

ETIC 2.0 設計・検証ノート03

AIが更新されたら、 安全確認も やり直すべきか

変わった機能だけでなく、
変わった暮らしを確かめる



Daiki Morimoto / Unknown Lab

ORCID: 0009-0009-6444-750X

Version 1.0 / 公開日 2026-09-21

未査読 / 著者確認・公開承認済み

独立レビュー未了 / CC BY 4.0

DOI: 10.5281/zenodo.22851849

30秒で読む

昨日まで使えていたAIが、今日も同じように使えるとは限らない。更新で答え方や画面が変わることもあれば、AIは同じでも、使う人や周囲の状況が変わることもある。

ただし、更新のたびにすべての試験を最初から繰り返せばよいわけでもない。必要なのは、前の安全確認がどの条件に支えられていたかを振り返り、変化の影響が届く範囲を確かめ直すことである。

本稿は、その範囲を考えるために「出力の妥当性」「人が使う過程」「本人の生活」の三つを分ける。これは既存の研究や安全管理の考え方を生活支援へつなぐ提案であり、この方法による健康改善や事故予防の効果は、まだ検証されていない。

1. いつもの画面が変わった朝

買い物へ出かける前に、タブレットで一日の予定を確認する人がいるとする。以前は、出発時刻の下に大きく「付き添いに連絡」と表示されていた。更新後は予定をまとめる機能が加わり、その連絡先が別の画面へ移った。

予定の時刻は正しい。アプリも正常に動いている。それでも、本人は連絡先を見つけられず、出かけるのをやめてしまう。

これは論点を示すための架空の場面であり、実際の事例報告ではない。ここで注目したいのは、AIが間違った答えを出したかどうかだけでは、この困りごとを説明できない点である。情報の置き場所が変わったことで、いつもできていた一連の行動が途中で止まっている。

開発側から見れば、機能を増やした更新かもしれない。支援者から見れば、画面の小さな変更かもしれない。しかし本人にとっては、一人で外出の準備を進められるかどうかに関わる変化となる。

では、安全確認はどこまでやり直すべきか。新しい機能の説明を読めば足りるのか。もう一度、本人と画面を開く必要があるのか。使い続ける判断をいったん見直すほどの変化なのか。

本稿でいう「再検証」は、こうした問いに答えるために、以前の確認結果を引き続き使えるかを調べることである。毎回大規模な臨床試験を行う、という意味ではない。逆に、更新通知を読んだだけで、以前の確認がすべて有効だと考えることでもない。

2. AIを変えなくても、確かさは変わる

安全確認をやり直すきっかけは、ソフトウェアの更新だけではない。

Davisらは、臨床予測モデルの確率と実際の転帰の対応が時間とともにずれる「校正のずれ」を検出する方法を検討した。米国退役軍人医療の事例では、2006年の235,548件の入院で学習したモデルを更新せず、その後5年間の1,205,457件の入院で追跡した。モデルの種類によって、ずれを示す警報が複数回生じた。[1]

ここで数えられているのは入院であり、それぞれ異なる人の数ではない。また、実データでは真の変化の開始時点が分からず、警報が事故や健康被害を直接示したわけでもない。この研究は、予測確率の監視方法の検討であって、再検証によって患者の健康が改善したことを証明した試験ではない。[1]

それでも重要な示唆がある。モデルを固定しておけば、確認した性能がそのまま保たれるとは限らない、ということである。

この知見を生活支援へ応用して考えるなら、更新履歴だけでなく、使用条件の変化にも目を向ける必要がある。例えば、以前は家族と一緒に使っていた人が一人で使うようになった場合、同じ画面でも、困ったときに助けを求めるまでの過程は変わる。この例は上記研究の結果ではなく、本稿の推論である。

「何も更新していないから大丈夫」という判断と、「更新されたから危ない」という判断は、どちらも変化の中身を飛ばしている。

3. 何が変わったのかを分けて考える

「AIが変わった」という言葉には、異なる変更が含まれる。次の表は、生活支援の場で変更を見落とさないための、本稿による整理である。正式な分類規格や検証済みの評価尺度ではない。

