

厚生労働科学研究費補助金
(労働安全衛生総合研究事業)
総括研究報告書

法学的視点からみた AI の活用に伴う現場管理上の課題と 現状の労働安全衛生法の法令上の課題の検討のための研究

研究統括者

近畿大学法学部教授

東北大学大学院医学系研究科客員教授（産業医学分野）

日本産業保健法学会代表理事

三柴 丈典

要約

1 産業における AI 導入が安全衛生にもたらすメリット

AI（Artificial Intelligence：人工知能）の職場利用は、一部の先進企業に限られず拡大し、化学、建設、製造、農業、輸送・物流、廃棄物・リサイクル、公共サービス等で活用例が確認された（英国）。

英国では実際のユースケースを調査した上で、以下の 4 類型に整理した。

- 1) 保全（Maintenance：危険箇所の点検の遠隔化、予知保全〔設備故障の兆候把握〕）、
- 2) 安全衛生マネジメント（災害データ解析、リスクアセスメント文書や教材の作成の支援等）、
- 3) 設備・プロセス制御（自律走行機器の自動運転や人検知による制御等、プラントの異常検知と自動調整等）、
- 4) 職業的モニタリング（Occupational Monitoring：PPE〔個人用保護具〕の装着確認、危険が接近した場合の警告、ウェアラブルによる熱ストレスや有害要因ばく露の監視等）、等。

ドイツ・フランスでは、災害データの解析、危険接近の検知、健康・負荷の把握、危険

作業代替等が、安全衛生上有効となり得る用途として関係機関文書等で列挙されている（ただし、現段階での実装は限定的）。

著者所属に米国の研究機関を含む文献や米国の専門団体が公表した声明のレビューからは、危険作業の代替（危険・重負荷作業を AI 装備の機械が担うこと）が進む場合、建設、製造、鉱業、輸送等の高リスク産業において、危険職種に就きやすい移民、マイノリティ等低所得者層の被害が減り、健康格差（Health Inequity：集団間の健康上の不公平）が緩和され得るとの推論の存在が示されている。

安全工学からは、上記の用途が災害抑制に資し得る因果経路（機序）が示される一方、誤作動が重大災害に直結し得る用途（制御直結など）ほど、冗長化、フェイルセーフ、検証等が必要になると示唆された。

産業医学からは、AI を用いた健康管理サービスが利用者の健康行動等を高め得る、予測 AI（健診等のデータの集約分析）により産業保健職等の業務負担を軽減し得る、生成 AI チャットボットが一次的な相談やセルフケア促進等の機能を果たし得る、表情、ウェアラブル等から従来の方法では拾いにくい疲労・ストレス増大や抑うつ傾向の兆候を拾える可能性等が示唆される一方、良質な実証研究（RCT 等）や費用対効果の検証が乏しく、導入判断には、業者による情報開示と継続評価が必要とされた。

2 産業における AI 導入が安全衛生にもたらすデメリット

AI が制御・判断に深く関与するほど、次のようなリスクが生じ得ることが指摘されている。もっとも、列挙される多くは、「こうなり得る」という危険の類型整理であり、その全てに実証結果を伴うわけではない。

1) deskilling（技能劣化：有事の介入技能・状況認識の低下）

AI が平常時の判断や制御を担い、人が異常時のみ介入する形が定着すると、非常時に必要な人の技能が衰え得る（英国）。

2) ブラックボックス性（判断過程が見えにくいこと）に伴う「無視」と「過信」

AI の判断過程を理解できないためにその判断を無視するリスクと、逆に過信するリスクが同時に生じ得る（英国）。

3) アラート疲労（warning/alarm fatigue：警告過多による重要警報の埋没）

警告が過多になると重要な警報が目立たなくなり、無視され得る（英国・安全工学）。

4) 自動化バイアス（Automation bias）

AI 等の提示を過信し、人が確認せずに従ったりする心理傾向を意味し、見逃し放置等に繋がり得る（安全工学）。必ずしも AI のブラックボックス性を前提とはしない。

必ずしもブラックボックス性を前提とせず、自動化バイアス

5) 学習データの偏りによるエッジケース脆弱性（稀だが重大なケースに対応する力）

データの偏り（網羅性の欠如）が例外的な重大事故への対応力の弱さに繋がり得る（英国・安全工学）。

6) モード錯誤 (Mode Confusion：自動モード誤認)

自動化の状態(どのモードで動いているか)を人が誤認し、誤操作等につながり得る(安全工学)。例えば、作業者が機械が手動状態にあると誤認して近づいたら自律走行状態にあり、接触災害に至る等。

7) 生成 AI の誤情報作成 (Hallucination)

生成 AI が誤情報をそれらしく生成し、誤判断を誘発し得る(安全工学・産業医学)。例えば、生成 AI が誤って安全装置の停止を提示し、作業者がそれに沿って誤操作し、災害に至る等。

8) ドリフト (Model Drift：環境変化による性能劣化)、偽陽性／偽陰性 (False Positive/Negative)

ドリフトとは、環境変化により AI が学習したデータと運用環境の間に格差が生じ、精度が劣化し得ること。偽陽性とは、基準に該当しないものを該当すると示すこと(誤警報など)、偽陰性とは、基準に該当するものを見逃すこと(安全工学・産業医学)。

9) CPS の誤判断

CPS (Cyber-Physical System：サイバー世界と物理的世界〔現実〕が連動するシステム)において、サイバー世界の AI がロボットや AGV (Automated Guided Vehicle：無人搬送車)等の挙動に関与する場合、その誤作動が挟まれ・衝突等の重大災害に直結し得る(安全工学)。生成 AI の Hallucination の場合と異なり、一般に人が介入し難い。

10) アルゴリズム管理 (Algorithmic Management) の安全衛生リスク

アルゴリズム管理 (Algorithmic Management) では、評価指標が生産性に偏ると、過密ノルマや休憩削減を誘発し、疲労蓄積、注意力の低下、安全手順の省略等を介して災害確率を押し上げ得る(安全工学)。

また、監視強化は心理社会的リスク(監視感による不安、孤立、裁量喪失、雇用不安等)を増幅し、ヒヤリハット(災害未満の出来事)の申告を萎縮させ得る(英国・心理・安全工学)。

11) サイバー攻撃

生成 AI フィッシング(偽指示。AI を使って上司の声を創作し、誤った指示を出す等)、敵対的攻撃 (Adversarial Example：入力細工による誤認の誘発。既にデータ学習済みの AI の運用時に、機械のセンサーに光をあてるなどして誤認させ、本来停止すべき場面で停止しないように誘導する等)、データ汚染 (Poisoning：学習データそのものの改ざん誤情報の混入による操作等。人が写っている写真に小さな模様(バックドア：隠し扉)を付けて、機械にそれがあれば安全と学習させ、普段は正常に動く機械に、操作者の思う時点で異常行動させる等)が AI の正常な機能を阻害し、その制御を受ける機械の誤作動等により、人間の安全性を脅かす(安全工学)。

3 メリット最大化・デメリット最小化のための物的・人的要件

以下はいずれも「実装済みの事実」ではなく、仮説（理論値）である。

3. 1 物的要件（技術・運用設計）

英国（HSE）および安全工学からは、AI は誤る前提で管理すべきとの前提で、以下の要件が示された。

1) サンプルング検証（スモールステップでの導入）、重大な判断における人の最終承認（human review）、AI を通じた情報漏れを防ぐためのセキュリティ対策、異常時の停止など安全寄りの設計を意味するフェイルセーフ（Fail-safe）、異種の組み合わせ（画像 AI と赤外線等）を含めた冗長保護（redundant protection：故障が生じた際に別回路が機能する仕組み）、ドリフト等に備えるための導入後の性能監視の組み合わせ（英国・安全工学）。

2) ISO 12100:2010（機械安全）に沿った階層制御（いわゆるスリーステップアプローチ：(i)本質的安全設計（危険が生まれにくい設計）→(ii)防護（ガード・停止機構等）→(iii)取扱い手順の教育等の人的措置）（安全工学）。

3) 学習型 AI を安全関連機能（停止・遮断等）に直接組み込むことの回避（機能安全 IEC 61508:2010〔SIL：Safety Integrity Level＝安全度水準〕の観点から証明困難が生じ得るため）（安全工学）。

4) ISO/IEC TR 5469:2024（AI と機能安全）が整理する AI 活用の 3 類型（①安全機能内部に組み込み〔AI が安全性の判断自体を行う〕、②AI 以外の安全機能で機械の安全を守る〔例えば自律走行機械で走行の最適化は AI が行っても、停止は光電センサ等の独立の安全回路が担う〕、③安全関連機能の設計・点検支援に AI を使う〔危険源の分析や手順書の作成等に AI を活用する〕）につき、いずれでも V&V（Verification/Validation：検証〔設計した通りに作られているか〕、・妥当性確認〔実際の利用条件下で目的を達成できるか〕）を行うこと（安全工学）。

5) 人が立ち入らない搬送専用通路でのみ AI 実装の自律走行機械を使用するなど、ODD（Operational Design Domain：運用条件の限定）を図ること（安全工学）。

6) 人検知 AI が人の接近で警報を出す場合に、距離が 2m 以内の場合や人と車両が同一レーンにいる場合にのみに警報を制限すること（簡易ルールによる門番機能の設置）により、AI 警報による疲労を抑えること（安全工学）。

7) ハード側での安全限界（速度上限、侵入禁止区域、最大負荷、最大作業時間等）を設定し、AI に誤判断があっても越えられないようにすること（安全工学）。

8) AI が関与する運用について、事故時の原因究明、導入後の性能監視や変更管理のため、トレーサビリティ（Traceability：追跡可能性）を確保するためのログ設計（入力データ、出力、警報履歴、モデル／設定のバージョン、更新履歴等を記録し、後から検証できるようにすること）（安全工学）。

9) AI のモデル更新や閾値変更などによる誤検知等の防止のための変更管理（Change Management）の社内ルール化（安全工学）。

10) 旧来の SOP（Standard Operating Procedure：標準作業手順）や緊急対応手順を AI

対応版へ改訂すること（停止権限の所在、連絡体制、ログの保存等）（安全工学）。

11) 災害・ヒヤリハット報告書に「AI 関与の有無」を記載し、関与時には必要なログ（入力データ、出力、警報履歴、モデル／設定のバージョン、更新履歴等の記録）を確実に入手できる手順を作成すること（安全工学）。

12) AIQM（機械学習品質マネジメント指針）や ISO/IEC 42001:2023（AIMS：AI マネジメントシステム）等を参照し、AI 利用について、組織として PDCA サイクルを回すこと（安全工学）。

13) AI の機能評価の KPI（Key Performance Indicator：うまくいっているかの測定指標）に安全・健康を組み込むと共に、KRI（Key Risk Indicator：危なくなっていないかの測定指標）で誤検知率（危険でないものを危険と判定する割合）、見逃し率（危険なものを安全と判定する割合）、拒否率（人が AI の提案を拒否した割合）、ヒヤリハット件数（AI の判断によって事故未満の出来事が生じた件数）、負荷指標（疲労やストレスなどの指標）等を継続的に監視すること（安全工学）。

3. 2 人的要件（教育・権限・文化）

1) AI リテラシー（特に AI の限界や誤り方の理解）、自動化バイアスの自覚、異常時の切替方法の教育訓練（安全工学）。

2) 人間によるチェック、特に現場による AI への異議（停止、手動切替等の例外ルール）の周知徹底（安全工学・心理）。

3) 導入プロセスへの労働者参加、定期的な振り返り（AI 導入後、一定周期で実績データと現場の声を点検し、設定や利用方法等を見直すプロセス）、心理的安全性（労働者が AI の関与について安心して異論を言える条件）の確保（心理・安全工学、英国）。

4 責任配分のあり方

英国では、HSWA（Health and Safety at Work etc. Act 1974：1974 年労働安全衛生法）が雇用主に対してゴール設定型（性能要件型：手段は名宛人に委ねる方式）で SFAIRP（So Far As Is Reasonably Practicable：合理的に実行可能な範囲で）を前提に、安全衛生と快適性を達成するよう義務づけており、たとえ職場に AI が導入されても、事業者がリスクアセスメントを行い、リスク低減を図る責任構造は維持されている（英国）。

HSE は市場監視当局（Market Surveillance Authority）として、AI 搭載製品の設計・供給段階も監視しており、「使用」と「供給」の二段階で監督している（英国）。

事業者中心の規律のみでは、上流で対応が必要な課題（仕様、限界、学習データの偏り等）への対応が難しいので、英国モデルには参照価値がある（英国）。

今後の焦点として、用途別ベンチマークの定義（* 人検知における逆光やしゃがみ姿勢や見逃し率、予知保全における機械個体差がある中での誤警報率などの機能を最低限担保する指標づくり）、評価・保証（assurance）手順の標準化（導入前の求める要件の定義→リスク分析→小規模からの実装でのテスト→導入判定→運用条件（ODD）の特定

→ODD 等に基づく監視→変更管理、等の手順)、competent person (有資格者概念) への AI の位置づけ (英国安全衛生法がリスクアセスメント等に求める competent person の要件を AI でどこまで代替させられるか)、サプライチェーン全体 (ベンダー [開発者]、データ提供者、システムインテグレータ、導入決定者、現場運用者等) での役割分担 (誰がどのような責任を負うか) の可視化、機械安全規格への組み込み (製造販売段階での AI 安全の確保) が挙げられている (英国)。

本研究 PT による日本国内での Web 調査での AI 導入決定経験者による回答では、AI 起因労災の責任の重さについて、「利用現場の責任者→訓練データ提供者→導入決定者→プログラマー」の順に考える傾向が示された (国内社会調査)。

AI の判断過程の説明責任を法定すべきとの意見は導入経験者の過半に達し、責任主体をプログラマーと訓練データ提供者とみる回答が多かった (国内社会調査)。

安全工学の立場からは、災害発生時の原因究明と迅速な再発防止を図る観点から、ログ設計 (後の検証を可能にするための入出力、設定、操作履歴等の記録方法の決定)、変更管理、責任分界 (誰が停止・復旧・報告・説明を担うか) を事前に明確化すべきことが示唆された (安全工学)。

5 AIWM (AI による労務管理)

この点について進歩的かつ具体的な示唆を行っているのは英国である (EU-OSHA もほぼ同様の示唆も行っており、英国も参考にしたと解される)。

少なくとも英国では、AIWM (AI based Worker's Management : 採用・配置・懲戒・解雇等の AI 管理) の領域では、安全衛生よりもプライバシーと公平性が主要論点になる。

英国では、AI によるイノベーションの推進を重視する政府の方針から、当面、横断的・包括的な AI 法 (horizontal AI law) を制定せず、個々の緩衝領域を持つ既存の規制当局が以下の共通原則を文脈依存 (context-specific) で実装する方針が採られている (英国)。

