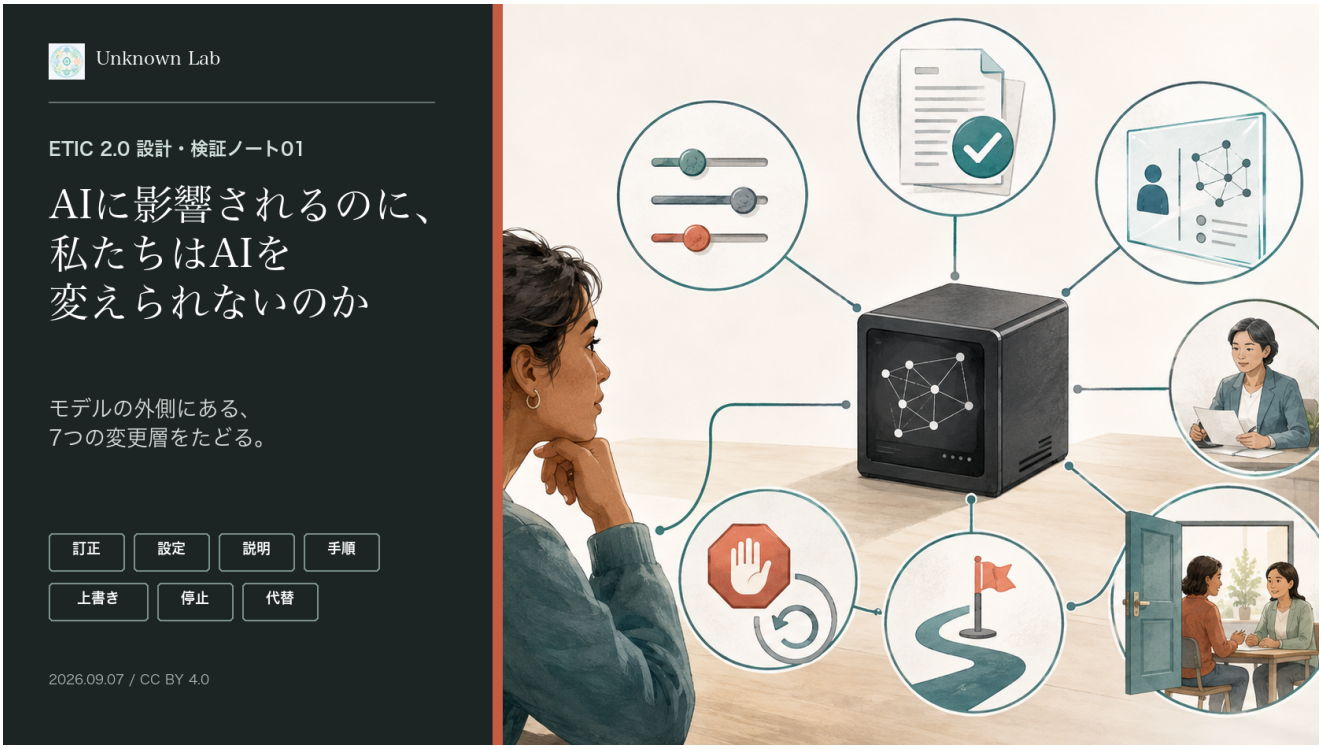
AIに影響されるのに、私たちはAIを変えられないのか

ケアの対象として考えるための「変更できる範囲」の見取り図



Daiki Morimoto / Unknown Lab / ORCID 0009-0009-6444-750X
Research Note / Version 1.0 / 公開候補日 2026-09-07
DOI: 10.5281/zenodo.22487001 / 未査読 / CC BY 4.0

はじめに

もし、健康アプリがあなたの生活を「活動不足」と判定したとする。

けれども、その日は体調を崩した家族を支えながら、家の中で何度も立ち上がり、食事を作り、連絡を取り続けていた。歩数には表れにくくても、その人にとっては大切で負担の大きい一日だった。

