

ETIC 2.0 設計・検証ノート02

AIを訂正・停止 できるのは誰か

「最後は人が責任を持つ」を、実行できる手順に変える

Unknown Lab / ETIC 2.0 設計・検証ノート02

AIを訂正・停止できるのは誰か

責任の名前だけでなく、実行できる権限を考える。



Daiki Morimoto / Unknown Lab

ORCID: 0009-0009-6444-750X

Version 1.0 / 2026-09-14

未査読 / CC BY 4.0

DOI: 10.5281/zenodo.22734208

著者合格: 2026-09-13 / 独立レビュー未了

30秒で読む

AIに間違いや困りごとを見つけたとき、「担当者に伝えます」という返答だけでは、生活は変わらない。訂正や停止を決める人、その操作を実行できる人、返答を待つ間の支援を引き受ける人まで、対応が繋がっている必要がある。

本稿では、維持管理の担当が曖昧になる問題を扱った研究、公的な安全管理資料、AI撤去後の職員調査を読む。その上で、本人の申出から実際の対応と返答までを確かめる見方を提案する。これは既存の安全管理をケアの場面へつなぐ未検証の整理であり、新しい様式を導入すれば事故が減ると実証した記事ではない。

1. 「伝えました」の先で止まってしまう

ここからは、論点を考えるための仮想の場面である。実在する製品や利用者の事例ではない。

一人暮らしの人が、在宅の見守りサービスを使っている。夜中に生活上の小さな動きが通知され、家族が何度も電話をかけてくる。本人は「必要なときに助けてもらえるのはありがたい。でも、毎晩起こされるのはつらい」と相談した。家族も、不要な通知は減らしたい一方、本当に必要な知らせまで失いたくないと考えている。

訪問した作業療法士は、どの時間帯に何が起きているかを一緒に整理した。しかし、サービスの通知条件を変更する権限はない。問い合わせ窓口は「担当部署へ伝えました」と返答する。それから数日、同じことが続く。

この場面で、誰か一人が不親切だったと決めつけるのは早すぎる。窓口には変更権限がなく、技術担当には生活上の困りごとが伝わっていないのかもしれない。管理者は契約上の変更手続きを確認中で、現場の専門職は返答を待つ間の担当が誰かを知らないのかもしれない。

前回の設計・検証ノート01 (<https://unknownlab.pages.dev/articles/ai-modifiability-care/>)では、AIモデルそのものを書き換えなくても、記録、設定、表示、個別判断、使い方などに変更できる部分があり得ることを整理した。今回はその続きである。変えられる部分を見つけた後に、誰がその変更を現実のものにするのか。

2. 責任と権限は、似ているようで違う

「このAIは人間が最終責任を持ちます」。この説明は重要だが、何を意味するかを分けておく必要がある。本稿では、少なくとも次の三つを区別する。

責任は、担当する仕事を果たし、その結果や理由に応答することである。ただし、道義的な責任、組織の職務、法律上の責任は同じものではない。意思決定の権限は、ある範囲の変更や停止を決めてよいことである。操作の権限は、その決定をシステム上で実行できることである。

例えば、専門職が「この人にはAIの提案を採用しない」と判断することと、施設全体のAIを止めることは別である。管理者が停止を承認しても、実行には契約先の技術担当が必要かもしれない。反対に、操作できる人が、本人の生活に関わる変更を一人で決めてよいとは限らない。

Elishは、自動化されたシステムを十分に制御できない人へ責任が偏る問題を、「moral crumple zone」という概念で論じた[1]。複雑な仕組みの失敗を、最後に近くにいた人だけの失敗として扱わないための警告である。医療での発生率を測定した論文ではなく、本稿では概念の範囲で参照する。

この区別は、専門職の責任をなくすためではない。責任を果たすために必要な情報、時間、決定権、操作手段が、どこにあるかを確かめるためである。

3. 開発する人はいても、使い続ける仕組みの担当がいらない

AIを導入するプロジェクトには、開発者、研究者、管理者、利用する専門職など、多くの人に関わる。しかし、人数が多ければ維持管理の仕事がすべて誰かに割り当てられる、とは限らない。

