

認識している。AI システムの成熟度とは、ユーザーまたは規制要件の形でシステムに課せられた制約に従って、AI システムが独自のメカニズムに基づいて制御および進化できる範囲を意味している。

私たちの成熟度モデルは制御設計に基づいており、これは、製品の品質と患者の安全を保護する制御を引き継ぐシステム機能である。また、システムの自律性に基づいており、自動的に更新を実行することで改善を促進する可能性を意味している。

AI アプリケーションの制御設計と自律性は、GxP 環境で実行するアプリケーションの能力を評価する上で重要な側面をカバーしていると私たちは考えている。したがって、ここでは、制御設計と自律性にまたがる 2 次元マトリックス (図 1 を参照) で成熟度を定義し、定義された AI 成熟度を使用してバリデーション活動の程度を識別できることを提案する。

本稿は、AI バリデーションに関するより大きなイニシアチブの一部として開発された。成熟度モデルは最初のステップである。実際、AI のバリデーションでは、データ管理やリスクアセスメントなどの他の多くのトピックを考慮する必要がある。基本的な成熟度モデルは、AI アプリケーションのリスクアセスメントに影響を与えるであろう。

本稿では、さまざまな制御メカニズムを備えた AI システムに必要なバリデーション活動と、批判的思考によって調査する必要があるさまざまな段階の自律性について詳しく説明する。目標は、自律性と制御設計の次元によって定義された領域全体 (図 1 を参照) にまたがって、類似したバリデーションニーズを持つクラスターを見つけることであった。

制御設計

表 1 は、制御設計の 5 段階を示している。

ステージ 1 では、アプリケーションは GxP プロセスと並行して実行され、データの整合性、製品の品質、または患者の安全に影響を与える可能性のある決定に直接影響を与えることはない。これには、製品にとって重要な環境で実際のデータを用いて実行されるアプリケーションが含まれる。アプリケーションは、オペレーターに推奨事項を表示する場合がある。GxP 関連情報を収集することができ、概念実証のためのパイロットがこの段階で開発される。

ステージ 2 では、アプリケーションはプロセスを自動的に実行するが、オペレーターによる能動的な承認が必要となる。アプリケーションが複数の結果を弾き出す場合、オペレーターがそのうちの 1 つを選択することができなければならない。4 つの眼の原則 (つまり、一方ではアクションのための独立した提案を行い、もう一方がチェックをする。) の観点から、システムは 1 対の両眼を掌握する。これは、人間のオペレーターによって受け入れられなければならない GxP 上の重要な出力を作成する。ステージ 2 アプリケーションの例は、オペレーターによって承認される必要があるレポートを作成する自然言語生成アプリケーションであろう。

ステージ 3 では、システムはプロセスを自動的に実行するが、オペレーターが中断および修正することができる。このステージでは、オペレーターは、AI アプリケーションによって提供される出力の拒否を決定するなど、操作中にシステム出力に影響を与えることができなければならない。実用的な例は、AI アプリケーションによって自動的に開始されたプロセスを手動で中断することであろう。

ステージ 4 では、システムは自動的に実行され、自身を制御する。技術的には、これは信頼できる範囲で実現可能で、そこではシステムが入力パラメータと出力パラメータが履歴データ範囲内にあるかどうかを自動的に制御することができる。入力データが明らかに定義された範囲外にある場合、システムは動作を停止し人間のオペレーターに入力を要求する。出力データの信頼性が低い場合、新しいデータを使用した再トレーニングを要求する必要がある。

ステージ 5 では、システムが自動的に実行されて自動的に自身を修正するため、出力を制御するだけでなく、既定の信頼性の値で出力を生成するために、変数の重み付けおよび、新規データの取得をすることで変更を開始する。

私たちの知る限り、現在、レベル 4 または 5 の医薬品製造システムは存在しない。それでもやはり業界がさらに経験を積むことで、アプリケーションがレベル 4 および 5 まで進化することを期待している。

自律性

自律性は 6 つの段階で表される（表 2 を参照）。ステージ 0 には、機械学習（Machine Learning ; ML）に基づかない複雑なアルゴリズムを備えた AI アプリケーションがある。これらのアプリケーションはアルゴリズムが固定されており、トレーニングデータに依存しない。バリデーションに関して、これらのアプリケーションは従来のアプリケーションと同様に処理することができる。

ステージ 1 では、ML システムはいわゆるロック状態で使用される。アップデートは、新しいトレーニングデータセットを使用した手動の再トレーニングによって実行される。システムは、学習によって生成された結果のメタデータを処理しないため、同じデータ入力は常に同じ出力の生成につながる。これは現在、非常に最も一般的な段階である。モデルの再トレーニングは、主観的な評価に従うか、定期的に行われる。

ステージ 2 では、システムはまだロック状態で動作しているが、アップデートは手動でトレーニングされたシステムが指示した後に実行される。この段階では、システムは生成された出力または入力データのメタデータを収集し、たとえば入力データ分布の明らかな変化に反応して、再トレーニングが必要であるか、または考慮されるべきであることをシステム所有者に提示する。