- 1) 安全性 : AI が誤作動した際の人や設備への危害を最小化すること。
- 2) 堅牢性 : データの乱れや環境変化、攻撃などがあっても AI 性能が維持されること。
- 3) 透明性 : AI の利用目的や範囲・限界、入力データ等が利用者に見えるようにすること。
- 4) 説明可能性 : AI が出した結論の理由を人が理解できるようにすること。
- 5) 公正 : AI の判断が属性による差別を生んだり、悪化させないこと。
- 6) 説明責任 : AI 利用の結果につき、利用主体である人が理由や是正策を説明でき、責任する義務を負うこと。
- 7) ガバナンス : AI の導入・運用・監査・改善をルールと体制で統制すること。
- 8) 異議申立て・救済可能性 : AI の判断に対して当事者が再審査、見直し、訂正等を求め、不利益を回復される可能性。

AI を用いた職場モニタリングについては、ICO (Information Commissioner's Office :

情報コミッショナー局。日本の個人情報保護委員会のような役割を担う）が、UK GDPR (EU GDPR を国内法化したもの) を根拠に、

- ・必要性・均衡性（目的達成に必要な限りで AI を利用し、AI を用いた監視やデータ収集を最小化すること）、
- ・透明性（AI の利用目的や範囲・限界、入力データ等が利用者に見えるようにすること）、

DPIA（Data Protection Impact Assessment：データ保護影響評価。AI で個人データを扱う前に、それによるリスクと対策を検討し、必要に応じて見直すこと）、

- ・人の見直し（human review）確保（AI の判断を人が見直せるようにすること）を軸としたガイダンスを発出している（英国）。

CDEI（Centre for Data Ethics and Innovation：データ倫理イノベーションセンター）は、採用における特定の属性への傾倒や差別の強化への注意、運用への監査（が可能な設計）の必要性を示唆している（英国）。

ACAS（Advisory, Conciliation and Arbitration Service：労使助言・調停・仲裁機関）は、労使対話と不服申立手続の重要性を示唆している（英国）。

国内文献レビューでは、AI を活用した採用の効率化のメリットが示される一方、過去データの偏りによる差別の強化や応募者の納得感低下が示されている（国内文献レビュー）。

日本 IBM での紛争事例では、同社が用いた Watson という人事評価 AI につき、労働組合より評価の公正性に疑義が生じ、最終的に労働委員会で成立した和解協定で、AI 評価項目や、不利益処遇時の AI 提案内容の開示・説明について合意されたことが示されている（国内文献レビュー）。

6 AI 導入に伴う労働者の非雇用化と求められる安全衛生法政策

企業での AI 導入は、①定型業務の自動化、②労働市場の二極化（中技能の事務・管理が縮小し、知識集約業務と低技能対人サービス等が残り、仕事と賃金が上下に分断する）、③プラットフォーム就業の拡大をもたらすため、非雇用化を促進し得る（国内文献レビュー）。

非雇用就業者（フリーランス）は多類型で一律に扱えないが、特にプラットフォーム就業では、AI／アルゴリズムが仕事の割当（dispatch）、単価決定（pricing）、業務評価と次の割当への反映（rating）、発注停止（deactivation）などを担う結果、名目上は独立自営業でも、「デジタル従属性（digital subordination：システムによる従属）」と言われる強い統制を受け得る（国内文献レビュー）。

非雇用就業者（フリーランス）の労働者性判断は、従来の人的従属性（personal subordination）と経済的従属性（economic dependence）の二軸で捉える傾向が見直され、管理（management：仕事の方法、時間、場所、条件を注文者が支配すること）、統合性

（integration：注文者の組織や工程に組み込まれること）、事業者性（business ownership：労働手段、損益負担、顧客開拓、価格交渉などの事業的要素の自律性）の視点で実態把握する必要性が示唆されている（国内文献レビュー）。

就業者を三層（①雇用労働者、②依存的自営、③真に独立した自営）に分類し、保護のレベルを調整すべきとの示唆もなされている（国内文献レビュー）。

日本の労働安全衛生法では、2025 年改正で、個人事業者ら（個人事業者や小規模事業の経営者であって作業現場で働く自然人）を意味する「作業従事者」概念が導入され、彼らにも、まとまった建設現場等を管理する元方事業者等の統括管理、注文者による物の安全確保、危険作業従事時の教育確保等の対象となることが法定された。しかし、安衛法の基本規定が「事業者による労働者保護」の構造を維持したまま、それに紐付く規則が作業従事者の保護を先行的に定めたため、両者間に“ねじれ”が生じている（国内文献レビュー）。

クラウドワーク／ギグワークのプラットフォーム事業者は、請負契約の注文者とそれに該当しない者に区別され、前者には法第 3 条が適用され、受注者の安全衛生に配慮するよう通達で示されたが、法第 3 条自体が訓示規定であり、法的拘束力はない。

今後の課題として、プラットフォーム事業者や大規模発注者など、「リスクを設計・制御できる主体」へ責任をどう配分するかが問われる（国内文献レビュー）。

また、個人事業者の就業条件の実質的保護は取引の安定と不可分であり、安全衛生・取引適正・社会保障を連結させた多層的保護の仕組みが必要になる。

A. 研究目的

本研究は、AI が安全衛生に与える影響を明らかにし、今後の法政策を展望することを目的としている。具体的には、

①産業における AI の導入が安全衛生にもたらすメリットとデメリット、

②前者を最大化し後者を最小化するための物的・人的な要件、

③従来の事業者を中心的な名宛人とする責任体系をどう修正すべきか（AI のプログラマー、学習するデータの作成者、導入の決定者、ユーザー（管理者）等の責任配分をどうするか）、

④①②を踏まえた③以外の現行の安全衛生規制の限界と改正提案、

⑤AIWM（AI based Worker's Management：労働者の採用、配置、懲戒、解雇等）に関する規制のありよう、

⑥AI の活用により非雇用化が進む場合、どのような労働安全衛生法政策が求められるか、

を明らかにすることを目的としており、この分担報告書は、そのうち英国で公表された文献を通じて①②③⑤と一部④に関する彼国の事情を明らかにすることを目的としている。

B. 研究方法

分担体制は、以下の通り。

法学者（三柴・井村（+池添）・佐々木・細川）は、調査対象国（英米独仏）の実態と対応に焦点に、上記①②③⑤と一部④につき調査する。

産業技術及び安全工学の研究者（藤原・平岡）は、産業機械や安全機器、安全衛生管理を焦点に、①②につき調査する。

産業医学及び人間工学の研究者（中澤・榎原）は、健康管理ツールを焦点に、①②につき調査する。

心理の研究者（島崎）は、心理的側面から①②につき調査する。

社会調査・統計学の研究者（西本）は、本研究プロジェクトが実施する企業関係者向け調査のうち、特に量的調査のデザインから分析までを担当する。

セイフティーグローバル推進機構（IGSAP）の梶屋氏は、AIを用いた自動運転機械に関する理論や国際規格、実装についてレクチャーすると共に、社会調査の対象となるAIを活用している企業を紹介する。

調査手法には、主に文献レビューの手法を用いた。対象文献には、学術書や論文よりも、政府機関のグリーンペーパーを積極的に採用した。

文献収集と報告書作成に際して、三柴及び島崎は、積極的にAIを活用した。まずは文献収集に際して上掲の報告書を探索させた。そこで示されたものと、筆者が独自にサーチしたものの原典を渉猟して内容を把握したうえで、調査課題について適切な回答が導かれるまで適当な指示を与えて出力を

調整し、その内容を筆者が査読して正確性を確認すると共に、適宜（大幅に）修文しつつ私見を加える手順を採用した。

C. 調査結果

1 法政策学的分析

1. 1 諸外国の安全衛生政策での対応

1. 1. 1 UK（三柴）

GB（英国本土）における職場のAI活用は、一部の先進企業を超えて拡大している。HSE（Health and Safety Executive）がまとめたユースケース報告では、化学・建設・製造・農業・輸送/物流・廃棄物/リサイクル・公共サービスなどでの活用事例が確認された。AIが産業横断で「現場の道具」になり始めていることが窺われる。

HSEは、職場におけるAIの使われ方を4領域に整理している。

第1は保全（Maintenance systems）であり、典型例は「危険な点検を遠隔化する」活用法である。ドローンに画像認識機能を搭載し、橋やクレーンの高所、タンク内部など人が入りにくい場所を点検する、設備データを学習したAIが故障の前兆を見つけ、点検方法や頻度を提案する等である。ここでAIは「遠隔点検係」として機能する。

第2は安全衛生マネジメント(Health and Safety Management)である。災害報告などの大量データをAIに解析させ、災害の傾向を抽出する。リスクアセスメント文書の下書きや教育教材（注意点を図解する動画など）の作成も支援し、ここでのAIは、「過去資料を読み解く参謀」といえる。

第3は設備・プロセス制御(Control of Equipment and Process Plant)である。自律走行ロボットの動きを調整したり、プラントの圧力・流量等の計測値の異常を検出して、早期の停止や介入につなげる。ここでのAIは「機械の見張り役」となる。

第4は職業的モニタリング(Occupational Monitoring)である。カメラ映像からPPE（個人用保護具）の装着を確認し、人と車両が危険距離に近づいた際に警告を発する。あるいはウェアラブル機器で熱ストレスや振動ばく露を見守る。ここでのAIは「現場の安全の監視役」を担う。

総じて、AIは「危険や手間が大きい部分を肩代わりする補助役」として、現場に浸透している。

では、こうしたAIを英国はどう規律しているのか。その骨格は、HSWA（1974年労働安全衛生法）が採用している「ゴール設定型（性能要件型）」の思想である。法律は「労働者の安全と健康を守る」という目標を示し、どの技術を用いてどう達成するかは事業者が選択する技術中立型の発想である。AIを生産や安全管理に導入しても、この基本原則は変わらない。事業者はリスクアセスメントを行い、SFAIRP（合理的に実行可能な範囲で）リスクを低減する責任を負う。

また、HSEは、職場での運用だけでなく、市場監視当局（Market Surveillance

Authority）として「AIを組み込んだ製品・機械が安全に設計・供給されているか」を監視する立場にある。つまり英国では、AIを「使う現場」と「販売される機械・製品」の二層で安全衛生専門機関が監督し、危険の芽を上下流から摘み取る構図を取っている。

政策全体のトーンを決定づけているのが、DSIT（科学・イノベーション・技術省）の白書が示す「プロ・イノベーション（innovation-friendly）」路線である。重要なのは、当面は新たな横断的・包括的なAI法（horizontal AI law）を制定せず、既存の規制当局が共通原則——①安全・セキュリティ・堅牢性、②適切な透明性と説明可能性、③公正、④説明責任とガバナンス、⑤異議申立てと救済可能性——を、各所掌分野の事情に合わせて実装するという「文脈依存（context-specific）」の設計をとっている点だ。具体的手法として、規制サンドボックス（新技術に対して既存の規制を一時的に緩和して行う実証実験）や様々な管轄領域で行われる規制全体の有効性を総合的に監視するモニタリング機能を採用している。また、まずはガイダンスで誘導しつつ、必要に応じて後から法制化する「段階的アプローチ」を採用している。技術進歩の速い領域において、過剰規制や実効性のない規制を避けるための現実主義といえる。

AIが労働安全衛生にもたらすメリットは明白である。危険箇所点検の遠隔化による危険の回避、予知保全、教育訓練の改善、個別化された健康管理など、いずれも現場の安全を底上げする可能性を広げる。しかし、以下のような現場視点でのデメリットや懸念もある。

①AIが制御を担い、人間が異常時のみ介

入する形が定着すれば、有事の際の技能や状況認識が衰える（deskilling）。

②判断過程が見えにくい AI のブラックボックス性は、理解不能ゆえの「無視」と、過度な「過信」という相反するリスクを同時に生む。

③AI による警告が過多になれば、アラート疲労（warning fatigue）を招き、重要警報が埋没する。

④学習データの偏りは、稀ではあるが重大な事故（エッジケース）への脆弱性につながる。

⑤専門人材の不足、初期投資やネットワーク環境の制約も大きな壁となる。

結局のところ、AI は「魔法の解決策」ではなく、「新しい失敗モードを持った道具」として扱う必要があるという結論に収れんする。

そこで英国が提案するのは、「AI が間違えることを前提とし、その間違いを許容・補完できる運用にする」ことである。具体策としては、サンプリング検証、重大な判断における人の最終承認（human review）、セキュリティ対策、フェイルセーフ、冗長保護（redundant protection）、導入後の性能監視といった「監視・停止・二重化・継続点検」のセットが強調される。AI を「便利な相棒」とするには、データ（品質と代表性）、人（理解と訓練）、運用（点検と是正）の三要素が不可欠と考えられている。

さらに、「事業者中心の規律だけでは不十分ではないか」という議論も生じている。AI は、開発者からエンドユーザに至るサプライチェーンで成立する。現場の事業者にのみリスクアセスメント責任を課しても、上流に偏在する情報の非対称性は解消しにく

い。そのため今後の焦点は、用途別のベンチマーク定義、評価・保証（assurance）手順の標準化、既存の「有資格者（competent person）」概念への AI の位置づけ、サプライチェーン全体での役割分担の可視化、そして機械安全規格への組み込みへと移っている。鍵となるのは「責任の所在を明確化し、現場の SFAIRP を実効性のある形で機能させること」である。

採用・人事評価・解雇などに関わる AIWM（AI による労務管理）では、安全衛生の領域より、プライバシーや公平性が前面に出る。

ICO（情報コミッショナー局）は、職場モニタリングについて、必要性・均衡性、透明性、データ保護影響評価（DPIA）の実施、および人の見直しの確保を軸にガイダンスを示している。その根拠は UK GDPR 等の既存法制である。また、CDEI（データ倫理イノベーションセンター）は、特に採用分野でのデータの偏りや差別に注意を促し、監査可能な設計を求めている。ACAS（労働関係助言・調停・仲裁機関）は、労使対話や不服申立て手順といった「手続の設計」による適正利用を図っている。

EU の労働先進国におけると同様に、GB でも、AI の活用に伴う心理社会的リスク——監視感による不安や仕事の喪失への懸念——は、重視されている。英国の基本的な対応方針は一貫しており、①透明性、②必要最小限の活用、③人の介在、④異議申立てと救済を「共通の最小公倍数」として積み上げる方針を採用している。

総じて、英国の AI 政策・実務は「イノベーションは阻害しない。ただし、AI は誤るという前提に立ち、責任と点検と救済の仕組

みをあらかじめ組み込む」という方針で貫かれている。AIを導入するか否かよりも、導入後に「いかに見張り続けるか」に注力している。

1. 1. 2 US（池添）

米国では、国家戦略としてのAI（人工知能）活用は進行している。しかし、雇用労働領域での活用の想定を中心は、再教育・技能訓練などの労働市場政策であり、労働安全衛生（Occupational Safety and Health：OSH）法政策でのAIへの対応や、積極的なAI活用戦略は、まだほとんど動いていない。以下では、その認識を裏付ける主に連邦政策の情報と、米国の学術・専門職団体で進んでいる先行議論の要点を整理する。

