主権的教育 AI の実装と演出家モデル

—32 回のリテイクが導く知的主権の育成—

山口伸也*1
Email: gachapinkingu@yahoo.co.jp
＊1: 長崎県立川棚高等学校

©Key Words　生成 AI，AI 言語設計，演出家モデル，情報セキュリティ，守破離

1. はじめに

生成 AI の急速な普及は，教育現場に二律背反の課題を突きつけている。大規模言語モデル（LLM）による個別最適化への期待の一方で，クラウド型 AI への情報入力による流出リスクが公教育の安全性を脅かしている。文部科学省は「初等中等教育段階における生成 AI の利活用に関するガイドライン（Ver. 2.0）」(1)を公表し国レベルの方向性を示したが，各自治体のセキュリティ規律との整合性の確保には現場での努力が必要である。

本研究の実践校が準拠する県立学校情報セキュリティ基本方針(2)，および同方針に基づき教育庁が定める一連の規定は，校務系と学習系ネットワークの厳格な物理的分離を規定し，Edge／Chrome のブラウザ分離と Microsoft Teams を基軸としたネットワーク二層構造という堅牢な県域インフラを整備している。本研究は，この県域インフラの安全性を大前提として受容した上で，それを現場の教育課程で安全かつ機能的に運用するための情報ガバナンス体制──信号機ルール，Teams ラボ，知能資産 23 点，教科別指導案等──を本校独自に構築・実装した実践の報告である。

本研究の問題意識は三点に集約される。第一に，県域インフラの堅牢性を前提としつつも，それを日々の多様な教育課程で機能的に運用するための，現場レベルの具体的な運用プロトコル（動線指針）を模索する段階にあること。第二に，AI を「正解の自動供給装置」と見なす認識に留まり，学習者の思考を研ぎ澄ます「砥石」として再定義する視座の欠如。第三に，個々の教員の属人的実践が組織的に持続可能な仕組みへと統合されていないことである。本稿は，これら三つの課題に対し 5 層からなる「組織的知性マネジメント」モデルを提示する。

生成 AI の教育利用に関する既往研究は，個別授業における対話活用とその学習効果の検証(3)，あるいはプロンプト技法の精緻化による出力の最適化(4)に焦点を置いてきた。本研究の差分は，組織的なセキュリティガバナンスと教育方法論を単一のアーキテクチャに統合した点，および出力の最適化ではなく，あえて過剰に生成させ人間が削ぎ落とす「演出家モデル」という逆方向の設計思想を採る点にある。

2. 県域インフラへの適合と現場ガバナンスの構築

2.1 ブラウザ分離と規律適合性の検証

長崎県教育委員会が定めた前述の基本方針(2)およびその下位規定は，校務アカウント（ドメイン A）を Edge，学習アカウント（ドメイン B）を Chrome で運用するブラウザ分離を規定し，クラウド AI による「情報のスープ化」──全データの横断的インデックス化による分離不能な混入──を物理的に遮断する堅牢なガバナンス基盤を提供している。令和 8 年 3 月のパイロット授業において，Copilot デスクトップアプリの認証時の挙動が Windows OS 上の校務セッションと密結合し，学習アカウントでのログインに際して追加の技術的検証が必要となった。筆者は前年度の Teams 運用で蓄積した知見から仮説を導出し，県が整備したブラウザ分離の枠組みを維持したまま，Chrome Web 版をプラットフォームとして選択することでログの継続性を担保した。

以下，本選択の規律適合性を検証する。Chrome Web 版 Copilot へのアクセスは，学習系セカンドドメイン（ドメイン B）のセッション内に完全に閉じており，校務系テナント（ドメイン A）のデータには一切接触しない。Edge が保持する校務セッションとの間にセッション境界が物理的に存在し，基本方針(2)とその下位規定が規定する校務・学習間の物理分離要件は完全に充足されている。「情報のスープ化」は論理的に遮断されており，本選択は県の定めるセキュリティ規律の精神と要件に整合する技術的適応である。国語科教諭の「前回の続きから表現を深化させたい」という要請は，技術管理と教育目標が「ログ」という一点で接続する契機であり，この観測基盤が第 4 章の実証の前提条件をなす。