Owensらは、AIの医療実装に関わる専門家21人へ面接し、導入後の監視や保守を調べた。参加者には臨床情報学、診療、計算機科学、法・政策の専門家が含まれ、目的に合う人を選ぶ方法と、紹介をたどる方法で集められている[2]。

研究が描くのは、AIが変化し得ることを知らない人たちの話ではない。変化や問題を認識していても、長く使い続けるための仕事が曖昧で、十分に支えられない状況である。同時に、現場の人が不具合に気づき、他分野の経験を使って修復や監視を工夫していることも報告されている。

これは小規模な質的研究で、参加者の自己報告に基づく。何割の施設で同じ問題があるか、役割を明確にすると患者の健康がどの程度改善するかは分からない。それでも、「人間が関わっている」とことと「必要な仕事の担当が確定している」ことを分ける手がかりになる。

4. 「医師が判断する」だけでは決まらないこと

日本では、2018年の厚生労働省通知が、AIを用いた診断・治療支援プログラムを利用する診療でも、診断・治療の主体は医師であり、医師が最終的な判断の責任を負うと整理している[3]。2026年2月の厚労省説明資料でも、この通知が示されている[4]。

ただし、これはAIを含むあらゆる仕組みについて、あらゆる責任を医師一人に集める表現ではない。この通知から、事業者の技術的責任がなくなる、

作業療法士がシステム全体を停止できる、医師が製品を自由に改変できる、という結論は導けない。

別の問題として、医療情報システムの安全管理には、医療機関と事業者の分担がある。厚労省の第7.0版ガイドラインでは、組織の体制、委託先との責任の範囲、技術的な対応の分担を明確にすることが扱われている[5,6]。こちらは情報セキュリティを中心とする安全管理であり、個々のAI出力の臨床的妥当性を判定する手順そのものではない。

診療上の判断、組織の運用判断、システムの技術的変更は、関係しながらも異なる仕事である。具体的な責任や権限は、法令、製品の位置づけ、契約、施設の規程によって確認する必要がある。本稿は個別事案の法的責任を判定しない。

5. 「止める」の対象を小さく、正確に言う

「このAIを止めたい」という言葉だけでは、何を止めるのかが分からない。ある人への一つの提案を使わないことなのか、通知の一部なのか、施設内での利用なのか、サービス全体なのかで、影響する人も必要な判断も変わる。

以下は、行為を混同しないための本稿の整理である。特定の職種へ新たな権限を与える表でも、すべての製品に機能があるという一覧でもない。

行為	変える対象	決定・実行の前に確認すること
記録を訂正する	誤った入力や記載	何を根拠に訂正するか、誰が更新するか、変更履歴をどう残すか
個別の提案を採用しない	一回の判断・支援方針	専門職の職務範囲、本人の意向、別の判断根拠、記録
利用範囲を制限する	対象、時間帯、機能、使う場面	条件を決める人、設定する人、制限からこぼれる必要な支援
一時停止・廃止する	機能、施設での利用、サービス	停止の範囲、実行者、連絡、代替、継続する仕事
旧版へ戻す	モデル・設定・ソフトウェア等の版	復帰できるか、依存関係やデータの整合性、旧版の安全性、再開条件

旧版へ戻す操作は、一般にロールバックと呼ばれる。しかし、以前動いていた版なら現在も安全、とは限らない。本稿では、安全な復帰先を確認できる場合の候補として扱う。データや連携先が変わっている場合もあり、単に「元へ戻す」と書くだけでは足りない。再検証の詳しい設計は、次の主題である。

6. 担当者名ではなく、対応がつながるかを見る

ここまでの資料を踏まえ、本稿は一件の申出を、受付、決定、実行、支援の引継ぎ、本人への返答まで追ってみることを提案する。これは Unknown Labによる統合案である。法令上の新しい権利や、検証済みの評価尺度ではない。

NISTのAIリスク管理枠組みは、役割と連絡経路の明確化、異議申立てや上書き、停止、非AIの代替手段、導入後の回復・変更管理をすでに扱っている[7]。本稿の案は、そのような既存事項を、本人の困りごとが実際に処理されるかという順序で読み直したものである。

