 年までに教師の契約更新の拒否や解雇に利用したため、多くの教師が VAM の使用を止めるための司法救済を求めた。）が存在する。

[12] 神戸市においては、2024 年 3 月、一定のルール下での AI の効果的かつ安全な活用を目的として AI 条例を制定するほか、文章要約、アイデア出し、プログラミングコードの生成等 AI の活用を進めている（AI 制度研究会（第 2 回）資料 3 参照）。

内閣府「AI制度度研究会 中間とりまとめ（案）」17～18頁 https://www8.cao.go.jp/cstp/ai/ai_kenkyu/5kai/shiryou1.pdf (last visited Mar. 15, 2025).

https://www8.cao.go.jp/cstp/ai/ai_kenkyu/ai_kenkyu.html

背景：日本国政府/地方自治体もAIを利活用する方針

具体的な制度・施策の方向性

■ **全般的な事項**（Ⅲ.1.）

AIの研究開発・実装が最もしやすい国を目指す

速やかな法制度化が必要
世界のモデルになるような制度

- **政府**の司令塔機能の強化、戦略の策定
  - 全体を俯瞰する司令塔機能強化
  - AIの安全・安心な研究開発・活用のための戦略（基本計画）の策定
- **安全性の向上等**
  - 国による指針（広島AIプロセス準拠）の整備、事業者による協力
  - 国による調査・情報収集、事業者・国民への指導・助言、情報提供等

■ **政府等による利用**（Ⅲ.2.）
- 適正なAI政府調達・利用 等

■ **基盤サービス等における利用**（Ⅲ.3.）
- 各業法等による対応 等

前掲「中間とりまとめ（案）」1, 17頁.

内閣府「AI制度研究会」神戸市提出資料

AI条例が想定するリスクへの対応

神戸スマートシティ

▶ AI条例が想定するリスクへの対応

- **事前に安全性を完全に担保するルールづくりは現実的ではない一方で、すべてを事後対策に委ねることも危険である**
- **行政が積極的にAIを活用していくには、事前に十分なリスクの洗い出しを行い、リスクが顕在化した際の対処方法を検討しておくことが重要ではないか**

○AI条例に基づくリスクアセスメントの概要

・対象　市及び受託事業者が次の業務にAIを活用するとき
①**市民の権利利益に影響を与える行政処分**（課税、各種給付認定など）
②基本計画等の計画策定（市の基本計画など）
③その他市民生活に重大な影響を与えるおそれがあるもの

・方法　AI事業者ガイドラインのワークシートを参考にしつつ、神戸市の独自項目を加えた**48項目のワークシート**に基づき実施

・特徴　リスク軽減のために、技術的な対策以上に**利用者の運用面での取り組みを重視**

7

神戸スマートシティ

資料3

神戸市におけるAI活用のためのルール整備

2024. 8 .23

神戸市企画調整局デジタル戦略部

https://www8.cao.go.jp/cstp/ai/ai_kenkyu/2kai/shiryou3.pdf (emphasis added).

☑不正確　☑不透明　☑説明無責任

- アルゴリズムやソフトウエアに依拠した不利益処分等が厳しく批判されたアメリカの先例に学ぼう

1. Houston Fed. of Teachers v. Houston Independent, 251 F.Supp. 1168 (S.D.Tex. 2017)． アルゴリズムの評価に従って教員を解雇した際の説明無責任と不透明性が違憲とされた事例。

2. Cahoo v. SAS Analytics Inc., 912 F.3d 887 (6th Cir. 2019)． 失業保険受給申請の詐欺を自動的に決定等するシステムの不正確性等が問題になった事例。

- **自動的/アルゴリズム的意思決定(ADM)の問題／行政処分・不利益処分の問題**

☑ 不透明　☑ 説明無責任　[☑不正確]

Houston Fed. of Teachers, Local 2415 et al.
対
Houston Independent School District事件

https://www.houstonpublicmedia.org/articles/news/2017/10/10/241724/federal-lawsuit-settled-between-houstons-teacher-union-and-hisd/ (last visited Mar. 16, 2025).

NEWS

Federal Lawsuit Settled Between Houston's Teacher Union and HISD

Now that EVAAS won't be used, the Houston Federation of Teachers hopes for a more hands on approach

Houston Public Media Staff | October 10, 2017, 2:36 PM

Listen

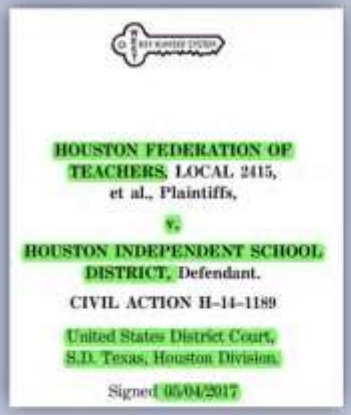
Members of the Houston Federation of Teachers.