米国の労働安全衛生は、基本的には1970年連邦労働安全衛生法（Occupational Safety and Health Act：OSH法）という監督取締法規により規制されており、連邦労働省の労働安全衛生局（Occupational Safety and Health Administration：OSHA）が所管している。もっとも、州は、連邦基準と同等以上の効果を連邦によって承認されれば、州独自の枠組み（いわゆる「州計画（State Plan）」）で代替させられる。つまり、連邦の安全衛生政策としてAIへの対応やAI活用戦略を図るにも、州との連携が求められる。

連邦のAI政策の大枠は以下の通り。すなわち、バイデン政権下では、雇用労働領域でのAI活用の方針が示されていたが、トランプ政権となって破棄され、同政権から新たに提示された「AI行動計画（American AI

Action Plan）」では、AIによる雇用事情の変化に対応するための労働者への再教育や技能訓練、労働者にそれらを実施する雇用者や受講する労働者らへの税制優遇の拡充、連邦労働省（DOL）の下に、AIが労働市場と労働者に与える影響（AI導入の広がり、雇用の置換〔仕事の減少〕や創出、賃金への影響など）を継続的に分析評価する「AI労働力研究ハブ（AI Workforce Research Hub）」の設立などが掲げられた。しかし、その焦点はあくま労働市場（雇用・能力開発）であり、労働安全衛生へのAI利活用は、少なくとも明示的には射程に入っていない。実際に、報告書執筆時点では、DOLからは、労働安全衛生を含め、DOLとしてのAI利活用方針は公表されていない。

他方で、研究者や実務家による研究や議論は進んでいる。2020年前後から2025年までの文献のレビューから、AI導入の利益と共に、次のような副作用が強く警戒されるという。

- ①差別（差別的評価）の新たな作出、
- ②プライバシー侵害（個人情報の不適切収集・利用）、
- ③監視社会化（監視が常態化する社会への変質）、
- ④誤判定に伴う新たな法的責任の発生、
- ⑤アルゴリズム（判断過程）のブラックボックス化による判定の予見可能性の低下、
- ⑥責任主体の不明確化。

求められる対策としては、以下のようなものが挙げられている。

- ①アルゴリズム設計の適正化（目的外利用が生じないようにする、差別を生まない補正機能の組み込み、ミスジャッジが生じた場合に安全側に機能する設計、判断過程

を人間が理解できるようにする設計等)、

②データ理解の適正化（データの適切な収集と解釈〔データの代表性の確保、非正規・外国人などに関する少数データの重視、データの適切な意味づけ等〕、限界を踏まえた利用等）、

③透明性（何ができて何ができないかの説明）の確保、

④人間の介入（Human-in-the-loop：人が最終判断に関与する仕組み）の確保、

⑤ベンダー（開発者）による管理、

⑥モデルの構築、運用、監視、データ収集の各段階における AI を利用する企業の説明責任（Accountability）の確保。

労働安全衛生の文脈では、AI の「両刃の剣」的性格がより強調される。

Fisher らのスコopingレビュー（Scoping Review：広範な先行研究を地図化するレビュー）は、AI が危険作業の代替を進めれば、建設・製造・鉱業・輸送など高リスク産業での健康格差（Health Inequity：集団間の健康上の不公平）が緩和され得るという。とりわけ低所得層、移民、マイノリティが従事しがちな危険職種ではメリットが大きい。他方、採用や勤務割当に AI を用いれば、統計的差別が働き、差別が増幅される危険もある。AI による代替（雇用喪失）は低技能にも高技能にも起こり得て、教育・訓練機会が乏しい地域や集団ではデジタル格差（Digital Divide：ICT 利用能力の格差）が拡大する。これらは一定の労働者層の立場を弱め、結果的に健康格差まで強化する。要するに、AI は安全性を上げ得るが、格差を悪化させ得ると解されている。

示唆されている対策は、社会・文化・経済的背景を織り込んだ設計（最初からその社

会で冷遇されている人々を適切に捕捉するよう指示する等）、第三者によるアルゴリズム評価（Audit：監査・点検）、データ学習の際の人口構成の反映（少数データの適切な評価）、すべての労働者への AI スキル獲得機会の提供等である。

Jetha らのシステマティックレビュー（Systematic Review：一定の基準に基づき、関連する文献を網羅的に検証するレビュー）は、AI を用いたツールが、実際に傷病（罹患率・死亡率）に対して測定可能な効果を持つかを検証した。しかし、レビューが設定した基準を満たす定量研究（Quantitative Study）は極めて少なく、例外的に基準を満たした有望事例として、頸・肩・腰の痛み等の筋骨格系症状を抱えるホワイトカラーに AI チャットボット（Chatbot：対話型プログラム）がストレッチや姿勢改善を促す介入を行い、12 週間で痛みスコアが有意に低下し、重度症状が大きく減ったとの研究が挙げられる。もっとも、実証研究があまりに乏しく、広範導入を推奨するには時期尚早であると共に、バイアス、プライバシー、格差等の倫理的・社会的課題の検討も必要と結論づけられている。

他方、専門職団体である ASSP（American Society of Safety Professionals：米国安全専門職協会）は、2024 年 9 月に AI は安全衛生を高め得るが新たな危険やプライバシー問題も生むとし、①信頼性（Trust：専門家の判断を補完するが置換しない）、②透明性（Transparency：能力と限界を共有する）、③公平性（Equity：格差や差別を増幅しない）、④プライバシー（Privacy：機微情報を保護する）という四原則を掲げた。その後、AI の安全衛生への影響を調べるタスクフォ

ース (Task Force: 特別作業部会) を設置し、職場、労働者、会員（安全専門職）への影響分析を行い、同団体の今後の戦略に反映させる予定だという。

以上を整理すれば以下の通り。

米国では、AI 利用に関する一般的な国家方針が示され、学者や専門家団体による研究や議論が公表されるようになったが、労働安全衛生“法政策”として AI 導入への対応や積極的活用に関する具体的な議論は進んでいない。学者らからは、「AI に危険作業や危険探知、健康のモニタリングや指導をさせつつも、人が関与することで労災や職業病を予防し、健康増進を図る」ことでメリットを得られる、という前向きな見取り図が共有されつつある。ただし、透明性、説明責任、人の介入（判断のチェックや修整可能性）、第三者評価、プライバシー保護、格差拡大の抑制等を同時並行させる必要性が示唆されている。

1. 1. 3 ドイツ（佐々木）

ドイツでは、職場での AI（Artificial Intelligence：人工知能）導入はまだ限定的だが、検討・計画段階の企業が多い。ifo（ifo 経済研究所）の企業調査（2023 年 6 月）では、AI を「活用中」13.3%、「導入計画中」9.2%、「検討中」36.7%となっている。ただし「導入」の内実（単発利用か、労働過程に定着しているか）は不明で、企業規模による差も想定される。製造業は関心が高い一方、建設業では約 6 割が「関心なし」と回答し、業種差も大きい。

以下、研究目的に記した課題ごとに判明

した情報を示す。

①安全衛生面でのメリット／デメリット

DGUV（ドイツ法定災害保険）は、危険接近検知から機械の即時停止をもたらす AI 搭載支援システム、Gefährdungsbeurteilung（リスクアセスメント）システムの半自動的な作成と更新、Arbeitsmedizin（労働医学）でのスクリーニング（労働医に繋げる一次的な皮膚の異常、疲労などの拾い上げなど）や自己モニタリング支援（ウェアラブルによる健康パラメータの監視など）を挙げている。BMAS（連邦労働社会省）傘下の BAuA（連邦労働安全衛生研究所）も、精度の高いストレス把握や身体負荷の計測、画像解析による有害物質検出支援、包摂的職場設計（inklusionsförderliche Gestaltung von Arbeit）などを示している。WEKA Media（B to B に特化した専門的な情報提供会社）は、早期のリスク検知、報告書の自動作成、安全衛生教育の設計、健康や安全管理、機械や諸設備の点検の自動化等の領域で有効とする。

他方、BAuA や WEKA Media 等からは、ストレス等の負荷の増大、AI が学習するデータの偏りによる判断の偏り、サイバー攻撃への脆弱性、判断のブラックボックス化、責任の所在の不明確化などのデメリットが指摘されている。

②メリット最大化・デメリット最小化の技術的・人的要件

物的には、十分なデータ基盤（精密にデータを捉えるセンサー等）と、専門家が監修し更新する安全衛生データベースが前提となる。KAN（労働安全衛生と標準化に関する委員会）が示すように、機械側の安全性を重視しがちだった既存の規格や標準を、人間

工学的な知見を踏まえて、人間との相性、適合性を考えて作成する必要がある。これは、AIが規格や標準に照らして安全衛生管理を行う状況では、特に求められる。人的には、AIの運用や監督者のAIリテラシーを高めるための教育体制が重要であり、IHK（商工会議所）による「KI（Künstliche Intelligenz：人工知能）-Beauftragter im Arbeitsschutz（安全衛生分野のAI担当者）」研修制度などの利用が望まれる。

③責任体系：AIに起因する災害の責任を誰に負わせるか

ドイツ法にはAIの定義にかかる定めがなく、ArbSchG（労働安全衛生基本法）にもAIに関する特則はないが、同法3条1項の一般的義務（事情を考慮して必要な安全衛生措置を講じ、適宜見直すべき義務）により、使用者はAI導入によって生じ得る新たなリスクを管理する義務を負う。5条が定めるリスクアセスメントでは、AIが支援ツールになる一方、AI自体もリスクとして評価対象となり得る。12条の教育義務には、AIの安全操作等の教育も含まれ得る。

しかし、雇用主のみへの義務づけでAI安全を実現できるわけではない。

連邦政府のAI戦略（2018年、2020年改訂）は、AIの発展と安全・信頼の確保を両立させることを基本方針として、ドイツ独自の標準の作成より、ISO/IECやEU規格などの国際標準を踏まえつつ、AI時代に耐えるような専門人材、組織体制、検証手法などの整備を課題としている。その際、AI基盤の製品・サービスでは、AIの学習能力ゆえに機械の安全基準が変化する可能性も指摘している。

しかし、未だArbSchGのAI向けの改正

などの法的対応は講じられていない。よって、AI活用上必要なサプライチェーン全体を捉えた対応は未だ図られていない。

⑤AIWMへの対応

AIWM（AI based Worker's Management：AIを活用した採用・評価・配置等の労務管理）について、安全衛生法制としての直接的対応は限定的である。他方、2020年のAI戦略改訂は、企業におけるAI利用が個人データ処理を拡大し得ることを踏まえ、従業員データ保護（employee data protection）や共同決定（Mitbestimmung：企業内共同決定）を含む規制領域での対応を論点として挙げている。あわせて、国際原則・標準化の文脈で、透明性（transparency）や意思決定の追跡可能性（traceability）を重要な要素と位置づけている。

1. 1. 4 フランス（細川）

フランスではAI（Artificial Intelligence：人工知能）をIA（intelligence artificielle）と表記するが、便宜上AIで統一する。

フランスでは、労働条件は産業別（職種別）労働協約を基礎として、職種区分（Ouvrier、ETAM〔employé, technicien, Agent de maîtrise〕、Cadre／Ingénieur）ごとに基準が設定される。規範設定は「法律主義」により、主に労働法典（Code du travail）で規制されるが、マクロン改革で進められた分権化により企業別協定の比重が増した。安全衛生は労働法典第4部に集約されており、労働医（médecin du travail）の強い権限、CHSCT（Comité d'hygiène, de sécurité et des conditions de travail：安全衛生労働条

件委員会。現在は大企業を除き CSE（社会経済委員会）に統合されている）の関与に特徴づけられる。

AI 導入の概況は、INSEE（L'Institut National de la Statistique et des Études Économiques：フランス国立統計経済研究所）／Eurostat（EU 統計局）調査で示され、従業員 10 人以上の企業の導入率は 2023 年 6%、2024 年 10%で、EU（European Union：欧州連合）平均（2024 年 13%）を下回る。250 人以上では、2024 年 33%、情報通信業は 42%と高い一方、製造業（7%）、建設業（3%）等では低い。用途は文書分析、機械学習によるデータ分析、意思決定の自動化や支援、文章生成が中心で、既成のソフトの購入が中心的な方法とされる。

安全衛生への AI 活用は、近年ようやく端緒についた段階だが、INRS（Institut national de recherche et de sécurité：国立労働災害・疾病予防研究所）報告から以下のような示唆を得られる。研究目的に記した課題ごとに述べる。

①安全衛生にもたらすメリット・デメリット

メリットは、事故・疫学データの大規模分析、現場データにもとづく危険検知と警告、自律・協働機械による危険作業代替と人間の負担軽減、ウェアラブル端末による健康リスク予防である。デメリットは、監視強化とプライバシー侵害、学習するデータの不足や偏りによる誤判断とブラックボックス化、自律機械の予測不能挙動や環境変化が生む事故、人間のストレス、データと知見の大企業への偏在である。

②メリット最大化及びデメリット最小化の要件

データ利用ルールの策定と判断の透明性の確保、学習するデータの品質管理、AI 人材と安全衛生人材の相互教育、労働者らへの導入前の情報提供と職業訓練、中小零細企業もデータの提供と利用が可能となるような制度設計等が挙げられる。

③AI に起因する災害の責任の所在

AI によるリスクアセスメントなどの AI 活用は使用者責任を免責しない。それだけに、導入プロセスや責任分担の可視化や監査が必要だが、フランスの労働安全衛生法制は、AI 活用を前提としたデータの取扱い規制を行っていない。すなわち、データ利用ルールの未整備、データ品質管理体制の未整備、データの大企業偏在、責任ルールの未整備等の問題がある。

⑤ AIWM（AI based Worker's Management：採用・評価・配置・懲戒・解雇等の AI 管理）

差別的な効果、解雇理由のブラックボックス化、個人情報保護が主要論点となっている。

労働者性（雇用関係該当性）をめぐっては、2016 年改革で、プラットフォーム就業者の集团的権利等が確認された。他方、2019 年モビリティ法は、プラットフォーム就業者について、労働者性の認定から離れて最低限の保護を与える方針を採用した。すなわち、プラットフォーム事業者に対し、就業者に対して、報酬や評価の仕組み、仕事の可否の自由の有無、安全対策、災害補償、代表との対話の枠組み等を Charte（シャルト：就業条件の「ひな形」）で呈示し実施する義務を課すと共に、事業者がこれらを実施しても労働者とみなさないこととした（実施の阻害要因を除去するため）。また、2022 年

