

VCDesignアーキテクチャの価値継続性に関する比較検証

VCDesign (Value Continuity Design) は「設計対象の価値を継続して守る」ことを重視したアーキテクチャ思想です¹。以下では、このVCDesignが従来手法と比べてどのように優れた構造を持ち、価値の継続という難題に挑んでいるかを3つの観点で比較検証します。それぞれ課題、現状のベストプラクティス、VCDesignのアプローチ、解決される領域の順に整理します。

1. システム設計の寿命・変更可能性に関する比較（クリーンアーキテクチャ・DDD など）

課題：システムは長期運用の中で環境変化や要件変更さらさらされ、初期の設計意図が失われがちです。時間経過とともに技術的・構造的負債が蓄積し、システムは「動いているのに価値を発揮できない」状態に陥ります²。これは既存構造では本来の価値を維持できなくなる瞬間、いわゆる価値断絶の発生につながり³、結果としてシステムの寿命を縮めてしまいます。また、当初の判断理由や文脈が失われると、変更時に誤った改修を行い価値を毀損するリスクが高まります⁴。

現状のベストプラクティス：システムの寿命延長や変更容易性を高めるために、業界ではクリーンアーキテクチャやドメイン駆動設計（DDD）などが採用されています。クリーンアーキテクチャでは関心事を明確に層分けし、ビジネスルールとインフラ技術を分離することで変更の影響を局所化します⁵。例えばUIやデータベースなど外部要素を内核から切り離すことで、個々のコンポーネントを他に影響させず交換・改修可能にし、システムの柔軟性と保守性を高めます^{5 6}。一方DDDはビジネスの核心ドメインに設計を集中させ、境界づけられたコンテキストごとにモデルを構築することで複雑さを抑制します。DDDによりソフトウェアを業務領域に密接に適合させ、論理的で適応性の高い構造を作ることで、変更要求に的確に対応できる柔軟性を確保します⁷。これらベストプラクティスは、層の独立性や依存関係逆転によって修正時の波及を抑え、長期的な保守容易性を実現する点で成果を上げています⁸。

VCDesignのアプローチ：VCDesignは従来のコード構造中心の手法とは異なり、時間を超えて受け継ぐべき判断基準や責任そのものを設計対象に含める点に特徴があります^{9 10}。具体的には、「どの判断がいつ誰によってなされたか」「その前提条件は何か」を機械可読な形で体系化し、将来に渡って参照可能にします¹¹。こうした判断の境界（どこから先を変えてはならないか）を明示した憲法的な層を設け、設計当初からシステムに組み込むのです¹²。VCDesignでは実装や機能ではなく、判断と責任の所在を中核的な設計要素とします¹³。これにより、人間の判断が時間とともに風化・散逸することを防ぎ、システム変更時にも当初の価値基準が参照・踏襲される仕組みを提供します^{4 14}。例えばVCDesign上のリファレンス実装であるBOA (Boundary-Oriented Architecture) では、事実・解釈（意味）・責任を厳密に分離してシステム構造に反映します¹⁵。これによって、時間経過で意味づけ（解釈）が変化しても事実データや責任の境界が崩れないようにし、組織変更や技術刷新下でもシステムが本質的価値を保ちやすくなります^{16 17}。

解決される領域：VCDesignは以上のアプローチにより、システム寿命を脅かす「価値の静かな喪失」を構造的に防ぎます¹⁸。クリーンアーキテクチャやDDDがコードの分離で変更容易性を高めるのに対し、VCDesignは価値の前提（判断基準）の分離と継承によって変更後も価値が損なわれないことを保証します^{10 19}。その結果、実装が自動生成で置き換わるような状況でも意図した価値は維持されうるとされています^{20 21}。もっともVCDesignも万能ではなく、環境や目的自体が大きく変わる場合（章の転換）には再設計が必要になる点は認識されています^{22 23}。しかし「判断の継承される設計」というVCDesignの構造

は、少なくとも従来手法では対応しきれなかった領域（設計判断の風化による価値毀損）を埋めるものであり、価値継続性の困難な課題に対し新たな解決の地平を開いています 24 25。