確認する点	記録や説明で確かめたいこと
何に困っているか	本人が変えてほしいこと、生活上の影響、いつまでに対応が必要か
誰が受け付けたか	窓口だけでなく、次の担当が引き受けたことを確認できるか
誰が何を決められるか	個別不採用、設定変更、利用制限、停止のどれを決める権限か
誰が実行するか	承認された変更を実施する人、実行した範囲と時刻、結果
間に合わないときはどうするか	待機中の担当、代替支援、上位・別経路への連絡、不在時の対応
本人に何が返ったか	変更したこと、変更できない理由、残る問題、再検討を求める経路

重要なのは欄を全部埋めることではない。「担当部署へ転送した」と「担当部署が引き受けた」は違い、「停止を承認した」と「実際に止まった」も違う。その間のどこで止まっているかを見つけることである。

同じ人が複数の役割を担っても構わない。小さな事業所に、仕事の数だけ新しい役職を作る提案ではない。逆に、多数の役職があっても、引受けと結果が分からないなら問題は残る。

7. 返事の期限と、生活の期限は違う

問い合わせには「数営業日以内に回答」といった目安があるかもしれない。しかし、冒頭の本人にとっては、今夜も眠れるかが問題である。窓口の標準日数に収まっても、生活上必要な時点に対応できない場合がある。

ここから導くのは、どんな申出にも即時変更を約束すべきだ、ということではない。すぐに変えるには検証が必要な場合もあり、本人と家族の希望が一致しない場合もある。変更を決めるまでの間に、何を保ち、誰が連絡し、どの時点で別の対応へ進むかを、変更そのものと分けて決める必要がある、という提案である。

その期限は、本稿が一律に24時間や3日と定められるものではない。影響の大きさ、切迫性、本人の事情、代替手段、実施体制によって、責任を持つ組織と専門職が確認すべき事項である。緊急の医療上の問題を、一般の問い合わせ窓口の返答待ちで扱う趣旨でもない。

8. 止めたら、誰かの仕事が消えるとは限らない

停止ボタンが機能すれば、その瞬間にすべてが解決するのだろうか。使っていたAIを外した後に、確認、連絡、記録などの仕事が別の人へ戻るかもしれない。逆に、過剰な通知への対応から解放される人もいるかもしれない。

Packerらは、英国NHSの複数拠点を持つ一つの組織で、胸部X線のAIトリアージを撤去した2か月後に職員へ調査した。187人中40人が回答し、回答率は21.4%だった。自由記述には、業務上の負担からの解放と、臨床的な価値や経路の効率を失ったという受け止めの両方があった[11]。

28人、つまり回答者の70%は、AIに患者への便益があったと認識していた。ただし、これは職員がどう感じたかを尋ねた結果である。患者の70%が改善したという意味でも、AIを残すべきだったと証明する数字でもない。

本稿で確認できたのは原要旨までである。低い回答率、単一組織、自己報告という限界があり、撤去の因果効果や長期の安全性は判断できない。それでも、停止後の影響を一種類の「負担」としてまとめず、誰に何が起こるかを確かめる必要を示唆する。

NISTのPlaybookも、システムの一部または全体を止める際、業務の混乱や下流への不利益を抑える手順、代替・バックアップ、記録、再稼働条件を考えるよう促している[8]。安全のために止める必要がある場合に停止を遅らせる理由ではなく、停止と支援の継続を一緒に準備するための視点である。

9. 既存の安全管理を、最初からやり直さない

こうした話を聞いて、「結局、昔からある安全管理ではないか」と感じる人もいるだろう。その指摘は重要である。

NHSの臨床安全に関する資料では、製造側のDCB0129と、導入・利用側のDCB0160を区別している。臨床安全責任者、危険源の記録、リスクをどう低減したかを示す証拠、導入から保守・廃止までの再評価が扱われる[9]。日本でそのまま義務になる制度ではないが、提供者と現場の双方が仕事を持つ、という具体例になる。

一方、NHSが2026年に公表した改訂検討資料は、2025年1〜7月の11のフォーカスグループで、AI統治、組織をまたぐ連携、導入後監視、資源の手引きなどを強化する必要が挙げられたと報告している[10]。これは制度が無意味という結果でも、改訂後の効果が確認されたという結果でもない。