記録が違えば、入力を直せるかもしれない。通知が多ければ、設定を変えられるかもしれない。提案が生活に合わなければ、専門職が採用しない判断をすることもできる。

では、判定の仕組みそのものが合っていないときはどうだろう。

誰に伝えればよいのか。いつまでに返事が来るのか。モデルを作り直せないなら、AIを使わない支援へ戻れるのか。そもそも、本人や支援者が変えられないものを、ケアの介入対象と呼べるのだろうか。

この問いは、AIの専門家だけのものではない。

ケアはこれまで、本人の心身機能だけでなく、住まい、道具、人の支え、制度、仕事や学校の環境を調整してきた。作業療法も、望む生活や作業への参加を支えるため、物理環境に加えて、製品・技術・支援、サービス、制度、政策を扱っている[20][21]。

AIが、何を見せるか、誰を優先するか、どの選択肢を勧めるか、専門職がどこへ時間を使うかに影響するなら、AIも生活環境の一部になり得る。

ただし、影響することと、変えられることは同じではない。

本稿では、AIを「書き換えられない黒い箱」か「自由に調整できる道具」かの二択にしない。AIモデルの外側にある変更の層を分け、本人や支援者が実際に介入できる条件を考える。

「AIを変える」は、一つの操作ではない

AIを変えると聞くと、モデルを再学習したり、プログラムを書き換えたりする場面を想像しやすい。

その技術的な調整は、ここではチューニングと呼ぶ。学習、パラメータ、判定の閾値、検索や生成の設定などを調整することだ。

しかし、ケアの場で変えられるのは、モデル本体だけではない。

例えば、次の操作はそれぞれ違う。

- 誤った生年月日や病歴を訂正する。
- 通知頻度や判定の基準値を設定する。
- AIの不確実性や別の選択肢が分かるように表示を変える。
- AIを使う場面、確認する人、再評価の条件を業務手順で変える。
- 個別のAI提案を採用せず、専門職が判断を上書きする。
- 判断に異議を申し立て、人に再検討を求める。
- 危険が疑われるときにAI利用を停止する。
- AIを使わない支援や別の製品へ切り替える。

モデル本体を変えなくても、その人が受ける結果を変えられる場合がある。反対に、「ご意見を送信できます」と表示されていても、仕様、判断、運用のどれも変わらない場合がある。

大切なのは、変えられる機能があるかではない。誰が、何を、いつまでに、実際に変えられるかである。

人が見ていても、変えられるとは限らない

医療AIでは「最終的には人が判断します」という説明がよく使われる。

しかし、人が画面を見ることと、その人がAIに逆らえることは同じではない。

健康領域のmeaningful human controlを調べたHilleらの系統的レビューは、42件を整理したうえで、健康領域で頑健な概念はまだ確立していないと結論した[1]。

van de Sandeらは2026年の論考で、意味のある監督には少なくとも4つの条件が必要だと提案した[2]。

- AIが何をでき、どこで間違うかを知るための知識。
- AIとは別に考え直せる時間。
- AIに同意しない判断をしてもよいという権限。
- AIの動作や結果を止め、変え、取り消せる介入手段。

この4条件は、効果が確立した評価尺度ではない。それでも、「人がいる」という表示だけでは監督の実効性を判断できないことを具体化している。

例えば、専門職に上書きボタンがあっても、押すたびに長い理由入力が必要で、忙しい時間帯には使えないかもしれない。上書きが評価上の不利益になる組織では、権限が画面上にあっても使いにくい。AIがすでに予約、振り分け、連絡を自動実行した後では、判断を取り消しても間に合わないことがある。

EU AI Actも、適用対象となる高リスクAIについて、人が出力を無視、上書き、反転し、必要時には停止できることや、監督者に能力、訓練、権限、支援を与えることを規定している[15]。これは日本の全AIサービスに直接適用される規則ではない。しかし、監督を「人が関わったか」ではなく、「結果を変える権限と手段があったか」で見る重要な例である。

