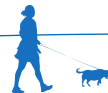


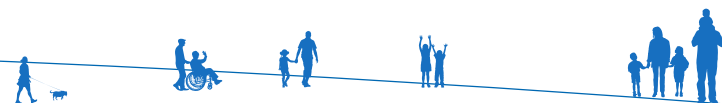
生成AI活用医療機器は 承認審査をどう変えるか！？

独立行政法人 医薬品医療機器総合機構(PMDA)
プログラム医療機器審査部
加藤健太郎

本発表は、発表者の個人的見解に基づくものであり、
独立行政法人医薬品医療機器総合機構の公式見解を示すものではありません。



生成AIを活用した医療機器の審査は 何が特殊か？



医療機器プログラムと汎用AIの違いについて

医療機器プログラム

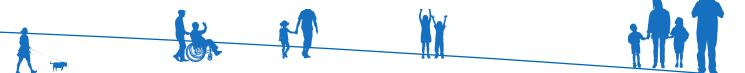
- ・ 医療機器としての目的性（疾病の診断、治療等に寄与するなど）を有しており、かつ、意図したとおりに機能しない場合に患者（又は使用者）の生命及び健康に影響を与えるおそれがあるプログラム（ソフトウェア機能）。
- ・ 提供するためには、医薬品医療機器等法に基づく製造販売承認や認証を事前に取得することが必要。

その他（汎用AIなど）のプログラム

- ・ 疾病の診断や予防、治療の目的を標榜せずに、提供されるプログラム。医療用途以外の一般的な目的で提供される、汎用AIなどが該当。

- 疾病の診断や予防、治療に用いる医療機器プログラムは、医療機器として、有効性や安全性が確認された上で提供されている。
- 一方、汎用AIなどのその他のプログラムは医療機器として承認・認証されたものではなく、疾病の診断や予防、治療の目的を標榜して提供することはできない。また、健康状態や疾病に関する質問をした場合の回答内容を含めたその性能は、医薬品医療機器等法に基づき、その妥当性が確認されたものではない。

https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/0000179749_00004.html



生成AIの活用範囲は多岐にわたり、既に一部利活用も進んでいます
何が規制されるべき製品であるか(医療機器該当性)は
別途慎重な検討が必要です

また、PMDAでは生成AIを活用した医療機器の経験はほとんどありません

本講演は、“**医療機器に該当した場合**”について
“**従来の審査経験に基づく**”
審査上どのような特殊性があるかを考察するものです。

審査で考えていること

医療全体への貢献

患者QoLへの貢献

患者アウトカムへの貢献

出力の臨床的な正しさ

技術的な品質

一般に、どの程度貢献するか
までは承認審査では求めない

臨床導入時のリスク

臨床的有用性

【診断支援系の評価例】

- ・ 臨床導入した診断成績の既存診断法との比較
- ・ 臨床導入し診断結果に基づき治療した際の効果・予後

【治療支援系の評価例】

- ・ 臨床導入による治療効果

製品性能（臨床性能）

【診断支援系の評価例】

- ・ 実使用を想定したデータ入力した際の感度・特異度

【治療支援系の評価例】

- ・ 実使用を想定した際の出カメッセージの適切性



“臨床導入を想定”するケースが
現実可能な範囲で想定できれば
従来の審査方針の応用は比較的容易



審査上課題となるような生成AIとは？

使用者からの自然文や音声による様々な入力に応じて
特定のタスクに依らない多様な回答や創造的なコンテンツを
生成する生成AIを活用したもの

- 多様な自然文・音声入力、使用者との相互的なやり取りが可能
- 特定のタスクに依らない多様かつ創造的なコンテンツの生成
- 最新の知識や入力情報をリアルタイムで取り入れた出力生成が可能

入出力の多様性が膨大

使用者の使い方の
影響が大きい

製品状態の固定が困難



生成AIの特性と評価上の課題

	特性	主な課題
運用経験に基づく学習性	パラメータの更新を伴うことなく製品の振舞いを調整することができる性質	<ul style="list-style-type: none">• どの時点の製品の評価を行うべきか• 調整され続ける製品の有効性・安全性をどのように管理するか
動的参照性	Web情報や外部データベースなど動的に変化する情報源をリアルタイムに参照して出力を生成する性質	<ul style="list-style-type: none">• 参照する情報が変化するために、同じ入力をしていても評価時点により出力が変化する• Web上にある情報を制限なく活用する場合には、誤情報等や偏った情報を取り込み、不正確な回答を生成する可能性がある
データバイアス	特に極めて大規模なデータを用いて構築されたモデルの場合、多量・多様な非標準化データを用いて学習をしているため、内在するバイアスの特定や制御がより難しくなる	<ul style="list-style-type: none">• バイアスがどのように表在化するか確認することが困難
出力の非決定性	同一入力に対しても毎回異なる出力が得られるという性質	<ul style="list-style-type: none">• ある入力に対して1回の出力だけを評価しても、その製品全体の出力を評価できていない
生成コンテンツに対する要解釈性	生成コンテンツの解釈（臨床への反映）が使用者に強く依存する	<ul style="list-style-type: none">• 自然文などの生成コンテンツにおける意味内容を客観性をもってどのように評価するか



使用上の課題

- 出力の正確性や妥当性を、使用者が判断できないことがある。また、ハルシネーションに気づけない場合がある。
- 人よりAIを信じてしまい、助言を過大評価するリスク(自動化バイアス)がある。
- 自由入力のため、適応外・禁忌・根拠の弱い入力が行われる可能性がある。
- 意図せず適応外のコンテンツ(回答)を生成してしまう可能性がある。

これを踏まえて何ができるか

市販前の
充実した評価



市販前の
限定的な評価

+

トータルライフサイクルアプローチ
市販後の
継続的管理

市販前評価

【臨床的有用性】

- ある時点・条件での有用性の評価
- (可能であれば) 開発者とは異なる第三者が作成した評価セットの利用

【製品性能 (臨床性能)】

- 異なる時点・条件での性能の評価 (正確性と安定性)

市販後の監視と報告

市販後の性能や有害事象の発生状況、ハルシネーションの発生状況等に関する情報を積極的に収集し、臨床上問題ないことの考察について定期的に報告

使用者への要請

生成AIに対するリテラシーが醸成するまでは、使用者要件や使用前研修を充実



その他の事項

特に、以下のような場合は、関連学会等のステークホルダーとの連携も重要になる。

- 従来の診療や治療のフローを変えるもの。
- 生成AIの出力の解釈や妥当性判定が明らかに困難であるもの。
- 自動化バイアスの促進により、医療者と患者のコミュニケーションに影響を与える可能性があるもの。



本内容は、**審査経験に基づくただのアイデア**です

より合理的な方法やアイデアは大歓迎です！

よりよい医療のため、よりよい開発環境実現のため
是非お知恵を貸してください！！

PMDA

健やかに生きる世界を、ともに

ご清聴ありがとうございました！
ご質問等ありましたらkato-kentaro@pmda.go.jpまで！

