生成AI駆動型R&Dにおける「生産性と多様性のパラドックス」：技術的メカニズムと解決策に関する包括的調査報告

Gemini 3 pro

エグゼクティブサマリー

現在、企業の研究開発（R&D）部門や技術発明の現場において、大規模言語モデル（LLM）をはじめとする生成AIの導入が急速に進んでいる。これらのツールは、個々の研究者やエンジニアの生産性を劇的に向上させ、アイデア出し（アイディエーション）の質を高める強力な触媒として機能している。しかし、Doshi and Hauser (2024) による先駆的な研究をはじめとする最新の文献は、この技術的恩恵の裏側に潜む重大なリスクを浮き彫りにした。それは、個人の創造性が強化される一方で、集団全体のアイデアの多様性が損なわれる「ホモジェナイゼーション（均質化）」現象である。

本報告書は、先行研究の詳細な検証を出発点とし、この現象が発生する技術的・認知科学的メカニズムを深堀りする。具体的には、LLMの学習プロセスにおける「人間によるフィードバックを用いた強化学習（RLHF）」が引き起こす「モード崩壊（Mode Collapse）」や、人間の評価データに含まれる「典型性バイアス（Typicality Bias）」が、いかにしてAIの出力を「高品質だが画一的」な領域に収束させるかを解明する。

さらに、エンジニアリング・デザイン、創薬、材料科学といった高度な技術分野における影響を分析し、単なるテキスト生成を超えた物理的・化学的探索空間における多様性喪失のリスクを評価する。最後に、Verbalized Sampling（言語化サンプリング）やマルチエージェント論争（Multi-Agent Debate）、進化的アルゴリズムを用いた最新の緩和策を提示し、R&D組織が「効率」と「革新」の両立を図るための具体的な戦略的ロードマップを提案する。

---

1. 先行研究の検証：Doshi & Hauser (2024) が示す「創造性のジレンマ」

Doshi and Hauser (2024) が *Science Advances* に発表した研究[1]は、生成AIが個人の創造性と集団の多様性に与える相反する影響（トレードオフ）を実証した点で、本分野における記念碑的な成果である。本章では、この研究の実験設計、評価指標、および得られた知見を詳細に再検証し、R&D文脈への適用可能性を検討する。

1.1 実験設計と因果推論

研究チームは、オンライン実験を通じて、生成AIの支援有無が短編小説の執筆に与える影響を因果的に分析した。参加者は、AIによるアイデア提供を受けるグループ（介入群）と、AIを使用しないグループ（対照群）にランダムに割り当てられた。ここで重要なのは、AIが完成品を出力するのではなく、あくまで「発想の種（シード）」を提供し、人間がそれを拡張する協働プロセスを採用した点である。これは、現在のR&D現場における「Copilot」的な運用形態を正確に模倣していると言える。

1.2 個人的創造性の向上（フロアの底上げ）

実験の結果、AIの支援を受けた参加者が作成した物語は、第三者の評価者によって「より創造的」「より良く書かれている」「より楽しめる」と判定された[1]。特筆すべきは、この効果が、もともと創造性のベースラインスコアが低い参加者において最も顕著であったことである。

参加者の属性	**AI**支援による影響	解釈
低創造性グループ	大幅な向上 **(↑↑)**	AIはスキルの不足を補完し、一定の水準まで質を引き上げる「レベラー（平準化装置）」として機能する。
高創造性グループ	限定的またはなし **(→)**	すでに高い能力を持つ個人に対しては、AIの提案が既存の能力を上回ることは稀であり、天井効果（Ceiling Effect）が見られる[1]。

この結果は、R&D組織において、若手や経験の浅いエンジニアがAIツールを使用することで、ベテランに近い品質の提案書や設計案を即座に作成できるようになることを示唆している。しかし、それは

同時に「平均への回帰」の始まりでもある。

1.3 集団的多様性の喪失（天井の低下）

個人の出力品質が向上した一方で、研究は集団レベルでの重大な副作用を検出した。AI支援を受けて書かれた物語群は、人間単独で書かれた物語群と比較して、**意味論的類似度（Semantic Similarity）**が有意に高かったのである [1]。

定量的評価指標：意味論的距離と埋め込み表現

多様性の計測には、自然言語処理（NLP）における「埋め込み表現（Embeddings）」が用いられた。具体的には、各物語を高次元ベクトル空間（例：BERTやS-BERTなどのモデルを使用）にマッピングし、物語ペア間のコサイン類似度（Cosine Similarity）を計算した。