説明、訂正、異議申立ては、それぞれ別の道である

AIについて説明を受けることは大切だ。

PlougとHolmは、AI診断へ実効的に異議を申し立てるには、使われたデータ、潜在的な偏り、性能、AIと専門職の分業についての情報が必要だと論じた[3]。

けれども、説明を受けられても、判断が変わるとは限らない。

日本の個人情報保護委員会は、行政機関等を対象とするQ&Aで、保有個人情報の訂正請求は事実でない内容を対象とし、評価や判断の内容は対象外になると説明している[17]。

これは、少なくとも次の3つを分けて設計する必要があることを示す例になる。

- 事実を訂正する道: 氏名、記録、入力、ラベルなどを正す。
- 評価を問い直す道: AI出力の根拠、偏り、適用範囲を再検討する。
- 判断を変える道: AIを用いた個別判断を上書き、再審査、停止する。

データを直せても、評価が変わらない場合がある。評価の問題を説明できても、最終判断を変える権限者へ届かない場合がある。

Cohenらは、すべてのAI判断を人が事前確認する方式とは別に、申立てがあった案件を人が再検討するhuman-on-appealを提案した[4]。個別事情や数値化されていない情報を人が加えられる可能性がある一方、著者ら自身も異なる方式を比較した直接的な実証証拠は少ないと述べている。

異議申立て窓口を作るだけでは足りない。申立てを支援する人、費用、期限、再検討者の権限、結果の通知、同じ問題を繰り返さない仕組みまで必要になる。

参加したことと、設計が変わったことは同じではない

「患者や市民の声を聞いた」という説明も、そのまま変更可能性の証拠にはならない。

Frostらは、医療AIに関する市民の見解を調べた62研究を整理した。47研究は参加者へ背景情報を提供していたが、事前知識の不足はなお研究上の限界とされていた。48研究は、市民の見解をAI統治へ生かす提言をしていた[5]。

これは、当事者にAIの専門知識がないから参加できない、という考え方を見直す材料になる。必要なのは、専門家と同じ技術知識を求めることではない。何が決まり、どんな不利益が起こり得て、どこを変えられるのかを、判断できる形で伝えることである。

一方、参加の時期と権限には大きな差がある。

Bainesらのデジタルヘルスに関する系統的レビューでは、10,540件から433報が採用された。患者は、変更への影響力が限られる最終段階のユーザビリティ試験へ参加することが多かった[6]。

Loftusらのレビューでは、約10,880件の医療AI研究のうち、コミュニティ関与を記載したのは21件、0.2%で、地域の関係者をアプリケーション設計へ含めた報告は1件だった[7]。

ここでの0.2%は、AIの影響を理解している人の割合ではない。医療AI研究の報告のうち、コミュニティ関与を記載した割合である。一般の人の理解度を示す数値に読み替えてはならない。

さらに、Dantasらが2026年に公表したがん領域の系統的レビューでは、40研究中32件、80%が設計・初期開発に集中し、統治・導入後監視への参加を報告した研究はなかった[8]。40研究の大半は一般的なデジタルヘルスで、AIだけを扱った割合ではない。それでも、参加を最初のアイデア出しだけで終わらせず、更新、監査、停止の段階まで続ける必要性が見える。

参加を評価するときは、人数や会議回数だけでなく、次を記録すべきである。

- 誰が議題を決めたか。
- どの提案が採用されたか。
- どの仕様、運用、評価項目が変わったか。
- 採用されなかった提案に、どんな理由が示されたか。
- 誰に上書き、停止、撤退を決める権限があったか。
- 決定後の結果が本人へ返されたか。

意見を言えることは、参加の入口である。決定を動かせることとは分けて確かめなければならない。

「あなたが変われればよい」という救済にしない

AIによる不利な判断を変える研究には、algorithmic recourseという領域がある。

Ustunらは、固定された線形分類器から望ましい結果を得るため、本人が変更可能な入力変数を変える能力としてrecourseを定式化した[9]。年齢のように変えられない属性と、収入のように変化し得る項目を分ける発想である。

しかし、望ましい結果になる入力例を示すことと、その人が現実に実行できることは同じではない。

Karimiらは、反実仮想の説明から、因果的な介入へ考え方を進める必要があると論じた[10]。後のレビューでは、recourseを、不利な自動判断を受けた人への説明と行動提案を扱う研究領域として整理している[11]。