2.2 信号機ルールと 23 点の知能資産

物理的隔離だけでは善意のヒューマンエラーを排除できない。筆者は「信号機ルール」（赤＝入力禁止，黄＝マスキング，緑＝推奨）を設計し，Teams ラボの 01 チャンネルにデジタル・プロトコルとしてデプロイした。教員が資料にアクセスする動線上に規律が必ず存在する「環境一体型規律」を実現し，ゼロトラストの原則を組織の共通言語として定着させた。02 チャンネルには全 23 点の教育資料を格納した。内訳は，教科別 AI 活用ロードマップ 10 点（国語・英語・社会・数学・物理・生物・生活・情報・保体・全校ドクトリン）と，教科別指導案 13 点（国語・英語・数学・生物・化学・物理・日本史・公民・世界史・生活総合・情報・保健・音楽）である。

3. 知の循環基盤と演出家モデル

3.1 AI の「砥石」としての再定義

本研究は AI を「知の砥石」として再定義する。砥石自体は刃物ではなく，研ぐ行為の主体はあくまで人間（学習者）である。従来の「作成 80%：検品 20%」から「作成 20%：検品 80%」への労力配分の構造的転換がこの再定義の核心であり，AI が素材を大量に生成し，人間が判断と選択に集中する構造を実現する。

従来のプロンプト設計論(4)は，AI を「一発必中の自動販売機」として扱い，手順（How）を逐一指定する細密命令（Befehlstaktik）に陥っていた。入力を精緻に整えれば完璧な出力が得られるという前提は，プロンプトの起草に人間のリソースを浪費させ，かつ出力への無批判な依存を助長する。本研究はモルトケの委任命令（訓令戦術：Auftragstaktik）(5)に着想を得て，この前提を根本から転換する。ただし，本概念は軍事理論の厳密な移植ではない。「目的のみを伝え，手段は委ねる」という制御原理を教育文脈へ翻案するための着想源（アナロジー）として位置づけるものであり，理論的な妥当性の検証対象は，あくまで後述する教育実践のデータである。

万能型紙の 4 要素（#役割・#命令・#条件・#参考情報）は，一回の入力で完成した出力を得るための固定的な型

ではない。それは，AI という砥石に対し人間が思考を働かせるための起点である。初手で大局的な「目的と価値観（What/Why）」のみを注入し，AI が自律的に返してきた粗削りな素材に対して，人間が「80%の検品（眼による精査）」と再質問（リテイク）を厭わず繰り返す。この対話の反復によって学習者の知的主権を漸進的に磨き上げる過程こそが，本制御原理の骨子である。

3.2 相互検証モデルと守・破・離

AI の同調バイアス（sycophancy）[6]──大規模な LLM ほど対話相手の選好に迎合した回答を返す傾向──を克服するため，役割を先鋭化させた複数のペルソナに討議させ，相互に出力を監視させる「相互検証モデル」を設計した。各ペルソナが固有の視点から他の出力を批判的に検討することで，迎合的応答とハルシネーションの発生を抑制する。さらに教育実装の局面では，「情報科のブレーキ」（ハルシネーション検出・情報分類）と「国語科のアクセル」（拡張型指示による生成能力の解放）を組み合わせる文理融合の構成とした。検証機能のみでは生成が萎縮し，生成機能のみでは妥当性の検証を欠く。両者の併用が STEAM 教育の理念を AI 活用において具体化する。この理論を守→破→離の 3 段階でスパイラル的に循環させる育成ロードマップとして実装した。

表 1　「守・破・離」育成ロードマップ

段階	目標	検品の深度
守	リスク検知・意図同期	Lv.1 意図の同期
破	デバッグ・論理監査	Lv.2 論理の検証
離	倫理・主権・統合	Lv.3 創造的判断

4. 国語科の実践とリテイクデータの検証

4.1 パイロット実践の設計

令和 8 年 3 月，国語科教諭との協働で高校 2 年生 21 名を対象に三行詩創作を実施した。教諭は型紙を用いず「ふくらませて」という素朴なアクセルで生徒を AI と向き合わせ，「最低 10 回は修正指示を出すこと」を条件とした。