*Houston Fed.*事件の概要

- ヒューストン市の教員の評価を、EVAASと呼ばれるアルゴリズムに算出させ、そこで最低評価になった教員が解雇や契約非継続となった。教員組合と解雇等された教員達（π）がEVAAS使用の恒久的な差止等を求めて、ヒューストン独立教区（◿）に対する訴えを提起。
- EVAASは訴外ベンダのSAS社が供給していて、そのソースコードや統計的手法等の情報は**営業秘密**であるとして、◿にもπにも開示せず、いわば「ブラックボックス」であった。しかし、ソースコード等の情報が開示されなければ、評価結果を再現できず、その誤りの有無をπが検証できないことが問題になった。
- 裁判所は、再現できなければ、手続的な適正手続保障違反に当たると指摘。
  - → SH: 求職者を選別するAIに対する批判と一致！
  - → SH: peer reviewに服して正確性が再現できなければ、怪しいという考え方。

☑ 不透明 ☑ 説明無責任 [☑不正確]

営業秘密 v. 透明性



HOUSTON FEDERATION OF TEACHERS, LOCAL 2415, et al., Plaintiffs,

v.

HOUSTON INDEPENDENT SCHOOL DISTRICT, Defendant.

CIVIL ACTION H-14-1189

United States District Court, S.D. Texas, Houston Division.

Signed 05/04/2017

VALUE-ADDED RATING	EVAAS ®TGI	RELATIONSHIP TO EXPECTED AVERAGE GROWTH
Well above	Equal to or greater than 2	Students on average substantially exceeded expected average growth
Above	Equal to or greater than 1 but less than 2	Students on average exceeded average growth
No detectable difference	Equal to or greater than -1 but less than 1	Students on average met expected growth
Below	Equal to or greater than -2 but less than -1	Students on average fell short of average growth
Well below	Less than -2	Students on average fell substantially short of expected average growth

> [W]ithout access to [the vender's] proprietary information—the . . . computer source codes, decision rules, and assumptions—EVAAS scores will remain a mysterious 'black box,' imperative to challenge.
>
> . . . .
>
> HISD teachers have no meaningful way to ensure correct calculation of their EVAAS scores, and as a result are unfairly subject to mistaken deprivation of constitutionally protected property interests in their jobs.

Houston Fed'n of Tchrs., Loc. 2415 v. Houston Indep. Sch. Dist., 251 F. Supp. 3d 1168, 1179 (S.D. Tex. 2017)(emphasis added).

平野 Human in the Loopの問題」48頁 https://www.soumu.go.jp/main_content/000899843.pdf @ 総務省 情報通信法学研究会」2023年9月6日

[☑ 不正確性]

"Algorithms are human creations, and subject to error like any other human endeavor."

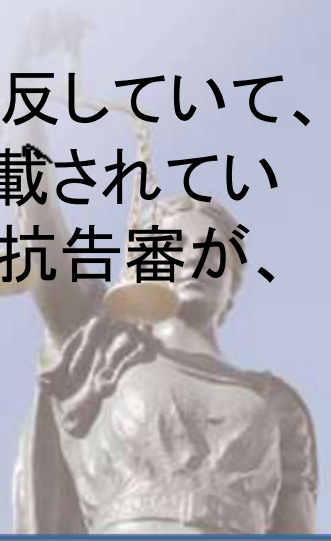
Houston Fed. of Teachers v. Houston Independent School District, 251 F.Supp.3d 1168, 1177 (S.D.Tex. 2017)(emphasis added).