ステージ 3 では、アップデートサイクルが部分的または完全に自動化され、半自律システムとなる。これには、トレーニングデータの選択と重み付けが含まれる場合がある。唯一の人間による入力は、個々のトレーニングデータポイントの手動バリデーション或いは

はトレーニングデータセットの承認だけである。

ステージ 4 とステージ 5 では、入力データに基づく独立した強化 ML によってシステムは完全に自律する。

ステージ 4 では、システムは完全に自動化され、定量化可能な最適化目標と明確に測定可能な評価基準を使用して独自に学習する。目標は、1 つの変数または変数のセットを最適化することで定義できる。生産において、変数は特定の反応の収率と選択性の最適化であろう。

ステージ 5 では、システムは入力データのみに基づいて、明確な測定基準なしで独自に学習し、そして任務遂行能力と戦略を自己評価し、その両方を人間が理解できる形で表現することができる。ユーザーのフィードバックと修正に基づいて学習する翻訳アプリケーションが実例であろう。突然ユーザーが別の言語で入力を修正し始めたとしても、長期的には、システムは新しい言語への翻訳を提供するであろう。

バリデーションレベル

成熟度レベルは、6 つの AI バリデーションレベルにクラスター化され（表 3 を参照）、自律性と制御設計の次元によって定義された領域に配置することができる（図 2 を参照）。AI バリデーションレベルは、システムの規制順守を高次元で達成するために必要な最小限の制御手段を表す。詳細な品質保証要件は、AI システムの使用目的とリスクポートフォリオを考慮して、分類に基づいて個別に定義する必要がある。

AI バリデーションレベル I のシステムは、製品の品質と患者の安全性（およびデータの整合性）に影響を与えない。したがって、バリデーションは必須ではない。それでもやはり、このカテゴリのアプリケーションでは、人的要因を過小評価してはならない。システムがアドバイスを提供するように設計されており、通常のプロセスと並行して長時間実行されている場合、オペレーターが批判的思考に基づいて結果を処理し、決定を正当化するためにこれらの結果を使用しないようにするための保護手段を講じる必要がある。

AI バリデーションレベル II のシステムは、ML に基づかない AI アプリケーションであるため、トレーニングは不要である。結果は純粋にコードベースで決定論的であるため、従来のコンピュータ化システムのバリデーション手法を使用してバリデーションすることができる。

AI バリデーションレベル III のシステムは、ML や深層学習などのメカニズムに基づいている。それらは、アウトプットを生成するためにデータを用いたトレーニングを必要とする。このカテゴリのシステムは、再トレーニングが実行されるまでロック状態で動作している。

バリデーションには、従来のコンピュータ化システムのバリデーションに加えて、データモデルと使用データに関連する AI 固有の対策を実行する必要がある。トレーニングデータの完全性をバリデーションする必要がある。開発に使用されたデータが特定の出力を

生成するのに適切であり、偏りや破損がないことを確認する必要がある。AI バリデーシ
ン文書は、次の側面をカバーする必要がある：

- データの全ての抽出、変換、読み込み（ETL）処理段階のリスク分析
- データ完全性への潜在的な影響に関するデータ変換の評価
- ラベルがどのように作成され、品質が保証されているかに関する手順

さらに、モデルの品質は、開発段階と運用段階でバリデーションする必要がある。開
発中に、次のことを確認する必要がある：

- 選択されたアルゴリズムがユースケースに適している
- トレーニングされたモデルが入力データに基づいて予想される結果を厳格に提供
している

運用フェーズでは、これらの追加の側面を考慮して定義する必要がある：

- モデルのパフォーマンスを監視するための適切な品質基準
- モデルのパフォーマンスに応じて再トレーニングを開始するために必要な条件

再トレーニングでは、モデル入力の入力構造が同じままであることが望ましい。それ
以外の場合は、開発フェーズの方法論的設定の新しい評価が必要になる場合がある。

システムがバリデーションされた範囲でのみ動作していることを確認するには、動作
中の入力データを監視する必要がある。さらに、このカテゴリと上位のシステムでは、異
なる入力データに基づく出力の生成について論理的根拠が明確でない可能性があるため、
透明性の問題が発生する。このため、結果までの経路の深い理解を提供するために利用可
能な全ての情報を可視化する必要があり、また、意思決定プロセスを検証し関係者に説明
と理論的根拠を提供するために説明可能性調査（AI を活用した意思決定プロセス、AI モデ
ル自体、およびその予想される影響と潜在的なバイアスを説明することで、AI アプリケー
ションへの信頼を構築することを目的としている）を実施する必要がある。

AI バリデーションレベル IV のシステムは、アップデートプロセスのさまざまな局面が
自動化されているため既に大きな自律性を継承しており、これには新しいトレーニングデ
ータの選択が含まれる場合がある。このため、運用中のモデルの品質を反映する重要業績
評価指標の制御に重点を置く必要がある。モデルの出力品質を監視して、検証された範囲
内にあることを確認する必要がある。さらに、システムが再トレーニングを必要とする場
合、または検証された範囲外で動作している場合の通知プロセスを確認する必要がある。