既存の手順には使える部分がある。その手順から、本人の申出、必要な権限、実行の結果、支援の継続まで確認できるなら、新しいETIC専用様式は不要かもしれない。記録が増えるだけで対応が変わらないなら、追加を見送る方が適切である。

10. 本人を、システムの保守担当にしない

冒頭の本人は、見守りAIを再設計したいと言ったわけではない。眠りを妨げられず、必要なときには助けてもらえる生活を望んでいる。ここを取り違えると、「AIについて学び、毎日の誤作動を分類し、複数の窓口を調整してください」という負担を、本人へ渡してしまう。

本人が分かる言葉で困りごとを伝えられ、その内容を技術的・組織的な対応へ翻訳する仕事を、支援側が引き受ける。そのような役割も必要ではないだろうか。これは本稿の規範的な提案であり、あらゆる現場での実施体制が確立しているという記述ではない。

家族の手助けも、無限に利用できる資源ではない。本人の希望と家族の不安は、片方をもう片方の代わりに記録しない方がよいだろう。意見が一致しないときは、その不一致を残したまま、説明や支援を通じて検討する必要がある。この記事の表だけで、本人の意思、代理判断、法的権限に関する個別問題を解決できるわけではない。

11. ETIC 2.0のどこを見直すか

考え方と重なる点

ETIC 2.0は、本人の生活を、人、技術、環境、制度の関係の中で考える枠組みである[12]。今回の資料を合わせて読むと、AIの統治も、利用者と一人の専門職だけでは完結しないという見方が得られる。ただし、外部研究がETIC 2.0を直接検証したわけではない。

ETIC 2.0が答えるべき問い

「人間の責任を保持する」「停止可能にする」という表現を置くだけで、何を確認したことになるのだろうか。実際に変更できる人までつながらない場合、そのAIをケアの変更可能な要因として扱ってよいのだろうか。停止した後の不利益が本人や家族へ移った場合、どのように評価すべきだろうか。

具体的な見直し案

今回の候補は、申出から対応と返答までの担当、権限、期限、代替を、実際の記録から追えるようにすることである。専門職がすべての操作を持つ形ではなく、必要な人へつながり、引き受けられ、結果が本人へ戻る形を考える。

これは未検証の提案である。比較するなら、責任者名だけを示す方法と、新しい表を導入する方法の二択にはいけない。すでにある臨床安全・

インシデント管理・連絡の記録で同じことが追える方法も、有力な比較対象である。

最初は仮想場面を使い、必要な担当を見つけられるか、引継ぎが抜けないか、記入と確認にどの程度の時間がかかるかを検討できる。その後に実際の運用を調べる場合も、未解決の申出、対応までの時間、本人にとっての解決、代替支援の中断、家族・職員の負担を分けて確認する必要がある。対人研究や現場の手順変更は、倫理・組織・専門職・当事者の確認を経る別工程である。

申出が減るだけでは成功としない。申出の入口が難しくなったのかもしれないからである。変更が増えるだけでも成功としない。不適切な上書きや不要な停止が増える可能性もあるからである。既存の方法より役立たず、負担だけが増えるなら、提案を修正するか、採用しないことが必要である。

12. 調査の範囲と限界

2026年9月13日までの公開情報を対象に、事前に中心質問、採否基準、反証条件を定め、原論文、政府・公的機関資料を優先して探索した。系統的レビューではなく、特定の問題を考えるための対象限定レビューである。検索の網羅性は主張しない。

経験的な根拠として使ったのは、少人数の専門家への質的研究と、単一組織の撤去後職員調査である。後者は原要旨の範囲に限られ、Elish論文も要旨・序論の概念定義を中心に参照した。公的資料は設計や管理の要件を示すが、実装されていることや臨床効果を保証しない。

また、日本の情報セキュリティ管理、英国の臨床安全、米国の任意のAIリスク管理枠組みは、対象と制度が異なる。本稿の統合案を、特定施設にそのまま導入できる手順書と解釈してはならない。独立した臨床・作業療法、研究方法、倫理・AI統治、当事者・家族レビューは別工程として残っている。