には、就業者の代表を選出し、同人が行政仲介のもとでプラットフォーム事業者と交渉し協定する枠組みが整備された。

総じて、フランスにおける安全衛生での AI の活用や議論の停滞は、そもそもの導入の遅れや、中小零細企業の多さ等と関連づけられる。

1. 2 人事労務管理と労働法一般の変化に関する国内文献レビュー（三柴）

日本の労働安全衛生法は、従来の災害防止や有害物質対策といった「狭義の安全管理」ととどまらず、過重労働に伴う疲労・精神的ストレスへの対応、生活習慣病の予防、さらには私傷病者の就労支援まで含む「健康確保の総合政策」へと射程を拡大してきた。健康リスクは、有害物だけでなく、長時間労働や労働密度、配置・評価・処遇、職場関係といった「働き方そのもの」からも生成される以上、安全衛生は労務管理（Worker's Management）や人事制度（HRM：Human Resource Management）の設計と切り離せなくなる。労働安全衛生法の施行通達（昭和47年9月18日基発602号）は、すでに「一般労働条件と安全衛生の密接な関連性」を指摘しており、AI時代にはこの連関が一層あらわになる。

AI（Artificial Intelligence：人工知能）が人事労務管理にもたらす変化は、①採用等の選別、②人事評価、③労務管理一般、④人事部員の認識、の4視点で整理できる。採

用では、ソフトバンクが新卒採用のエントリーシート選考に AI（IBM Watson）を導入し処理時間を大幅に削減した例のように、定型作業の高速化という利益が確認できる。一方で、AI は過去データの偏りを増幅し得る。アマゾン社の採用 AI が女性差別的な評価傾向を示した事例が象徴するように、効率化と同時に「バイアス再生産」の危険を抱える。さらに国内実証でも、AI 面接を経験した応募者は「人間のみの面接より個性が正しく評価されない」と感じ、納得感が下がる傾向が示されており、AI が当事者の主観的信頼を毀損しうる点を看過できない。

人事評価では、勤怠・業績データに加え、ウェアラブル（wearable device：装着型センサー）で得た心拍数等の生体データを組み合わせて AI 解析する手法が現れ、「客観性・公平性の向上」が期待される反面、アルゴリズムの不透明性や監視の強化が不信や労使関係の悪化につながり得る。日本 IBM では、AI による評価・賃金調整をめぐる紛争が発生し、都労委での和解協定において、AI 評価項目の提示や、不利益処遇時に AI 提案内容の開示・説明が合意された。これは、AI の推論過程を完全に可視化できなくても、「何を見ているか」「不利益局面で何が提案されたか」の可視化により、ブラックボックスの害を低減する動きとも解し得る。

労務管理一般については、JILPT（労働政策研究・研修機構）の国際比較調査が、AI 導入後のタスク変化を、①完全自動化、②部分的自動化、③補完的効率化、④新規タスク創出、の4類型に整理している。注視すべきは④で、勤怠管理やシフト編成などの従前からある定型的な作業は AI により迅速化される一方、AI 監視（AI による労働者の

働きぶり等の監視）やデータ分析（AI 導入で自動化された勤怠管理や生産管理などから得られる大量のデータを整理・解析し、業務効率化、人員配置の最適化、安全衛生リスクの把握等に活用する業務）といった新しいタスクが創出され、労務管理の質に変化をもたらしている。これらは AI の手を借りつつ、人間が統括する。

労働者向けの Web 調査でも、仕事のパフォーマンスや管理の公平性、職場の安全性が改善したという回答が一定程度みられるが、業種・職種により、メリットとデメリットの分布に差異がある。人事部員調査では、AI の強みは定型業務の効率化や数値処理の正確性に集中する一方、「人間らしさや個性の尊重」「職務上の矜持 (professional pride)」は、AI では代替困難と見られている。

では、AI・デジタル化は「働き方」自体をどう変えるか。大内教授が指摘するように、情報化社会の労働では、「情報(information)を収集し、分析 (analysis) し、その結果に基づき行動 (action) する」ことが重要になる。特にデジタル産業では知識付加価値の創出が中核となり、場所や時間の拘束が弱い働き方が展望される一方、社会的保護の設計も雇用契約に固定せず、協同組式的枠組み(中小企業等協同組合法等)を活用して支える選択肢があるとされる。また、規制の形としては、基準を示しつつ逸脱の余地を残すナッジ型 (libertarian paternalism) の設計が有力になり得る。

健康管理についてはウェアラブルと AI を用いたヘルス・テック (health tech) がコラボヘルス(collabo-health)と連携し得る。また、AI 時代には、労働時間管理と健康確保を連携させる必要がある。報告者は、2005

年に疲労・ストレスという「質的観点」から労働時間規制を再設計する提案を行ったことがある。これは、長時間労働が一律に有害なのではなく、仕事内容や適性との相性で健康影響が変わるなら、ストレスチェック等を用いて仕事の質を調整する方向を示唆するものだった。ストレスチェックは後に法制度化されたが、量(時間数)依存から質(仕事の密度・負担)へと比重を移す発想はなお未完であり、ウェアラブルと AI の普及は、技術的にはその実装可能性を高める。

AI は非雇用化を進める可能性が高い。その背景は、①定型業務の自動化、②労働市場の二極化(「高スキルの知識集約型業務」と「低スキルのサービス業務」の需要拡大と、「中程度の技能を要する事務・管理業務」の縮小)、③プラットフォーム就業 (Platform Work) の拡大、の 3 点と解される。

非雇用就業者(フリーランス)は専門職型・自営型・業務委託型・フランチャイズ型・非雇用テレワーク型・デジタルプラットフォーム型など多様であり、同列に扱えない。とりわけプラットフォーム就業では、AI/アルゴリズムが仕事の割当・単価・評価・停止を担うため、名目は独立でも実態は強い統制が働くという逆説が生じる。よって、従来の「労働者」概念(雇用契約に結びついた人的従属性中心)だけでは保護の輪郭が定まりにくい。

そこで、諸外国の法制度や判例法理、日本の労働法学説は、労働者性(労働者該当性)の判断枠組みを、人的従属性 (personal subordination: 指揮命令への服従の程度))と経済的従属性 (economic dependence: 特定発注者への〔事業ないし生計〕依存の程度) の二軸で捉え直し、さらに、①管理

（management：仕事の仕方や条件の指示・監督）、②統合性（integration：事業の仕組みへの組込み）、③事業者性（business ownership：設備・損益負担・顧客・価格交渉力等の自立性）という三視点で実態を読む必要を指摘するようになってきている。

日本の判例・実務は、人的従属性を軸に、指揮命令、報酬の労務対償性、労務代替性、事業者性などを総合考慮してきたが、プラットフォームでは指示主体が人ではなくシステムであり、人的従属性の“目印”だけでは従属の実像が見えにくい。プラットフォームでは、(1) 割当（dispatch：ログイン状態・位置情報・完了率等で仕事割当の優先度が自動判定される）、(2) 単価（pricing：動的価格で交渉余地が乏しい）、(3) 評価（rating：スコアが次の割当や単価に反映される）、(4) 停止（deactivation：条件に達するとアカウント停止＝取引排除）という一連の仕掛けが「デジタル従属性（digital subordination）」を生んでいる。そこで、就業者を大きく、①雇用労働者（employee）、②依存的自営（dependent self-employed／雇用類似就業者）、③真に独立した自営（genuinely self-employed）という三層に整理し、保護の濃淡を設計すべきだと指摘されている。

日本の法学者による比較法研究では、EU、フランス、イギリス、アメリカ、ドイツの動きが調査されている。ほぼ共通して「形式より実態」へ重心が移りつつ、立証構造（推定）と適用領域（最低賃金・労災・安全衛生等）を調整して“中間地帯”を拾い上げる傾向が指摘されている。具体的には、EU ではプラットフォーム就業をめぐり雇用推定やアルゴリズム管理の透明化が図られ、イギリス

では employee／worker／self-employed の三類型のうち worker に最低賃金・労働時間・平等などの基礎的保護が及ぶ構造が採られ、Uber 事件では契約文言より実態の統制を重視して worker 性が肯定された（他方で Deliveroo 事件では労務代替性が重視され、組合法上の労働者性が否定された）。アメリカでは法律ごとに判断枠組みが異なるが、連邦法では、FLSA における経済的実態テスト（economic realities test：「名目上の契約形態ではなく、実際の経済的関係が労働者をどの程度“雇用者に依存させているか”」を判断するもの）や NLRA の common law test（伝統的な「指揮命令関係（control）」を中心に雇用関係を絞り込むテストであり、経済的実態テスト（economic realities test）よりも「雇用」と認定される範囲が狭くなる）が議論を牽引している。州法レベルでは、カリフォルニア州最高裁判所で示された ABC テストが議論を牽引している。これは、A：指揮命令から独立しているか（Absence of control）、B：使用者の通常業務とは異なる業務か（Business of the employer）、C：独立した事業を営んでいるか（Customarily engaged in an independently established business）、の3要件をすべて満たさない限り労働者と推定するものである。さらにプラットフォーム運転手については、independent contractor と位置づけつつ、最低報酬、健康保険の補助、対物・対人保険、一定の差別抑制や安全対策等を適用する方針を採る動きもある。ドイツでは、雇用労働者の外側に雇用類似者（arbeitnehmerähnliche Person）を置き、依存的自営に部分的保護を与える。総じて、人的×経済的従属性の二軸を基礎にしつつ、

実態判断を精緻化し、推定や保護領域の広げ方で“中間地帯”を救済する試みと理解できる。

では、非雇用化に対する日本の法政策的対応の到達点はどうか。労働関連法ではフリーランス新法の新設やガイドラインの公表がなされてきたが、立法上の包括的対応は限定的だった。ただし安全衛生については、2025年改正で「作業従事者」概念が導入され、個人事業者にも元方事業者らによる統括管理の対象化、注文者による物の安全確保の対象化、危険作業に従事する場合の教育の確保などが法定された。もっとも、安衛法の基本規定（法第20条から25条の2のうち第25条の2以外）は依然として「事業者による労働者保護」の構造を維持しており、特に規則レベルで個人事業者の保護を図った個別規定と本法の基本規定の間に“ねじれ”が生じている。このねじれを解消し、プラットフォーム事業者や大規模発注者など「リスクを設計・制御できる主体」へいかに責任を配分するかが次の課題になる。

なお、個人事業者の就業条件の実質的保護は契約（取引）の安定と不可分であり、法的にはフリーランス新法、下請法、家内労働法、中協法等がその基盤を支える。今後は、安全衛生・取引適正・社会保障を連結させ、AI時代の多層的保護体制を構築することが求められる。

2 産業技術及び安全工学的分析（藤原・平岡）

2.1 AIの定義（平岡）

AI（Artificial Intelligence：人工知能）の活用を労働安全衛生（OSH：Occupational Safety and Health）の文脈で論じるには、まず「AIとは何か」を定義する必要がある。AIの捉え方は、歴史的にも技術的にも一様ではなく、時代、社会で指す意味も揺れてきた。

AIの定義論は、まず「そもそも機械に知能は持てるのか？」という哲学的な問いから始まった。1940-50年代は、その問いを、数学や工学の言葉で表現しようとした。アラン・チューリング（Alan Turing）は1950年に、機械が賢いかどうかを「人間と区別できない会話ができるか」で判定するチューリング・テスト（対話で人と見分けられない機械を知能とみなす基準）を提案した。さらに彼は、「最初は子どもの頭のようなプログラムを作って教育すればよい」とし、学習が鍵になることを直感していた。

同じ頃、周辺分野でもAIにつながる土台が形成された。

ノーバート・ウィーナーのサイバネティクスは、状況への応答の仕組みが賢さの基本だ、という考え方である。たとえば、室温を測定して、暑ければ冷やし、寒ければ止めるエアコン。人間も、体温を測って、暑ければ汗をかく。こうした結果に基づく調整の繰り返し（フィードバック）が、生物にも機械にも共通する賢さの原理だと説明した。

ジョン・フォン・ノイマン（John von Neumann）は、「コンピュータとは何か」を分かりやすく示した。現代のコンピュータは、基本的に、①作業の手順（プログラム）と、②材料（データ）を同じ場所に格納し、③順番に取り出して実行する仕組みで動いている。ノイマンはこの“基本形”を示した。さらに彼は、「機械が自分のコピーを作れるか」という、未来の自動化を先取りする発想も考案した。

クロード・シャノン（Claude Shannon）は、「情報」の扱い方を体系化した。たとえば、電話や無線での情報送信で、どれだけの量を送れるのか、ノイズがあっても正しく伝えるにはどうすべきか等を、数学的に整理した。さらにチェスで、次に取り得る手を列挙し、その先の展開を考えて最善策を選ぶ方法を示した。

つまり AI は、コンピュータがどう動くか、情報をどう扱うか、どのように答えを探すか等の土台の上に作られてきた。

1950 年代後半から 60 年代は、いわゆる第 1 次 AI ブームだった。ここで主役になったのが記号処理型 AI（GOFAI：Good Old-Fashioned AI）で、「人間の考えをルールや論理式（記号）に書き下ろして、推論エンジンで解く」という発想だった。例えば、ルール 1：もし「高熱」かつ「咳」なら「風邪の可能性」、ルール 2：もし「高熱」かつ「発疹」なら「麻疹の可能性」、ルール 3：もし「麻疹の可能性」なら「隔離して医師へ連絡」のようなルールを設定し、患者データ（高熱、発疹…）を記号として入れると、ルールに従って推論して結論を導く。こうした「知識＝ルールの集合」という発想が

GOFAI である。

1956 年のダートマス会議で「Artificial Intelligence」という言葉が掲げられ、GPS（General Problem Solver：一般問題解決器）のような試みが登場する。その要は“手段—目的分析（means—ends analysis）”である。これは、現状と目標の差を見つけ、その差を縮める操作を選ぶモデルであり、例えば、別の部屋にある「箱」を入手するが目的なら、その差を縮めるには「箱のある部屋へ行く」必要があり、そのためには「ドアが開いている」必要があるので、GPS は、ドアを開ける、そのため鍵を使うなどのサブ目標を作る、などの階段状の考え方を採用する。

しかし、現実世界は曖昧で例外だらけで、ルールを全部書き尽くすのは難しい。また、複雑な課題ほど試すべき組合せが爆発的に増える（組合せ爆発）ため、計算が追いつかない。データから学ぶニューラルネット系（Connectionism：神経回路を模した学習）もあったが、当時のパーセプトロン（Perceptron：初期の学習モデル）は表現力に限界があるなどの課題があって批判を受け、研究熱が冷え込み、「AI の冬」につながったとされる。

1980 年代の第 2 次 AI ブームでは、再びコンピュータを「ルールで賢くする（≡人間に近づける）」方向が復活したが、以前より実用的になった。知識工学（Knowledge Engineering：専門家の知識をルールとして整理し、推論で使う）が注目され、エキスパートシステム（専門家の判断を模したルールベースのシステム）の実用化が進んだ。MYCIN（感染症の診断・処方支援）のように高い診断性能が報告された例も現れた。

