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

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

Sept. 6, 2023

総務省_情報通信法学研究会

目次

はじめに

バトー：「人形使い」の一件以来、変なんだって総合評価のレポートに書いといたでしょうが？読んでねぇのかよ。

部長、自分の脳をいじくらせてる電脳医師の人格を疑ったことは？

荒巻：電脳医師は定期的な精神鑑定を義務付けられとるし、公安関係の医師には身辺調査も入れとるわ。

[尤も] それを実行するのも同じ人間だ [がな . . .]。

バトー：疑い出せばキリがないか [. . .]。

映画「攻殻機動隊Ghost in the Shell」（松竹, 1995年）

Human-in-the Loop

循環」の中の〈考え/判断〉は、O-O-D-A ループの②と③を合わせたものと理解することもできるので、参考になろう。また②の「情勢判断（方向づけ）」は、トリノ大学（出典公表時）のパガロ教授がロボットの構成要素の第１要素として挙げている「インタラクティヴ［ネス］」(interactiveness)、すなわち環境情報の刺激に反応して優先度を変更すること、に近い概念であると理解できよう。

ところで O-O-D-A ループの要点は、このループを敵方（ミグ戦闘機パイロット）よりも早く回すことによって、敵よりも常に早く（F-86 戦闘機パイロットが）主導権をとり続けることにあった[★116]。民生品のロボットにおいても、ヒトより早く O-O-D-A ループを回すことができるロボットの方が、ヒトよりも優れた能力を発揮することが期待されよう。

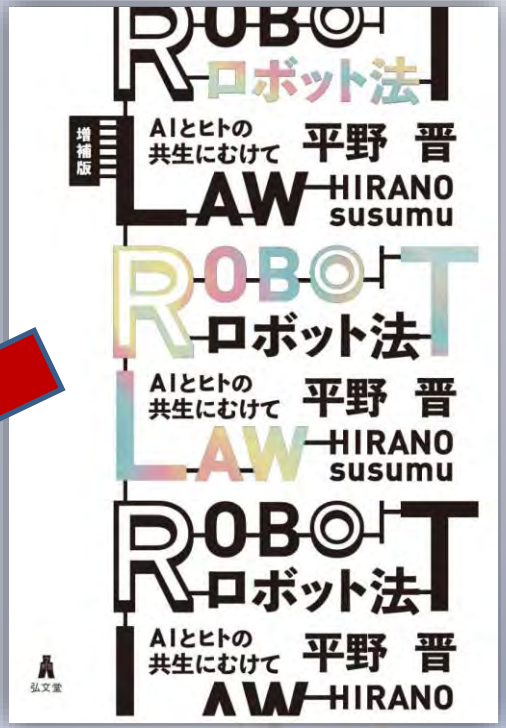
3「ループ」の中のヒトの介入

ロボット兵器の文脈で、「ループ」の中のヒトの介入の程度によって自律性の高低を把握する概念が存在する[★117]。すなわち、以下の通りである。

①「human *in* the loop」な兵器	：標的選択および交戦はヒトが指令する
②「human *on* the loop」な兵器	：標的選択および交戦はヒトの監視下で行い、ヒトが兵器の行動を停止できる
③「human *out of* the loop」な兵器	：ヒトの介入なしに標的選択および交戦できる

この３分類において③を「完全自律型兵器」と捉えうることはすぐに理解できる。が、しかし、ヒトが監視する②「human *on* the loop」な兵器であっても、実際にヒトが介入してそのロボットの行動をヒトに奪い返せる場合は限定的であるという指摘もある。なぜならロボットの意思決定はほんの一瞬のナノセカンド——nano-second——で行われてしまうばかりか、その決定に至った情報源は監視者にはアクセス不能であるから、ヒトは事実上「*out of* the loop」に置かれ、「*on* the loop」な兵器が結

84　第３章　ロボットの定義と特徴

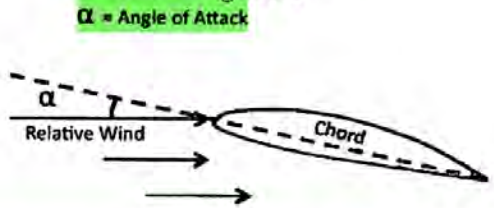


拙書『ロボット法（増補版）』84頁（弘文堂, 2019年）.

Human-on-the Loop

The purpose of MCAS is to prevent an inadvertent aerodynamic stall during certain, relatively unusual flight conditions. To non-pilots, the word "stall" may suggest something wrong with the aircraft's engines. (My children, who are learning to drive stick-shift, frequently stall the car in that sense.) An aerodynamic stall, on the other hand, occurs when the wing exceeds its critical angle of attack. Angle of attack (AOA) is the angle between the chord line (the straight line connecting the leading and trailing edges of the wing) and the relative wind.[16]

Figure 1



Exceeding the critical angle of attack causes the airflow around the wing to separate and create such an excess of drag over lift that the wing can no longer function to keep the airplane on the desired flight path. The result is a loss of control if

[12] *See* Mac McClellan, *Can Boeing Trust Pilots?*, AIR FACTS (Mar. 11, 2019), https://airfactsjournal.com/2019/03/can-boeing-trust-pilots [https://perma.cc/YZ6X-DT2J].

[13] *See, e.g.*, Brian Palmer, *Boeing vs. Airbus*, SLATE (July 11, 2011), https://slate.com/news-and-politics/2011/07/do-pilots-have-a-preference-between-boeing-and-airbus.html [https://perma.cc/S3D3-M45W].

[14] *See id.*

[15] Yaw damping to prevent Dutch roll is a more familiar intervention in the flight control system and is part of every modern jetliner. *See* Bjorn Fehrm, *Boeing's Automatic Trim for the 737 MAX Was Not Disclosed to the Pilots*, LEEHAM NEWS (Nov. 14, 2018), https://leehamnews.com/2018/11/14/boeings-automatic-trim-for-the-737-max-was-not-disclosed-to-the-pilots [https://perma.cc/43TN-UWG9]. Leeham News is a respected aviation industry blog.

[16] *See* H.H. HURT, JR., AERODYNAMICS FOR NAVAL AVIATORS 22 (rev. ed. 1965).

TECHNOLOGICAL SOLUTIONS TO HUMAN ERROR AND HOW THEY CAN KILL YOU: UNDERSTANDING THE BOEING 737 MAX PRODUCTS LIABILITY LITIGATION

W. BRADLEY WENDEL*

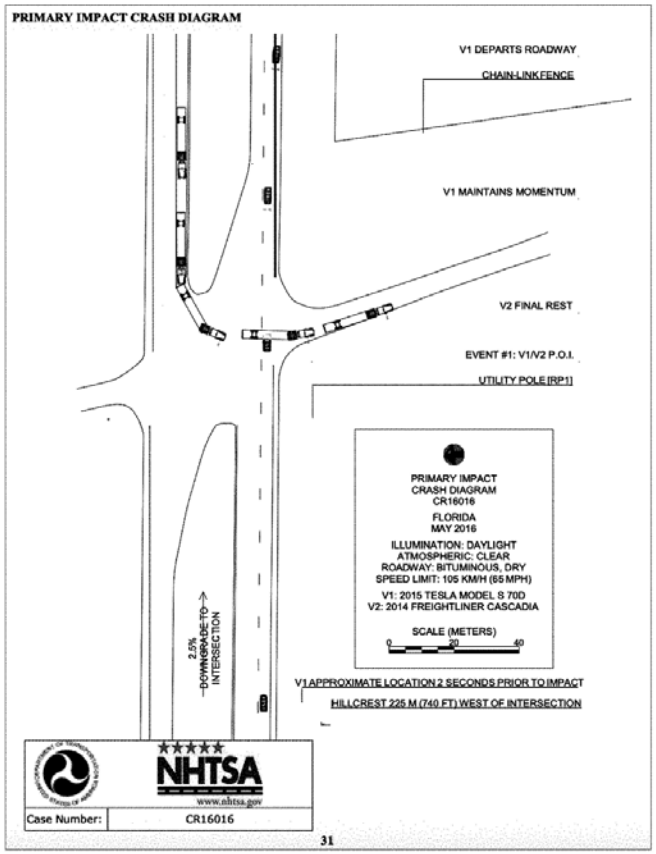
W. Bradley Wendel, Technological Solutions to Human Error and How They Can Kill You: Understanding the Boeing 737 Max Products Liability Litigation, 84 J. AIR L. & COM. 379, 383 & Fig.1 (2019).

"HMI"

human-machine interface

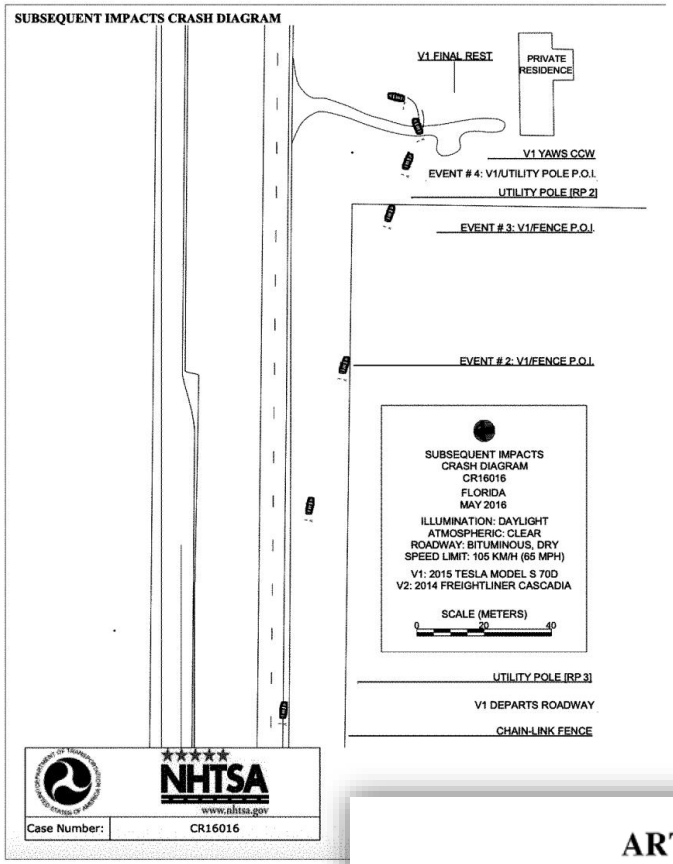
Human-on-the Loop

Figure 2: NHTSA Accident Reconstruction, Joshua Brown (1 of 2)
(a) Reconstruction Based on Sensory Input Data Taken by Tesla Autopilot System[476]



476. NAT'L HIGHWAY TRAFFIC SAFETY ADMIN., U.S. DEP'T OF TRANSP., SPECIAL CRASH INVESTIGATIONS: ON-SITE AUTOMATED DRIVER ASSISTANCE SYSTEM CRASH INVESTIGATION OF THE 2015 TESLA MODEL S 70D 31(2018), https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812481 [https://perma.cc/8KYQ-GCN3].