ここで「ふくらませて」という指示の技術的特性を定義する。3.1 節で批判した細密命令が，ハルシネーションを排し一発で正確な出力を得るための固定的な型として機能するのに対し，「ふくらませて」は LLM の持つ確率的な言語生成能力を極限まで解放する拡張型コマンドである。この指示は AI に対し，語彙の選択肢を意図的に広げ，修辞・比喩・情景描写を過剰に重畳させた，極めて冗長度の高い言語素材を生成させる。その出力には必然的にノイズ──学習者の意図とは乖離した表現──が大量に混入する。

この設計は，3.1 節の制御原理と正確に対応している。「ふくらませて」は大局的な目的（What：表現を豊かにせよ）のみを注入し，具体的な手順（How：どの語彙を，どの構文で）を一切指定しない。これは委任型制御の教育的実装に相当する。教諭はこの一語によって，AI の出力精度を高めるのではなく，生徒の前に，選び取るべき素材とそぎ落とすべきノイズが混在する広範な言語素材を提示させた。生徒に求められたのは，この素材に対して 80%の検品（眼による精査）を繰り返し，自己の内面にある固有の情景や感情との整合性を問い続ける行為である。

なお本稿では「リテイク」を，生徒が AI に送信した発話のうち，修正指示・問い返し・完了判断を含む一連の対話ターンとして操作的に定義する。これは，委任型制御において対話の各ターンが意図同期の構成単位をなすという 3.1 節の枠組みに基づく。「どう思う」という問い返しは AI に判断を委ねて素材を引き出す行為であり，「満足した」という完了判断は自己の表現意図との一致を確認した主権的判断の表明である。いずれも外的条件を離れて生徒が対話を主導した契機であり，リテイク過程の正当な構成要素として計数する。後述の回数は，この定義に基づき全生徒について同一基準で集計したものである。

4.2 定量データと同調バイアスの否定

表 2　三行詩創作の定量データ（N=21）

指標	数値	教育的意味
AI 初期出力の採用率	0%	全員が自己の表現意図を優先
修正指示の平均回数	約 12 回	反復的な意図の同期（中央値 10 回）
最大リテイク回数	32 回	主権的態度の持久的発揮（生徒 E）
最少リテイク回数	3 回	高解像度の初期設計（生徒 W・4.4 参照）
労力配分の転換	—	作成 20%：検品 80%の実現

「教諭が最低 10 回と命令したから生徒が従っただけではないか」という同調バイアスの批判に対し，リテイク履歴を質的に分析する。最大値 32 回は生徒 E と生徒 H の 2 名に観察されたが，生徒 H の記録は完成作品と総リテイク数のみで修正過程の対話ログが残されていないため，本稿では修正指示の全過程が検証可能な生徒 E を分析対象とする。生徒 E は「恋心」という主題から出発し，32 回のリテイクを重ねた。前半は「明るい詩に変えて」「恋をしている相手への気持ちがとても強い情景」という主題設定・雰囲気指定が中心であり，教諭の「10 回」条件はこの段階で充足される。しかし中盤以降，生徒 E は条件を大幅に超過しながら「2 連目を全部変えて」「1 連目の 3 行目をこの気持ちに変えて」という構造操作や，「ニヤけているがきもいから別の表現で」「君だとかっこつけているから別の表現で」という精密な語句調整へ推移し，「もっと男子高校生っぽくして」という自己の表現意図への自発的こだわりを発露させた。

この変容の因果メカニズムを，4.1 節で定義した「ふくらませて」の技術的特性から解明する。AI が生成した極めて冗長度の高い言語素材には，LLM の確率分布に基づく平均的な修辞──恋愛主題に対する定型的な感傷表現や汎用的な比喩──が大量に含まれる。生徒 E はこの平均的な表現群と，自己の内面に存在する固有の感情との間に，決定的なズレ（違和感）を検知した。中盤以降の精密な構造操作・語句調整は，このズレを自発的にデバッグしようとする行為──すなわち，AI の平均的な語彙を自己の固有表現で上書きする反復的な意図同期のプロセスである。