しかしこれも限界を示した。システムの賢さが知識ベースのルールに依存するため、作るのも更新するのも膨大な手間がかかる。条件が変わるたびに修正せねばならず、柔軟性に欠ける。維持コストの重さが障害となり、第2次ブームも失速していった。

2000年代後半からの第3次AIブームの最大の特徴は、コンピュータが「自ら大量のデータを食べて賢くなる」方向へ舵を切ったことである。これをデータ駆動型（データから規則性を学ぶ方式）への転換という。背景は、ビッグデータ（大量データ）が蓄積され、GPU（画像処理を目的に発展した高速計算装置）が普及して、計算力が一気に伸びたことである。これにより機械学習（データから予測や分類の仕組みを作る技術）が再び実用の主役になり、とくにディープラーニング（Deep Learning：多層ニューラルネット）という「層を重ねた学習モデル（「小さな判断役」を何段も並べ、前の段の結果を次の段が受け取って、だんだん複雑な特徴を捉えること。線の把握→形の把握→物体の把握というように、最後に全体として判断する）」が中心に躍り出た。

もっとも、いきなり深層学習だけが現れたわけではない。ディープラーニング（Deep Learning）が広まる前も、会社の現場では「表形式データ」（例：売上・年齢・作業時間など行と列に並んだデータ）を使って予測や判定をするAIがよく使われていた。代表例は次の通り。

① 決定木（Decision Tree）：

「もし～なら～」という条件分岐（チャート式）で答えを導く方法。

② ランダムフォレスト（Random

Forest）：

決定木をたくさん作り、最後は多数決で結論を出す方法。決定木の“偏り”をカバーして安定した判定を導きやすい。

③ SVM（Support Vector Machine）：

データを2つに分類するための基準を探索する方法。

たとえば「不良品／良品」「危険／安全」を明快に分けるラインを探す方法。

④ 勾配ブースティング（Gradient Boosting）：

精度が不十分な「判定器」の機能を、結果のフィードバックにより向上させていく方法。

これらは、表形式データに強く、長い間、実務の標準的な道具だった。

その後の潮目の変化の象徴が、2012年のImageNet（ImageNet Large Scale Visual Recognition Challenge：大量の画像を使って、何が写っているかを当てる国際コンテスト）だった。ここでは、深層CNN（Convolutional Neural Network：畳み込みニューラルネット）が大活躍した。深層CNNとは、画像を「小さな部分」から順にみて（「線や角」→「目やタイヤのような部品」→「顔や車のような全体」）、その特徴を段階的（階層的）に捉えるディープラーニングの仕組みである。以後、画像認識ではディープラーニングが事実上の標準になった。

「主流の交代」は、画像以外の分野にも拡大した。

音声認識（人の声を文字にする技術）では、以前はHMM（Hidden Markov Model：時系列での音の状態の移り変わりから文字を確率で推定する古典的モデル）が中心だ

ったが、2010年代には深層学習方式で精度が大きく改善した。

自然言語処理（人間の使う言語を理解し処理する技術）でも、Transformer（Self-attention：文章中の重要な単語に重みづけし、文脈を捉えるモデル）が登場し、更にBERT（Bidirectional Encoder Representations from Transformers：前後の文脈を同時に参照して単語や文の意味を捉えるモデル）が普及すると、従来主流だったRNN/LSTM（文章を順番に読み進めながら情報を保持するニューラルネット）の系統より文脈理解の精度が大きく向上した。

このように、第3次AIブームでは、「大量データ」、「高い計算力」、「深い学習モデル」という組み合わせにより、画像・音声・言語といった複数分野で同時に性能の底上げと技術の主流交代が引き起こされた。

そして、2022年以降、深層学習の延長線上で社会の景色を一気に変えたのが生成AI（Generative AI：文章や画像など“新しい内容”を作るAI）である。従来のAIの機能が、主に「分類する」「予測する」だったのに対し、生成AIは、「文章を書く」、「要約する」、「翻訳する」、「コードを書く」、「画像を作る」など、一般人も直接価値を感じられる機能を具備した。中核技術はやはりTransformerで、GPT系列（Generative Pre-trained Transformer：大規模言語モデルの系譜）やChatGPTが代表例である。規模（パラメータ数＝ある数値と別の数値の関係の深さや、どの数値を重くみるかなどに関するモデル内部の調整ネジの数）が桁違いに大きく、記憶できるパターンの量と細かさが増えるので、情報の微妙な違いを判

別できたり、出力品質があがり、過去のブームを上回る社会的インパクトをもたらしている（第4次ブームと呼ぶ識者もいる）。

AIの歴史は欧米の研究者や企業の話として語られがちだが、日本の研究者もその土台となる重要な貢献をしている。たとえば甘利俊一は、情報幾何学（Information Geometry：確率モデル（「こうなりそうだ」という予測の式）を“地形”のように見なしで考える方法。モデルごとに、同じデータを見て、どんな分布を想定し、どんな確率を割り当てるかの違いを「距離」や「方向」で表し、どの方向に動かせば予測が改善するかを見える化する。結果的に、学習（パラメータ調整）や推定（最適なモデル選び）を、スムーズに進めるための理論）を体系化し、自然勾配法（Natural Gradient：学習によって、モデルの内部パラメータ＝調整値をより効率的、安定的に改善するための更新法）を提案するなど、機械学習の「学び方」を支える理論を提供した。これは目立つアプリや製品より、学習アルゴリズム（Learning Algorithm：データから規則性を学ぶ手順）の背骨を作る仕事であり、後の研究でも繰り返し参照されている。

また福島邦彦は、ネオコグニトロン（Neocognitron：画像を段階的に見て特徴を捉える階層型の視覚認識モデル）を1980年に提案し、位置が少しずれても同じ物体として認識しやすい仕組みを示した。これは、現在のCNN（Convolutional Neural Network：畳み込みニューラルネット。画像の特徴を「部分→全体」と段階的に抽出する深層学習モデル）の先駆けとされる。重要なのは、ニューラルネット研究が世界的に下

火だった「AI の冬」(研究資金や関心が冷え込んだ時期)にも、この研究を続けた点である。

日本の国家プロジェクトである第五世代コンピュータ計画 (ICOT: 知識処理・論理推論を中心に“コンピュータのスマート化”を目指した国家研究プロジェクト、1982-1992 年) があった。商業的な成功はともかく、研究者や研究コミュニティの育成、基盤技術の蓄積という面で貢献した。

また、ホップフィールドとヒントンがニューラルネットの基礎的発見につき、2024 年のノーベル物理学賞を受賞したことは、AI が「応用技術」に留まらず、自然科学の根本的テーマとして位置づけられたことを示している。

こうした多様な系譜のため、「AI」の定義も文脈次第で変わる。哲学的には「人間の知的活動を模倣する技術」、工学的には「知的活動の一部を計算機で実現する技術」、社会的には「深層学習などの機械学習や生成 AI」を指しやすい。ISO/IEC 22989:2022 (AI 概念と用語) は AI システムを「目的に基づきコンテンツ・予測・推奨・決定等を出力する工学的システム」と定義するが、実務では狭義、広義の定義が混在する。

労働安全衛生のルール策定に際して悩ましいのは、AI が毎年のように姿を変えることである。ある年は深層学習が注目されるが、別の年には生成 AI が注目される。もし法律や指針が「この技術 (例: 深層学習) だけを AI と呼ぶ」と限定すると、技術が少し変わっただけで規制が古くなり (陳腐化)、抜け穴が生じ易い。さらに現場では、中小企業ほど古典的なルールベース (if-then: もし

A なら B) や統計モデルなど「堅実で導入しやすい手法」を使うことが多く、狭義の AI 定義だと制度の射程から外れてしまう。

そこで、政策論の土台として「操作的定義」(実務的に使える柔軟な定義) の採用が考えられる。AI を「データ・ルール・アルゴリズム (計算手順) に基づき、人間の意思決定・判断・作業を補完または代替する情報処理技術」と広く捉える方法である。ここで重要なのは、技術名ではなく「何をしているか (機能)」で AI を捉える点である。この定義に入るのは、機械学習 (データから予測や分類の仕組みを作る手法) だけではなく、統計的予測 (回帰・時系列分析: 過去データから将来を読む)、ルールベース (if-then で判断する仕組み)、最適化 (最善の計画の選択によるシフトや配車の最適化)、画像・音声・自然言語処理 (認識、理解の技術)、そして生成 AI (文章や画像など新しい内容を作る AI) までを、同じ枠で扱える。

この「広い定義」により、政策上の 3 つの利点が生まれる。第一に、技術が変わってもルールが使えること (技術中立) になり、将来の変化にも対応しやすい。第二に、責任配分を議論し易くなる。AI が関与する場面では、プログラマー (作った人)、データ作成者 (学習材料を整えた人)、導入決定者 (使うと決めた人)、ユーザー = 管理者 (現場で運用する人) など、関係者が複数になる。狭い定義だと技術ごとの責任の切り分けになり易いが、機能ベースの定義なら「誰がどう判断し、どう使ったか」という共通の物差しで整理できる。第三に、AI による労務管理 (AIWM: AI based Worker's Management = 採用・評価・配置等を AI で支える管理) や、自動化による非雇用化 (雇用契約の外で

働く人が増える）といった、働き方の変化がもたらすリスクを、特定技術に限定せず包括的に捉える前提が整う。

要するに、「AI＝特定の先端技術」と見なすより、「人の仕事や判断に影響を与える情報処理」として広く捉えた方が、安全衛生の制度設計にフィットする。

2. 2 AI のメリット・デメリット、メリット最大化・デメリット最小化の技術的・人的要件―産業技術的観点から―（藤原）

①産業における AI 導入が安全衛生にもたらすメリットとデメリット

職場で用いられる AI（Artificial Intelligence：人工知能）は、生成 AI（大規模言語モデル等）による助言・文書作成、画像認識による危険検知、予測モデルによる異常予兆検出、ロボット・AGV（Automated Guided Vehicle：無人搬送車）等の制御、アルゴリズム管理（Algorithmic Management：仕事配分・評価・監督の自動化）まで多様である。

メリットは、危険作業の代替や監視の高度化により、災害・健康障害を減らし得る点にある。たとえば、反復・有害作業を機械が担えば身体的負傷は減りうるし、設備データを解析して故障の兆候を早く捉えれば事故の芽を潰しやすい。

一方で、AI が「人の判断」や「働き方そのもの」に介入すると、新しい危険が生まれ

る。典型は自動化バイアス（Automation Bias：AI の提案を過信して誤りを見逃す）で、部分自動化でも起きる。AI 任せが続けば人間の技能劣化（Skill Degradation）やモード錯誤（Mode Confusion：自動運転等のシステムがどの自動モードで動いているかを誤認し、誤操作などを招くこと）が生じ、非常時に手動介入できない。警報が多すぎれば警報疲労（Alert/Alarm Fatigue）で重要な警報が無視される。生成 AI では幻覚（Hallucination：もっともらしい誤情報の生成）が、誤手順・誤判断を誘発する。学習データの偏りや代表性（適切な実態の反映）の欠如、運用中のモデルドリフト（Model Drift：環境変化による実態との乖離）は、誤検知（False Positive）／見逃し（False Negative）を増やし、物理世界に接する AI ではそのまま事故に直結する。

AI が現場の意思決定だけでなく、ロボットや搬送車など「物理世界を動かす装置」の制御に入り込むと、“重大災害”リスクが生じる。AI は挙動が確率的かつ複雑なので、どうしてもエッジケースに弱かったり、期待を外す動きを伴う。ここで問題になるのが CPS（Cyber-Physical System：サイバーフィジカルシステム。センサーで現場（物理世界）を計測し、その結果をデータ化して AI 等が解析・判断し、機械が現実作用する統合系）である。CPS では、AI 搭載ロボットや自律移動体（AGV：Automated Guided Vehicle＝無人搬送車、AMR：Autonomous Mobile Robot＝自律走行搬送ロボット等）の誤動作、人の侵入や危険状態の検知の遅れによる挟まれ・衝突などの災害が生じ得る。さらに、Fail-safe（フェイルセーフ：異常時に安全側＝停止等へ倒す設

計)が不全になると、異常状態でも機械が動き続け、被害が拡大しやすい。

アルゴリズム管理（Algorithmic Management：AI等が仕事の配分、作業ペース、評価、監督を自動化する管理）では、こうした“物理系の危険”と別経路のリスクが生じる。アルゴリズム管理は人の作業効率を上げるが、評価指標が生産性に偏ると、過密ノルマや休憩削減が誘発されやすい。その結果、疲労の蓄積、注意散漫、安全手順の省略が連鎖し、累積的に災害確率を押し上げる。ここでのポイントは、危険源が機械そのものではなく、「働かせ方（労働強度・裁量の奪われ方）」として現れることである。

さらに、AIによる監視強化がもたらす心理社会的リスクの増幅も指摘される。AIによる常時監視は、ストレス（追い立てられている感覚）、社会的孤立（上司や同僚との対話が減る）、裁量喪失（現場判断が許されない）を通じて安全文化を弱め得る。とくに、ヒヤリハット（事故には至らないが危険だった出来事）の報告は、本来は再発防止の基盤だが、監視下で「報告すると評価に不利」と感じれば萎縮して報告が減少する。結果として、改善の回路が細り、同種の危険が繰り返されやすくなる。

情報セキュリティ上の攻撃が安全衛生に波及する経路も無視できない。生成AIフィッシング（Generative-AI phishing：生成AIで上司や取引先を装い、もっともらしい偽指示を作って誤操作させる手口）により、安全装置の停止や危険作業の実行が誘導される可能性がある。敵対的攻撃（Adversarial Example：画像にステッカー等を貼ってAIの認識を意図的に誤らせる入力細工）では、人を検知したり、停止の標識があるのにす

り抜けて機械が止まらない、といった危険が生じる。データ汚染（Poisoning：学習データを改ざん・混入し、AIの判断精度を落としたり、特定状況で誤作動する“裏回路”を仕込むこと）も、安全機能を骨抜きにし得る。要するに、AIが安全の一部を担うほど、サイバー空間の攪乱が物理災害として現れ易くなるため、セキュリティと安全衛生を一体視する必要が高まる。

②メリット最大化・デメリット最小化のための技術的要件と人的要件

AIの職場導入に際して最初に注視すべきは、AIの性能より災害防止の道筋である。安全の分野では、ハザード（Hazard：危険源）と、リスク（Risk：その危険が実際に起きる可能性と重さ）を分けて考える。そしてISO 12100:2010（機械安全の基本）に沿って、①そもそも危険が生まれにくい設計（本質的安全設計）を最優先し、②それでも残る危険はガードや停止機構などの防護で塞ぎ、③最後に「注意書き・手順」などの人的対策で支える、という手順を講じる。要するに、人間が「注意しなくても危険が起きにくい」構造を先につくるという考え方である。