- コサイン距離が近い（類似度が高い）： 物語のプロット、語彙、テーマが似通っている。
- コサイン距離が遠い（類似度が高い）： 物語が互いに異質で、多様なアイデアが含まれている。

AI支援群では、物語間の平均距離が縮小しており、参加者が互いに独立して作業していたにもかかわらず、生成されたコンテンツが「似たり寄ったり」の結果となった。これは、AIが学習データ内の「最も確からしいパターン」や「成功しやすい物語の構造（貴族の舞踏会、探偵の謎解きなどのクリシェ）」を全ユーザーに対して一様に提案したためであると考えられる。

1.4 最近の追試と「Sui Generis」スコア

この知見は、その後の追試によってさらに強化されている。例えば、LLM生成コンテンツの物語レベルでの有用性を測定するために開発された **"Sui Generis"**（独自性）スコア を用いた研究では、GPT-4やLLaMA-3などのモデルが生成する物語において、特定のプロット要素（Plot Elements）が頻繁に繰り返される「パターンの反響（Echo）」現象が確認された [4]。人間が書く物語が高い独自性を保つのに対し、LLMは「驚き」の要素が欠如した、予測可能な組み合わせに終始する傾向がある。

---

2. ホモジェナイゼーション（均質化）の技術的メカニズム

なぜ、インターネット上の膨大なテキストデータを学習したはずのLLMが、極めて狭い範囲の出力に収束してしまうのか。R&Dの文脈でこの問題を解決するためには、その技術的根源を理解する必要がある。ここでは、**RLHF**（**Reinforcement Learning from Human Feedback**）、モード崩壊、および認知バイアスの観点から深堀りする。

2.1 RLHFの代償：典型性バイアスと報酬ハッキング

現代のLLM（ChatGPT, Claude, Gemini等）の性能を支えているのは、事前学習（Pre-training）後のアライメント工程、特にRLHFである。しかし、Zhang et al. (2025) の研究は、この工程こそが多様性喪失の主犯であることを突き止めた [5]。

典型性バイアス **(Typicality Bias)**

RLHFでは、人間のアノテーター（評価者）がモデルの複数の出力候補をランク付けし、そのデータを基に報酬モデル（Reward Model）を学習させる。しかし、人間には認知心理学的なバイアスがある。

- 処理流暢性 **(Processing Fluency):** 人間は、見慣れた、理解しやすい、典型的なテキストを「高品質」と判断する傾向がある。
- バイアスの注入**:** アノテーターは無意識のうちに、斬新だが難解なアイデアよりも、平凡だが流暢なアイデアを好む。これにより、報酬モデルは「典型性（Typicality）」を最大化するように訓練される [7]。

モード崩壊 **(Mode Collapse)** の数理

事前学習済みのベースモデル（Base Model）は、次トークンの予測において広大な確率分布を持っている。しかし、RLHFによるチューニングを経ると、モデルはこの分布の「裾（Tail）」部分—すなわち、確率は低いが創造的な表現—を切り捨て、確率密度のピーク（Mode）周辺にリソースを集中させるようになる。
これをモード崩壊と呼ぶ。GAN（敵対的生成ネットワーク）における同名の現象と同様に、LLMは「安全で、確実に報酬が得られる」特定のスタイルや回答パターンに固着し、温度パラメータ（Temperature）を上げてもこの強力な引力圏から脱出できなくなる 9。結果として、何百万人のユーザーが異なるプロンプトを入力しても、モデルは似たような「AI構文」や「優等生的な回答」を出力する

に至る。

2.2 アンカリング効果 (Anchoring Effect) と推論の硬直化

モデル内部の推論プロセスにおいても、ホモジェナイゼーションを促進するメカニズムが働いている。Zhou et al. (2025) は、LLMにおけるアンカリング効果の実証研究を行った [11]。

- **現象:** プロンプトの初期に提示された情報（アンカー）が、その後の生成プロセス全体を強力に拘束する。
- **メカニズム:** 「Causal Tracing（因果追跡）」技術を用いたモデル内部の解析により、アンカーとなる単語や概念が、モデルの初期層（Early Layers）における特定のアテンションヘッドを活性化させ、それが最終層に至るまでのトークン選択の探索空間を劇的に狭めていることが判明した [11]。
- **R&Dへの影響:** 技術者が「既存の製品Aを改良せよ」とプロンプトした場合、モデルは「製品A」の特徴に過剰にアンカリングされ、製品Aの概念を完全に覆すようなラディカルなイノベーション（例：構造そのものの廃止）を提案できなくなる。人間であれば可能な「前提を疑う」思考が、アーキテクチャレベルで阻害されている可能性がある。