10 回の条件を満たした後に 22 回もの追加リテイクを自発的に行った事実は，外的命令への服従では説明できない。生徒 E にとってノイズは「排除すべきバグ」ではなく，自己の内面にある言葉を浮上させるための砥石として機能した。この「ノイズの教育的逆用」──ハルシネーションを知の砥石へと転化する過程──こそが，拡張型コマンドが 32 回の自発的リテイクを駆動した技術的・言語学的な要因であり，受動的な利用者から能動的な編集主

体への移行を示す事例である。

4.3 ノイズを砥石として活用した生徒 B

最大値（4.2）と中間層（4.4）がリテイク回数に基づく分析であるのに対し，本節では一回の対話内部における精錬の過程を，回収した対話ログに即して定性的に記録する。生徒 B は「雨上がりの朝／靴下が濡れた感覚」から出発した。AI は「未来の気配に触れたような」という大仰な修辞を生成したが，生徒 B は「なんか違う」と直感的に応じ，「やっぱ登校の描写なくして」と指示を精錬させた。AI のノイズが「自分は何を求めているのか」を問い直す砥石として機能した過程が，対話ログに明確に記録されている。

4.4 中間層の成熟パターンと対抗事例

生徒 E は最大値の事例であり，分析を極値のみに依拠させることはできない。21 名の多数を占める中間層（リテイク 6〜11 回）の対話ログには，共通する成熟パターンが観察された。初期段階では「チームワーク系の言葉にしてほしい」「淡さを加えてほしい」という抽象的な形容詞による指示が主であるが，対話を重ねるにつれ「二連を削除して」「最後の三文をカット」「行を短く切ってほしい」等の具体的な構造操作へと指示が精錬されていく。抽象から構造操作へのこの推移は外部から教示されたものではなく，AI との反復対話の中で内発的に生じた学習であり，生徒 E に観察された変容が特異な一例ではないことを示す。

一方，対抗事例も存在する。最少リテイク回数は生徒 W の 3 回であった。生徒 W は「もっと短く」「もたもたをのんびりに変えて」「時間がないという言葉を入れて」という，いずれも具体的かつ構造的な 3 回の指示で作品を完成させた。これは修正以前の段階で表現意図を高い解像度で言語化できていたことを示唆するが，他方で，検証を途中で切り上げた早期の妥協の可能性も排除できない。すなわち，リテイク回数の多寡それ自体は能力の高低を直接反映せず，修正指示の質的内容と完成作品における表現意図の実現度を総合的に評価する方法論の開発が今後の課題である。なお本実証は 21 名・1 教科・1 単元のパイロットに限定され，守段階に相当する活動の報告であることを明記する。

5. 教員研修の二極化とローカル AI への拡張

5.1 専門家検証と「知の検品」

23 点の資料は各教科の専門教諭に個別対面で提示し査読を依頼した。国語科教諭は内容的に使えると納得し，教頭（数学専門）は内容的におもしろいと評価した。社会（歴史）科教諭は強い関心を示し，その後数回の授業に取り入れている。物理科教諭の「従来の延長線にあり，従来からその方法は使われているので，どうだろう」という慎重な反応は，新規性への留保として記録されるべきものである。他方でこの反応は，AI との対話で構成した物理教材が，実験による検証・理想条件と現実の乖離の認識等の知的伝統を，現場の教員が既知と感じる水準で再現していたことの傍証ともなる。この開発・検証プロセス自体が「作成 20%：検品 80%」という演出家モデルの組織的実践である。

5.2 発想のキャズムとマイクロ・サクセス

数学科教諭の「活用の発想が出ない」という言葉は，活用層と未経験層の間の溝──「発想のキャズム」──を端的に示す。図書館の開館日数算出のような日常業務でのデータ処理が「マイクロ・サクセス」として機能し，「自分にもできるかもしれない」という予感を与えて壁に亀裂を入れる。しかし現時点では，その発生と共有の双方にアーキテクト（筆者）の介入が不可欠である。

5.3 ローカル AI：主権的処理基盤の展望

教務主任から，年度後半の入試業務（成績・調査書・合否判定等の機密性の高い大量データの処理）を見据え，「機密性を保持して校内で処理できる AI」への打診があった。これは信号機ルールの赤（入力禁止）を正確に理解した組織中枢が，クラウド AI 依存という構造的脆弱性の物理的排除を自発的に求めた帰結である。筆者は，Llama-3-ELYZA-JP-8B 等の日本語対応オープンソース LLM をインターネット非接続のスタンドアロン PC 単体で運用する技術的可能性を提示し，動作検証に着手した。外部に送信しないという物理的原理が情報分類の判断負荷そのものを解消する点に，この主権的処理基盤の本質がある。導入は予算措置・管理体制を含む組織的意思決定を要する段階にある。