この考え方に照らすと、重大災害に直結する「安全制御」（例：人が近づいたら止まる、挟まれそうなら止まる）に、学習型AIをそのまま活用するのは危うい。学習型AIは現実の多様な状況で“ゼロ誤り”を保証しにくい上、誤りが生じた際に、異常が生じているか、なぜ生じたかを、機械側も人間側も即座に判断できないことがあるからだ。機能安全の国際規格 IEC 61508:2010（Functional Safety：機能安全）は、SIL（Safety Integrity Level：安全度水準）とい

う考え方にに基づき、「安全機能の確実性」を厳しく求めるが、AI にはその証明が難しい場合が多い。従って、AI を安全の中核ではなく、“補助する側”（監視・診断・保全など）に置き、なおかつ Fail-safe（異常時に安全側＝停止へ倒す）を明確にする設計が求められる。結局、AI の性能より既存の安全原則を重視すべきということである。

なお、ISO/IEC TR 5469:2024（AI と機能安全に関する技術報告）によれば、AI の活用法には大別して以下の 3 通りある。すなわち、①安全機能の内部に AI を組み込む、②AI が動かす機械の安全を“AI 以外の安全機能”で守る、③安全関連機能(safety related function) の設計や点検の支援に AI を使う一という整理である。この規格自体は、そのいずれを推奨しているわけでもないが、安全の実効性という点で、①は困難を伴い、②は現実的で、③は比較的良い方法だが万全ではなく、プロセス管理（作業手順に基づき PDCA サイクルを回す管理方法）等を要するとされる。結局、いずれを採用しても、導入前や更新時などに安全効果を確認する V&V（Verification/Validation：検証・妥当性確認）を実施する必要がある、AI を活用すれば安全が担保されるわけではない。

次に重要なのが、AI の“使いどころ”を絞ることである。これを ODD（Operational Design Domain：運用条件の限定）と言い、たとえば AGV(Automated Guided Vehicle：無人搬送車) なら、「人が入らない通路だけ走らせる」、「照明が暗い区域には入れない」、「混雑時は自動運転を止める」などの条件づけを意味する。AI を過信すると、想定外の条件で見落としや誤作動が起きる。

そのうえで、「AI が外した場合の網」とし

て、工学的な防護を講じる。

たとえば人の検知を画像 AI に頼るなら、同時に赤外線センサーやセンサーマットなど従来技術も併設し、どちらか一方でも人を検知すれば止まる“二重の安全網”にする（異種冗長）。また、AI が出す警報を簡単なルールで一次チェックしてから警報を出す「安全の門番」を置くと、誤報による警報疲労を抑え易い。加えて、AI が“最適”と判定しても越えられない安全限界（速度上限、侵入禁止区域、最大負荷、最大作業時間など）をハード側で設定することで、災害確率を大きく引き下げられる。

人的要件は、AI が間違ふ前提で人が関わる仕組みに他ならない。具体的には、AI リテラシー（AI の限界・苦手・誤り方を知る）、自動化バイアス（AI を過信して自身での確認を怠る心理）の自覚、異常時の切り替えの訓練などを指す。さらに重要なのが、現場の人間が AI に異議を言える運用である。AI の指示が危険だと感じたら、止める、上司に申告する、手動に切り替えるなどの“例外ルール”が整っていないと、現場は盲従か潜脱かになり、安全文化が崩れる。AI を導入しても、人間には「監督・検証・例外処理・是正」という新しい仕事が増える。医療でも、誤った AI 提案を混ぜたシミュレーションで「流されない訓練」が行われている。

③残余リスク

対策を尽くしても、AI 固有の不確実性は残る。未知のバグや想定外の相互作用、環境変化による性能劣化、敵対的攻撃、そして人間側の心理（過信や萎縮*）はゼロにならない。AI が統計的推定に依拠する以上、誤検知や見逃しを完全になくすことは困難であ

り、残余リスクを前提に「停止基準」、「手動切替手順」、「復旧手順」、「再発防止のためのログ解析」を組み込む必要がある。さらに、統計的に安全性が改善しても、AI による管理や判断に労働者が納得できない場合、抵抗やルール不履行（指示の無視、抜け道を探す、監視を避ける、ヒヤリハットを隠す等）が別の安全リスクを生む可能性も指摘されている。

＊萎縮：ここでは、AI による監視や評価を恐れて失敗、異論、ヒヤリハットを言い出せず、無難に従う心理。

④現場管理上の課題と現行法制の限界

AI が意思決定に介入すると、事故時の責任分担（開発者、データ提供者、導入決定者、運用者、現場管理者）が曖昧になりやすい。ブラックボックス化は原因究明を難しくし、再発防止や被害者救済を遅らせる。したがって、トレーサビリティ（Traceability：追跡可能性）を確保するログ設計、変更の管理（Change Management：モデルの更新や設定変更の統制〔管理者による意思決定や変更に伴うリスクの共有など〕）、説明可能性（XAI：Explainable AI）の確保といった現場管理上の課題に対し、現行法制は対応できてきない。

以下のような運用上の課題への対応も法定する必要がある。

先ず、AI を導入する以上、それに対応した作業手順の整備が必要である。たとえば、AI 装備のロボットや監視 AI が誤作動したとき、誰がどの画面を見て、どのボタンで止め、どこに連絡し、どのログを保存するのかが決まっていなければ、現場は混乱し、初動が遅れる。従って、まず既存の SOP

（Standard Operating Procedure：標準作業手順）や緊急対応マニュアルを、AI が介入する前提で書き換える必要がある。具体的には、AI が出すアラートや異常表示の意味、AI を手動に切り替える条件、AI の判断に疑義があるときの例外ルート（誰が止める権限を持つか、誰が最終判断するか）を、手順の中に明記する必要がある。

ヒヤリハット（災害には至らないが危険事）や災害が起きたときの調査手順も変わる。従来は現場の聞き取りと設備点検が中心だったとしても、AI が絡むと「AI が何を見て、どう判断し、どう動いていたか」を確認しないと原因特定できない。そこで、災害・ヒヤリハット報告書には、「AI が関与していたか」をチェック項目として追加し、関与している場合、必要なログ（AI の出力、入力データ、警報履歴、モデルや設定のバージョン等）を確実に入手、確認できるよう手順を決めておく。原因解析チームには AI エンジニア（モデルやデータの専門家）を加え、「現場の事実」と「AI 側の挙動」を突き合わせて再発防止策を練られる体制とする必要がある。

こうした運用は部署横断で統一する必要がある。AIQM（AIST 機械学習品質マネジメント指針：機械学習システムの品質を管理する枠組み）などを参照して、使用するデータ、精度の評価方法、変更管理（Change Management：モデル更新や設定変更の条件やそれに伴うリスクの共有、承認権者などの管理）を社内ルールで明示する必要がある。モデルや閾値を少し変えるだけで、誤検知や見逃しの頻度が変わり、現場の安全余裕が変化し得るため、変更管理は特に重要である。

さらにレベルアップを図るなら、ISO/IEC 42001:2023（AIMS：AI マネジメントシステム）等を用い、方針決定（安全とAI の両立方法等）、責任分担（誰が最終責任を負うか等）、資源（人員、教育、監査等）、継続改善（災害やヒヤリハット事例を踏まえた手順とモデルの修正等）を組み込んだ仕組みを、組織として設定し、回すことが望まれる。EU AI Act（Regulation（EU）2024/1689：EU の AI 法）は、高リスク AI（製品安全、バイオメトリクス、重要インフラ、教育、雇用などでの利用）に対してこうした QMS（Quality Management System：品質管理体系）を求めている。

⑤優先的な実行課題

第一に、安全関連機能での安易な AI 依存を避け、AI は支援用途に限るようにすると共に、フェイルセーフと異種冗長で安全網を確保する。第二に、人間中心設計と説明（人間が理解する）可能性を確保し、人間の警報疲労や技能劣化を前提に人間の訓練や演習を設計・実施する。第三に、アルゴリズム管理では、AI による監視と人間の自律性のバランスを取り、AI の機能を測るための KPI（Key Performance Indicator：業績指標）に人の安全・健康を組み込み、業務の過密化等を抑止する。第四に、ISO/IEC 23894 や NIST AI RMF 等に沿ったリスクマネジメントを現場で実施し、AI の誤検知率や見逃し率、AI による提案を人間が拒否した割合、AI による自動運転でヒヤリハットが生じた件数、AI によるアルゴリズム管理等で生じた負荷指標等を KRI（Key Risk Indicator：重要リスク指標）として継続監視し、モデル更新と運用改善を回し続ける。

2. 3 AI のメリット・デメリット、メリット最大化・デメリット最小化の技術的・人的要件—安全工学的観点から—（平岡）

①産業における AI 導入が安全衛生にもたらすメリットとデメリット

産業現場では、AI（Artificial Intelligence：人工知能）が、「危険を見つける」、「起こりそうな異常を予見する」、「物や人の動作の記録をとる」、「ヒヤリハットや災害を記録する」、「危険防止の手順を作る」役回りとして活用され始めている。ここでいう AI は、機械学習（データから規則性を学ぶ手法）だけでなく、統計学習（データの傾向や相関から規則性を推定し、分類、予測、異常検知などに使う手法。機械学習より統計による推論を重視するが、区分は曖昧）、ルールベース（人が決めた規則で判断する仕組み）、最適化（制約の中で最も良い計画を選ぶ計算）まで含む広い概念である。

現場のメリットは三つに整理できる。

第一に、作業計画や配置を調整により危険場面を招く人間の過密や過重負荷を減らせる。

第二に、環境や設備の異常を早く把握し、事故や故障の芽を早めに摘める。

第三に、手順を策定して教育に用いることで人々の安全行動を定着させやすい。

一方、デメリットは「使い方を誤ると負担が増える」点に集約される。たとえば警報が

多すぎれば“警報疲れ”で重要アラートが埋もれる。AI の判断プロセスが説明されないと、現場は無視するか過信するかの両極に振れ、結果として安全が脅かされる。さらに AI が個人の行動や体調に関するデータを扱う場合、監視感や不信が強まると、ヒヤリハット（事故には至らない危険状況）の申告が萎縮し、安全文化が弱る。

したがって、AI のメリットは導入の方法に強く左右される。

②技術事例の分類の考え方（3つの軸）

産業現場での AI の活用にも様々な種類がある。そこで、実在する 154 事例（IPA＝独立行政法人情報処理推進機構の『DX 白書 2023』に掲載された事例）を中心に、補足のため仮想事例 55 件も参考にしながら、AI 活用を 3 軸で整理した。ポイントは、「AI が何をしているか」、「安全衛生への貢献」、「誤作動がもたらす危険の大きさ」にある。

第 1 の軸：AI がしている仕事（技術機能）

AI の役回りは大きく 3 つに分けられる。

認識：センサー（カメラ、音、温度、ウェアラブル等）による知覚。

例：ヘルメット未着用の検出、危険な接近の検出、異常音の検知。

予測：過去の記録（ログ＝作業・設備の履歴データ）から生じそうな危険を予測。

例：設備故障の予兆、熱ストレス上昇の予測、混雑・繁忙の予測。

生成：対応の手順・教材・書類などの作成。

例：手順書や KY（危険予知）の要約、点検票、訓練コンテンツの作成。

第 2 の軸：安全衛生への貢献（OSH 領域）

OSH（Occupational Safety and Health：

労働安全衛生）における AI 活用は A～J の 10 領域に分けられる。

たとえば、A：危険源の検知と除去（機械の停止や遮断など）、B：行動監視（危険行動への警報での注意喚起など）、C：環境監視（作業環境測定など）、D：予知保全、E：計画の最適化、F：疲労の推定、G：教育、H：記録と申告（安全点検の結果や是正措置の記録、ヒヤリハットの報告など）…という具合に、安全衛生への貢献を見える化するための分類である。

第 3 の軸：AI の誤作動がもたらす危険度（Risk Tier）

AI の誤作動の影響度は、以下の 4 段階に分けられる。

R4：AI の誤りが、そのまま重大災害に直結し得る（例：止まるべき機械が止まらない）。

※「制御直結」＝AI の判断が停止・遮断・減速などの機械制御に直につながる状態。

R3：最終制御は人や二重の安全装置が担うが、誤作動が事故確率を大きく上げる。

R2：誤判定が過重労働や健康悪化を招き、中長期的に災害疾病リスクを上げる。

R1：教育・記録のミス等で効果が落ちるが、直接の急性危害は小さい。

この 3 軸への分類から、同じ「AI 導入」でも、認識 AI で人を検知して機械を止める（＝R4 になりやすい）のと、生成 AI で点検記録の下書きを作る（＝R1 になりやすい）のとは、必要な確認や二重化（冗長対応）の必要性がまるで違うことが分かる。

ここで重要なのは、AI が安全の中核（“止める・遮断する”などの制御）に近づくほど、

間違いの代償が大きくなるということである。この前提を踏まえれば、「どの用途にどれだけ厳しい要求を課すべきか」を整理しやすくなる。

③どの領域で導入が進んでいるか（A～Jの分布の意味）

154 件の実装事例を A～J の 10 領域に振り分けると、AI を導入し易い用途が見えてくる。結果として多かったのは、E（作業計画等の最適化）42 件、次いで H（安全点検の結果や是正措置の記録、ヒヤリハット報告などの支援）36 件、F（健康や疲労の状態などのモニタリング）20 件である。

このうち E は、配車、搬送、在庫、ルーティング、シフトなどを最適化して、人間にとって「無理のない段取りを組む」AI である。H は、安全点検の記録やヒヤリハット報告などを収集整理して、後で確認できる記録を残すための AI である。F は、疲労、熱ストレス、転倒リスクなど、人の体調や負担（ヒューマンファクター＝人間の状態や特性が安全に与える影響）を推定して注意喚起する領域で、人が「倒れる前に気づく」AI と言える。

一方で少なかったのは、A（危険源の検知と排除）5 件、D（設備などの予知保全／異常検知）8 件、I（緊急時対応〔異常検知と初動の案内等〕、避難〔安全な退避経路の案内等〕、レジリエンス〔混乱下での機能維持と復旧機能〕）3 件である。

A は、AI の判定が機械等の停止、遮断、減速などの制御に直結する（＝制御直結）領域で、誤作動が重大災害に結びつきやすい。D は、設備やインフラの故障の前兆を捉える領域で、センサーの設置やデータ整備（故

障データと計測値の関係性の維持）、モデル変更時の再学習・閾値調整など、有効性を維持するためのコストがかかる。I は、災害・異常時の判断や避難に関わるため、誤誘導の代償が大きく、導入に慎重になりやすい。