Figure 3: NHTSA Accident Reconstruction, Joshua Brown (2 of 2)
(a) Further Reconstruction Based on Sensory Input Data Taken by Tesla Autopilot System[477]



477. *Id.* at 32.

Bryan Casey, Robot Ipsa Loquitur, 108 GEO. L.J. 225, 285-86 & Figs. 2-3 (2019).

ARTICLES

Robot Ipsa Loquitur

BRYAN CASEY*

Towards the Human-out-of-the Loop

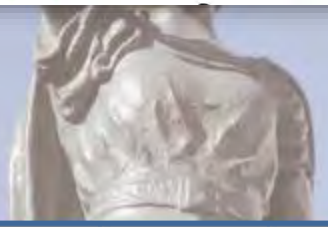
As recently as five years ago, talk of removing humans from the vehicle driving loop entirely—as was the case in the Cruise accident—might not have sat well with many automakers.[187] Even for companies in hot pursuit of driverless capabilities, the end goal looked less like Cruise's vision of automation and more like Tesla's. Instead of closed-loop systems, companies were aiming for an open-loop design paradigm, which the industry called incremental autonomy.[188] Under this philosophy, the "dynamic driving task" would not be handed over to robots entirely. Instead, incremental advancements in technologies such as "adaptive cruise control, lane keeping assist, [and] pedestrian recognition" would iteratively improve over time, allowing machines to take over more and more of the driving task.[189] Whenever intervention was required, though, robots would still need to hand off control to human drivers, thereby obliging humans to serve as ever-vigilant backups.[190]

Casey, Robot Ipsa Loquitur, supra, at 248.

Towards the Human-out-of-the Loop

This iterative vision has mostly faded, however. The reason is straightforward: although an incremental approach sounds great on paper, an increasingly robust body of evidence has shown that "humans are for the most part horrible back-ups."[191] Studies have demonstrated that human operators are particularly ill-suited to long-term supervision tasks and tend to develop a false sense of confidence when automating technologies function for long periods without issue.[192] Confidence, in turn, begets complacency; operators' "attention wanders, and they often begin to doze."[193] It's this phenomenon that got Tesla into hot water with the NTSB in the aftermath of Joshua Brown's fatal collision.[194] Now that the evidence against incremental autonomy is all but incontrovertible, most companies are trying to skip it entirely by racing straight for closed-loop robots.[195]

Id. at 249 (emphasis added).



Human-out-of-the Loop [and the Skynet?!]: ロボット兵器

AUTONOMOUS WEAPONS: HOW EXISTING LAW CAN REGULATE FUTURE WEAPONS

*Charles P. Trumbull IV**

Charles P. IV Trumbull, Autonomous Weapons: How Existing Law Can Regulate Future Weapons, 34 EMORY INT'L L. REV. 533, 543 (2020).

The difference between "autonomous" and "fully autonomous" weapons is not always clear, but it is significant.[69] Weapon systems that can select and engage targets once activated may nevertheless incorporate significant degrees of human judgment or control in the targeting process. Similar to a driver that programs an autonomous vehicle to drive to a "grocery store" without having a particular destination in mind, a human operator could theoretically instruct an autonomous weapon only to attack "tanks." In both cases, the machine can exercise autonomy in determining how to achieve the objective set by a human operator. The definitions of fully autonomous weapon systems, by contrast, do not contemplate a role for the human operator in defining the parameters of the targets to be attacked. Rather, such a weapon might be understood to search for and attack whatever the weapon itself determines to be a military objective, perhaps based on some understanding of the commander's general intent or strategy.

The military advantages outlined in Section A make it inevitable that weapons with significant degrees of autonomy will be deployed on the battlefield. The employment of autonomy in warfare is likely to mirror how autonomy has been utilized in civilian sectors. Humans will assign to machines those tasks that machines can perform more effectively than humans.[98] Routine tasks that entail if/or decision-making are most likely to be automated. Similarly, tasks that require decision-making speeds that exceed human capabilities will increasingly be given to machines.[99] By contrast, tasks that require contextual awareness, common sense, creativity, or abstract decision-making will be performed by humans for the foreseeable future.[100] The introduction of weapons with autonomous features will almost certainly be "deliberate and incremental" as militaries seek to ensure the safety and effectiveness of this technology before deploying it on the battlefield.[101] Militaries will also seek to maximize the strengths of machines and humans by designing weapons systems that collaborate with soldiers.[102]

Human-in-the Loop: ロボット兵器

Id. at 548, 559.

force.[182] A factual determination required by IHL may be whether an individual is an enemy combatant or "directly participating in hostilities," rendering him a legitimate target.[183] These factual determinations are often based on observable and objective evidence. An individual wearing an enemy uniform and carrying a weapon may easily be designated an enemy combatant. In other circumstances, however, this determination may not be so clear. Assessing whether an individual is directly participating in hostilities may require a more contextualized assessment of the individual's conduct. A shepherd carrying an AK-47 may not raise suspicions in countries where it is common to carry semi-automatic weapons. This shepherd may be deemed hostile though if he also carries a satellite phone and is positioned on a trail leading to a village occupied by enemy forces. A degree of professional intuition is often required in making these contextualized determinations.

Human-in-the Loop: ロボット兵器

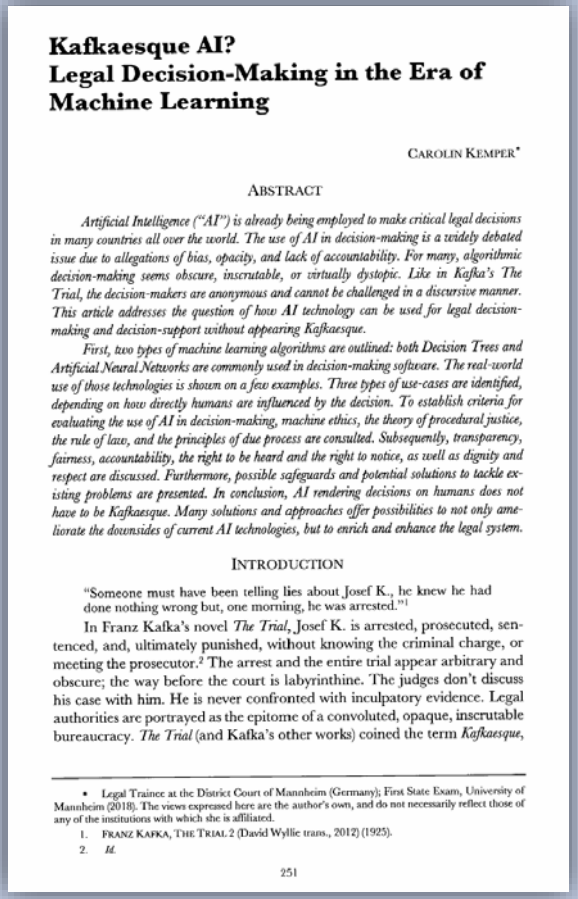
By contrast, autonomous weapons may never reach human capabilities in making determinations that require contextualized judgments, such as when targeting objects that have both civilian and military uses. Whether an object constitutes a legitimate military objective depends on whether it makes an "effective contribution to military action" and whether its "total or partial destruction, capture or neutralization, in the circumstances ruling at the time, offers a definite military advantage."[264] A pick-up truck parked at Walmart would almost certainly be a protected civilian object. This same truck parked near a known terrorist camp, however, could be a legitimate military objective. Similarly, a bridge used predominantly for civilian purposes may become a military objective if it would be a viable escape route for enemy forces in an upcoming assault. Determining whether dual use objects can be attacked requires contextualized judgments, based on a strategic or tactical understanding of the specific operating environment, that machines may never be capable of making.

Id. at 576.

Algorithmic/Automated/ Automatic Hiring: AI採用

- そもそも一生を左右するような重大な決定を機械に委ねることには違和感が。
  - 映画「ブレードランナー2049」(Warner Bros. 2017)に於ける主人公の「K」(KD6-3.7 OR "Joe")が、任務後に白い狭い部屋でThe Post Traumatic Baseline Test (emotional devianceを測る試験)の受験を強要されるような非人間性。
  - フランツ・カフカの小説『審判』で、主人公の「ヨーゼフ K」が、理由も説明されないままに逮捕され死刑執行されるような不条理感。

Algorithmic/Automated/ Automatic Hiring：AI採用



Kafkaesque AI?
Legal Decision-Making in the Era of Machine Learning

CAROLIN KEMPER*

ABSTRACT

Artificial Intelligence ("AI") is already being employed to make critical legal decisions in many countries all over the world. The use of AI in decision-making is a widely debated issue due to allegations of bias, opacity, and lack of accountability. For many, algorithmic decision-making seems obscure, inscrutable, or virtually dystopic. Like in Kafka's The Trial, the decision-makers are anonymous and cannot be challenged in a discursive manner. This article addresses the question of how AI technology can be used for legal decision-making and decision-support without appearing Kafkaesque.

First, two types of machine learning algorithms are outlined: both Decision Trees and Artificial Neural Networks are commonly used in decision-making software. The real-world use of those technologies is shown on a few examples. Three types of use-cases are identified, depending on how directly humans are influenced by the decision. To establish criteria for evaluating the use of AI in decision-making, machine ethics, the theory of procedural justice, the rule of law, and the principles of due process are consulted. Subsequently, transparency, fairness, accountability, the right to be heard and the right to notice, as well as dignity and respect are discussed. Furthermore, possible safeguards and potential solutions to tackle existing problems are presented. In conclusion, AI rendering decisions on humans does not have to be Kafkaesque. Many solutions and approaches offer possibilities to not only ameliorate the downsides of current AI technologies, but to enrich and enhance the legal system.