---

3. エンジニアリング・デザインにおける「生成AI固着 (Fixation)」

Doshi & Hauserの研究は文章作成に関するものであったが、Chiarello et al. (2024) らは、同様の現象がエンジニアリング・デザインの領域でも発生していることを指摘している。これは**「デザイン固着 (Design Fixation)」**のデジタル版と言える [13]。

3.1 認知科学的固着 vs アルゴリズム的固着

伝統的なデザイン研究において、デザイン固着とは、デザイナーが既知の解決策や先行事例に囚われ、新しいアイデアが出せなくなる認知現象を指す（Jansson & Smith, 1991）。生成AIもまた、これと類似した、しかしより深刻な固着を示す。

固着のタイプ	起源	特徴	影響範囲

人間のデザイン固着	個人の経験、直近の事例	意識的な努力やブリーインストーミングで打破可能。	個人または小規模チーム
生成AIのデザイン固着	訓練データの統計的偏り	アルゴリズム的に最適化されているため、打破が困難。	AIを利用する全ユーザー（産業全体）

3.2 CADと設計意図の標準化

具体的なツールへの実装例として、Autodesk社の「AutoConstrain」のようなAI機能が挙げられる。これは、手描きのスケッチからユーザーの「設計意図」を推測し、自動的に拘束（平行、垂直、同心円など）を適用するものである[14]。

- **利点:** 作業時間の短縮。
- **リスク:** AIは「一般的な設計意図」を学習しているため、統計的に稀な、しかし意図的な「非対称性」や「不規則性」をノイズとして修正してしまう可能性がある。エンジニアが特異な形状を作ろうとしても、AIがそれを「標準的なブラケット形状」に補正しようとすることで、設計全体が平均的な幾何学形状に収束していく[14]。

3.3 意味論的創造性と幾何学的限界

また、現在のLLM（マルチモーダルモデル含む）は、テキストによる意味論的な提案（例：「軽量化のためにハニカム構造を採用する」）には長けているが、物理的な幾何学的生成（例：3D空間内で新規のトポロジーを構築する）には限界がある[15]。LLMは形状を「言葉」や「コード」で記述するため、言語化しにくい複雑な曲面や有機的な結合部を持つデザインよりも、言語記述が容易な（名前のついた）標準形状を優先して生成する傾向がある。これが、ハードウェア設計における多様性のボトルネックとなっている。

---

4. 科学的発見（創薬・材料科学）におけるリスクと機会

ホモジェナイゼーションの影響が最も深刻なのは、失敗のコストが高く、探索空間が無限に近いサイ

エンスの領域である。

4.1 創薬：ケミカルスペースの縮小と「偽の多様性」

新薬開発においては、数百万の化合物候補から有効なものを探索する。生成AI（Generative Chemistry）は、この探索を加速することが期待されているが、ここでもモード崩壊が致命的な問題となる。

- トークンフィルタリングの弊害: 分子をSMILES記法（文字列）として生成する際、モデルは低確率なトークンを排除する（Top-k sampling等）。しかし、化学において「低確率なトークン」は、これまでにない新規な官能基や結合様式を意味する場合がある。これを排除することで、AIは「既知の薬に似た分子」ばかりを量産する [16]。
- 有効性と新規性のトレードオフ: 温度パラメータを上げて多様性を出そうとすると、化学的に無効な（合成不可能な）分子が増える。逆に、有効性を重視すると、既存の特許空間に抵触するような類似化合物に収束する。この「Goldilocks Zone（最適領域）」の狭さが、AI創薬の課題となっている [17]。

4.2 材料科学と「AI Scientist」の自律ループ

Sakana AIなどが提唱する「The AI Scientist」のような自律型研究エージェントは、仮説立案から実験（シミュレーション）、論文執筆、査読までを全自動で行うシステムである 19。
ここで懸念されるのは、自己評価ループによる質の低下である。

- 査読エージェントのバイアス: 生成された論文を評価するのもまたLLMである。もし評価用LLMに典型性バイアスがあれば、定説を覆すような革新的な論文を「低評価」とし、既存のパラダイムに沿った無難な論文を「高評価」とする [21]。これが何サイクルも繰り返されると、科学的発見のプロセスそのものが保守化し、局所最適解（Local Optima）にスタックする。