この活用度合いの分布が示すのは、AI を機械等の安全作動（「止める」「遮断する」等）より先に、異常に繋がる情報の整理と運用に使い始めているという傾向である。理由は明快で、E（作業計画等の最適化）や H（安全点検結果や是正措置の記録、ヒヤリハット報告などの支援）、F（人の健康や疲労指標のモニタリング）は、既存設備を大きく改造せずとも導入できる場合が多く、比較的短期に効果（作業の段取りの改善、記録の負担の軽減、体調不良の早期把握）が見えやすい。つまり導入のコストパフォーマンスが良いように見える。

逆に A や D は、AI が現場の安全機能や設備の核心に近づくほど、重厚な誤作動対策（冗長化、フェイルセーフ、復帰手順）が求められ、既設設備の改修等も必要になりやすい。つまり導入のコストパフォーマンスが良くないように見える。

言い換えれば、AI の導入は、先ずは「働き方の改善（E）」、「証跡の整備（H）」、「人の不調の早期把握（F）」から始め、そこで得たデータと運用経験を土台にして、「機械の停止等の危険源の除去（A）」、「予知保全（D）」へ進む余地を残している。すなわち、A や D の導入に必要となるデータや設計を、E・H・F の段階から少しずつ準備しているとも言える。

④主要領域のメリット・デメリットをどう読むか

AI が“止める”側（安全機能の中核）に近いほど安全効果も大きい、誤動作の代償も大きい。“支える”側（作業計画の最適化、点検結果など安全関係の記録、健康パラメータの監視など安全機能のサポート）なら導入しやすいが、運用設計が悪いと、却って正常化のための調整などで人の負担が増える。

A（危険源の検知と排除：ハード・インターロック系）

メリット

AI が危険を知覚した瞬間に、機械の停止・減速・遮断などの制御アクションを取る。インターロック（interlock：危険時に機械が自動停止する安全連動）が働くため、判断→停止が速い。重傷・死亡につながりやすい事象（R3～R4：誤作動時の影響が大きい層）の確率と影響を、同時に下げられる可能性がある。

デメリット

見逃し（止めるべき時に止めない）が致命傷になりうる。一方、誤停止（危険でないのに止める）は生産への影響が大きく、「止まりすぎるから切る」という誤った運用に流れやすい。従って、AI の性能に頼り過ぎず、結局は、冗長化（同じ危険を二重の検知で拾う）、自己診断（センサーや AI が壊れていないかを自分で検知する）、フェイルセーフ（Fail-safe：異常時は安全側＝停止へ倒す設計）、停止後の安全復帰手順（安全確認して再起動する手順）といった安全の原則を反映させる必要がある。

B（行動監視：ソフト・ガード系）

メリット

危険領域への作業者の立入り、接近、危険姿勢、保護具未着用などを見つけて警報を出し、災害を“手前で止める”。機械設備を大きく改造せずに導入でき、広範囲をカバーできるので、導入し易い。また、人の危険行動が起き易いポイントが可視化され、機械設備のレイアウトや配置の改善にもつながりやすい。

デメリット

警報が多すぎるとアラート疲れで「鳴っても無視」が常態化する。また、人の行動をモニタリングするので、丁寧な同意、監視感の軽減、目的外利用の防止など、J（監視、プライバシー、公平性）領域のルールを伴わないと、現場の反発を招き易い。

C（作業環境リスク監視）

メリット

粉じん・ガス・騒音・照度・WBGT（暑熱指数：熱中症リスクの指標）など、環境要因を連続的にモニタリングして基準からの逸脱を知らせる。時系列でばく露（有害要因にさらされる量）管理でき、迅速な換気や作業中断などの判断ができる。データが残るので、監査や法令対応の証拠（トレーサビリティ：追跡可能性。災害予防の場面、災害発生後の原因究明の場面、法令遵守監査の場面で、後から「手順を守った／守らなかった」を確認できる形で記録を残すこと）もできる。

デメリット

ドリフト（測定モデルが現実から少しずつ乖離し、測定値が少しずつ狂う）などのセンサーのずれ、設置条件の影響などで誤判定が生じ得る。閾値（どこから危険とするか）の寛厳の設定も難しい。定期的な校正や

設置位置見直しを行う必要がある。

D（設備・インフラの予知保全／異常検知）

メリット

D は、振動、温度、電流、画像などから機械等の劣化や故障の前兆を掴み、事故になる前に、停止、警報等の介入を行う。突発故障の減少を通じ、安全と稼働率、コスト減を同時に達成し易い。高所や狭隘部などの危険な点検作業も減らせる。

デメリット

機械ごとの個体差（同じ機械でも癖が違う）、季節変動、負荷条件の変化などから誤警報や見逃しが生じる。従って、一度の設定では完遂せず、再学習、閾値調整、監査を含む継続運用（MLOps 的運用＝モデルを育て続ける運用）が前提になる。

E（作業計画・負荷設計の最適化）

メリット

配車、動線、シフト、要員配置などの作業計画を最適化し、作業者の負荷を減らす。焦りや無理が減れば、接触、転倒、ヒューマンエラーなどの危険が減るので、根本的な安全対策になる。

デメリット

コストや効率ばかりを追求すると安全のための余裕が削られ、逆に事故が増える。当初から、安全制約（休憩時間、同時作業禁止、最大人員密度、連続作業時間など）を組み込む必要がある。

F（健康状態・疲労・熱ストレス等の推定）

メリット

疲労、熱ストレスなどの健康パラメータをモニタリングし、休憩提案などで体調起

因の災害を防ぐ。暑熱現場などでは即効性が高い。

デメリット

体調や健康データは個人情報の中でもセンシティブで、目的限定（何のために使うか）、最小化（必要最小限だけ集める）、保存期間、第三者提供の条件設定など J 領域のガバナンス（統治ルール）が求められる。また、現場の人員や工期などの根本条件と両立しないと形骸化したり、そもそも実行されなくなる。

H（手順遵守・記録・申告の支援）

メリット

安全点検の記録の自動化、ヒヤリハット等の報告書の下書き、申告の分類（ヒヤリハット等の報告内容を、原因・作業・場所・危険種別などのカテゴリに自動/半自動で振り分けること）などで、守るべき手順を“証拠として残すこと（災害予防の場面、災害発生後の原因究明の場面、法令遵守監査の場面で、後から「手順を守った／守らなかった」を確認できる形で残すこと）を助ける。是正が早くなり、監査対応の負担も減る。C や D のデータとも結びつき、改善の循環を作りやすい。

デメリット

生成 AI（*）を使う場合、ハルシネーション等が混じる可能性がある。そこで、レビュー手順（誰が最終確認するか）と責任分界（AI の出力を誰が責任をもって承認するか）を先に決め、出力を鵜呑みにしない運用を制度化する必要がある。

* 文章・画像など新しい内容を作る AI。他に、認識したパターンに照らして異常、故障などを析出する認識／検知 AI、時系列デ

ータから故障などを予知する予測 AI などがある。

⑤運用上の要点（結論の骨格）

現場で AI 導入が進んでいる領域を見ると、まず目立つのは、以下の E・H・F である。

E＝作業計画・負荷設計の最適化（配車、シフト、動線など「段取り」を整える）

H＝記録・申告の支援（点検記録、報告書、申告の下書きなど「証跡」を整える）

F＝健康状態・疲労・ヒューマンファクター評価（疲労や熱ストレス等を推定し「不調の芽」を拾う）

この 3 つに運用が偏る理由は、多くの場合、既存の機械設備を大改造しなくても、まず「情報の扱い方」を変えることで始められるからである。言い換えると、生産の現場は、AI をいきなり「機械を止める側」で活用するより先に、働き方（計画）と、記録と監視、手順の支援での活用から入っている。

ただし、重大災害の防止効果という観点では、以下の少数活用領域に大きな伸びしろがある。

A＝危険源の検知と排除（停止・遮断などの制御に直結する“ハード・インターロック系”）

D＝設備・インフラの予知保全／異常検知（早めに兆候をつかんで介入する）

これらは、うまく機能すれば、重大災害の確率を直接下げられる。しかし、A は誤作動が重篤災害につながり易いので保守的運用に偏り易く、D はデータ整備やメンテナンス（特に更新・監査）に手間がかかる。そのため現実には、検証を重ねつつ段階的に活用の比重を高める方法が現実的である。

では、どのような運用が「失敗しにくい」か。骨格は以下の四本柱になる。

①リスクに応じた要求水準の設定

誤作動時クリティカル度（Risk Tier：誤作動が安全に与える影響の大きさ）で、R4（最も重大）に近い用途ほど、冗長化・検証・手順・承認を厳しくする。逆に R1（教育・記録など急性危害が小さい用途）では、人の最終確認を前提に、過度に保守的な仕組みにしない。

②個人データ取扱いの明確化

B（危険行動等の監視）や F（疲労や健康指標のモニタリング）のように、位置や行動、健康データを取り扱う場合は、本人同意、利用目的の特定（何のために使うか）、利用の最小化（必要最小限だけ取得し、利用する）、保存期間、第三者提供の有無や条件を、分かりやすく定める。

③責任分担の整序

AI が関わると事故が起きても、「誰の判断のせいかわからない」が不明確になり易い。導入の決定者、現場管理者、運用担当、開発・保守担当のそれぞれが、誤作動時にどのような対応責任（連絡、停止、復旧、報告等）を負うかを決めておく。責任が曖昧だと、事故後の原因究明も是正も遅れる。

④参加型の設計と定期的な振り返り

AI は導入後の運用やメンテナンスが重要な意味を持つ。現場の声（警報が多すぎる、提案が非現実的、記録が負担等）を拾い、定期的に設定や手順を直す。また、労働者の参加を前提にしないと、隠れた逸脱や萎縮（ヒヤリハットを報告しない等）を生みやすい。

以上をまとめると、現状では導入し易い E（作業計画の最適化）・H（記録・監視によ

る手順の支援）・F（健康指標のモニタリング）から広がっているが、重大なリスク防止効果の観点では A（危険感知によるインターロック等）・D（予知保全等）の伸びしろが大きい。いずれの領域でも、①用途の危険度に応じた寛厳の調整、②個人データの取扱いの透明化、③関係者の分担責任の規定、④現場参加での設計と運用が、メリットを伸ばしデメリットを抑える最短ルートになる。

3 産業医学・人間工学的分析 (中澤・榎原)

職場で AI（Artificial Intelligence：人工知能）が使われる機会が増え、産業保健にも影響が及びつつある。JILPT（独立行政法人労働政策研究・研修機構）の労働者向けの調査では、調査対象のうち AI が職場で使われているとの回答は 12.9% だった。

メンタルヘルス領域では、AMED の事業として、日本産業衛生学会が中心となって DHT（Digital Health Technology：アプリ等のデジタルによる健康管理）に関する予防介入指針が策定されたが、その中で、生成 AI（Generative AI：文章などを作る AI）を使う場合の運用ガイドラインや、効果のエビデンス蓄積が課題として挙げられている。さらに LLM（Large Language Model：大規模言語モデル、ChatGPT の中核であり、大量の文章の学習により、文章の途中でも次に来る語を予測する。AI の言語能力を飛躍的に高めた）の登場で、その医療・健康サービスへの応用に関する研究も増えている。

そこで、日本の産業保健で「AI を用いた健康管理サービス」がどの程度存在し、どんな課題があるかを整理した。

ウェブ検索での抽出の結果残った 135 サービスの内訳は安全関連が多く（82 件）、健康管理に直接関係するサービスは 58 件程度だった。他方、経産省が職域向けに周知を図っているメンタルヘルス・サービスの選択支援ツール（ウェルココ）に掲載された 109 プログラム中 AI 利用が明確だったのは 2 件のみだった。これら（135 サービス＋2 プログラム）がいう「AI」は、①予測 AI（機械学習等でリスクを推定する）と②生成 AI（チャットボット等で対話・文章生成する）を混在させていた。

期待されるメリットとして、健康意識の向上、うつ症状の軽減や欠勤減少などが文献で示されている。AI 特有の利点としては、予測 AI が健診・ストレス等のデータを集約・分析することで、人事・産業保健職の負担を軽減し得ること、生成 AI チャットボットがセルフケア促進や一次相談の窓口になり得ること、表情・音声・ウェアラブル等から従来拾いにくかった指標を捉えられる可能性が挙げられた。

しかし、様々な課題がある。

第一に、「本当に効果があるのか」を確認する研究が足りない。それを確かめる代表的な方法が RCT（Randomized Controlled Trial：ランダム化比較試験。参加者を“くじ引き”のように無作為に「サービスを使う群」と「使わない群」に分け、結果を比べる方法）だが、AI を使った産業保健アプリでは、良質な RCT が少ない。確認できた例も、運動支援型の AI 健康プログラムで筋骨格系の症状（腰痛や肩こりなど）の改善を見たとい

った、ごく限られたものだった。

第二に、費用対効果を判断する研究も不足している。産業保健では、医療費だけでなく、労働逸失も関心事項となるので、ROI（Return on Investment：投資利益率）や、欠勤（休む日数）だけでなく、プレゼンティーズム（presenteeism：出勤はしているのに体調不良や不調で能率が落ちる状態）まで含めて経済的効果を評価する必要がある。しかし、AI を使う産業保健サービスに絞って費用対効果を検証した研究は、現時点ではほとんど見当たらない。

第三に、使い方のルールが整っていない。英国では HSE（Health and Safety Executive）が、職場で AI 利用方法に関するガイダンスを出している。一方で日本では、産業保健・安全衛生に特化した AI ガイドラインがまだ作成されていない。医療・ヘルスケア分野の生成 AI ガイドラインは存在するが、対象範囲が限られており、産業保健サービスにそのまま適応しない。

第四に、AI の種類ごとに注意点が異なる。予測 AI（過去データから将来を推定する AI）では、偽陽性／偽陰性が問題になる。偽陽性は「本当は危険ではないのに危険と判定されること」で、偽陰性は「本当は危険なのに大丈夫だと判定されること」である。さらに、学習データ（AI が勉強した元データ）と、対象集団の特徴が違っていると、精度が急に落ちることがある（いわゆるドリフト）。このため、TRIPOD+AI（予測モデル研究が編み出した報告チェックリスト。どんなデータで学んだか、どの程度の的中率か等を見える化する基準）などに沿って、性能を管理する必要がある。

生成 AI（Generative AI：文章・画像など

を作る AI。チャットボット相談などで使われる）については、評価軸として、正確性（内容が正しいか）、信頼性（情報管理、間違いで損害が出たとき場合の責任者の特定）、共感性（相談者の気持ちに沿った返答ができるか）、性能（動作の安定性など）などが挙げられている。

以上を整理すると、AI 活用を標榜する産業保健サービスは増えているが、概ね、効果、安全性、費用対効果を裏付ける科学的評価と、その結果の公開（情報開示）を伴っていない。今後は、これらを測る客観的な指標を整え、ウェルココ（経産省が周知を図っている職域向けの心の健康サービス選択支援ツール）のように、利用者がサービス同士を比較できる形での公開を促す必要がある。さらに、AI は進歩が速いのでため、導入時点評価だけでなく、導入後の追跡評価枠組み（アップデートや運用変更も含めた継続チェック）が必要だと解される。