INTRODUCTION

"Someone must have been telling lies about Josef K., he knew he had done nothing wrong but, one morning, he was arrested."[1]

In Franz Kafka's novel *The Trial*, Josef K. is arrested, prosecuted, sentenced, and, ultimately punished, without knowing the criminal charge, or meeting the prosecutor.[2] The arrest and the entire trial appear arbitrary and obscure; the way before the court is labyrinthine. The judges don't discuss his case with him. He is never confronted with inculpatory evidence. Legal authorities are portrayed as the epitome of a convoluted, opaque, inscrutable bureaucracy. *The Trial* (and Kafka's other works) coined the term *Kafkaesque*,

* Legal Trainee at the District Court of Mannheim (Germany); First State Exam, University of Mannheim (2018). The views expressed here are the author's own, and do not necessarily reflect those of any of the institutions with which she is affiliated.

1. FRANZ KAFKA, THE TRIAL 2 (David Wyllie trans., 2012) (1925).
2. *Id.*

リチャード・A・ポズナー

法と文学

第3版

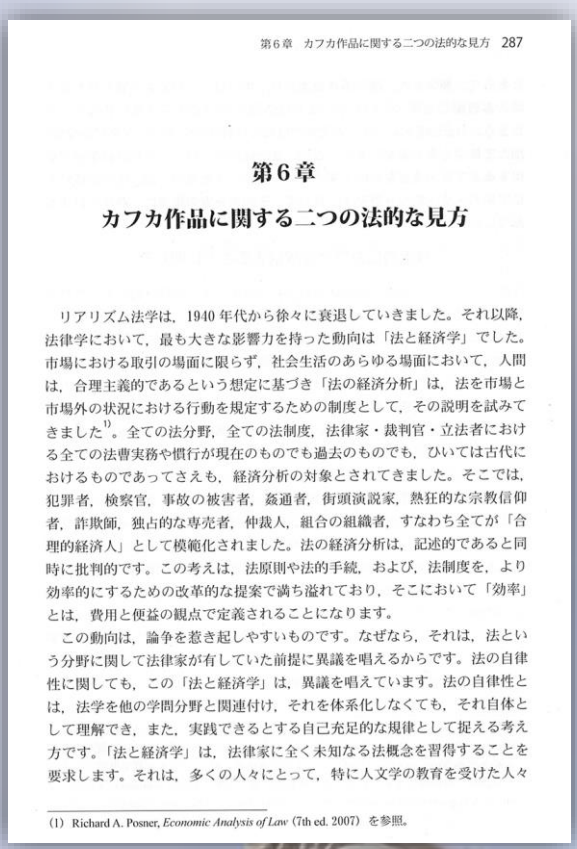
（上）

［監訳］

平野 晋

坂本真樹・神馬幸一訳

木鐸社



第6章

カフカ作品に関する二つの法的な見方

リアリズム法学は，1940年代から徐々に衰退していきました。それ以降，法律学において，最も大きな影響力を持った動向は「法と経済学」でした。市場における取引の場面に限らず，社会生活のあらゆる場面において，人間は，合理主義的であるという想定に基づき「法の経済分析」は，法を市場と市場外の状況における行動を規定するための制度として，その説明を試みてきました[1]。全ての法分野，全ての法制度，法律家・裁判官・立法者における全ての法曹実務や慣行が現在のものでも過去のものでも，ひいては古代におけるものであってさえも，経済分析の対象とされてきました。そこでは，犯罪者，検察官，事故の被害者，姦通者，街頭演説家，熱狂的な宗教信仰者，詐欺師，独占的な専売者，仲裁人，組合の組織者，すなわち全てが「合理的経済人」として模範化されました。法の経済分析は，記述的であると同時に批判的です。この考えは，法原則や法的手続，および，法制度を，より効率的にするための改革的な提案で満ち溢れており，そこにおいて「効率」とは，費用と便益の観点で定義されることになります。

この動向は，論争を惹き起しやすいものです。なぜなら，それは，法という分野に関して法律家が有していた前提に異議を唱えるからです。法の自律性に関しても，この「法と経済学」は，異議を唱えています。法の自律性とは，法学を他の学問分野と関連付け，それを体系化しなくても，それ自体として理解でき，また，実践できるとする自己充足的な規律として捉える考え方です。「法と経済学」は，法律家に全く未知なる法概念を習得することを要求します。それは，多くの人々にとって，特に人文学の教育を受けた人々

(1) Richard A. Posner, *Economic Analysis of Law* (7th ed. 2007) を参照。

リチャード・A.ポズナー『法と文学 第3版』（平野晋監訳，坂本真樹＆神場幸一訳，2011年）

Carolin Kemper, Kafkaesque AI? Legal Decision-Making in the Era of Machine Learning, 24 U.S.F. INTELL. PROP. & TECH. L.J. 251 (2020).

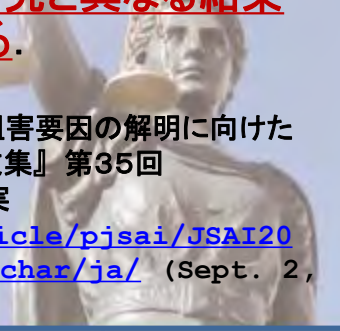
Algorithmic/Automated/ Automatic Hiring：AI採用

AIに面接されることに納得しない候補者は多い[]．．．．．候補者が採用選考に納得が得られないことで，内定を出しても内定を辞退することにつながることや不合格であった候補者が企業にクレームをつけることにつながる可能性がある．

本研究では、面接官がAIと人間で面接を行う場合よりも，人間のみで面接を行う場合の方が，個性が適切に評価されている実感が高いことが支持された．．．．また，．．．個性が適切に評価されている実感は納得感を高めることも支持された．．．．これらの結果より，企業がAI面接に人間を関与させた場合，人間のみの面接における納得感の方が高いことが示唆された．従って，面接を行う際は人間のみで行う方が好ましいと考えられる．これに付随して，候補者の技能評価を目的とする適正検査や，書類選考段階において，企業が効率化等を図ることを目的としてAIを用いる際には，新たに検討の必要がある．

．．．，就職差別の解消を促し，採用における機械均等を促すためにAIによる公正な採用選考を行う必要があり，候補者の納得感に繋がる[ママ]．しかし，実験結果から，AIを面接に用いることにより，候補者は納得感が低くなることがわかった．AI面接は評価が公正であり，候補者の納得につながると期待されていたが，候補者が結果についてどう感じるかという視点で見ると，必ずしも候補者の納得につながっていると断言できない．

以上から，現状の公正な採用選考の考え方は，候補者がどう感じるかという点が考慮できていないため，今後，真に公正な選考とはどういうものなのか検討していく必要がある．ただし，サーベイのシナリオやAI面接の利点・実績を認知させることにより，本研究と異なる結果が得られる可能性は十分ある．



山下芳樹 et al.,「AI人事採用における納得阻害要因の解明に向けたサーベイ実験」in 人『工知能学会全国大会論文集』第35回(2021)1,3頁(実験は2021年2月15～18日に実施)https://www.jstage.jst.go.jp/article/pjsai/JSAI2021/0/JSAI2021_4H2GS11c04/_article/-char/ja/ (Sept. 2, 2023)．

AI採用の背景

- 大量の応募を、効率的に、早く、処理したい。
- 人（面接員）によるバラツキのある判断を統一的で公平な判断にしたい。

AI採用の諸問題

1. 人間性・尊厳への侵害

2. 正確性を欠くAI利活用

See, e.g., Courtney Hinkle, The Modern Lie Detector: AI-Powered Affect Screening and the Employee Polygraph Protection Act (EPPA), 109 GEO. L.J. 1201 (2021).

Humanity & Dignity

- 例えば人を個として判断せずに、自動的な分類化によって判断することが尊厳に反するという批判:

> Automatically making decisions based on what categories an individual falls into--that is, what correlations can be shown between an individual and others--can fail to treat that individual as an individual.[] If algorithmic decision-making does not allow individuals to proclaim their individuality ("I may look like these other people, but I am not in fact like them"), then it violates their dignity and objectifies individuals as their traits, rather than treating an individual as a whole person.[] Both decisional discretion and individual process rights are, under this reasoning, necessary not just to prevent error but to adequately recognize and respect individuality.[]

Margot E. Kaminski, Binary Governance: Lessons from the GDPR's Approach to Algorithmic Accountability, 92 S. CAL. L. REV. 1529, 1542 (2019)(emphasis added).

Humanity & Dignity

- 例えば因果関係ではなく相関関係に依拠する仕組みが尊厳に反するという批判:

> [O]ne could even further argue that dignity interests require that causation and not mere correlation is found prior to launching action.[] With such additional disclosure, individuals can obtain sufficient insight to the process and how it relates to their lives.

Tal Z. Zarsky, Transparent Prediction, 2013 U. Ill. L. Rev. 1503, 1548 (2013)(emphasis added).

Humanity & Dignity

"Dehumanization of ~~Killing~~ Hiring"

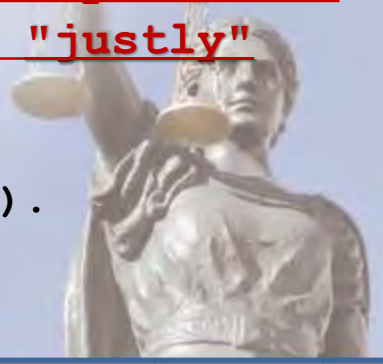
- ロボット兵器是非論からの類推

AUTONOMOUS WEAPONS: HOW EXISTING LAW CAN REGULATE FUTURE WEAPONS

Charles P. Trumbull IV[*]

A related ethical argument is that giving machines the ability to make ~~life-and-death~~ hiring-and-rejection decisions disregards the human dignity of ~~combatants~~ applicants.[] . . . "[T]he central argument here is that it matters not just if a person is ~~killed and injured~~ rejected but *how* they are ~~killed and injured~~ rejected."[] Human intent needs to be linked to the outcome of ~~an attack~~ a decision in order to preserve moral accountability and the ability to determine whether an individual was "justly" ~~killed~~ [rejected].[]

Trumbull, supra, at 552-53 (revision and emphasis added).