☑ 不正確性　☑ 説明無責任

Cahoo et al.
対
SAS Analytics Inc., et al.事件

Cahoo事件の概要

- ミシガン州失業保険庁(UIA: Unemployment Insurance Agency)の採用した失業保険受給者の詐欺を自動的に判断するMiDAS (Michigan Integrated Automated System)が、**大量(93%)の誤判断・偽陽性(faulse positive)**を下した(robo-adjudicated)。
- その被害者であるの原告達(π)が、MiDASを監督・使用していた同省職員達被告(⊿)に対して提訴。
- 訴状では、⊿が合衆国憲法上の適正手続保障に違反していて、〈資格に基づく免責〉に値しない旨の主張が十分記載されているとして、原審の判断を支持した控訴裁判所による抗告審が、本件。



国際規範との整合性

人工知能関連技術の研究開発及び活用の推進に関する法律案（AI法案）

法案の概要	目的	**国民生活**の向上、**国民経済**の発展
	基本理念	経済社会及び**安全保障上重要** ➡ 研究開発力の保持、**国際競争力**の向上 基礎研究から活用まで総合的・計画的に推進 **適正な研究開発・活用**のため**透明性の確保等** **国際協力において主導的役割**
	AI戦略本部	**本部長：内閣総理大臣　構成員：全閣僚** 関係行政機関等に対して必要な協力を求める
	AI基本計画	研究開発・活用の推進のために政府が実施すべき施策の基本的な方針等
	基本的施策	研究開発の推進、施設等の整備・共用の促進 人材確保 教育振興 **国際的な規範策定への参画** 適正性のための**国際規範に即した指針**の整備 **情報収集、権利利益を侵害する事案の分析・対策検討、調査** **事業者・国民への指導・助言・情報提供**
	責務	国、地方公共団体、研究開発機関、事業者、国民の責務 関係者間の連携強化 **事業者は国等の施策に協力しなければならない**
	附則	**見直し規定**（必要な場合は所要の措置）

内閣府「第127回国会」「人工知能関連技術の研究開発及び活用の推進に関する法律案」「概要」 https://www.cao.go.jp/houan/217/index.html (last visited Mar. 14, 2025).

AI法案：

国際規範に即したガイドライン整備を国の義務化

（適正性の確保）

第十三条

国は、人工知能関連技術の研究開発及び活用の適正な実施を図るため、国際的な規範の趣旨に即した指針の整備その他の必要な施策を講ずるものとする。

（強調付加）

内閣府「AI制度研究会」の冒頭提出講師資料

内閣府「AI戦略会議（第11回）・AI制度研究会（第1回）※合同開催」https://www8.cao.go.jp/cstp/ai/ai_senryaku/11kai/11kai.html (last visited Mar. 15, 2025).

Transparency and explainability (Principle 1.3)

This principle is about transparency and responsible disclosure around AI systems to ensure that people understand when they are engaging with them and can challenge outcomes.

> *AI Actors should commit to transparency and responsible disclosure regarding AI systems. To this end, they should provide meaningful information, appropriate to the context, and consistent with the state of art:*
>
> - *to foster a general understanding of AI systems, including their capabilities and limitations,*
> - *to make stakeholders aware of their interactions with AI systems, including in the workplace,*
> - *where feasible and useful, to provide plain and easy-to-understand information on the sources of data/input, factors, processes and/or logic that led to the prediction, content, recommendation or decision, to enable those affected by an AI system to understand the output, and,*
> - *to provide information that enable those adversely affected by an AI system to challenge its output.*

Rationale for this principle

The term transparency carries multiple meanings. In the context of this Principle, the focus is first on disclosing when AI is being used (in a prediction, recommendation or decision, or that the user is interacting directly with an AI-powered agent, such as a chatbot). Disclosure should be made with proportion to the importance of the interaction. The growing ubiquity of AI applications may influence the desirability, effectiveness or feasibility of disclosure in some cases.

Transparency further means enabling people to understand how an AI system is developed, trained, operates, and deployed in the relevant application domain, so that consumers, for example, can make more informed choices. Transparency also refers to the ability to provide meaningful information and clarity about what information is provided and why. Thus transparency does not in general extend to the disclosure of the source or other proprietary code or sharing of proprietary datasets, all of which may be too technically complex to be feasible or useful to understanding an outcome. Source code and datasets may also be subject to intellectual property, including trade secrets.

Explainability means enabling people affected by the outcome of an AI system to understand how it was arrived at. This entails providing easy-to-understand information to people affected by an AI system's outcome that can enable those adversely affected to challenge the outcome, notably – to the extent practicable – the factors and logic that led to an outcome. Notwithstanding, explainability can be achieved in different ways depending on the context (such as the significance

Therefore, when AI actors provide an explanation of an outcome, they may consider providing – in clear and simple terms, and as appropriate to the context – the main factors in a decision, the determinant factors, the data, logic or algorithm behind the specific outcome, or explaining why similar-looking circumstances generated a different outcome. This should be done in a way that allows individuals to understand and challenge the outcome while respecting personal data protection obligations, if relevant.