変化する場所	例えば何が変わるか	まず確かめたいこと
モデル	同じ相談に対する推奨内容	以前と異なる答えが、重要な判断を変えないか
データ	入力項目、測定機器、対象となる人	想定した情報が、今も同じ意味で得られるか
設定	通知の頻度、表示対象、利用範囲	許された設定変更が、必要な情報を隠さないか
画面・説明	ボタンの位置、言葉、確認手順	本人が理解し、意図した操作にたどり着けるか
支援の体制	利用場所、付き添う人、引継ぎ	困ったときの連絡と代替りの支援が機能するか

一つの更新が、複数の場所に影響することもある。モデルの変更によって通知が増えれば、答えの性質だけでなく、確認する人の負担も変わる。画面だけを変えたつもりでも、本人が注意を向ける順序や、支援者へ相談するタイミングが変わる可能性がある。

だから、再検証の出発点は「新しい版を入れたか」ではなく、「前に確かめた条件のうち、何が変わったか」となる。

なお、この整理は、本人や支援者がモデルの内部や事業者のデータを書き換えられるという前提を置かない。手元で変更できない部分は、提供者への確認や、利用方法の調整、別の支援の検討につなぐ必要がある。設定画面にある機能と、製品の設計や承認範囲を変える行為も同じではない。

4. 毎回、最初から試す必要はない

変更に応じて確認範囲を決める考え方は、ETIC 2.0独自の発明ではない。

米国FDAのAI医療機器に関する変更管理計画の指針では、予定する変更、変更を開発・検証・実装する方法、影響評価をあらかじめ記述する。承認等を受けた計画に沿う変更について、個別の申請を繰り返さずに実施できる場合があるが、これは検証自体を不要にする仕組みではない。[2]

人が機器をどう使うかに関するFDAの指針も、変更の直接・間接の影響を検討し、追加の検証の要否と範囲をリスクに基づいて判断する考え方を示している。[3]

日本にも、医療機器の変更計画を事前に確認するIDATEN制度がある。関連通知とQ&Aでは、計画に含める変更や、その品質・有効性・安全性を説明するための資料などが扱われている。[4,5]

これらは、それぞれの国の医療機器に関わる制度・指針である。一般向けアプリにそのまま適用される共通の法的義務ではなく、支援者が機器を自由に改変してよいという根拠でもない。個別製品の変更可否や必要手続は、製品の位置づけと適用制度に沿って確認する必要がある。

本稿がここから受け取るのは、変更を一律に扱わず、確認する範囲と理由を結びつけるという考え方である。

例えば、支援で使わない紹介ページの誤字修正と、毎日の連絡ボタンの移動を、同じ確認で済ませる理由はない。一方、見た目だけの変更だから軽微だと決めつけることもできない。文字の大きさや確認画面の順序が、その人にとって重要な操作を左右する場合があるからだ。

小さな変更なら小さな確認で済むことがある。ただし、その判断は、変更の名称ではなく、影響を受ける行動と、失敗した場合の重さに基づくべきである。

5. 全体の成績だけでは、その人の困りごとは見えない

平均的な性能が維持されていても、利用者の一部では状況が異なるかもしれない。

Subbaswamyらは、AIモデルが相対的に苦手とする患者群をデータから見つける枠組みを報告した。検討には一つの医療ネットワークの5病院、60,998件の診療エピソードを用い、全体の性能評価だけでは見えにくい部分集団の差を調べている。これはモデル更新の前後を比較した試験ではなく、別の医療ネットワークや日常の警報運用にそのまま当てはまるとも限らない。[6]

この研究から、高齢者なら必ず性能が悪い、あるいは個人ごとに統計モデルを作るべきだ、とは言えない。ここでの要点は、「全体では良好」と「自分たちが支援する対象でも妥当」を区別することである。

生活支援の場で考えると、説明を聞くだけで操作できる人と、文字を拡大して何度も確かめる人では、同じ画面変更の意味が違うかもしれない。確認すべき相手を「平均的な利用者」だけに絞らず、影響を受けそうな人や使い方から選ぶ。この方向づけは本稿の提案であり、人数や選び方の妥当性は別途検証が必要となる。

少人数で問題が見つからなかったことも、「全員に安全」の証明にはならない。まれな不具合や、その場にいない利用者の困難は残る。確認できた範囲と、まだ分からない範囲を一緒に残すことが大切である。