正確性を欠くAI利活用

1. “algorithmic bias” 等

2. “automation bias” 等 ／
   HITL: Human-in-the Loop

Algorithmic Bias等

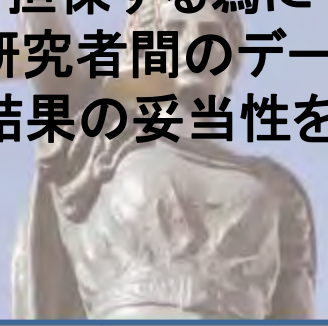
Algorithmic Bias等

- 過去に選抜した者の履歴書を使用している限りは、過去の差別を永続させる。 E.g., Sonderling et al., infra, at 22.
  - E.g., アマゾン社のAI採用の例.
- 偏見のあるデータを学ぶと、差別的効果が生じるおそれがある。Id. at 23; 竹地 at 95.
- アルゴリズムを作成するのは人間であり、その人間の偏見等がアルゴリズムに入り込む。竹地, supra, at 94; Kim, infra, at 876.
- 〈アルゴリズムの偏見〉は、実は背後に居る人が偏見を起こしている事実を忘れさせてしまうことが問題。最初にアルゴリズムのトレーニングに使うデータを決めるのも、トレーニング・データの中のどのデータが重要なのかを決めるのも、関係性のある要素と関係性の無い要素を決めるのも、ヒトなのである。従って、暴走したアルゴリズムの責任は、怪物フランケンシュタインの責任をフランケンシュタイン博士が負うべきであるのと同様に、その創作者が負わねばならない。Ajunwa, Paradox, infra, at 1707. See also 251 F. Supp. 3d, infra, at 1171 ("Algorithms are human creations, and subject to error like any other human endeavor.").

Keith A. Sonderling et al., The Promise and the Peril: Artificial Intelligence and Employment Discrimination, 77 U. MIAMI L. REV. 1 (2022);竹地潔「人工知能による選別と翻弄される労働者」『富山経済論集』第65巻2号91,95頁(2019年12月); Pauline T. Kim, Data-Driven Discrimination at Work, 58 WM. MARY L. REV. 857 (2017); Ifeoma Ajunwa, The Paradox of Automation as Anti-Bias Intervention, 41 CARDOZO L. REV. 1671, 1705 (2020).

Algorithmic Bias等

- AIは、大量のデータをデータ・マイニングし、応募者の特徴と仕事上の将来の成功との間の相関関係(correlation)を見つけ出すのみなので、因果関係(causation)があるとは限らない;データ・マイニングは、関係性の理由を気に掛けない; 仕事遂行能力とは無関係な、全くの偶然性で生じる相関関係を見出して、これを使って将来を予測してしまう。 Sonderling et al., supra, at 24 & n.125; Kim at 874-75.
- データ・マイニングは因果関係の有無を気に掛けないけれども、その結果に基づいて決定を下す使用者にとっては因果と相関の違いは重要である。Kim, supra, at 881.
- 仮説に基づいて統計を分析する社会科学者とは異なって、データ・マイニングは「理論に基づかない」("atheoretical");大量のデータの中から関連性のある要素を見つけ出すだけである。理論に基づかないから、依拠したデータの代表性が正しいかの評価が難しく、モデルの正しさを担保する為に適切な変数が含まれていたか否かを評価することも難しい。研究者間のデータの交換もされず、モデル構築の際の選択も不透明だから、結果の妥当性を他者が検証出来ない。 Kim, supra, at 879-80.



Algorithmic Bias等

- 仮にFBで「らせん状のフライドポテト」(curly fries)を好きなことと高い知性との相関関係が指摘されたらば、業務遂行上知性の高さが重要であるとする企業がこの相関関係に従って採用判断するおそれがある。実際、FBに於いて、**「母親であることが好き」("I Love Being a Mom")に「いいね」("liking")をクリックする者は知性の低さを予測させる**とした例があるけれども、この相関関係に基づく選抜は明らかに女性に対する差別的効果(disparate impact)型の雇用差別に当たる。Kim, supra, at 880-81 & n.95.
- 更に、応募者が仕事で**成功する2つの要素は、高校で「ラクロス」をやっていて、かつ「Jaredジャレッド」という名前であるとアルゴリズムが判断**したことは、トレーニング・データの良し悪しがシステムを左右する例として有名。 Lori Andrews & Hannah Bucher, Automatic Discrimination: AI Hiring Practices and Gender Inequity, 44 Cardozo L. Rev. 145, at 154 (2022).

Algorithmic Bias等

履歴書をAIに選別させる仕組みについては:

- 望ましい技能や資格に関連するキーワード(依頼企業が通常は選択)で履歴書をレビューし、応募書類をフィルタリングしてヒトによる面接段階に進める者を選別する。
- キーワードで応募者が落とされることが知れ渡って、対策記事も多い中、そのようなゲームの仕組みを知らない有能な候補者が面接前に落とされる問題がある。
- 文脈を考慮に入れないので、"code"をキーワードに入れていた場合に応募者が"programming"と書いていた場合には落とされてしまうけれども、ヒトが読めば文脈から落とされないという問題がある。
  - 例えば、職歴に一定の空白期間があれば自動的に落とされてしまうところ、人が読めばそれは女性が結婚・出産・育児等に専念する為の空白期間であることが文脈・全体から読み取れるので、落とさずに有用な応募者を面接に進めて審査することが出来る。

Calli Schroeder, Ben Winters & John Davisson, We Can Work It out: The False Conflict between Data Protection and Innovation, 20 Colo. Tech. L.J. 251, 267-69 (2022).

Algorithmic Bias等

ビデオ面接等により顔認識・音声認識を採用時に使用することについては:

- 応募者のビデオ面談[上の回答]を、成績の良い従業員の回答と比較して評価する。Solderling et al., supra, at 29.
- 表情、アイ・コンタクト、スピーチ・パターン、言葉の選択等の要素を機械学習で分析する。Id. at 29.
- それらを基に、AIが、応募者の仕事との適合性("fit for jobs")や企業文化との相性("compatibility with the culture of the organization")を予測すべく評価する。Id. at 30.
- 人種や性別との関連性で不当に評価されるとして、人権団体から批判されている。Id. at 29.
- 顔の表情を採用判断の考慮に入れることは、そのような特徴と職場に於ける成功との間の因果関係が科学的に立証されていないので、疑問が抱かれている。Ifoema Ajunwa, An Auditing Imperative for Automated Hiring Systems, 37 HARV. J.L. & TECH. 621, 637 (2021); Ajunwa, Paradox of Automation, at 1703.
- H社のアルゴリズムは、候補者の認知能力、心理的特徴、感情的知性、及び社会的才能を判断できると主張するけれども、如何に判断できるのかという問いに同社は答えず、おそらくは同社自身にもどうして判断できるのかが分からず、その不透明性ゆえに監査が著しく難しく、証拠に基づかない論争("a game of they-said, we-said.")に終わっている。Schroeder et al., supra, at 272.
- なおH社は、2019年1月に、facial analysisに対する激しい抗議とFTCへの訴え提起に応じて、同サービスを停止したけれども、音声分析は継続している。Id. at 272; Courtney Hinkle, The Modern Lie Detector: AI-Powered Affect Screening and the Employee Polygraph Protection Act (EPPA), 109 GEO. L.J. 1201, 1203-04 n.4 (2021).

Algorithmic Bias等

ビデオ・ゲームで採用審査を行う場合については:

- ゲームから発見した応募者の様々な特徴が如何に点数化され、それが如何に仕事遂行上の成功に繋がるのかについて不明(不透明)であり、相関関係が因果関係を示すとは限らないし、応募者個人の能力と諸価値とを適切に評価できる証拠もないから、最良な応募者の採用に繋がらない。
  Andrews & Bucher, supra, at 184-90 (特に188-90).

  - プレイヤーがウエイターに成って、高い不満足度を示す客のアバタの表情を読み取って食事を[優先して]給仕することにより危険評価能力を試すと云われるゲームは、これにより感情知能(EQ: emotional intelligence)やその他の特徴を測れたり、マックの従業員から、ゲームとは無関係と思われる医師や投資銀行家等々さえをも含む多種多様な職種の仕事遂行能力を予測できる、という実証研究を欠いている。Id. at 189.
  - ゲームによる応募者評価の妥当性(validity)を証明する、利害関係のない独立した者による研究が存在しない。 Id.

Algorithmic Bias等

- 従業員にゲームをさせたデータをベースラインな特徴として、応募者のゲーム結果と比較させた場合、従業員が例えば白人の中年男性で占められていた場合、ゲームの結果分析は業務遂行能力を表す特徴を示すよりも寧ろ従業員の統計的人口構成上の特徴を示すことに成ってしまうという偏見のおそれがある。 Schroeder et al., at 271. See also Andrews & Bucher, supra, at 179. See also Savage & Bales, supra, at 225 (text accompanying note 143)(同旨).

- But see, David D. Savage & Richard Bales, Video Games in Job Interviews: Using Algorithms to Minimize Discrimination and Unconscious Bias, 32 A.B.A. J. LAB. & EMP. L. 211, 221, 222, 224-25 (2017)(personal performanceに着目するゲームによる採用は差別を生まないと主張しつつも、注意深く行わないと差別的な危険がある、と以下のように指摘している)(emphasis added).

> As with any type of employment hiring practice, discrimination may occur if proper precautions are not taken.

Algorithmic Bias等

厚生労働省 労働政策審議会労働政策基本部会「報告書～働く人がAI等の新技術を主体的に活かし、豊かなsh将来を実現するために～」令和元年[2019年]6月

https://www.mhlw.go.jp/stf/newpage_05463.html (last visited Sept. 2. 2023)の指摘：

> **（2）AIによる判断に関する企業の責任・倫理**
>
> AIの情報リソースとなるデータやアルゴリズムにはバイアスが含まれている可能性があるため、AIによる判断に関して企業が果たすべき責任、倫理の在り方が課題となる。例えば、HRTechでは、**リソースとなるデータの偏りによって、労働者等が不当に不利益を受ける**可能性が指摘されている。このため、AIの活用について、企業が倫理面で適切に対応できるような環境整備を行うことが求められる。特に働く人との関連では、人事労務分野等においてAIをどのように活用すべきかを労使始め関係者間で協議すること、**HRTechを活用した結果にバイアスや倫理的な問題点が含まれているかを判断できる能力を高めること、AIによって行われた業務の処理過程や判断理由等が倫理的に妥当であり、説明可能かどうか等を検証すること等が必要**である[]。他方、AI等を活用することにより、人間による業務判断の中にバイアスが含まれていないかを解析することもできるため、技術革新が人間のバイアスの解消に資する可能性もあるという指摘もあり、今後、こうした面からもAI等の活用が期待される。

Id. at 10 (emphasis added).