https://oecd.ai/en/dashboards/ai-principles/P7 **(emphasis added)**

AI諸原則 ／ AIガイドライン：日本が国際標準を主導

G20 2019.6.
- デジタル経済大臣会合

OECD 2018.9. ～ 2019.5.
- 理事会勧告 2019.5
- AI専門家会合～2019.2

内閣府 2018.4. ～ 2019.3.
- 人間中心のAI社会原則[検討]会議

総務省 2016.1. ～ 現在

- AIネットワーク社会推進会議 **2016.10.～現在**
  - OECDに於けるカンファレンスを総務省と共催／参加。
  - AIネットワーク社会推進フォーラムに於ける国際シンポを総務省主催／参加。
  - カーネギー平和財団のカンファレンスを日本大使館後援／総務省参加。
  - OECD技術予測フォーラムに於ける発表。

✓ 高松 香川G7情報通信大臣会合（高市早苗総務大臣）

- **AIネットワーク化検討会議 ～2016.6.**

AIにおける国際的な議論への貢献

G７情報通信大臣会合（高松、2016年4月）

・高市総務大臣（当時）からの提案※:

“G7各国が中心となり、OECD等国際機関の協力も得て、AIネットワーク化が社会・経済に与える影響や、**AI開発原則の策定**等AIネットワーク化をめぐる社会的・経済的・倫理的・法的課題に関し、産学民官の関係ステークホルダーの参画を得て、国際的な議論を進める”

⇒ 参加各国からの賛同を得る

※提案に先立ち、叩き台として、８項目のAI開発原則を配付。

Distributed material

Proposal of Discussion toward Formulation of AI R&D Guideline

Referring OECD guidelines governing privacy, security, and so on, **it is necessary to begin discussions and considerations toward formulating an international guideline consisting of principles governing R&D of AI to be networked** (“**AI R&D Guideline**”) as framework taken into account of in R&D of AI to be networked.

Proposed Principles in “AI R&D Guideline”

1. **Principle of Transparency**
   Ensuring the abilities to explain and verify the behaviors of the AI network system
2. **Principle of User Assistance**
   Giving consideration so that the AI network system can assist users and appropriately provide users with opportunities to make choices
3. **Principle of Controllability**
   Ensuring controllability of the AI network system by humans
4. **Principle of Security**
   Ensuring the robustness and dependability of the AI network system
5. **Principle of Safety**
   Giving consideration so that the AI network system will not cause danger to the lives/bodies of users and third parties
6. **Principle of Privacy**
   Giving consideration so that the AI network system will not infringe the privacy of users and third parties
7. **Principle of Ethics**
   Respecting human dignity and individuals’ autonomy in conducting research and development of AI to be networked
8. **Principle of Accountability**
   Accomplishing accountability to related stakeholders such as users by researchers/developers of AI to be networked

講師が近年着目するリスク (cont'd)

- [「不適切な」AI利用]
- AIへの過度な依存
  - ✓ 内閣府「人間中心のAI社会原則」
  - ✓ 平野『AIとヒトの共生にむけて』

「人間中心のAI社会原則」

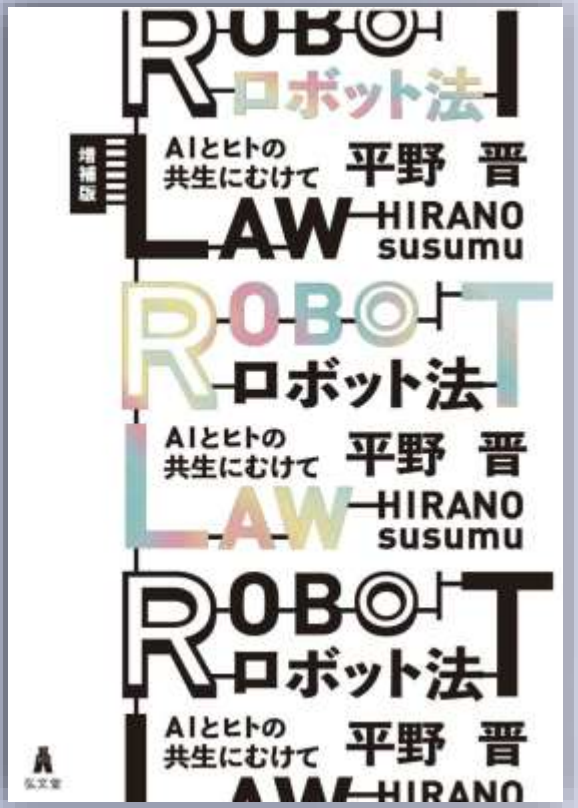
. . . 。AI が活用される社会において、人々が AI に過度に依存したり、AI を悪用して人の意思決定を操作したりすることのないよう、我々は、リテラシー教育や適正な利用の促進などのための適切な仕組みを導入することが望ましい。

