提言書

医療分野における生成 AI 導入に伴うセキュリティの重要性について

2026 年 5 月 7 日
医療 AI プラットフォーム技術研究組合
一般社団法人 Health ISAC Japan

【提言1】
医療機関等における生成 AI の導入においては、単発の実証や部門最適の機能導入をめざすのではなく、経営層の指示のもと、組織横断的なルールに基づき、求められる安全管理水準に応じた技術・運用上のリスク管理対策、教育、契約管理、利用不可時の業務継続計画も含めた、全体最適型のセキュリティガバナンスを整備すべきである。

【提言 2】
医療機関等における生成 AI の安全・安心な活用を本格的に実現するために、官民のインセンティブを組み込み、共通基準のもと、評価・監査、情報共有、導入支援等を段階的・継続的に実施可能とする、生成 AI プラットフォームの構築が不可欠である。

1. 提言の目的

医療 DX の進展に伴い、病院・クリニック、薬局・介護施設等（以下、「医療機関等」）では、生成 AI を業務改善に生かそうとする動きが急速に広がっている。医療 DX がめざすのは、診察・治療・文書作成・請求・地域連携・研究開発までを通じた全体最適であり、生成 AI はその有力な構成要素の一つである。

特に、患者ごとの専門的判断と意思決定が求められ、さらに記録や説明に多くの時間を要する医療機関等では、その効果は大きいといえるが、その効果を継続的かつ安定的に維持するためには生成 AI 環境のセキュリティの確保が不可欠である。

なぜなら医療分野における生成 AI の誤作動や停止は、患者安全と医療継続の問題に直結するリスクを抱えるものだからである。

そのため、医療 AI プラットフォーム技術研究組合、及び一般社団法人 Health ISAC Japan は、医療機関等における生成 AI におけるセキュリティの重要性についての共同提言を行うものとする。

2. 生成 AI 導入の意義と課題

現時点で、医療機関等における生成 AI の導入・活用状況を考慮した場合、その実益は、法定又は実務上作成が求められる文書の下書き、要約、転記、表現調整等の負担を大きく軽減する点にあるといえる。

例えば、①診察音声やメモから診療録、退院時サマリー、診療情報提供書のドラフト作成、②看護記録や申し送り、カンファレンス要約の整理、③検査結果や治療計画を患者向けにわかりやすい説明文へ構成する、④服薬指導や生活習慣改善、通院フォローアップのメッセージ案の作成等、様々な場面で生成 AI の技術が有効である。

医療従事者が最終確認を行う前提に立てば、文書・記録作成時間の短縮を通して、医療業務の効率化や医療従事者の負荷低減に大きく寄与することになるといえる。

こうした生成 AI は医療情報システムの周辺のシステムとして用いられている一方、システム内部で診療・治療の判断、トリアージ、読影、治療方針の提案等に関与するため、患者安全に直結するリスクを持つ製品もすでに登場し始めている。製品が実現する機能に応じて求められる生成 AI 上の安全管理水準は異なるが、いずれも患者情報や医療情報を取り扱い、医療提供体制の一部を構成する以上、一定水準のセキュリティ確保は不可欠である

3. 生成 AI のセキュリティの重要性、及び後回しにすべきでない理由

医療分野における生成 AI は、独立したアプリケーションとして扱うべきではなく、医療情報システムの一部として扱うべきである。入力した医療情報が意図せず外部へ再利用されれば機密性が崩れ、指示の混入や知識基盤の汚染によって誤った文章や危険な助言がもっともらしく生成されれば完全性が崩れ、クラウド障害や認証障害、ランサムウェア等で利用できなくなれば可用性が崩れる。

セキュリティが不十分であることは、単に生成 AI のアプリケーションが危険にさらされるという問題ではない。医療従事者が期待する本来の機能が知らぬ間に損なわれているにもかかわらず、そのことに気づかないまま医療従事者が AI を利用し続けかねないという、患者をリスクにさらす医療安全管理上の課題として考えなければならない。とりわけ、システム内部で診療・治療等に関与する生成 AI については、そのリスクの程度はより深刻である。

なお、注意すべき点は、生成 AI を特定の医療業務で使うという個別最適の目的で推進させるリスクである。個別最適のもと、各診療科・各部ごとに別々のセキュリティルールが導入されていくと、途中から組織としての全体最適の観点から、入出力制御、監査、認証、ログ管理、契約管理、利用不可時の業務継続等、必要となるセキュリティルールを統一的に確立することが困難となる。