Automation Bias等/ HITL

Automation Bias等/HITL

- algorithmic bias等への対策として、AIの勧告・予想・判断等を「最後は人が判断している」という主張が見受けられる。
- すなわち、Human-in-the Loop （HITL）。
- しかし、HITLはautomation bias等ゆえに機能しないという指摘が多く見受けられる。
- 更には、システムを正当化する「口実」に使われるという厳しい批判も。

Automation Bias等/HITL

The cognitive system's engineering literature has found that human beings view automated systems as error-resistant .[] Operators of automated systems tend to trust a computer's answers.[]

. . . . Studies show that human beings rely on automated decisions even when they suspect system malfunction.[] The impulse to follow a computer's recommendation flows from human "automation bias"--the "use of automation as a heuristic replacement for vigilant information seeking and processing."[] Automation bias effectively turns a computer program's suggested answer into a trusted final decision.[]

Under the influence of automation bias, workers will likely adopt a computer's suggested eligibility determinations . . . . In this respect, little meaningful difference exists between a mixed system and its fully automated counterpart.

Automation bias may become increasingly acute in the twenty-first century as our regulatory rules become increasingly intricate. As a general matter, those who view themselves simply as data processors will lose their motivation to learn the rules applied by computers.[] This may be especially true where human operators believe that an automated system is better equipped than they are to master a wide swath of complicated rules. Over time, human operators may lose the skills that would allow them to check a computer's recommendations. As the public's demand for government services grows, and as policy becomes more complicated, human operators may be increasingly forced to trust automated systems.[]

Danielle Keats Citron, Technological Due Process, 85 WASH. U. L. REV. 1249, 1271-72 (2008) (emphasis added).

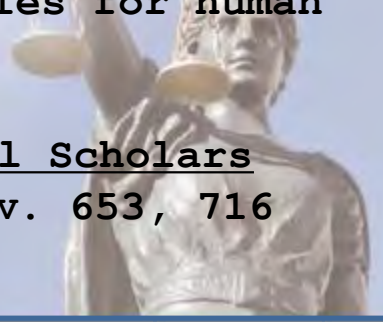
Automation Bias等/HITL

> Automation bias occurs when humans ascribe excessive value to automated decisions or predictions, often ignoring contradictory information. This in turn limits the extent to which having a "human in the loop" can actually improve algorithmic decisions. . . . .

Alice Xiang, Reconciling Legal and Thechnical Approarchs to Algorithmic Bias, 88 TENN. L. REV. 649, at 661 n.47 (2021).

> Many think that the best way to ensure fairness or justice is to inject a human into the decision-making process, perhaps with the veto power to override the inanimate counterpart. We are worried that if we simply thrust the human at the output end of the running model, there is very little she can do to root out bias. The human becomes a rubber stamp for the machine, providing nothing more than a cosmetic reason to lull ourselves into feeling better about the results. There might be better, more productive roles for human oversight elsewhere in the process.

David Lehr & Paul Ohm, Playing with the Data: What Legal Scholars Should Learn about Machine Learning, 51 U.C. DAVIS L. REV. 653, 716 (2017)(emphasis added).



Automation Bias等/HITL

- たとえ判断するのはヒト・裁判官であっても、そのヒトは自動装置を信頼してしまう偏見が働く問題の指摘は以下：

> The assessments are not understood by the judges who apply them, let alone the public at large. In the words of a New York Times article: "no one knows exactly how COMPAS works.[] Algorithmic risk assessments are akin to "an anonymous expert [that the defendant] cannot cross-examine."[] These concerns are exacerbated due to the weight people give to technology in our society today.[] Judges may be "likely to assume that quantitative methods are superior to ordinary verbal reasoning."[] This is a form of automation bias.[] Automation bias can easily change a technological suggestion into a "final, authoritative decision."[] The technology, then, can work to anchor decisions in technological certainty that is both improper and inappropriate.[]

Leah Wisser, Note, Pandora's Algorithmic Black Box: The Challenges of Using Algorithmic Risk Assessments in Sentencing, 56 Am. Crim. L. Rev. 1811、1824 (2019)(emphasis added). .

Automation Bias等/HITL

> Although inserting a "human in the loop" may appear to satisfy legal and philosophical principles, research into sociotechnical systems demonstrates that people and technologies often do not interact as expected . . . . / . . . . [T]he vast majority of research suggests that people cannot reliably perform any of the desired oversight functions. This first flaw leads to a second flaw: human oversight policies legitimize the use of flawed and unaccountable algorithms . . . . Thus, rather than protect against the potential harms of algorithmic decision-making . . . , human oversight policies create a regulatory loophole: it provides a false sense of security in adopting algorithms and enables vendors and agencies to foist [押し付ける] accountability for algorithmic harms onto lower-level human operators.

Ben Green, The Flaws of Policies Requiring Human Oversight of Government Algorithms, 45 COMPUT. L. SEC. REV. 1, 2 (2022) (emphasis added).

HITL機能不全に鑑みて

- HITLが機能しないことが明らかになった以上は、「人が最終判断しています」というallegationは説得力に欠ける。[See 次葉]
- アメリカ法学の視座からは、根拠・証拠に欠けるallegation(NOT argument)は、conclusiveであるとして認容されない。
- evidence-based policyの視座からも、更なる証明(後掲の説明責任や開示・透明性)、又は代替案等が必要。

採用AIに於けるHITLの機能不全

Rather than take for granted that people can effectively oversee algorithms, policymakers must empirically evaluate whether any proposed forms of human oversight are actually effective. Given the empirical evidence demonstrating the limits of human oversight, the default assumption should be that human oversight is likely to be ineffective, unless proven otherwise. The burden should therefore fall on [enities] proposing human oversight of algorithms to provide affirmative evidence that this mechanism actually improves outcomes and addresses concerns about algorithmic decision-making. This requirement remedies the first flaw of human oversight policies, in which human oversight is presented as a safeguard against algorithmic harms despite minimal empirical evidence that it actually provides reliable protections.

Green, supra, at 15(emphasis added).

採用AIに於けるHITLの欠点に対する対策方針

- そもそもAIには不得手な判断の場合には、AIによる機械的判断を避けるべき。
- AIが不得手な判断の場合とは、大局観や文脈を含めた解釈が必要な場合。
  - 前掲，履歴書を選別するAIの場合。
  - 前掲，ロボット兵器に関する指摘。
- ∴配慮が必要。

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。**

採用AIに於けるHITLの欠点に対する対策方針

表 1：法の二律背反

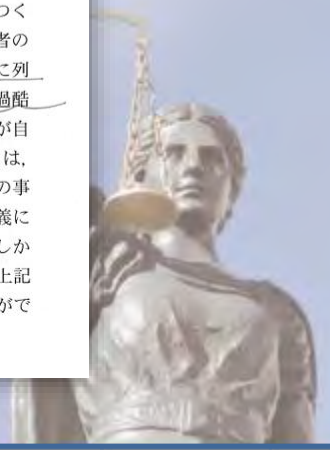
法治主義	人治主義
形式主義	現実主義
法	政治
コモン・ロー	エクイティ
法	慈悲
法	正義
規範	裁量
規範	模範
規範	根本原理
コモン・ロー上の準則	エクイティ上の法諺
本来的な規範（Per se rule）	合理性の規範
論理	政策
厳格	柔軟
正しい答え	良い答え
実定法	自然法
先例による判断	仲裁
裁判官	イスラーム教における裁判官，陪審
厳格責任	過失責任
契約の客観主義的解釈	契約の主観主義的解釈
客観性	主観性
非人間性	人間主義
原則主義的	結果志向的
正当性	必要性
権利	権力
制定法	判例法
制定法	憲法
逐語解釈的	非逐語解釈的
厳格な解釈	緩やかな解釈
文言	精神
裁判官による法の発見	裁判官による法の創造

合法性に関するイーグルトンの見解は『ヴェニスの商人』や『尺には尺

を』を誤解していることによります。この誤解は，一掃し得るものです。シャイロック，アンジェロ，イザベラ，ひいてはイーグルトンが法を体現化する者であり，ポーシャやウィーン公が法ではないものを体現化しているという理解は，間違っています。両者共に法という位置付けにおいて，前者の四人は，両極端の一方の端を表していて，後者の二人は，もう一方の端に近い位置にいるにすぎません。上記「表 1」の左側には，法学，哲学，心理学，制度に関する複数の異なる単語が列挙されています。それらは，法を抽象的なものとします。すなわち，実際に，法を執行し，紛争に判断を下す責任を負う人間から分離独立した概念を表しています。これら左側に列挙されている用語は，裁量と人的要因を最小化し「規範化された状態」や「法尊重主義」を最大化する要素を示唆しています。ここにおいては，専門職業意識，論理，厳格な規範，明確な区別，実定法，「解決困難な事件」（これは，通常の意味における難しい事件を意味するものではありません。理屈と心情とに折り合いが着かない厳しい結果に至る事件を意味しています）が強調されます。また，その事件における固有の状況，情状，紛争当事者の個性を排除することが強調されます。

裁判官が，より多くの裁量を有するようになれば，裁判官の行為を監視し，その腐敗，無能な仕事ぶりを発見することが困難になってくるのは確かです。もし，クローディオに対する死刑判決の取り消しをアンジェロが裁判官としての義務に違反するものと考えていたのならば，彼がイザベラに性交渉を強要することもなかったでしょう。しかし，この司法の腐敗を許す裁量の危険性と厳格な規範を文字通り解釈し，頑なに適用することから生じる費用は，相殺し合う関係にあります。そして，その費用は，実際に，高くつくものとなるでしょう。文明社会は，今まで純粋なかたちでの法尊重主義者の見解を受け入れてきませんでした。あらゆる文明社会は「表 1」の右側に列挙されている手段の一部，または，全部を用いて，厳格な法尊重主義の過酷さを和らげています。法とは，規範による統治の技術であり，それ自身が自動的に執行される機械装置ではありません。規範を厳格に執行することは，その趣旨を超えるものです（「順法闘争をする」とは，被用者が使用者の事業を中断させてしまう意味もあります）。過度の法尊重主義に対し，正義に関する過度に純粋な裁量を認める制度は，両者共に原始的な社会の中にしか見出せません。成熟した社会には，厳格な法と裁量が混在しています。上記の表に記載されている全ての事項を現代のアメリカ法の中に見出すことがで

ポズナー『法と文学（上）』前掲206～07頁.



採用AIに於けるHITLの欠点に対する対策方針

> Long concerned with the unfairness of formalistic application of legal rules, the principles and practices of equity allow otherwise unfair decisions to be adapted to individual circumstances.[]– Decisions in equity can permit courts to look beyond factors the law ordinarily considers, to think about fairness in a particular set of circumstances. Similarly, the subject of an automated decision should be able to explain why an algorithm's framing is not the full picture and to introduce individualizing, sometimes mitigating, factors an algorithm has not considered.