（強調付加）

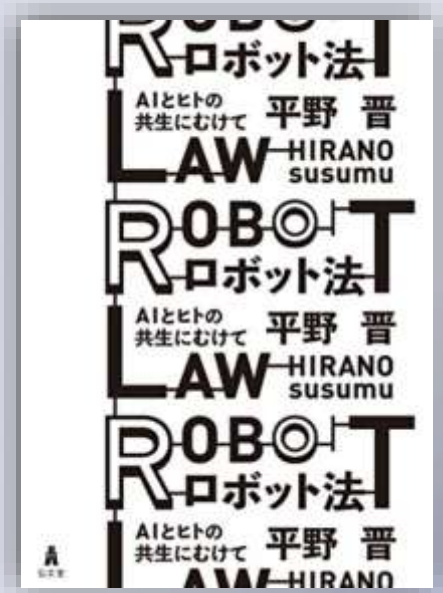
内閣府「人間中心のAI社会原則」8頁（2019年3月29日）内の「1）人間中心の原則」
https://www8.cao.go.jp/cstp/ai/aigensoku.pdf (last visited Mar. 15, 2025).

「AIとヒトの共生に向けて」

平野晋『ロボット法：AIとヒトの共生にむけて（増補第2版）』（弘文堂，2024年5月）



同『ロボット法：AIとヒトの共生にむけて』（弘文堂，2019年10月）



同『ロボット法：AIとヒトの共生にむけて』（弘文堂，2017年11月）

背景:「AIアイちゃんがそう言ったから」が通用しない場合もある(?!)

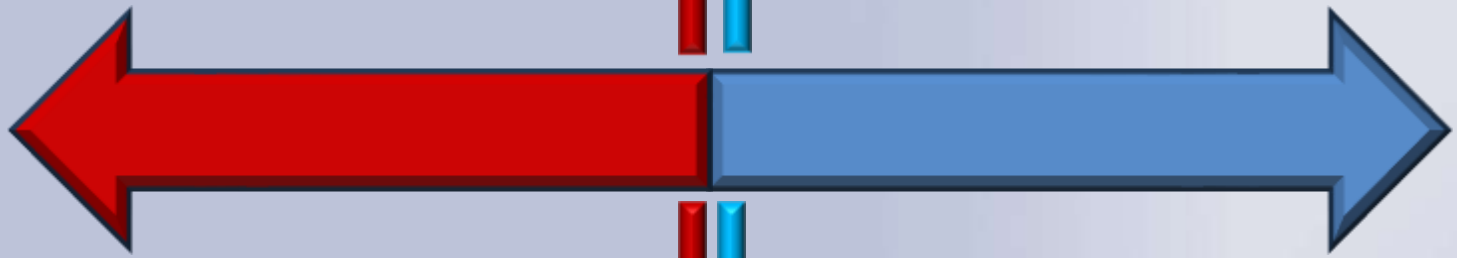
AIの方がヒトよりも正しいという誤解(偏見)

- ✖ AIは客観的だから。
- ✖ AIは数値に基づいているから。
- ✖ AIは偏見に左右されないから。

人間中心の／信頼に値するAI

- 人間疎外のAI
- 信頼に値しないAI

- 人間中心のAI
- 信頼に値するAI



- 制御不能
- 予見不能
- 不透明
- 説明無責任
- 不正確
- 不公正
- 不適合

including,
but not limited to, . . .

including,
but not limited to, . . .

- 制御可能
- 予見可能
- 透明
- 説明責任
- 正確
- 公正
- 適合

☑ 不適合

- 文脈が読めない。大局を理解しない。
  - 紛争地域に於いてAK-47を持った老人は戦闘員か？
  - 紛争地域に於けるデュアル・ユースなピックアップ・トラックは非戦闘用か？ (*1)
- データ化されていない情報は使えない。(*2)
- 裁量が必要な判断には使えない(次葉)。
  - 「制限速度60km/h」⇔「安全運転をする義務」
  - プロ・テニスに於ける「Hawk Eye」利用 (*3)

    SH: 50-50を達成した大谷翔平選手の次の打者に対して15秒以内に投球しなかったマイケル・バウマン投手に対して、ピッチ・クロックを宣言しなかった審判の判断。

(*1) Charles P. IV Trumbull, Autonomous Weapons: How Existing Law Can Regulate Future Weapons, 34 EMORY INT'L L. REV. 533 (2020).
(*2) 大湾先生 御発表資料, 前掲, at page 17.
(*3) Ehan Lowens, Note, Accuracy Is Not Enough: The Task Mismatch Explanation of Algorithm Aversion and Its Policy Implications, 39 HARV. J. L. & TECH. 259 (2020).