2. AI品質向上のための Human-in-the-loop (HITL) 導入との比較

課題： AIシステム単独では出力品質にばらつきや誤りが生じ、人間の意図や倫理から逸脱する危険があります。高度なAIであっても**人間の監督なしでは自己の誤りに気付かず暴走しうるため** 26、信頼性確保のためには人間によるフィードバックや判断介入が欠かせません。完全自動化されたモデルは、一見高効率でも不適切な判断や偏った結果を**確信をもって量産**してしまう恐れがあり 27、そのまま実運用すれば大きなリスクを招きます。

現状のベストプラクティス： そこで近年は**Human-in-the-Loop（人間をループに組み込む）**手法がAI品質向上の鍵とされています。具体的には、モデルの学習段階で人間がフィードバックする**強化学習（人間のフィードバック付き）**や、推論段階で人間が結果をレビュー・承認するプロセスが広く採用されています。HITLにより**アルゴリズムのスピードと人間の判断力を組み合わせ**て運用することで、AIはデータ処理を、人間は意味解釈と価値判断を担当し、お互いの長所を活かす形になります 28。例えば大規模言語モデルでは人間が出力内容を評価・調整しながらモデルを調教し、継続的に品質を向上させています。また自動運転や医療AIでも、重要局面では人間のオペレーターや専門家が介入・確認する仕組みが取られています。「AIに全て任せきりにしない」ことで、AIシステムは**現実とのズレを人間のフィードバックで修正し続ける**ことができ、結果として精度と安全性が担保されるのです 29 30。HITLは単にミスを検出するだけでなく、人間からの継続的な教示によってAIを理想的な方向に育てていく**品質改善のループ**でもあります 31 32。

VCDesignのアプローチ： VCDesignにおいては、人間を介在させるこの考え方がアーキテクチャの根幹に組み込まれています。特に**責任のクローズ（Responsibility Closure）**の概念により、AIが出した提案や判断（解釈）をそのまま最終決定として実行することを禁止し、必ず明示的な承認プロセスを通す構造になっています 33 34。VCDesign上のBOA実装では**責任の境界**に「AIゲートキーパー」とも言える **RCA（Responsibility Closure Agent）**を配置し、AIの出力に対し以下のいずれかの意図を明示させます 35 36：

- **ACCEPTED:** 出力に責任を持てるとAI（システム）が判断し、人間に代わって署名する
- **DENIED:** 前提条件を満たさないため出力を採用せず棄却する
- **UNKNOWN:** 判断に必要な情報が不足しており可否を決められない

AIが自動決定を下せない場合（DENIEDやUNKNOWN）には、**必ずプロセスを停止または人間にエスカレーション**される規約になっています 37。このように**あいまいなまま処理を先に進めない**仕組みを徹底することで、AIが誤判断をそのままシステムに反映することを防いでいます 38。さらにVCDesignの **Resolution Protocol（RP）**では、AIによる「意味上の結論（Meaning）」を**人間の判断による承認（Resolution）に昇格させるには一定の手続きを踏む**ことを定めています 39。具体的には「構造的チェック（事前条件の確認）」「責任の明示的受容（誰か人間がその決定を引き受け署名する）」「記録の保存（追跡可能性の確保）」という3要件を満たさなければ最終決定とみなさないとしています 40。これらの仕組みにより、VCDesign環境下ではAIの判断結果が人間の目を経ずに独走することは構造上あり得ません。設計プロセス上も、**P（前提）-D（実行/表示）-Δ（差分/異常）-R（解決）**という責任クローズループが定義されており 41 42、いかなる未解決のΔ（判断保留事項）が残っている限り設計完了とは認めないとすることで 43、人間の関与すべき判断抜け漏れをチェックリスト的に洗い出すようになっています。