13. ケアを残すために、仕組みを変える

冒頭の場面に戻ろう。本人が知りたいのは、会社の組織図のすべてではない。誰かに伝えれば話が進むのか、今夜はどうすればよいのか、変更できないなら別の方法を一緒に考えてもらえるのか、ということである。

「最後は人が責任を持つ」を実質のある説明にするには、その人が何を決められ、何を実行でき、できない部分を誰へつなげるかが必要である。さらに、止めた後にも本人との支援の接点を残せるかを考えなければならない。

AIを訂正・停止できるのは誰か。その答えは、一つの肩書きでは決まらない場合がある。けれども、だから曖昧でよいわけではない。一人の申出が、実際の対応と返答まで届く。その経路を確かめることから、考え始められる。

参考文献

[1] Elish MC. Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction. *Engaging Science, Technology, and Society*. 2019;5:40–60. <https://doi.org/10.17351/ests2019.260>

[2] Owens K, Griffen Z, Damaraju L. Managing a “responsibility vacuum” in AI monitoring and governance in healthcare: a qualitative study. *BMC Health Services Research*. 2025;25:1217. <https://doi.org/10.1186/s12913-025-13388-z>

[3] 厚生労働省医政局医事課長. 人工知能 (AI) を用いた診断、治療等の支援を行うプログラムの利用と医師法第17条の規定との関係について. 医政医発1219第1号. 2018年12月19日. <https://www.pmda.go.jp/files/000227450.pdf>

[4] 厚生労働省健康・生活衛生局・医薬局. AIを活用したがん検診について. 規制改革推進会議 第9回健康・医療・介護ワーキング・グループ資料1-7. 2026年2月12日. p.2. https://www8.cao.go.jp/kisei-kaikaku/kisei/meeting/wg/2501_02medical/260212/medical09_0107.pdf

[5] 厚生労働省. 医療情報システムの安全管理に関するガイドライン 第7.0版 経営管理編. 2026年6月. <https://www.mhlw.go.jp/content/10808000/001716291.pdf>

[6] 厚生労働省. 医療情報システムの安全管理に関するガイドライン 第7.0版 システム運用編. 2026年6月. <https://www.mhlw.go.jp/content/10808000/001716295.pdf>

[7] National Institute of Standards and Technology. Artificial Intelligence Risk Management Framework (AI RMF 1.0). NIST AI 100-1. 2023. <https://doi.org/10.6028/NIST.AI.100-1>

[8] National Institute of Standards and Technology. AI RMF Playbook: Manage, MANAGE 2.4. <https://airc.nist.gov/airmf-resources/playbook/manage/> (2026年9月13日参照)

[9] NHS England. Digital clinical safety assurance. Version 1.2, 3 March 2025. <https://www.england.nhs.uk/long-read/digital-clinical-safety-assurance/>

[10] NHS England. National review of clinical risk management standards DCB0129 and DCB0160: supporting information. 29 June 2026. <https://www.england.nhs.uk/long-read/national-review-of-clinical-risk-management-standardsdcb0129-and-dcb0160-supporting-information/>

[11] Packer J, Dean G, Storey M, Malamateniou C, Shelmerdine SC. Responsible Artificial Intelligence Off-Boarding in Radiology: Staff

Perspectives on Decommissioning and a Proposed
WithdrawalFramework. Journal of the American College of Radiology.
Publishedonline 6 June 2026. <https://doi.org/10.1016/j.jacr.2026.06.004>

[12] Morimoto D. エコテクノロジカル・ インテグラティブ・ ケア
(ETIC)2.0, Version 2.1改訂版 - AI時代の作業療法とケア生態系設計.
Zenodo; 2026. <https://doi.org/10.5281/zenodo.21848456>

著者・ 利益相反・ 位置づけ

著者はETIC 2.0の提案者である。そのため、自説に都合のよい解釈を採用
する可能性に注意し、既存枠組みとの重複、反例、採用しない条件を記し
た。調査、整理、初稿にはAI支援を使用している。本稿は一般的な研究・
教育情報であり、個別の医療判断、法的判断、製品の設定変更を指示する
ものではない。

執筆: Unknown Lab 編集部 (AI支援) / 人物著者: Daiki Morimoto / 著者確
認: 合格 (2026年9月13日) / 独立した専門職・ 当事者等の確認: 未了