6. おわりに：砥石が映し出す主権的意志

6.1 5 層構造の統合的評価と実証スコープ

表 3　組織的知性マネジメントの 5 層構造

層	構成要素	機能
第 1 層 物理的基盤	ブラウザ分離（Edge＝校務，Chrome＝学習）	情報のスープ化の物理的防止と観測基盤
第 2 層 環境一体型規律	信号機ルール＋検証済み資料 23 点	ゼロトラスト原則の環境実装
第 3 層 知の循環基盤	Teams ラボ＋専門監査員モジュール	成功・失敗・検品技法の組織的循環
第 4 層 主権的処理基盤	ローカル AI（スタンドアロン運用）	機密情報を外部に出さない処理（検証段階）
第 5 層 ドクトリン	実施ガイドブック（守・破・離スパイラル）	規律と教科裁量の統合（初期運用段階）

本研究は，表 3 に示す 5 層からなる組織的知性マネジメントの設計と試行を報告した。5 層は相互依存であり，いずれか一つが欠けても全体が機能しない。ただし，本研究の実証スコープには明確な限界がある。本稿のデータによって直接検証されたのは第 1〜3 層──ブラウザ分離による観測基盤，環境一体型規律と専門家検証済みの知能資産，知の循環基盤──であり，第 4 層（ローカル AI）および第 5 層（デジタル・ドクトリン）は，アーキテクチャの設計・初期運用の段階にとどまる。これらの本格的な効果検証は，今後の継続的な実践と縦断的な観測に委ねられる。

6.2 「発想のキャズム」とマイクロ・サクセスの堆積

最大の課題は，6.1 節の実証スコープと並んで，教員の認識の壁──「発想のキャズム」──であった。この壁は知識の不足ではなく体験の不在に起因するため，卑近な「マイクロ・サクセス（小さな成功体験）」の堆積によってのみ突破しうる。ただし，発生と共有の双方でアーキテクト（筆者）の介入が不可欠であった点は構造的課題であり，各教科へのファシリテーター育成とローカル AI 定着による属人的依存からの脱却が今後の最重要課題である。

6.3 主権的意志の再発見

0%の初期採用率と 32 回のリテイク，そして中間層に共通する指示の成熟が示したのは，AI の有用性ではない。AI という砥石を前にしたときに学習者の内部から立ち現れた「自分の言葉で語りたい」という主権的意志である。教育の役割は，この意志を喚起し，それを研ぎ澄ます環境を整えることにある。本研究の実践報告が，AI を主権的に制御する教育モデルの設計と検証に，一つの具体的な事例を加えるものと考える。

参考文献

(1) 文部科学省：“初等中等教育段階における生成 AI の利活用に関するガイドライン（Ver. 2.0）”，文部科学省（2024）．

(2) 長崎県教育委員会：“県立学校情報セキュリティ基本方針”，長崎県教育委員会．https://www.pref.nagasaki.jp/bunrui/kanko-kyoiku-bunka/kyoikukikannado/kyoikuiinkai/sec-policy/ （参照 2026-06-26）．

(3) 後藤勝洋，松浦執：“生成 AI とクリティカルシンキングの相互作用―教育現場における ChatGPT の応用と影響に関する研究―”，コンピュータ＆エデュケーション，Vol.56，pp.68-74（2024）．

(4) Schulhoff, S., Ilie, M., Balepur, N., et al. : “The Prompt Report: A Systematic Survey of Prompt Engineering Techniques”， arXiv preprint arXiv:2406.06608（2024）．

(5) 片岡徹也編著：『モルトケ』（戦略論大系 3），芙蓉書房出版（2002）．

(6) Perez, E., Ringer, S., Lukošiūtė, K., et al. : “Discovering Language Model Behaviors with Model-Written Evaluations”，arXiv preprint arXiv:2212.09251（2022）．