☑ 不適合

平野「Human in the Loopの問題」39頁

採用AIに於けるHITLの欠点に対する対策方針

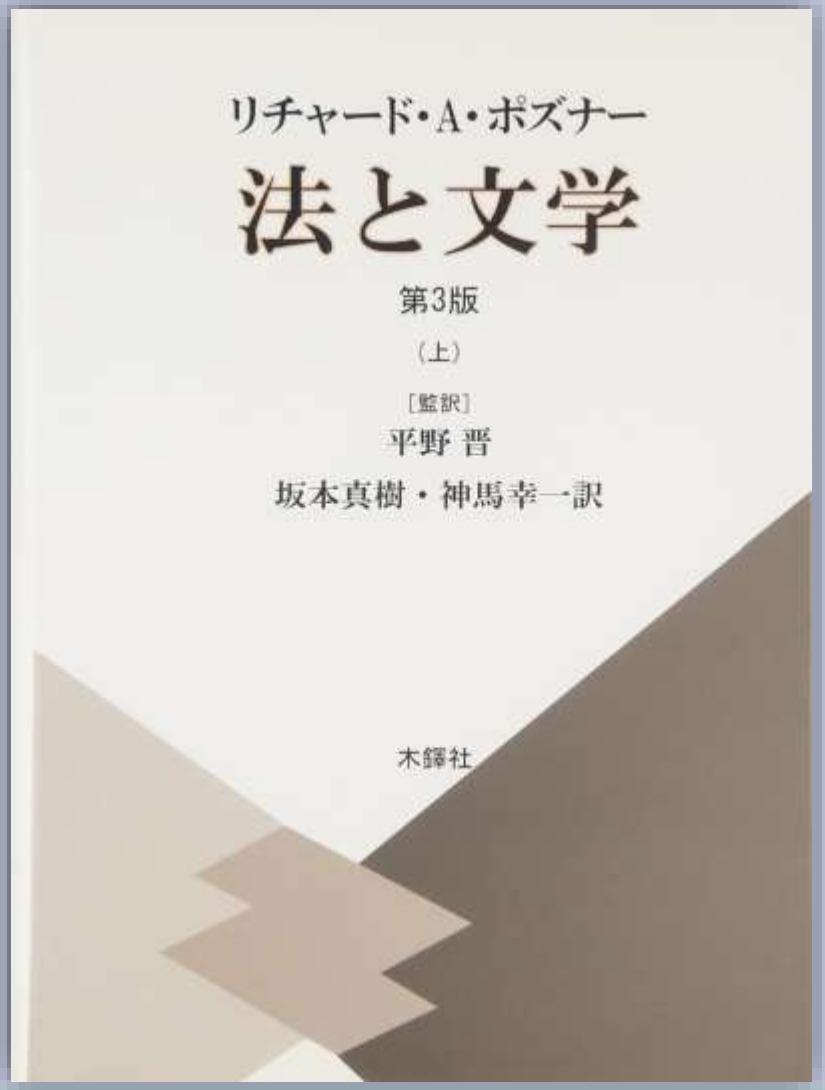
- 更にAIが不得手な判断の場合とは、事前の**rules**が適している場合ではなく、**standards**の当てはめによる個別的判断が適している場合。

Green, supra, at 12-13; See also Margot E. Kaminski, Binary Governance: Lessons from the GDPR'S Approach to Algorithmic Accountability, 92 S. CAL. L. REV. 1529, 1542-43 (2019).

- 〈法的安定性〉v.〈**個別具体的妥当性**〉
- common law v. **equity**
- **そもそも採用活動は、後者に該当するのではないか。**

Ben Green, The Flaws of Policies Requiring Human Oversight of Government Algorithms, 45 COMPUT. L. SEC. REV. 1, 12-13 (2022).

法的安定性 v. 具体的妥当性



See リチャード・A・ポズナー著／平野晋監訳『法と文学（上）』 206～07頁（2011年）.

Thank you !!! ;-)

INFORMATION TECHNOLOGY
& LAW
ICHIGAYA TAMACHI LINK