進化的アプローチによる解決策：**ShinkaEvolve**

この問題に対し、Sakana AIは **"ShinkaEvolve"** という進化的アルゴリズムを導入している [23]。

- 集団の維持: 単一の最適解を求めるのではなく、多様なエージェントの「集団（Population）」を維持する。
- 新規性探索: 単に性能が良いだけでなく、「既存の解といかに異なるか」を評価関数に組み込

む（Novelty Search）。これにより、一時的に性能が低くても、将来的にブレークスルーにつながる可能性のある「異質な解」を生き残らせる。これは生物進化の多様性維持メカニズムをAI開発に応用した好例である。

---

5. 多様性を回復するための技術的介入

以上の分析から、R&Dにおける生成AIの活用は、単に「プロンプトを入力して待つ」だけでは不十分であり、多様性を能動的に設計（エンジニアリング）する必要があることが明らかになった。以下に、最新の研究で提案されている具体的な解決策を示す。

5.1 Verbalized Sampling (言語化サンプリング)

Zhang et al. (2025) が提案する **Verbalized Sampling (VS)** は、モデルの再学習を必要としない、推論時の画期的な手法である [5]。

- **手法:** 通常のプロンプト（例：「コーヒーについてのジョークを言って」）ではなく、**「コーヒーについてのジョークを5つ生成し、それぞれの確率（自信度）を割り当ててリスト化せよ」**と指示する。
- **原理:** モデルに「答え」ではなく「分布」を出力させる。通常、モデルは内部で確率計算をしていても、出力時には最も確率の高い1つ（最頻値）を選んでしまう。しかし、「リスト化して確率を言え」と命じられると、モデルは自身の確率分布の「裾（Tail）」部分にあるアイデアも探索し、言語化して提示するモードに切り替わる。
- **効果:** 実験では、この手法を用いるだけで、創造的タスクにおける多様性が約2倍に向上し、かつ品質（人間による評価）も維持された。これは、典型性バイアスを回避する「認知バイパス」として機能する [7]。

表1: サンプリング手法による多様性と品質の比較

手法	メカニズム	多様性スコア	リスク	適用推奨場面
Greedy (貪欲法)	最も確率の高いトークンを選択	低	非常に低い（退屈）	定型業務、コード補完
高温度サンプリング	確率分布を平坦化しランダム	中～高	高（幻覚・破綻）	ブレインストーミング（要選別）

	性を注入			
Verbalized Sampling	分布そのものを言語化して出力	高 **(2x)**	低い（論理性を維持）	**R&D**アイデア出し、特許案作成

5.2 マルチエージェント論争 (Multi-Agent Debate)

単一のモデルではなく、複数の異なるペルソナ（人格）を持つエージェント同士を議論させる手法である [25]。

- 実装例: 「革新的なアイデアを出すエージェントA」に対し、「保守的な批評家エージェントB」と「市場アナリストエージェントC」を対立させる。
- 効果: エージェントBからの批判を受けることで、エージェントAは初期のアンカリング（思いつき）から脱却し、より洗練された、あるいは全く異なる角度からのアイデア（Rebuttal）を出さざるを得なくなる。この社会的相互作用のシミュレーションが、単体のモデルでは到達できない探索領域へのジャンプを引き起こす。

5.3 タスク定着型機能的多様性 (Task-Anchored Functional Diversity)

単に「見た目が違う」アイデアを出すだけでは不十分である。R&Dにおいては「機能的に異なる」解決策が必要である。

- 新しい評価指標: 文面の違い（意味論的多様性）ではなく、解決アプローチの違い（機能的多様性）を測定するフレームワークが提案されている [28]。
- 適用: 例えば、「橋を架ける」課題に対し、「吊り橋」と「アーチ橋」は構造的に異なるが、「橋」という機能では同じである。「トンネル」や「ドローン輸送」といった機能的に異なる解を、埋め込み空間上で識別し、報酬を与えるシステムを構築する。

---

6. R&D組織への戦略的提言

技術的分析を踏まえ、R&Dマネジメント層は以下の戦略的転換を図るべきである。

6.1 プロンプトエンジニアリングから「多様性エンジニアリング」へ

これまでの「良いプロンプト」の定義は、「正解を素早く引き出すこと」であった。これからのR&Dにおける定義は、「探索空間を最大限に広げること」にシフトすべきである。