ただし、これらの研究は信用判断などを主に扱い、医療や作業療法で健康を改善する方法として検証されたものではない。

もう一つ注意が必要である。

AIの判断を変えるために、「体重を減らしてください」「もっと働いてください」「利用履歴を増やしてください」と本人だけに変化を求めれば、モデル、データ、組織の基準、社会的な障壁を変えないまま、負担を本人へ移すことがある。

変更可能性は、本人の特徴だけを探すことではない。

本人、専門職、組織、開発者、調達者のうち、誰が何を変えるべきかを比較することである。

「いつか変えられる」では遅いことがある

異議申立てが正しく受理されても、返答が半年後なら、今週の退院先、明日のサービス利用、今回の給付判断には間に合わない。

De Toniらは2025年、行動に時間がかかり、その間に環境やデータ分布が変わると、提案されたrecourseが無効になり得ることを示した[12]。これは医療AIの申立て期間を調べた研究ではないが、変更可能性に時間を含める重要性を示す。

ケアでは、同じ一か月でも意味が違う。

- 緊急の治療判断では数分が長い。
- 退院支援では数日が重要になる。
- 福祉用具の調整では試用と再評価に数週間かかることがある。
- 組織の調達変更やモデル再検証には数か月以上必要な場合がある。

したがって、「技術的には変更できます」だけでは足りない。その人の生活上の期限までに間に合うかを確認する必要がある。

間に合わない場合は、個別判断の上書き、一時停止、人による対応、非AI経路など、別の層を先に動かさなければならない。

AIの外側にある7つの変更層

以上を、AIを含むケア環境の7つの層へ整理する。

層	変えられる可能性があるもの	主な方法	変わらないかもしれないもの
1 記録・入力	本人情報、記録、ラベル、利用データ	訂正、追加、削除、利用範囲の指定	モデルの評価基準、最終判断
2 設定・閾値	通知、基準値、検索・生成設定	設定変更、技術的チューニング	学習済みモデル全体、契約
3 表示・説明	不確実性、警告、選択肢、根拠	画面改修、説明追加	出力自体、意思決定権限
4 利用方法・業務手順	使う場面、確認者、再評価条件	手順変更、研修、役割分担	ベンダーのモデル、上位方針
5 個別判断	AI提案の採否、再検討	上書き、異議申立て、人の再審査	同じ問題を生むモデル・制度
6 停止・復旧	AI利用、更新、旧版への復帰	一時停止、利用制限、ロールバック、廃止	すでに生じた不利益
7 代替・調達	非AI経路、別製品、契約、監査条件	選択、切替え、契約変更、政策参加	すぐには変えられない市場・制度

NIST AI RMF 1.0とPlaybookは、実行可能な非AI代替、AIの上書き・切離し・無効化、導入後の上書き記録と監視を扱う[13]。これは任意利用の枠組みで、日本の法的義務ではない。それでも、停止や代替を例外的な「故障時対応」ではなく、最初から設計する考え方を示す。

WHOは、健康AIの設計、導入、利用の中心に倫理と人権を置き、影響を受ける個人や地域に対する説明責任と応答性を求めている[14]。日本では、経済産業省のAI事業者ガイドライン第1.2版が、開発者、提供者、利用者によるライフサイクル全体の統治を扱う横断的な参照になる[19]。どちらも、個別製品の安全性や効果を保証する資料ではない。

変更後の確認も必要である。

FDAのPCCP最終ガイダンスは、AI対応医療機器について、予定された変更、変更を開発・検証・実装する方法、影響評価を結びつけるよう推奨する[16]。日本でもPMDAのIDATENが、規制対象医療機器の変更計画を扱う[18]。

これらは一般のAIサービスすべてに適用される制度ではない。しかし、AIは一度評価して終わる固定物ではなく、モデル、データ、設定、画面、使い方が変わるたびに、以前の安全性や公平性をそのまま引き継げない可能性がある。