Kaminski, supra, at 1542(emphasis added).

採用AIに於けるHITLの欠点に対する対策方針

- そもそも採用活動が、個別具体的妥当性を審査する場合であることに鑑みると、そこでAIを利活用することは本来望ましくないはず。
- それでもAIを利活用したければ、実質的には機能しないHITLを単に形式的に取り入れるのではなく、FATの原則を遵守できる仕組みを取り入れるべき。

FAT

- Fairness : 公正/公平
- Accountability : 説明責任
- Transparency : 透明性

✓F:「人が最終判断しています」だけでは**公正**が保たれない。

✓A:「人が最終判断しています」だけでは**説明責任**を果たしていない。

✓T:「人が最終判断しています」だけでは**不透明**。

high stakesな場合

- キュウリの選別にAIを利活用する場合にまでも厳しいFATが求められる訳ではない。

> "[T]rustworthiness is relative to the stakes of the decision: decisions that involve higher stakes associated with erroneous predictions require a higher standard for validating algorithms"

Green, supra, at 13.

high stakesな場合

Table 1 – The appropriate roles for human and algorithmic decision-making, based on the need for discretion within the given decision and the algorithm's trustworthiness.

		Need for Discretion	
		Low	High
(Relative) Trustworthiness of Algorithm	Low	1) Primarily or solely human decision-making, with algorithms involved to the extent that rigorous research demonstrates benefits.	2) Solely human decision-making.
	High	3) Primarily or solely algorithmic decision-making.	4) Primarily or solely human decision-making, with algorithms involved to the extent that rigorous research demonstrates benefits.

Green, supra, at 13.

連邦雇用機会均等委員会委員の共著論文

ARTICLES

The Promise and The Peril: Artificial Intelligence and Employment Discrimination

KEITH E. SONDERLING, BRADFORD J. KELLEY & LANCE CASIMIR*

* The Honorable Keith E. Sonderling is a Commissioner on the U.S. Equal Employment Opportunity Commission ("EEOC"). Before joining the EEOC, he served as the Acting Administrator and Deputy Administrator in the U.S. Department of Labor's Wage and Hour Division ("WHD"). Bradford J. Kelley is Chief Counsel to Commissioner Sonderling. He previously served as a senior policy advisor in WHD. Lance Casimir is an Attorney Advisor to Commissioner Sonderling. The views and opinions set forth herein are the personal views or opinions of the authors and do not necessarily reflect views or opinions of the EEOC or any Commissioner.

Keith E. Sonderling, Bradford J. Kelley & Lance Casimir, The Promise and the Peril: Artificial Intelligence and Employment Discrimination, 77 U. MIAMI L. REV. 1 (2022).

U.S. Equal Employment Opportunity Commission

Languages

Search

About EEOC | Employees & Job Applicants | Employers / Small Business | Federal Sector | Contact Us

Home » Keith E. Sonderling, Commissioner

Keith E. Sonderling, Commissioner

th E. Sonderling was confirmed by the U.S. Senate, with a bipartisan vote, to be a mmissioner on the U.S. Equal Employment Opportunity Commission (EEOC) in 2020. til January of 2021, he served as the Commission's Vice-Chair. His term expires July of 4.

or to his confirmation to the EEOC, Commissioner Sonderling served as the Acting and puty Administrator of the Wage and Hour Division at the U.S. Department of Labor. ore joining the Department of Labor in 2017, Commissioner Sonderling practiced bor and Employment law in Florida. Commissioner Sonderling also serves as a fessional Lecturer in the Law at The George Washington University Law School, ching employment discrimination.

ce joining the EEOC, one of Commissioner Sonderling's highest priorities is ensuring t artificial intelligence and workplace technologies are designed and deployed sistent with long-standing civil rights laws. Commissioner Sonderling has published merous articles on the benefits and potential harms of using artificial intelligence-based technology in the workplace and speaks globally on these emerging issues.

Immediately before his confirmation to the EEOC, as Deputy and Acting Administrator of

U.S. Equal Employment Opportunity Commission, Keith E. Sonderling, Commissioner https://www.eeoc.gov/keith-e-sonderling-commissioner (last visited Aug. 28, 2023).

FATの具体例：連邦雇用機会均等委員達の2022年論文から

C. *Transparency and Explainability*

Transparency and explainability are two very important concepts that foster algorithmic reliability, trust, credibility, and a general understanding of AI systems.[438] Transparency promotes the visibility of processes, the accessibility of systems, and the reporting of meaningful information.[439] Explainability fosters trust in the process.[440] Neither is possible, though, if the user of the AI does not first understand the data on which it relies. And a lack of either or both can result in algorithmic systems that are difficult to control, monitor, correct, and defend.[441] This is the commonly cited "black box" issue.[442] In a similar vein, the absence of transparency, accountability, and understandability threatens to undermine any benefits offered by AI and machine learning technologies.

Employers should explore efforts to promote transparency and explainability surrounding their use of AI in employment decisions as the fruits of the technology and the underlying problems posed by innovation continue to develop.[443] To do so, they should aim to

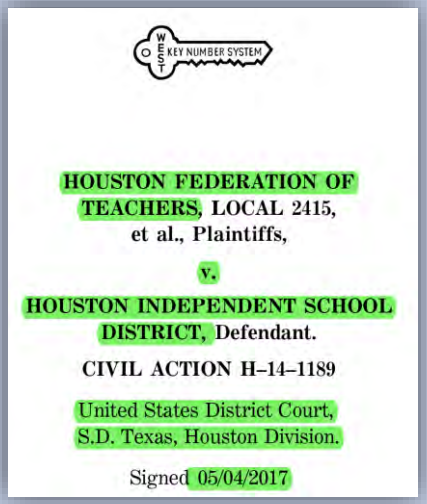
Id. at 77-78.

provide meaningful information appropriate to the context. Specifically, employers should inform applicants about what data is used and how it is used in the hiring process.[444]

Transparency and explainability also empower those affected by an AI system to understand the outcome. They enable those adversely affected by an AI system to challenge its outcome based on plain and easy-to-understand information on the factors and the logic that served as the basis for the prediction, recommendation, or decision. In the same vein, employers should define and assign roles to their HR professionals that ensures HR and management understand their responsibilities in relation to the company's use of AI in employment decision-making.[445] Actively pursuing transparent and explainable applications of AI fosters trust among employers seeking to prudently and lawfully implement this technology as well as job applicants and employees who are subject to it.[446]

Transparency and explainability require open, detailed, and clear communication. Employers should provide applicants and employees with robust notice that explains, at a minimum, what technologies are being used, for what purpose, how they work, the specific information that is collected, to whom it will be disclosed, how it will be used, and how long it will be retained.[447] Employers should also explain how access to any information collected will be controlled and any other safeguards for the information. Notices for AI technology involved with employee performance should also include these details, while also explaining anticipated benefits to the employees such as ways it will enhance their performance or make their work easier to accomplish. Ultimately, effective notice will allow candidates and employees to make informed choices about whether to participate in the activity, and whether to seek a reasonable accommodation or alternative arrangement.[448]

营业秘密 v. 透明性



WEST KEY NUMBER SYSTEM

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).

Bibliography

Bibliography

- Ifeoma Ajunwa, The Paradox of Automation as Anti-Bias Intervention, 41 Cardozo L. Rev. 1671 (2020).
- Ifoema Ajunwa, An Auditing Imperative for Automated Hiring Systems, 37 Harv. J.L. & Tech. 621 (2021).
- Loi Andrews & Hannah Bucher, Automatic Discrimination: AI Hiring Practices and Gender Inequality, 44 Cardozo L. Rev. 145 (2022).
- 「ブレードランナー2049」(Warner Bros. 2017).
- Bryan Casey, Robot Ipsa Loquitur, 108 Geo. L.J. 225 (2019).
- Danielle Keats Citron, Technological Due Process, 85 Wash. U. L. Rev. 1249 (2008).
- Malika Dargan, Comment, Model Act for Algorithmic Models: A Regulatory Solution for AI Used in Hiring Decisions, 13 Houston L. Rev. Onine 50 (2023).
- EU, AI Act.
- Lindsey Fuchs, Hired by Machine: Can a New York City Law Enforce Algorithmic Fairness in Hiring Practices?, 28 FORDHAM J. CORP. & FIN. L. 185 (2023).
- EU, GDPR Art.22.

Bibliography

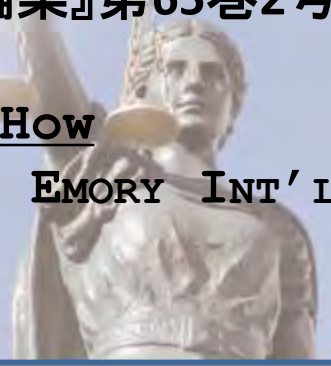
- Ben Green, The Flaws of Policies Requiring Human Oversight of Government Algorithms, 45 COMPUT. L. SEC. REV. 1 (2022).
- Courtney Hinkle, The Modern Lie Detector: AI-Powered Affect Screening and the Employee Polygraph Protection Act (EPPA), 109 GEO. L.J. 1201 (2021).
- 平野晋『ロボット法（増補版）』84頁（弘文堂, 2019年）
- Houston Fed'n of Tchrs., Loc. 2415 v. Houston Indep. Sch. Dist., 251 F. Supp. 3d 1168 (S.D. Tex. 2017).
- THE ILLINOIS ARTIFICIAL INTELLIGENCE VIDEO INTERVIEW ACT.
- フランツ・カフカ『審判』.
- Margot E. Kaminski, Binary Governance: Lessons from the GDPR's Approach to Algorithmic Accountability, 92 S. CAL. L. REV. 1529 (2019).
- Brittany Kammerer, Hired by a Robot: The Legal Implications of Artificial Intelligence Video Interviews and Advocating for Greater Protection of Job Applicants, 107 IOWA L. REV. 817 (2022).