解決される領域： VCDesignの構造は、HITLの考え方を徹底的に制度化することで**AI品質と信頼性を根本から支える仕組み**を提供します。人間がループに入る従来手法では運用面の努力や文化に依存する部分も大きいですが、VCDesignでは**アーキテクチャレベルで人間の介在を強制**するため抜け漏れがありません 34。これにより「気付かぬうちにAIが暴走する」「人間のチェックを経ていない仕様外の振る舞いが発生する」といった事態を原理的に防止します。特に「**AIに責任を持たせず、人間が最終責任を持つ境界を動かさない**」という方針は、AIシステムの暴走を防ぐ安全装置として機能します 44。その結果、たとえAIが高速に誤った推論を

しても自律的に実行されることはなく、常に人間の判断によって現実への反映がコントロールされます。これはHITL導入が目指す「AIに人間の良識と判断を組み合わせる信頼性を高める」という目的を、設計論として具現化したものと言えます。もっとも、このアプローチは人間の関与を前提とするため即時自動処理より反応が遅れるケースもあります。しかし**信頼できるAI**を追求する上では不可避のトレードオフであり、VCDesignはスピードよりも**価値の堅持と誤動作防止**を優先して設計がなされています¹⁷。その結果得られる人間とAIの協調的なプロセスは、単に人手でAIを補佐する以上に、**システム全体が常に人間の価値観と現実**に照らして最適化され続けるという意義を持ちます^{45 32}。VCDesignはこの領域に踏み込むことで、「AIの品質向上策」としてのHITLを超え、**AIと人間の共同による価値継続的な最適化**を実現しようとする構造と言えます。

3. 自律AIにおける倫理・安全・可観測性・人間の理解可能性に関する比較

課題: AIが高度に自律的になるにつれ、システムの倫理性・安全性を如何に担保するかが重大な課題となっています。ブラックボックス化しがちなAIの意思決定過程を人間が**観測（オブザーバビリティ）**し、理解・説明できるようにすることも求められます。例えば、AIが差別的な判断をしていないか（倫理）、異常な動作で人間社会に害を及ぼさないか（安全）、内部で何が起きているか監視できるか（可観測性）、そして出力や判断の理由を人間が理解・納得できるか（説明可能性）が重要視されています。近年の事例でも、不透明なAIの判断ミスが大きな事故や不信感につながったケースは少なくなく、**AIの価値を継続して享受するには信頼性と説明責任の確保が不可欠**です^{46 47}。

現状のベストプラクティス: こうした課題に対し、AI業界ではいくつかの対処アーキテクチャや原則が模索されています。まず**倫理性**については「AI Ethics by Design」の考え方が提唱されており、システム開発の初期から倫理的风险を評価・低減する枠組み（倫理フレームワークの策定、バイアス検出・緩和手法の導入、透明性の確保など）を組み込むことが推奨されます^{48 49}。**安全性**の面では、AIに対する安全境界やフェイルセーフを設ける設計が取られます。具体的には、**形式手法による検証**でAIが安全制約を破らないことを証明したり、冗長系を用意して一方の暴走時に強制停止させるなどのアーキテクチャが考案されています⁵⁰。例えば自動運転車では、メインAIとは別に安全モニタが走行を監視し、異常時には自動ブレーキを作動させる仕組みがあります。また**可観測性（オブザーバビリティ）**については、AIモデルの挙動を継続的にモニタリングするプラットフォームや手法が登場しています。モデルの入力データ分布や出力傾向を追跡し、ドリフトやバイアスの兆候を検知することで問題を早期に発見する狙いがあります^{51 47}。**説明可能性**に関しては、XAI（Explainable AI）の研究が進み、モデルの決定理由を人に説明するための手法（例えば特徴量の寄与度を示すSHAP値や、生成系AIではプロンプトトレースや理由文の出力など）が実用化されつつあります⁵²。さらに重要なのは**人間のガバナンスと監査**で、AIの意思決定プロセスや結果を記録・監査し、必要に応じて人間が介入できるガバナンス体制（AI倫理委員会の設置や運用ガイドライン策定など）をアーキテクチャに組み込む動きも見られます^{53 54}。総じて、「**AIを野放しにせず、人間の目と安全網を組み込む**」ことが最近の自律AIシステム設計における合意事項と言えるでしょう^{55 56}。