4 心理学的分析（島崎）

AI 導入が職場にもたらす「心理社会的リスク」を、①個人、②行動、③組織の 3 段階で整理した。出発点は、AI の影響が単に仕事の効率化にとどまらず、労働者の不安、期待、孤立、同調などの感情を通じて、人々の安全行動や組織の意思決定にまで波及し得るという見取り図である。労働法は AI 導入による雇用や労働条件の変化を扱うが、心理学は、その変化がメンタルヘルスや行動に与える影響を検討し、それなしに導入の成否は評価できないと考える。

個人レベルでは、AI はメリットとデメリットを同時にもたらす。

メリットとしては、単調な作業を AI が代替することで認知的負荷（頭の疲れ）が減り、より創造的で意味のある仕事に集中しやすくなる可能性が挙げられる。AI が研修を個別最適化（一人ひとりに合わせる）ことすれば、スキル開発も進む。さらに AI の支援は困難な作業も可能になること等で自己効力感（self-efficacy：自分はできる、という感覚）や職務満足度を高め得る。

一方、デメリットとしては、「AI の存在を強く意識すること（AI awareness）」自体がストレス源になり得ることが挙げられる。すなわち、AI に合わせた仕事の方法やペース、WIF（Work-Family Interference：仕事が生計生活を圧迫する状態）、職務不安（job insecurity：AI に仕事を奪われる不安）などにより、感情的疲弊（emotional exhaustion：消耗して燃え尽きに近い状態）が高められる、という経路が報告されている。さらに AI 統合（AI の業務への組み込み）は、場合によっては、創造的自己効力感（creative self-efficacy：創造性に対する自信）を低下させ得る一方で、職務自律性（job autonomy：仕事のやり方を自分で選べる自由度）を増やし得る、といった二面性も示される。つまり AI は「裁量が増えるのに、自信が揺らぐ」という矛盾した心理を生み得る。

行動レベルでも、功罪両面があり得る。

ポジティブ面では、AI が人々のタスク分担や働きぶりのフィードバックを支援することで、チームの協働が円滑化し得る。また AI を人々の成長支援に活用することで、エンゲージメント（仕事への前向きな関与）や

イノベーション行動が伸びる可能性がある。

しかしリスクとして、CWB（Counterproductive Work Behavior：逆機能的行動。サボり、規範違反など組織に損害を与える行動）が問題になる。

まず、AI と協働する仕事が増えると、人と話す機会が減って「孤立感（loneliness）」が強まることがある。孤立すると気持ちがすり減り（疲弊）、やがて CWB（Counterproductive Work Behavior：逆機能的行動＝サボり、規範違反等）に傾きやすい、という連鎖が報告されている。これは AI の誤用、濫用、情報漏洩、安全手順の形骸化などを通じて職場の安全にも波及し得る。

次に、人間が AI の出力に迎合する sycophancy（追従傾向）をもたらし得る。AI がこう言うなら正しいだろう、と自分の判断を抑えて反論や検討をしなくなる。すると、AI が間違っているにもかかわらず気づきにくくなり、誤情報や偏った見方がそのまま広がりやすい。

そして、AI の利用者が同じような情報を検索したり、質問傾向を偏らせるなど、レコメンド（recommend：推薦）機能が強い情報環境では、エコーチェンバー（echo chamber：特定の意見や好みに合う情報ばかりが回答される状態）が起きやすい。すると、組織の議論が同じ方向に偏り、反対意見やリスクの兆候が見えにくくなり、「集団的盲点」を大きくする可能性がある。

さらに、AI が“正解っぽい案”を常に出す結果、人が自分で考えたり試行錯誤する意欲（創造性への挑戦）が弱まることもある。

以上から、AI を補助に置き、人が異議を言える、別案を出せる、最終判断できる設計

（主体性を残す設計）が重要になる。

組織レベルでは、AI 導入は、公平性、裁量、信頼、文化に影響する。

メリットとしては、定型業務の自動化で人間が戦略的業務に集中でき、効率と競争力が高まる可能性がある。導入プロセスに従業員を参加させればコミットメント（組織への関与）が高まり、支援的風土が育つともされる。採用等で AI が人間の偏見を減らせば、公平性と多様性が向上し、定着にも寄与し得る。

他方、デメリットとしては、雇用不安と格差拡大、人間の裁量の縮小によるエンゲージメント低下、心理的安全性（psychological safety：失敗や異論を言っても罰されない安心感）の低下が挙げられる。AI による監視やアルゴリズム評価が強まると挑戦的行動が抑制され、学習と革新が停止し得る。さらに内部者（insider）による悪用（AI システムを使ったデータ改ざんや機密窃取）も、心理要因と結びついた組織リスクとして指摘される。ILO（International Labour Organization：国際労働機関）や EU-OSHA（European Agency for Safety and Health at Work：欧州安全衛生機関）の文書も、AI と心理社会的リスクを組織単位で扱う必要を繰り返し述べている。

以上から、AI は「性能の良し悪し」だけで評価できず、人の心理資源（不安・疲弊・孤立・自信）を媒介して、その行動と組織文化を変える“社会的存在”として捉える必要がある。

対策として、対個人ではメンタルヘルス・リテラシー（心の健康の知識）やストレス対処教育、孤立を防ぐソーシャルサポート（上司・同僚の支援）とキャリア支援を講じる。

対行動面では、CWB を生みにくい孤独回避の支援体制、AI への迎合やエコーチェンバーを生むような教習データの偏りを抑える設計など、多様な視点をもたらず取り組みを図る。

対組織面では、人間中心の価値を基本として、人の心理的安全性を守りつつ、AI の判断経路の透明性と公平性（性別・年齢など特定属性で不利な結果が偏って出ないこと。設計や教習データで調整する）を担保する制度設計を行う。

結局のところ、AI の真価は「AI の性能」ではなく、人・技術・制度が調和する形で導入されるかにかかっている。

5 社会調査結果の分析（西本）

同じ AI (Artificial Intelligence：人工知能) でも、国や文化によって受け止め方は変わる。そこで日本の労働者を対象に、AI の利用状況、労働安全衛生（OSH：Occupational Safety and Health＝職場の安全と健康）上のリスク認識、心理反応、対応策の実態や、運用と責任の仕組みに関する意見等を全国 Web 調査で把握した（N=1000、2025 年 6 月実施）。調査票は、建設業・製造業で AI 導入に関わる実務者への事前インタビューと、英国 HSE（Health and Safety Executive：英安全衛生当局）等による先行調査の知見を踏まえて設計された。

実装については、AI に関する「関心は高いが実装は限定的」な傾向がみられた。企業としての AI 導入は「本格活用＋一部導入」で約 2 割程度にとどまり、「導入予定なし」がほぼ半数、「分からない」も約 2 割あった。

個人側でも「AI をまるで知らない」層は 1 割程度にとどまる一方、業務で AI を使っている人は約 2 割にとどまり、利用目的は RPA（Robotic Process Automation：定型事務を自動化する仕組み）や翻訳、文字起こしなど、業務の効率化が中心だった。安全衛生や人事のような「人に関わる領域」での利用は限定的で、導入中・導入予定の技術としては生成 AI（Generative AI：文章などを作る AI）が半数程度と存在感を増していた。

AI 導入による OSH リスクについては、強い懸念を示す層が少数で、多くは判断保留だった。「AI を過信して人が確認を怠る」、「人間の注意力の低下」、「AI の判断の偏り、不正確性」への懸念は 1～2 割台にとどまる一方、各項目で「どちらともいえない」が約 3 割、OSH リスクをもたらすようなタイプの AI（安全監視 AI、意志決定支援 AI、生成 AI など）を使っていない」が約 4 割を占め、現場での接触の薄さが評価の曖昧さにつながっているとみられた。ただし、AI に関与が深い層（導入に関わる、設計・運用に触れる等）ほど、AI 導入による OSH リスクを高く見積もる傾向があり、とくにブラックボックス化（Black box：AI の判断理由が見えにくいこと）や、AI 任せで人の安全衛生感性が鈍る危険を強く意識していた。

興味深いのは、AI 関与が深い人物ほど AI を「危ない」と見る一方で、「健康にプラス」とも見る傾向が窺われた点である。すなわち、AI 関与が強い層ほど、AI 導入が、業務の効率化、危険業務の代替、疲労やストレスの評価、AI が相談相手になるなどにより、心身の健康に良い影響を与えると考える割合が高く、心理社会的リスク（心理・人間関係・職場文化に関わるリスク）や監視による

プライバシー侵害への懸念は、強くなかった。

AI 導入による OSH リスクへの対応策は、「組織内での人手」が中心で、外部監視の活用はあまり実施／想定されていなかった。人による最終確認、AI に任せない業務の区分、現場教育、リスク発生時の手順整備については、「ある程度／十分に実施」が 3-4 割前後である一方、外部専門家の監査・レビューは 2 割程度にとどまり、自己点検が主流だった。リスク対応上、今後必要な仕組みとしては、説明可能性（Explainability：判断根拠の見える化）、現場適合性のトライアル（導入前に現場で試して合うか確かめる）、AI に関わる人の責任分担の明確化、専門家連携がいずれも 4 割前後で支持され、必要性の共有認識はあるが、実際の体制整備は過渡期であることが窺われた。教育は重要とされながら、十分実施されていない実態も窺われた。

AI に「どこまで任せるか」については、分野で差が出た。安全衛生や人事のように人命や重要な権利に絡む領域では、「最後は人が判断」が基調だった一方、自律走行搬送ロボット、自動運転フォークリフトなどの技術領域では AI 判断を受け入れる回答が増える。ただし全体としては中立（どちらともいえない）の回答が多い。実際、そうした自動運転機器を AI に委ねることについて AI 利用経験者では賛成が多い一方、導入決定経験者では賛成が少なく、知識と責任感が慎重さを生む可能性が示唆された。人事判断を AI に委ねることは、AI 関与が強い層でも賛否が拮抗した。

AI 利用に伴う責任配分については、ルールによる明確化を求める声が強い。AI 起因

の労災が起きたときの責任決定ルールを求める回答が多く、導入決定経験者は、責任の重さを「利用現場の責任者→訓練データ提供者（学習用データを用意した者）→導入決定者→プログラマー」の順と考える傾向が窺われた。AI の判断過程の説明責任を法定すべきとの意見も、導入経験者で過半に達し、責任主体はプログラマーと訓練データ提供者とする見方が強かった。他方、導入コストを重視する回答も、導入決定経験者で約 8 割にのぼった。

最後に、今後 5 年の AI による業務代替見通しは不確実だった。「分からない」が約 3 割で最多で、全面代替を想定する層は少なく、限定的自動化を見込む層が中心だった。自動化が進む場合の社会的対応（関係者の再教育・リスクリング、職務代替に対する所得保障、フリーランス保護、若年層支援）を「必要」とする回答は 3～4 割台にとどまり、「どちらともいえない」が約半数で、必要性の認識はあっても、具体像は共有されていない。

総括として、導入は進むが運用整備が後追いという構図が確認された。

AI 導入の状況は企業や人によって大きく違うため、理想的な運用を一律かつ一気に求めるのは現実的でない。軌道に乗せるには、段階を踏んでガバナンス（導入・運用を支えるルールと体制）を整えるのが現実的である。具体的には、まず、①現場トライアル（実際の職場で小さく試し、現場に合うか・危険が出ないか確かめる）から始め、次に、②役割と責任の明確化（日常的には誰がどこまで決定し、最終判断は誰が行い、事故時に誰が説明するかを決める）を行う。そのうえで、③教育と手順の整備（AI の限界、

過信防止、異常時対応の手順を現場で共有する）を進め、運用が本格化すれば、④外部レビュー（社外の専門家による点検・監査で内輪の見落としを減らす）を取り入れ、最後に、⑤関係人材の育成（運用・監査・改善を継続して回せる人材を確保する）へつなげる。

ただし、この順序は「全員に同じペースで当てはめる」のではなく、AI への関与の深さに応じて、重点を変えて実装することが求められる。たとえば中小企業で AI にほとんど触れていない層では、いきなり外部レビューよりも、①トライアルと、最低限の②役割分担、③手順づくりが中心になる。大企業でも、実務経験が少ない現場層には、③教育と手順の整備や、現場適合性の検証を厚めにすべきである。中小企業の経営層のように、関心は高いが実務経験が薄い層では、まずは①トライアルとコスト感の把握を優先し、成功パターンが見えた段階で、②責任分担や運用体制の整備へ進む方がスムーズだろう。一方、大企業で実際に導入・運用に関わっている実務層は、既に「使っている」からこそ、説明可能性（判断根拠を示せること）や過信の抑制、責任ルール、外部レビューといった高度運用の課題が前面に出る。このように、関与の程度に合わせて実装の重点をずらしながら、全体として安全に AI 活用を広げる方向が提案される。

DE. 考察及び結論

要約所掲の通り。

F. 研究発表

1. 論文発表

三柴丈典. 2025. 「安全衛生規制の変革：Industry 4.0 時代における戦略的対応」産業保健法学会誌 4(1), pp.265-280 (Takenori Mishiba. 2025. Transforming Occupational Health and Safety Regulation: Strategic Pathways in the Era of Industry 4.0. *Journal of Work Health and Safety Regulation* 3(2). 151-169 の第 2 次出版)

2. 学会発表

Takenori Mishiba. The Future of OHS Regulation in the AI Era: A Japanese Perspective

Online International Conference “OHS Policy in the AI Era” (organized by JAOHL and Japan NIOSH)

July 10, 2025 – Online

日本の厚生労働省、労働安全衛生総合研究所のほか、EU-OSHA、HSE、Germany 建設業労災保険組合が参加したオンライン国際カンファレンス。

報告者が企画統括した。

カンファレンスレポートは、

Journal of Work Health and Safety Regulation 4(2)に掲載

https://www.jstage.jst.go.jp/article/jaohlev/4/2/4_cor.25-008/_article/-char/en

同会議とカンファレンスレポートは、EU-OSHA の WEB サイトに掲載

<https://osha.europa.eu/en/oshnews/exploring-osh-policy-ai-era-conference-report-out-now>

Takenori Mishiba. The Future of OHS Regulation in the AI Era: Results of the Social Survey Conducted in Japan
Safety, Health and Well-being at the World Expo 2025 in Osaka, Japan
July 16, 2025 – INTEX Osaka, Japan

Takenori Mishiba. Legal Framework for Occupational Mental Health in Japan
Safety, Health and Well-being at the World Expo 2025 in Osaka, Japan
July 18, 2025 – INTEX Osaka, Japan

G. 知的所有権の取得状況

1. 特許取得

該当せず。

2. 実用新案登録

該当せず。

3. その他

該当せず。

H. 引用文献

分担研究報告書の総括なので該当せず。