変更できることと、変更してよくなったことも別である。

変更可能性を確かめる8つの質問

7つの層のどこかに機能があるだけでは、介入可能とは言い切れない。

そこで、変更可能性を次の8問で点検する。

質問	確認すること
1. 何を変えるのか	データ、設定、表示、手順、判断、停止、契約のどれか

質問	確認すること
2. 誰に実権限があるか	相談窓口ではなく、決定・実装できる人や組織
3. 本人からの入口があるか	訂正、申立て、共同決定、代理、専門職判断の経路
4. どう変えるのか	操作、手続、必要資料、費用、支援
5. 必要な期限に間に合うか	健康・生活上の決定より前に変更できるか
6. 元に戻せるか	停止、取消し、旧版、以前の手順への復帰
7. 変更後を確かめるか	性能、安全、公平性、負担、作業参加への影響
8. 別の道があるか	非AI、別製品、別担当、対面、紙、電話など

この8問は、複数分野の資料をもとにUnknown Labが作った統合案である。妥当性や信頼性が確認された尺度ではない。

同じAIでも、答える人の立場によって結果は変わる。

本人は通知設定を直接変えられても、モデルの更新はできない。専門職は個別判断を上書きできても、契約を変えられない。組織は利用を停止できても、ベンダーの学習データを修正できない場合がある。

そのため、変更可能性を次の4段階で記録する。

- 直接変更可能: 本人や支援者が、その場または明確な手続で変えられる。
- 条件付き変更可能: 組織や開発者の承認・技術作業を経て変えられる。
- 間接的に影響可能: 意見、共同設計、異議、調達、政策参加で影響し得るが、変更は保証されない。
- 現在の介入範囲外: 権限、入口、期限、方法、代替が足りず、今の課題には間に合わない。

この分類はAIの固定的な性質ではない。契約、役割、法域、費用、障害への配慮、組織資源によって移動する。

作業療法が加えられる問い

作業療法は、本人だけを変えることから始めない。

机の高さを変える。道具を変える。手順を分ける。家族の支え方を調整する。通勤経路を変える。制度上の支援を探す。本人の能力と、環境の要求の両方を見ながら、望む作業へ参加できる条件をつくる。

AOTAは、作業療法実践者が環境を選択、作成、変更、利用し、製品・技術・支援、サービス・制度・政策も扱うと説明している[20]。Kirschnerらのスコーピングレビューも、作業参加へ影響する環境・文脈要因を詳細に分けることが、評価と介入を具体化すると結論した[21]。

ここから「AIは作業療法で自由に変更できる」とは言えない。

言えるのは、AIが参加の条件を変えているなら、本人の努力不足として片づけず、AIを含む環境のどの層が変更可能かを問う余地があるということだ。

作業療法が加えられるのは、技術的に変更できるかだけではない。

- その変更は、本人が望む生活や役割につながるか。
- 安全を高める一方で、外出や選択を狭めないか。
- 家族や専門職へ新しい負担を移していないか。
- 使えない人のための別経路があるか。
- 変更までの待ち時間が、その人の生活に何を起こすか。

技術の変更可能性を、生活上の意味と結び直す問いである。

ETIC 2.0への更新候補

ETIC 2.0は、AIを含むケアを、本人、家族、専門職、組織、技術、データ、制度の関係として考える設計・点検仮説である[22]。検証済み尺度や臨床ガイドラインではない。

今回の調査から、ETIC 2.0へ次の更新候補が生じる。

考え方と重なる点

- AI性能だけでなく、生活、作業参加、環境、制度、支援を見る。
- 説明、訂正、上書き、停止、代替を設計に含める。
- 本人の負担だけでなく、家族、専門職、組織への負担移転を見る。

ETIC 2.0が答えるべき問い

- 誰の立場で、AIを含む環境のどこが変えられるのか。
- 意見を送れることと、実際に決定が変わることをどう区別するか。
- 変更が生活上の期限に間に合わないとき、何を先に上書き・停止するか。
- 変更後に、安全、公平性、負担、作業参加をどう再評価するか。