2000 年代初頭の情報革命の中で医療情報システムが急速に浸透した状況のもと、ここ数年来のサイバー攻撃リスクの高まりのなかで、セキュリティ機能を後付けで組織全体に統一的に導入しなければならなくなったことは、医療機関等に大きな混乱をもたらしている。そのような混乱を生成 AI の導入において再び繰り返してはならない。そのためにも個別導入を進めるまえに、組織として安全に生成 AI を利用・管理するための全体設計をまずは検討することが重要である。

ただし、包括的なセキュリティルールの整備には相応の時間を要するため、まずは最低限のルールを定め、運用の成熟状況に応じて段階的に要件を拡張し、組織全体として最適化されたルールの確立をめざすことが推奨される。

4. 医療機関等で想定される生成 AI 関連のインシデント例

医療機関等において生成 AI が直接関係する重大インシデントの報告例は現時点では限定的である。しかしながら、生成 AI 自体がはらむセキュリティ上の既知の問題を医療分野に敷衍して考えた場合、以下に示すようなセキュリティインシデントの発生が容易に想定される。

- 会話履歴や属性情報の露出（クロステナント漏洩、アクセス制御不備等による被害例）
  クラウド型の対話 AI で、他者の会話タイトルや一部属性情報が別利用者に表示される不具合が発生することで、診療要約、相談履歴、患者説明文の漏えいにつながる。
- 公開設定不備によるデータベース露出（クラウド設定ミス、バケット露出等による被害例）
  AI の運用基盤や保存領域の設定ミスにより、チャット履歴、秘密情報、接続情報が外部から閲覧可能となることで患者情報だけでなく、運用の中枢まで同時に危険にさらされる。
- 指示注入や参照文献汚染による誤生成（プロンプトインジェクション、RAG ポイズニング等による被害例）
  外部文書や参照先に悪意ある指示が混入し、退院時サマリー、看護記録、患者説明文に誤情報や不適切な推奨が紛れ込むことで、現場に訂正負担と混乱がもたらされる。
- 画像、PDF 等を含む外部文書への植え込み指示による有害出力（マルチモーダル注入、敵対的サンプル等による被害例）
  画像や PDF 等に人間には見えにくい形で指示が埋め込まれ、読影補助や説明文の生成が誘導されることで、入力が正しく見えていても AI だけが誤った応答を返す事態が発生する。

- 障害時の一括停止・一括誤生成 （サプライチェーン攻撃や API 障害等による被害例）
  AI 文書支援認証障害や外部 API 停止、モデル更新不良を起こすと、複数部署で同時に記録作成や患者説明が止まり、あるいは同種の誤文書が大量に生成されることで、後追い訂正と患者対応が必要になる。

5. 生成 AI 導入に伴う基本的なセキュリティ対策、及び中長期的な展望について

医療機関が生成 AI を安全・安心に活用するためには、医療 AI プラットフォーム技術研究組合「医療・ヘルスケア分野における生成 AI 利用ガイドライン（第 2 版）」をもとに、最低限のセキュリティルールとして、次の事項を導入前に検討・決定すべきと考える。

① 組織として正式に利用を認める生成 AI の選定と範囲及び用途の明確化
② 入力した情報が外部に漏れないようにするための未然防止策の検討
③ 生成 AI の不適切な使い方を検知するためのモニタリングの仕組みの検討
④ 生成 AI が利用不可となった場合でも対患者業務を継続できる代替手順の検討

なお、これらの検討・決定には理事長、院長や事務長などの経営層の理解ある関与が求められる。医療機関等のセキュリティ施策の多くは現場の努力のみでは実行することが難しく、ヒト・モノ・カネの融通も含め、経営層の関与が不可欠である。セキュリティの不備が医療経営の継続性を損なうサイバー攻撃事案が多く発生するとおり、今やセキュリティリスクは経営リスクの一つである。セキュリティ対応は一過性の要請ではなく、不可逆的な課題である。セキュリティ対応を後回しにしてサイバー攻撃を受けてから苦労するよりも、先回りした対応を経営層主導のもとで組織的に行うことが費用対効果の面からも推奨される。

一方で、生成 AI の安全な活用を個々の医療機関の自助努力のみに委ねつづけることには限界がある。官のインセンティブとしては診療報酬での評価や継続的な補助制度、民のインセンティブとしては融資優遇、保険料割引、安全なデータ利活用への対価を組み合わせ、セキュリティを事前実装し、監査・訓練・改善を段階的に継続し改善できる生成 AI プラットフォームの構築を中長期的にはめざしていくべきであり、我々はその実現に向けた活動を今後とも実施していく考えである。

以上