AI バリデーションレベル V のシステムは、より優れたプロセス制御を備えている。し
たがって、運用中はより強力なシステム制御を行う必要がある。これは、定義されたテス
トデータセットを使用して定期的に再テストすることで実現できる。さらに、バリデーシ
ョンフェーズでは、自己制御メカニズムをバリデーションする必要がある。

AI バリデーションレベル VI のシステムは、自己学習システムである。近い将来、継続
学習システムの制御のための戦略が利用可能になると予想される。このカテゴリにおいて、

システムの規制順守を保証するために現在利用できるバリデーシヨンの概念はない。

要約すると、このフレームワークは、運用中に AI システムを制御する組織の負担と、ますます自律的になる AI システムを保証するために増加するバリデーシヨン活動を容易にする技術要件はトレードオフであることを説明しており、それはフレームワークの下位レベルでより顕著となる（図 2 を参照）。

成熟度測位と動的経路

これまでに概説したフレームワークに従うことにより、AI システムの制御設計は次の側面に関してサポートされる：

運用プロセスに人間の制御をどれくらい組み込むべきかについて決定する必要がある。

初期設計

AI 制御モードの初期設計では、批判的思考に基づいて人間の制御を運用プロセスにどれくらい組み込むべきかについて決定する必要がある。たとえば、初回の開始では、自律性が低く制御が多いモードを選択することができ、それゆえに技術的な枠組みに関する要件が軽減されるが、運用上の負担は大きくなる。この決定は、特定の AI システムの使用意図とリスクポートフォリオ、および AI システムの設計と保守全般に関する企業の経験に基づいて批判的に決定する必要がある。本稿では、AI に対する固有のリスクアセスメントのしくみについては取り上げていない。

動的経路

AI システムが確立されたら、バリデーシヨン活動、販売後の監視、およびリスク評価の更新からの結果と、運用上の負担と技術的な負担のバランスをビジネスの観点から考慮して、制御設計と成熟度範囲での位置が依然適切であるかどうか継続的に評価する必要がある。この評価は、設計をどちらの方向にも向けることができ、たとえば、新たに特定されたリスクを考慮して、制御設計を強化する（人間による制御を強化し、自律性を低下させる）か、もしくは、より厳格な技術的管理措置を伴って、AI システムの自律性を拡大する可能性もある。

管理者は、システムの自律性と人間による制御に関して、ライフサイクル全体で AI ソリューションを精力的に推進するための戦略的手段として成熟度モデルを検討できるかもしれない。

例（図 3 を参照）：

1. 企業は、既存の GxP 関連プロセス（AI バリデーシヨンレベル I）と並行して、特定のユースケースでの AI システムの有用性について調査することを決定する。
2. AI システムの導入が成功した後、ロック状態の操作モードかつ全ての事例を制御しながら（AI バリデーシヨンレベル III）、AI システムは GxP プロセスを引き継ぐ

必要がある；同時により厳密な技術的および機能的バリデーション活動が導入される。

3. AIシステムの付加価値をさらに拡張すると、制御設計は必ずしも全ての事例を制御しないモードへ変更される；まだロック状態のモードで動作しているため、AIバリデーションレベル III が適用される。ただし、GxPプロセスの重要性を考えると、事例のサンプルチェックなどのさらなる制御が導入される可能性がある。
4. 特定のユースケースでAIシステムに関して十分な経験を積んだ後、自律性が向上することでシステムが必要な再トレーニングを知らせるようになる（AIバリデーションレベル IV）。
5. システムの自律性をさらに拡張することで、トレーニングプロセスは、AIシステムのしくみ、つまり再トレーニングの実行方法に重点を置くようになるが、そのような新しいバージョンのアクティブ化は、依然として人間のオペレーターによって検証される（依然としてAIバリデーションレベル IV）。
6. ソリューションの成長経路の最終段階として、制御ステージ4が選択されることでシステムが自身を制御するようになる（AIバリデーションレベル V）。

結論

AI/ML 固有の規制がないため、業界および開発パートナーは、適切な数のバリデーション活動を決定する他の方法を必要としている。本稿で説明する AI 成熟度モデルは、AI モデルの自律性と制御設計の段階に基づいてバリデーションレベルを区別する論理的根拠を提供している。この成熟度モデルは、さらなる議論の出発点であり、製薬業界における AI/ML に基づくアプリケーションのバリデーションのための包括的なガイドラインの基礎に成ると考えている。私たちのモデルは、他のライフサイエンス業界での応用に大きな可能性を秘めていると確信している。

本文以上

表 A 制御設計ステージ

表 B 自律性ステージ

表 C バリデーションレベル

図 1 成熟モデル

図 2 ビジネスの意思決定に焦点を当てたバリデーションレベルの表現

図3 自律性と制御のフレームワーク内での動的 AI システム経路の例