具体的な見直し案

技術・データ統治の設計項目へ、変更可能性・介入境界を加える。

7つの層、8つの質問、4段階分類を候補とし、本人、家族、専門職、開発者、運用者、調達担当者が同じ事例を評価できるかを確かめる。既存のAIガバナンス手法より重要な問題を見つけられるか、追加負担に見合うかも比較する。

この更新候補は、外部資料がETIC 2.0を直接支持した結果ではない。Unknown Labによる統合推論であり、今後の内容妥当性、評価者間一致、比較、外部検証が必要である。

まだ言えないこと

- 本人や支援者は、AIモデルを自由にチューニングできる、またはそうすべきである。
- 変更経路を作れば、健康、生活、作業参加が改善する。
- 共同設計を行えば、本人の意見が実際の仕様へ反映される。
- 人が最終判断者なら、AIは安全である。
- データを訂正すれば、AI評価や組織判断も訂正される。
- 異議申立て窓口があれば、実効的な救済がある。
- EU、米国、日本の制度が、すべてのAIサービスに同じように適用される。
- 本稿の7層、8問、4段階が、検証済みの国際標準である。
- ETIC 2.0が既存の作業療法モデルやAI統治枠組みより優れている。

結論

AIに影響されるのに、私たちはAIを変えられないのか。

答えは、単純な「変えられる」「変えられない」ではない。

多くの人は、AIモデル本体を直接書き換えられない。しかし、記録・入力、設定・閾値、表示・説明、業務手順、個別判断、停止・復旧、代替・調達には、変更できる層があり得る。

ただし、窓口やボタンがあるだけでは不十分である。

何を変えるのか。誰に実権限があるのか。本人から届くのか。必要な期限に間に合うのか。元に戻せるのか。変更後を確かめるのか。別の道があるのか。

この問いに答えられない部分は、現在の介入範囲外として可視化する必要がある。

チューニングは重要である。けれども、モデルの調整だけが変更ではない。本人の生活へ届くまでの複数の層を、誰が実際に動かせるかを設計することが重要である。

AIを変える力とは、全員がプログラムを書くことではない。

自分に影響する仕組みについて、訂正し、問い直し、上書きし、止め、別の道を選べるようにすること。そして、その声を実際の変更と検証へ結ぶことである。

参考文献

[1] Hille EM, Hummel P, Braun M. Meaningful Human Control over AI for Health? A Review. Journal of Medical Ethics. 2026;52(e1):e50-e58. doi:10.1136/jme-2023-109095

[2] van de Sande D, Economou-Zavlanos N, van Genderen ME. Meaningful oversight of medical AI beyond human in the loop. npj Digital Medicine. 2026;9:569. doi:10.1038/s41746-026-02971-1

[3] Ploug T, Holm S. The four dimensions of contestable AI diagnostics: A patient-centric approach to explainable AI. Artificial Intelligence in Medicine. 2020;107:101901. doi:10.1016/j.artmed.2020.101901

[4] Cohen IG, Babic B, Gerke S, et al. How AI can learn from the law: putting humans in the loop only on appeal. npj Digital Medicine. 2023;6:160. doi:10.1038/s41746-023-00906-8

[5] Frost EK, Bosward R, Aquino YSJ, Braunack-Mayer A, Carter SM. Facilitating public involvement in research about healthcare AI: A scoping review of empirical methods. International Journal of Medical Informatics. 2024;186:105417. doi:10.1016/j.ijmedinf.2024.105417

[6] Baines R, Bradwell H, Edwards K, et al. Meaningful patient and public involvement in digital health innovation, implementation and evaluation: A systematic review. Health Expectations. 2022;25(4):1232-1245. doi:10.1111/hex.13506

[7] Loftus TJ, Balch JA, Abbott KL, et al. Community-engaged artificial intelligence research: A scoping review. PLOS Digital Health. 2024;3(8):e0000561. doi:10.1371/journal.pdig.0000561

[8] Dantas C, Reigada C, Cabrita M, et al. Co-research in the development of AI and digital health tools for cancer management and care: A systematic review. BMC Health Services Research. 2026;26:844. doi:10.1186/s12913-026-14620-0