6. 三つの問いで、更新前後をつなぐ

ここまでの整理を、冒頭の買い物場面に戻してみよう。

本稿では、更新後の確認を次の三つの問いに分ける。互いに関係しているが、一つが良好だからといって、残りも良好だとは判断しない。

出力は妥当か

予定の日時、連絡先、案内の内容に問題がないか。必要な情報が抜けたり、想定外の推奨が加わったりしていないか。ここでは、支援で実際に必要な情報を確認する。

ただし、本人や現場の支援者が、モデル全体の性能を検証できるわけではない。提供者が示す変更内容や評価結果と、現場で確認できる内容の境界を明らかにする必要がある。

人は意図したとおりに使えるか

本人が連絡先を見つけ、誰に連絡するかを理解し、必要な操作を終えられるか。支援者が横から答えを教え続けて完了した場合と、普段の支援条件で完了した場合も区別したい。

この問いによって、予定の情報が正しくても、生活上の手順は途切れうるが見えてくる。使えなかった理由を、本人の能力だけに帰さず、画面や説明の変化にも求められる。

本人の生活につながっているか

操作が終わった先で、本人の望む買い物の準備ができたか。外出を諦めることが増えていないか。見かけ上は自立して操作できても、家族や支援者の確認作業が増え続けていないか。

これは、短い操作確認だけでは十分に分からない問いである。必要に応じて、その後の生活を確認める期間や方法を考える。ただし、外出できたかどうかは天候や体調、交通手段にも左右される。前後の変化がすべて更新によるものだと決めつけてはならない。

三つの問いは、製品としての適合性、現場での使用、生活への意味を混ぜないための整理である。生活の変化まで視野に入れることを提案するが、個々の利用者には無制限の観察や記録を求めるものではない。本人に説明し、必要な情報に絞り、既存の支援記録で足りる場合はそれを使う。

7. 確かめ直すために、何を残すか

前の状態が分からなければ、「今回の更新で何が変わったか」も判断しにくい。

とはいえ、新しい帳票を何枚も増やすことが目的ではない。本稿では、既存の記録で、少なくとも次のつながりを追えるか確認することを提案する。

変更前は、誰が、どのような支援を受けながら、何をしていたか。今回、何が変わったか。どの人や行動に影響しそうか。何をどの条件で確かめたか。その結果、どう対応したか。そして、残った不確かさをいつ、誰が見直すか。

買い物の例なら、「更新後も問題なし」だけでは足りない。連絡先の位置が変わったこと、普段の条件では本人が見つけれなかったこと、許された範囲で表示や案内を調整したこと、その後の外出準備がどうなったかまでつながると、次の支援者にも判断の理由が伝わる。

重要な操作への影響が分からない場合は、分からないまま利用範囲を広げないという選択もある。ただし、AIを止めれば必ず安全になるわけではない。連絡手段や必要な支援まで失われるなら、停止そのものにも不利益がある。提供者や担当専門職と、代わりの方法を含めて判断する必要がある。

また、以前の版へ戻すことが常に可能、あるいは安全とは限らない。古い版に既知の問題がある場合もある。本稿は、医療上必要な利用やセキュリティ更新を独断で止めるよう勧めるものではない。

一般的なAIリスク管理でも、運用中の評価や利用者からのフィードバックは重視されている。[7] ETIC 2.0で問いたいのは、それらを実施したという記録だけでなく、本人が生活続ける条件まで確認できているか、という点である。

8. ETIC 2.0に何を持ち帰るか

考え方と重なる点

AIの性能を一度測って終わりにせず、使われる条件や時間の変化とともに見直す考え方は、既存研究や指針にすでにある。変更管理や継続監視そのものを、ETIC 2.0の新規性として主張することはできない。

ETIC 2.0の立場からは、人、環境、技術の関係を、固定された配置としてではなく、変化する支援の条件として見る必要がある。更新は技術だけに起きるのではなく、その人が行動できる環境の変化にもなる。この接続に、本稿の整理の意味がある。

ETIC 2.0が答えるべき問い

生活の観点を加えることで、通常の変更管理では見落とす問題を実際に見つけられるのか。単に質問と記録が増えるだけではないか。

この問いに答えないまま、三つの視点を推奨様式として広めるべきではない。既存の手順ですでに、本人の重要な活動と支援の継続まで扱っているなら、新しい名称や書式を加える必要はない。