Bibliography

- Pauline T. Kim, Data-Driven Discrimination at Work, 58 Wm. Mary L. Rev. 857 (2017).
- 厚生労働省「公正な採用選考の基準」(抄)
- 厚生労働省「募集・求人業務取扱要領」(抄)
- 厚生労働省「労働者に対する性別を理由とする差別の禁止等に関する規定に定める事項に関し、事業主が適切に対処するための指針」(抄)
- 厚生労働省 労働政策審議会労働政策基本部会「報告書～働く人がAI等の新技術を主体的に活かし、豊かなsh将来を実現するために～」(抄)令和元年[2019年]6月.
- David Lehr & Paul Ohm, Playing with the Data: What Legal Scholars Should Learn about Machine Learning, 51 U.C. Davis L. Rev. 653 (2017).
- Maryland Code, Labor & Employment § 3-717.
- New York City's Automated Employment Decision Tool Law (AEDT Law, Local Law Int. No. 1894-A.
- OECD, AI Principles 1.3.
- リチャード・A.ポズナー『法と文学 第3版』(平野晋監訳, 坂本真樹＆神場幸一訳, 2011年).
- 三部裕幸 「資料2: EUのAI規制法案の概要と、欧州での反応」 8-9頁 in 総務省・AIネットワーク社会推進会議(第20回)＋AIガバナンス検討会(第16回)合同開催(2022年2月8日).

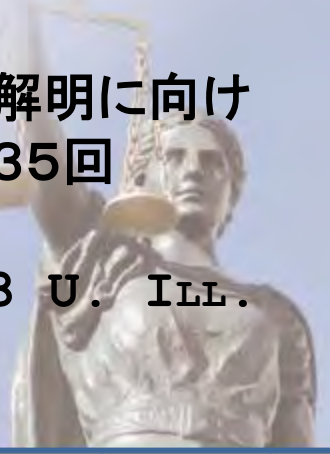
Bibliography

- David D. Savage & Richard Bales, Video Games in Job Interviews: Using Algorithms to Minimize Discrimination and Unconscious Bias, 32 A.B.A. J. LAB. & EMP. L. 211, 221 (2017).
- Keith E. Sonderling, Bradford J. Kelley & Lance Casimir, The Promise and the Peril: Artificial Intelligence and Employment Discrimination, 77 U. MIAMI L. REV. 1 (2022).
- Jonathan L. Sulds , 1 New York Employment Law §9.02 in New York Employment Law, Second Edition.
- Paul J. Sweeney, NYC Will Be Watching: Is Your Hiring Program Compliant?, 29 No. 6 N.Y. EMP. L. LETTER 1, June 2022.
- 竹地潔「人工知能による選別と翻弄される労働者」『富山経済論集』第65巻2号91,95頁(2019年12月).
- Charles P. IV Trumbull, Autonomous Weapons: How Existing Law Can Regulate Future Weapons, 34 EMORY INT'L L. REV. 533 (2020).



Bibliography

- W. Bradley Wendel, Technological Solutions to Human Error and How They Can Kill You: Understanding the Boeing 737 Max Products Liability Litigation, 84 J. Air L. & Com. 379 (2019).
- Leah Wisser, Note, Pandora's Algorithmic Black Box: The Challenges of Using Algorithmic Risk Assessments in Sentencing, 56 Am. Crim. L. Rev. 1811 (2019).
- Alice Xiang, Reconciling Legal and Thechnical Approarchs to Algorithmic Bias, 88 Tenn. L. Rev. 649 (2021).
- 山下芳樹 et al.,「AI人事採用における納得阻害要因の解明に向けたサーベイ実験」in 人『工知能学会全国大会論文集』第35回 (2021年).
- Tal Z. Zarsky, Transparent Prediction, 2013 U. Ill. L. Rev. 1503 (2013).



Thank you !!! ;-)

INFORMATION TECHNOLOGY
& LAW
ICHIGAYA TAMACHI LINK

Attachments

FAT実装例

Fairness: AIの誤謬を無くすこと。

- algorithm創作時に於ける公正と、事後にも定期的な監査(audit)による修正の継続。
  - ➢ **欠点を発見して修正する為には監査が必要。** E.g., Sonderling at 80.
  - ➢ 自主的内部監査だけでは不十分なので、中立的機関による外部監査も必要。 Ajunwa at 660.
- HITLを使う場合--meaningful human oversight--には、以下(例示)が必要:
  - ➢ **監視員の訓練**
    - **COMPAS判決でも裁判官の教育の必要性を指摘。** Wisser, supra, at 1827-28.
  - ➢ 監視員の疲労対策
  - ➢ 事前にhuman-algorithm collaborationのexperimental evaluationの実施。
- **ベンダに対しても公正さを等を要_要求。** Sonderling at 62-63.

Accountability: AI利用についての説明責任を果たすこと。

- 事前(ex ante)に説明すること、真の選択権(opt out)を付与すること。
- 事後(ex post)に説明すること、異議申し立てに応じること。 E.g., OECD AI原則1.3
- 立法例:Ill, MD, NYC, EU_AI-Act, [GDPR]
- GDPRは判断が「機械によってのみ」"solely"な場合にしか適用されないことが欠点。
- OECD AI原則1.3も、不利益を被った場合にのみ適用されると読める欠点(「総合判断だった」や「最後は人が判断している」等々の言い訳で逃れられるloop-hole あり)

Transparency: 開示すること。

- 分かり易い説明が必要。
- トレイニング・データ、及び開発過程[関係者への]開示。
- 場合によってはソースコードもIn camera reviewのような仕組みで開示要。
  - ➢ Houston Teachers事件; Wisser at 1813; Id. at 1830 (公的機関等へのソースコード開示を提案).
- 中立的存在による監査結果の開示。See Andrews & Bucher at 189-90; Ajunwa at 660 (**SOX法の例も鑑みて監査人の独立性も重視**).

州制定法

- THE ILLINOIS ARTIFICIAL INTELLIGENCE VIDEO INTERVIEW ACT (2020年発効)
  - 要:事前説明と同意,守秘義務, 要求から30日以内に複製物を破棄.
- MARYLAND CODE, LABOR & EMPLOYMENT § 3-717 (2020年発効)
  - 権利放棄書のない限り面接時の顔認識使用禁止.

Andrews & Bucher, supra, at 195; Kammerer, supra, at 834-36.

NYC条例

New York City's Automated Employment Decision Tool Law (AEDT Law又は NYC AI Hiring Ordinance), Local Law Int. No. 1894-A. (2023年発効)

- AEDT(次葉*1)は以下を遵守しない限り使用禁止:
  - ✓ ツール[が使われる事実]と、ツールが使われる仕事の資格と特徴を、応募者に対して10事業日よりも前に通知すること;
  - ✓ 応募者に[AEDT使用以外の]代替手段を認めること;
  - ✓ 収集したデータに関する情報の提供; 及び
  - ✓ 使用の一年前以内に独立した監査人による偏見監査(bias audits)に服すこと、かつ最新のその監査結果の要約等をウエブ上で公表すること(次葉*2)

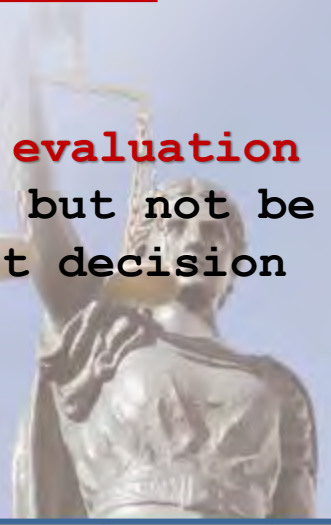
Paul J. Sweeney, NYC Will Be Watching: Is Your Hiring Program Compliant?, 29 No. 6 N.Y. Emp. L. Letter 1, June 2022; Malika Dargan, Comment, Model Act for Algorithmic Models: A Regulatory Solution for AI Used in Hiring Decisions, 13 Houston L. Rev. Onine 50, 55-56 (2023); Fuchs, supra at 212; Jonathan L. Sulds , 1 New York Employment Law §9.02 in New York Employment Law, Second Edition.

NYC条例

- (*1) AEDT means "a computational process, derived from machine learning, statistical modeling, data analytics, or artificial intelligence, that issues simplified output, including a score, classification, or recommendation, that is used to substantially (*i) assist or replace discretionary decision making for making employment decisions that impact natural persons." N.Y.C. Admin. Code §§20-870 (emphasis added).
  - (*i) 以下のような批判がある: "[B]y limiting its reach to automated tools that "substantially assist or replace discretionary decision-making" it would allow employers to avoid audits by arguing that they do not give "substantial weight" to AI results."

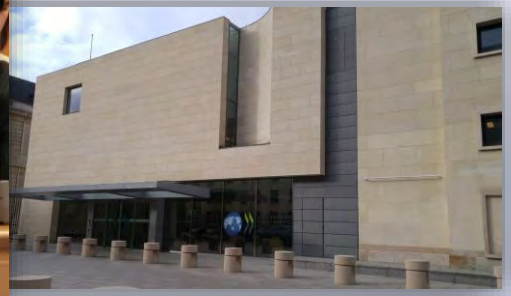
- (*2) A "bias audit" is defined as "an impartial evaluation by an independent auditor," which "shall include but not be limited to the testing of an automated employment decision tool to assess the tool's disparate impact on persons . . . . N.Y.C. Admin. Code §§20-870.

(emphasis added).



OECD AI Principles

OECD・AI専門家会合：
AI expert ***G***roup at the ***O***ECD
(**AIGO**：エイゴ／エイ・アイ・ゴー)



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,*
- *to make stakeholders aware of their interactions with AI systems, including in the workplace,*
- *to enable those affected by an AI system to understand the outcome, and,*
- *to enable those adversely affected by an AI system to challenge its outcome based on plain and easy-to-understand information on the factors, and the logic that served as the basis for the prediction, recommendation or decision.*

EU_AI Act

- 雇用(含，採用審査等々)に於けるAI利用は、ハイ・リスクなAIシステムに該当。
- 様々な規制の対象。

三部裕幸「資料2： EUのAI規制法案の概要と、欧州での反応」8-9頁 in 総務省・AIネットワーク社会推進会議(第20回)＋AIガバナンス検討会(第16回)合同開催(2022年2月8日)
https://www.soumu.go.jp/main_sosiki/kenkyu/ai_network/02iicp01_04000285.html (last visited Sep. 4, 2023) .

- 本発表の関心は特に、〈人による監視〉(human oversight)の要件 → 次葉

EU_AI Act

ANNEX III

High-risk AI systems pursuant to Article 6(2) are the AI systems listed in any of the following areas:

. . . .

4. Employment, workers management and access to self-employment:

(a) AI systems intended to be used for recruitment or selection of natural persons, notably for advertising vacancies, screening or filtering applications, evaluating candidates in the course of interviews or tests;

(b) AI intended to be used for making decisions on promotion and termination of work-related contractual relationships, for task allocation and for monitoring and evaluating performance and behavior of persons in such relationships.