VCDesignのアプローチ: VCDesignはその設計哲学上、上記の倫理・安全・可観測性・理解可能性の課題に正面から応える構造を備えています。まず**倫理・安全**について、VCDesignは「AIを制御して安全にする」のではなく、「**AIが責任の始まりを勝手に決めることを許さない**」ことで安全性を担保する姿勢をとります⁴⁴。すなわちシステム内でAIが関与できる領域（Meaningまで）と、人間のみが責任を負う領域（Resolution/Responsibility）を明確に線引きすることで、AIが人間の倫理的判断領域を侵犯しないようにしています^{57 58}。この**境界の設計**により、たとえAIが不適切な提案を行っても自動実行されることはなく、必ず人間のレビューを経るため倫理的・安全面の制御が効きます^{34 38}。次に**可観測性**の点では、VCDesignでは判断のプロセスと根拠が常に**記述され追跡可能**になっています⁵⁹。責任ある判断が下された際には人間が明示的に署名し記録されるため、後から「なぜその決定に至ったか」「誰が承認したか」をログやマニフェストから辿ることができます^{60 61}。さらにVCDesignの機械可読な設計文書（責任マニフェストなど）は、人間とAI双方の参照点となる**単一の真実源**として機能し、システムが前提とする価値基準を誰もが監視・確認できる形で提示します^{19 62}。これはいわばシステムの**憲法**であり、可観測性と透明性を高める重要な要素で

す。最後に**人間の理解可能性（説明可能性）**について、VCDesignのFact/Meaning/Responsibilityの三分割は、人間がシステムの動作を理解する助けとなります。AIが参照した**事実（データ）**と導出した**解釈（意味）**、そして最終的に下した（または人間が下した）**決断**が明確に分離されているため、それぞれを個別に検証・説明できます⁶³⁶⁴。例えば「AIがある提案をしたが採用されなかった」場合でも、提案内容（Meaning）自体は記録され人間が吟味できますし、採用しなかった理由も決定プロトコル上で明示されます。これにより**システムの振る舞いを後から人間が再現し、納得のいく形で説明できる**のです。加えて、BOAアーキテクチャでは**長期的な説明可能性の維持**が要件に含まれており⁶⁵、時間が経って組織メンバーが入れ替わっても判断のトレーサビリティやシステムの意図が理解できるよう設計されています⁶⁶⁶⁷。

解決される領域： VCDesignは以上のように、**自律AI時代に求められる諸課題を体系的にカバーする設計上の構造**を備えています。倫理・安全面では、人間の責任領域を侵さない境界設計とエスカレーション機構により、AIが暴走して重大な害を及ぼすリスクを大幅に低減します⁴⁴³⁴。可観測性・説明性の面では、判断過程の明示的な記録と責任の所在の可視化によって、従来ブラックボックスになりがちだったAIの挙動を監査可能なものにします⁴⁰⁶¹。結果としてVCDesign上のシステムは、変化し続けるAIであっても「**正しく動いているか**」「**おかしな偏りが出ていないか**」を継続監視し、**価値基準に反する動きがあれば人間が介入して修正できる状態を保ちます**⁶⁸⁶⁹。もっともVCDesign自体は「何が倫理的に正しいか」までは定義しません⁷⁰。倫理原則そのものの設定やデータの質の担保は別途人間のガバナンスに委ねられます。しかしVCDesignは「**倫理・安全のガバナンスが機能するための土台**」を提供しており、AIに関わる価値基準をシステム内に頑強に組み込む点で他のアプローチにない強みを持ちます⁷¹⁷²。総じて、VCDesignアーキテクチャは自律AIに不可欠な倫理性・安全性・透明性を支える構造上の工夫を凝らし、**システムが価値を継続して提供しうる条件（信頼と責任の枠組み）を守り抜く**ことに主眼を置いています⁷³。