確認の負担も評価対象となる。支援者が追加記録に時間を取られ、本人と確かめる時間が減るなら、本来の目的から離れてしまう。

具体的な見直し案

ETIC 2.0の今後の更新候補として、技術の利用を検討する際に「確認した条件」と「その条件が変わった時の見直し方」を一対で記述することを提案する。これを採用済みの仕様や検証済みの介入とは扱わない。

検証するなら、まず架空の変更事例を用いて、既存の点検手順と、三つの問いを補った手順を比べる方法が考えられる。重要な見落としを拾えるか、不要な中断を増やさないか、所要時間と記録負担はどうかを、あらかじめ定めて比較する。採点基準も、提案者だけで決めず、当事者や現場の専門職を含めて検討する必要がある。

追加の問いが判断を改善せず、負担だけを増やすなら、採用しないか、さらに簡略化する。机上の事例で役立っても、現場での安全性や健康への効果が証明されたことにはならず、その先の検証が必要となる。

9. 「安全だった」を、今の生活につなぎ直す

本稿は、2026年9月20日までに確認した原論文と公的資料を用いる対象限定の検討であり、網羅的な系統的レビューではない。採用した実証研究は臨床予測モデルに関するもので、一般的な生成AIや在宅支援へ、その結果を直接移すことはできない。三つの問いや記録のつながりは、本稿の未検証の提案である。

それでも、更新を考える入口は明確にできる。すべての変更を危険視することでも、最新版ならよいと考えることでもない。以前「使える」と判断したときの条件が、今も成り立っているかを確かめることである。

買い物の予定が正しく表示されることは大切だ。しかし、その予定を見た本人が、必要な連絡をして、望む外出へ進めることとは同じではない。

更新後に確かめ直したいのは、新しい機能だけではない。その機能と一緒に変わった、暮らしの進め方である。

参考文献

[1] Davis SE, Greevy RA Jr, Lasko TA, Walsh CG, Matheny ME. Detection of calibration drift in clinical prediction models to inform model updating. *Journal of Biomedical Informatics*. 2020;112:103611. <https://doi.org/10.1016/j.jbi.2020.103611>

[2] U.S. Food and Drug Administration. Marketing Submission Recommendations for a Predetermined Change Control Plan for Artificial Intelligence-Enabled Device Software Functions. Guidance for Industry and Food and Drug Administration Staff. 2025-08-18. <https://www.fda.gov/media/166704/download>

[3] U.S. Food and Drug Administration. Applying Human Factors and Usability Engineering to Medical Devices. Guidance for Industry and Food and Drug Administration Staff. 2026-08-03. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/applying-human-factors-and-usability-engineering-medical-devices>

[4] 厚生労働省医薬・生活衛生局医療機器審査管理課長. 医療機器の変更計画の確認申請の取扱いについて. 薬生機審発0831第14号. 2020-08-31. <https://www.pmda.go.jp/files/000236900.pdf>

[5] 厚生労働省医薬局医療機器審査管理課. 医療機器、人工知能関連技術を活用した医療機器、プログラム医療機器の変更計画の確認申請に関する質疑応答集 (Q&A) について. 2023-12-22. <https://www.pmda.go.jp/files/000266198.pdf>

[6] Subbaswamy A, Sahiner B, Petrick N, Pai V, Adams R, Diamond MC, Saria S. A data-driven framework for identifying patient subgroups on which an AI/machine learning model may underperform. *npj Digital Medicine*. 2024;7:334. <https://doi.org/10.1038/s41746-024-01275-6>

[7] National Institute of Standards and Technology. AI Risk Management Framework (AI RMF): Core, MEASURE. AI RMF 1.0関連ウェブ資料. 2026-09-20閲覧. <https://airc.nist.gov/airmf-resources/airmf/5-sec-core/>

著者・利益相反・位置づけ

著者はDaiki Morimoto、所属はUnknown Lab、ORCIDは0009-0009-6444-750Xである。ETIC 2.0の提案者として、自らの枠組みに関する知的な利害がある。

本稿は設計・検証ノートであり、査読済み論文、診療指針、製品の適合性評価、法的助言ではない。文献整理、原稿作成、引用・書式点検、概念図制作にAIを利用した。図は架空の場面であり、実際の製品や臨床効果を示さない。著者確認と公開承認は2026年9月20日に完了した。独立した学術的査読を受けたものではない。

本版はVersion 1.0である。ライセンスはCC BY 4.0、DOIは10.5281/zenodo.22851849である。