(emphasis added)

EU_AI Act

Article 14
Human oversight

1. High-risk AI systems shall be designed and developed in such a way, including with appropriate human-machine interface tools, that they can be effectively overseen by natural persons during the period in which the AI system is in use.
2. Human oversight shall aim at preventing or minimising the risks to health, safety or fundamental rights that may emerge when a high-risk AI system is used in accordance with its intended purpose or under conditions of reasonably foreseeable misuse, in particular when such risks persist notwithstanding the application of other requirements set out in this Chapter.
3. Human oversight shall be ensured through either one or all of the following measures:
   (a) identified and built, when technically feasible, into the high-risk AI system by the provider before it is placed on the market or put into service;
   (b) identified by the provider before placing the high-risk AI system on the market or putting it into service and that are appropriate to be implemented by the user.

(emphasis added)

EU_AI Act

4. The measures referred to in paragraph 3 shall enable the individuals to whom human oversight is assigned to do the following, as appropriate to the circumstances:

(a) fully understand the capacities and limitations of the high-risk AI system and be able to duly monitor its operation, so that signs of anomalies, dysfunctions and unexpected performance can be detected and addressed as soon as possible;

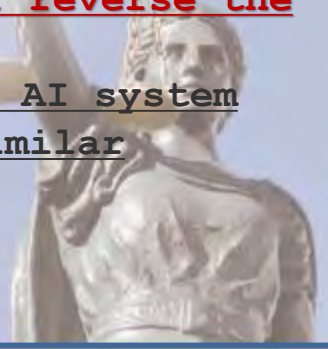
(b) remain aware of the possible tendency of automatically relying or over-relying on the output produced by a high-risk AI system ('automation bias'), in particular for high-risk AI systems used to provide information or recommendations for decisions to be taken by natural persons;

(c) be able to correctly interpret the high-risk AI system's output, taking into account in particular the characteristics of the system and the interpretation tools and methods available;

(d) be able to decide, in any particular situation, not to use the high-risk AI system or otherwise disregard, override or reverse the output of the high-risk AI system;

(e) be able to intervene on the operation of the high-risk AI system or interrupt the system through a "stop" button or a similar procedure.

(emphasis added)



EU_AI Act

5. For high-risk AI systems referred to in point 1(a) of Annex III, the measures referred to in paragraph 3 shall be such as to ensure that, in addition, **no action or decision is taken** by the user on the basis of the **identification resulting from the system unless** this has been **verified and confirmed by** at least **two natural persons**.

_____.

Annex III, point 1(a):

1. **Biometric identification and categorisation of natural persons**:

(a) **AI systems intended to be used for the 'real-time' and 'post' remote biometric identification of natural persons**,

(emphasis added)

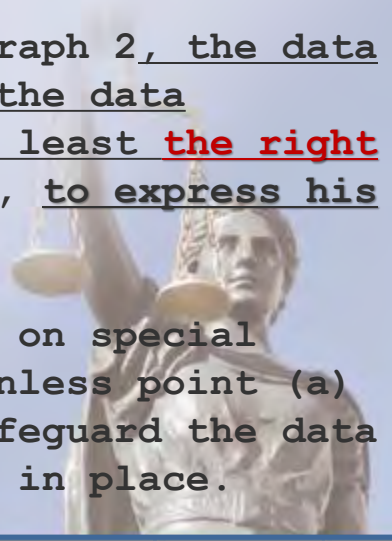
GDPR Art. 22

1. The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.

2. Paragraph 1 shall not apply if the decision:
   (a) is necessary for entering into, or performance of, a contract between the data subject and a data controller;
   (b) is authorised by Union or Member State law to which the controller is subject and which also lays down suitable measures to safeguard the data subject's rights and freedoms and legitimate interests; or
   (c) is based on the data subject's explicit consent.

3. In the cases referred to in points (a) and (c) of paragraph 2, the data controller shall implement suitable measures to safeguard the data subject's rights and freedoms and legitimate interests, at least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision.

4. Decisions referred to in paragraph 2 shall not be based on special categories of personal data referred to in **Article 9**(1), unless point (a) or (g) of **Article 9**(2) applies and suitable measures to safeguard the data subject's rights and freedoms and legitimate interests are in place.



厚生労働省 募集・求人業務取扱要領（抄）

（職業安定法第五条の五、指針第五）

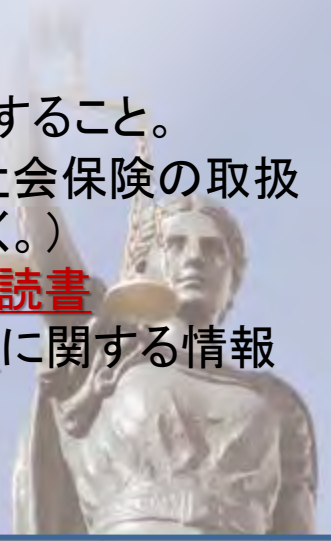
（ロ） 募集主及び募集受託者は、その業務の目的の達成に必要な範囲内で当該目的を明らかにして労働者の個人情報を収集することとし、次に掲げる個人情報を収集してはならない。ただし、特別な職業上の必要性が存在することその他業務の目的の達成に不可欠であって収集目的を示して本人から収集する場合はこの限りではない。

① 人種、民族、社会的身分、門地、本籍、出生地その他社会的差別の原因となるおそれのある事項
② 思想及び信条
③ 労働組合の加入状況

①から③までについては、具体的には、例えば次に掲げる事項が該当すること。

（ｉ） ① 関係家族の職業、収入、本人の資産等の情報（税金、社会保険の取扱い等労務管理を適切に実施するために必要なものを除く。）
（ii） ② 関係人生観、生活信条、支持政党、購読新聞・雑誌、愛読書
（iii） ③ 関係労働運動、学生運動、消費者運動その他社会運動に関する情報

(emphasis added).



厚生労働省 労働者に対する性別を理由とする差別の禁止等に関する規定に定める事項に関し、事業主が適切に対処するための指針（抄）

第３　間接差別（法第７条関係）

１　雇用の分野における性別に関する間接差別

⑴　雇用の分野における性別に関する間接差別とは、①性別以外の事由を要件とする措置であって、②他の性の構成員と比較して、一方の性の構成員に相当程度の不利益を与えるものを、③合理的な理由がないときに講ずることをいう。

⑵　⑴の①の「性別以外の事由を要件とする措置」とは、男性、女性という性別に基づく措置ではなく、外見上は性中立的な規定、基準、慣行等（以下第3において「基準等」という。）に基づく措置をいうものである。

⑴の②の「他の性の構成員と比較して、一方の性の構成員に相当程度の不利益を与えるもの」とは、当該基準等を満たすことができる者の比率が男女で相当程度異なるものをいう。

⑴の③の「合理的な理由」とは、具体的には、当該措置の対象となる業務の性質に照らして当該措置の実施が当該業務の遂行上特に必要である場合、事業の運営の状況に照らして当該措置の実施が雇用管理上特に必要であること等をいうものである。

⑶　法第７条は、募集、採用、配置、昇進、降格、教育訓練、福利厚生、職種及び雇用形態の変更、退職の勧奨、定年、解雇並びに労働契約の更新に関する措置であって、⑴の①及び②に該当するものを厚生労働省令で定め、⑴の③の合理的な理由がある場合でなければ、これを講じてはならないこととするものである。厚生労働省令で定めている措置は、具体的には、次のとおりである。（均等則第２条各号に掲げる措置）

(emphasis added).

厚生労働省 労働者に対する性別を理由とする差別の禁止等に関する規定に定める事項に関し、事業主が適切に対処するための指針（抄）

イ　労働者の募集又は採用に当たって、労働者の身長、体重又は体力を要件とすること（均等則第2条第1号関係）。

ロ　労働者の募集若しくは採用、昇進又は職種の変更に当たって、転居を伴う転勤に応じることができることを要件とすること（均等則第2条第2号関係）。

ハ　労働者の昇進に当たり、転勤の経験があることを要件とすること（均等則第2条第3号関係）。

2　労働者の募集又は採用に当たって、労働者の身長、体重又は体力を要件とすること（法第7条・均等則第2条第1号関係）

(1)　均等則第2条第1号の「労働者の募集又は採用に関する措置であつて、労働者の身長、体重又は体力に関する事由を要件とするもの」とは、募集又は採用に当たって、身長若しくは体重が一定以上若しくは一定以下であること又は一定以上の筋力や運動能力があることなど一定以上の体力を有すること（以下「身長・体重・体力要件」という。）を選考基準とするすべての場合をいい、例えば、次に掲げるものが該当する。

（身長・体重・体力要件を選考基準としていると認められる例）

厚生労働省 労働者に対する性別を理由とする差別の禁止等に関する規定に定める事項に関し、事業主が適切に対処するための指針（抄）

イ　募集又は採用に当たって、身長・体重・体力要件を満たしている者のみを対象とすること。

ロ　複数ある採用の基準の中に、身長・体重・体力要件が含まれていること。

ハ　身長・体重・体力要件を満たしている者については、採用選考において平均的な評価がなされている場合に採用するが、身長・体重・体力要件を満たしていない者については、特に優秀という評価がなされている場合にのみその対象とすること。

⑵　合理的な理由の有無については、個別具体的な事案ごとに、総合的に判断が行われるものであるが、合理的な理由がない場合としては、例えば、次のようなものが考えられる。（合理的な理由がないと認められる例）

イ　荷物を運搬する業務を内容とする職務について、当該業務を行うために必要な筋力より強い筋力があることを要件とする場合

ロ　荷物を運搬する業務を内容とする職務ではあるが、運搬等するための設備、機械等が導入されており、通常の作業において筋力を要さない場合に、一定以上の筋力があることを要件とする場合

ハ　単なる受付、出入者のチェックのみを行う等防犯を本来の目的としていない警備員の職務について、身長又は体重が一定以上であることを要件とする場合

３　労働者の募集若しくは採用、昇進又は職種の変更に当たって、転居を伴う転勤に応じることができることを要件とすること（法第７条・均等則第２条第２号関係）

（１）．．．．

．．．．

厚生労働省「公正な採用選考の基本」

厚生労働省「公正な採用選考の基本」
https://www.mhlw.go.jp/www2/topics/topics/saiyo/saiyo1.htm (last visited Sept. 2, 2023).