おわりに

以上の比較から、VCDesignアーキテクチャは「**価値の継続**」という困難な目標に対し、既存のベストプラクティスや最新アーキテクチャが解決しようとしてきた課題を包括的に捉え直していることが明らかです。システム設計の寿命・変更容易性の面では、従来の層化やドメイン重視の手法に加えて**意思決定の継承性**を設計に織り込み、技術的变化を超えて価値を保つアプローチを提示しました。AI品質の面では、人間を巻き込むHITL手法を構造レベルで制度化することで、品質と信頼性を継続的に確保する仕組みを構築しました。さらに自律AIの倫理・安全・可観測性の面では、境界の明確化と責任の見える化によって、AI時代のシステムに必要なガバナンス基盤を提供しています。それぞれの比較から見てきたのは、VCDesignが単なる理想論ではなく、**現在直面している課題を解決し得る具体的な構造を備えている**という点です。その構造は人間とAIの新たな協調関係を支えるものであり、変化し続ける環境下でもシステムが「**本来達成すべき価値を途切れさせない**」よう設計されています⁷³。もちろん最終的な成果はこの設計哲学をどう実装し運用するかにかかっていますが、VCDesignが提示する原則と構造は、価値の継続という難題に対する有力な解答となりうるでしょう。

参考資料： VCDesign公式サイトおよびGitHubリポジトリ⁷⁴⁷⁵、関連する哲学ドキュメント⁷⁶⁷⁷、BOAアーキテクチャのREADME⁷⁸⁷⁹等を参照して構成しました。各種ベストプラクティスやAI倫理の知見については業界ホワイトペーパーや記事⁵⁵⁰²⁸を参考にしています。

1 4 9 11 13 18 44 70 73 76 77 PHILOSOPHY.md

<https://github.com/VCDesign-org/vcdesign-core/blob/ef9e268d053659a06dd456f1373dfb632baa9eea/docs/philosophy/PHILOSOPHY.md>

2 3 20 21 22 23 glossary.jp.md

<https://github.com/VCDesign-org/vcdesign-core/blob/ef9e268d053659a06dd456f1373dfb632baa9eea/docs/glossary.jp.md>

5 6 8 Comparison Table: Hexagonal, Clean, Onion, and Traditional Layered Architectures · GitHub

<https://gist.github.com/EliFuzz/8ab693db36ff33ead1445a43c3f0ef7e>

7 Domain-Driven Design (DDD) Overview and Best Practices - Taazaa

<https://www.taazaa.com/domain-driven-design-ddd-overview-and-best-practices/>

10 14 19 24 vcdesign_core.yaml

https://github.com/VCDesign-org/vcdesign-core/blob/ef9e268d053659a06dd456f1373dfb632baa9eea/specs/vcdesign_core.yaml

12 15 16 17 25 34 35 36 37 38 39 40 59 60 61 65 66 67 71 75 78 79 README.md

<https://github.com/VCDesign-org/boa-core/blob/1f2f2c46a4769261115f09a7c76b188edf5e0d38/README.md>

26 27 28 29 30 31 32 45 Human-in-the-Loop: Why It's the Real Secret Behind Great AI | by Coursiv | Nov, 2025 | Medium

<https://medium.com/@coursiv/human-in-the-loop-why-its-the-real-secret-behind-great-ai-089fe12cc264>

33 57 58 63 64 74 From Coders to Legislators: Designing Boundaries in the Age of AI | by Hironobu Arakawa | Jan, 2026 | Medium

<https://medium.com/@arakawa.hiro/title-from-coders-to-legislators-designing-boundaries-in-the-age-of-ai-d729f2a079cc>

41 42 43 62 index.md

<https://github.com/VCDesign-org/vcdesign-core/blob/ef9e268d053659a06dd456f1373dfb632baa9eea/docs/protocols/rcl/index.md>

46 47 51 68 AI Observability: Key to Reliable, Ethical, and Trustworthy AI | Lakera – Protecting AI teams that disrupt the world.

<https://www.lakera.ai/blog/ai-observability>

48 49 50 52 53 54 55 56 69 72 AI Ethics – Part II: Architectural and Design Recommendations - Architecture & Governance Magazine

<https://www.architectureandgovernance.com/uncategorized/ai-ethics-part-ii-architectural-and-design-recommendations/>