[9] Ustun B, Spangher A, Liu Y. Actionable Recourse in Linear Classification. Proceedings of the Conference on Fairness, Accountability, and Transparency. 2019:10-19. doi:10.1145/3287560.3287566

[10] Karimi AH, Schölkopf B, Valera I. Algorithmic Recourse: from Counterfactual Explanations to Interventions. Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. 2021:353-362. doi:10.1145/3442188.3445899

[11] Karimi AH, Barthe G, Schölkopf B, Valera I. A Survey of Algorithmic Recourse: Contrastive Explanations and Consequential Recommendations. ACM Computing Surveys. 2022;55(5):95. doi:10.1145/3527848

[12] De Toni G, Teso S, Lepri B, Passerini A. Time Can Invalidate Algorithmic Recourse. Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency. 2025:89-107. doi:10.1145/3715275.3732008

[13] National Institute of Standards and Technology. Artificial Intelligence Risk Management Framework (AI RMF 1.0). NIST AI 100-1. 2023. doi:10.6028/NIST.AI.100-1; AI RMF Playbook: Manage: https://airc.nist.gov/airmf-resources/playbook/manage/; Measure: https://airc.nist.gov/airmf-resources/playbook/measure/

[14] World Health Organization. Ethics and governance of artificial intelligence for health. WHO; 2021. ISBN 9789240029200. Official publication: https://www.who.int/publications/i/item/9789240029200

[15] European Union. Regulation (EU) 2024/1689, Articles 14, 26, 86. Official text: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689

[16] U.S. Food and Drug Administration. Marketing Submission Recommendations for a Predetermined Change Control Plan for Artificial Intelligence-Enabled Device Software Functions. Final Guidance. August 2025. FDA-2022-D-2628. Official guidance: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/marketing-submission-recommendations-predetermined-change-control-plan-artificial-intelligence

[17] 個人情報保護委員会. 個人情報の保護に関する法律についてのQ&A（行政機関等編）Q5-9-3. 令和7年12月追加. 公式Q&A: https://www.ppc.go.jp/all_faq_index/faq7-q5-9-3/

[18] 独立行政法人医薬品医療機器総合機構. 医療機器の変更計画確認手続制度（IDATEN）関連情報. 2026年9月6日確認. 公式ページ: https://www.pmda.go.jp/review-services/drug-reviews/about-reviews/devices/0039.html

[19] 経済産業省. AI事業者ガイドライン第1.2版. 2026. 公式ページ: https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/20260331_report.html

[20] American Occupational Therapy Association. Occupational Therapy Practice Framework: Domain and Process, Fourth Edition. American Journal of Occupational Therapy. 2020;74(Suppl 2):7412410010. doi:10.5014/ajot.2020.74S2001; OT's role in shaping contexts and environments: https://www.aota.org/advocacy/everyday-advocacy/ots-role-in-shaping-contexts-and-environments

[21] Kirschner L, Doyle NW, Desport BC. Uncovering the Obstacles: A Typology of Environmental and Contextual Factors Affecting Occupational Participation: A Scoping Review. American Journal of Occupational Therapy. 2023;77(1):7701205040. doi:10.5014/ajot.2023.050043

[22] Morimoto D. ETIC 2.0 Version 2.1. Unknown Lab; 2026. doi:10.5281/zenodo.21848456

利益相反・限界・更新

著者はETIC 2.0の提案者であり、自己評価上の利益相反がある。本稿は査読済み論文ではなく、独立した統計、倫理、作業療法・臨床、AI・データ統治、当事者・家族レビューをまだ受けていない。

本稿はPRISMA準拠の系統的レビューではない。複数分野の概念を7層、8問、4段階へ対応づけた部分はUnknown Labの統合推論であり、国際合意や規制要件ではない。変更経路の存在は、健康、生活、作業参加、安全、公平性の改善を意味しない。

AI、規制、ガイドライン、ETIC 2.0の版は変化する。本稿は2026年9月6日までに確認した公開資料に基づく。