- 推奨アクション: チーム内で「Verbalized Sampling」のテンプレートを標準化する。また、単一のモデル（例：GPT-4のみ）に依存せず、Claude 3やLlama 3など、異なる学習背景を持つモデル群（アンサンブル）を併用し、それぞれの「バイアスの違い」を利用してアイデアの幅を確保する[29]。

6.2 知財戦略と「先行技術」の衝突リスク

競合他社も同じ基盤モデルを使用している可能性が高いため、生成AIに依存した発明は、他社と内容が重複する（Prior Art Collision）リスクが高まる。

- 推奨アクション: 生成されたアイデアをそのまま特許出願するのではなく、人間が意図的に「ノイズ」や「非連続な飛躍」を加える工程（Human-in-the-Loop）を必須とする。また、AIが生成したアイデア群の「類似度マップ」を作成し、クラスターの中心（ありがちなアイデア）ではなく、辺境（Outlier）にあるアイデアを戦略的に選択してリソースを投下する。

6.3 チーム構成：AIを「平均化装置」にしないために

人間のチームメンバーがAIの提案に過剰に依存すると、「AIグループシンク（集団浅慮）」が発生する[30]。

- 推奨アクション: 会議やブレインストーミングの初期段階ではAIの使用を禁止し、人間の発散的思考（Divergent Thinking）を先行させる。その後、AIを「批判者」や「拡張者」として導入するプロセス設計を行う。また、AIに対し「最もあり得ない解決策を提示せよ」といった、典型性バイアスを逆手に取る指示を組み込む。

結論

Doshi & Hauser (2024) が提起したパラドックスは、R&Dにおける生成AI活用の核心的な課題を突いている。生成AIは、個人の能力を底上げし、業務効率を劇的に改善する一方で、放っておけば組織全体の創造性を「高品質な平均」へと収束させる強力な引力を持つ。

技術的な深堀りにより、この現象はLLMの学習構造（RLHF、典型性バイアス）や推論特性（アンカリング）に根ざした不可避な特性であることが明らかになった。しかし、Verbalized Samplingやマルチエージェントシステム、進化的アルゴリズムといった新たな手法を適切に実装することで、この「多様性の罠」を回避することは可能である。

今後の技術発明・研究開発において、勝者となるのは「最も高性能なAIを使う組織」ではない。「AIの均質化圧力を理解し、それを乗り越えて異質なアイデアを組織的に育成できるシステムを構築した組織」である。AIは「正解」を出すマシンではなく、無限の探索空間への入り口として再定義されるべきである。

引用文献・情報源

本報告書は、以下の調査資料および先行研究に基づき作成された。

- [1] : Doshi, A. R., & Hauser, O. P. (2024). *Generative AI enhances individual creativity but reduces the collective diversity of novel content*. Science Advances.
- [13] : Chiarello et al. (2024). *Design Fixation in Generative AI*.
- [5] : Zhang et al. (2025). *Verbalized Sampling: How to Mitigate Mode Collapse and Unlock LLM Diversity*.
- [11] : Zhou et al. (2025). *An Empirical Study of the Anchoring Effect in LLMs*.
- [4] : *Sui Generis score for quantifying narrative uniqueness*.
- [19] : Sakana AI. *The AI Scientist* & *ShinkaEvolve*.
- [16] : Generative AI in Drug Discovery & Mode Collapse studies.
- [25] : Multi-Agent Debate frameworks.
- [14] : Generative AI in CAD & Engineering Design.

引用文献

1. Generative AI enhances individual creativity but reduces the ..., 11月 26, 2025にアクセス、https://discovery.ucl.ac.uk/10195027/1/Generative%20AI%20enhances%20individual%20creativity%20but%20reduces%20the%20collective%20diversity%20of%20novel%20content.pdf
2. Generative AI enhances individual creativity but reduces the collective diversity of novel content - University of Exeter, 11月 26, 2025にアクセス、https://ore.exeter.ac.uk/articles/journal_contribution/Generative_AI_enhances_individual_creativity_but_reduces_the_collective_diversity_of_novel_content/29809991/1/files/56856236.pdf
3. Ideation with Generative AI—in Consumer Research and Beyond - Harvard Business School, 11月 26, 2025にアクセス、https://www.hbs.edu/ris/Publication%20Files/Ideation%20with%20Generative%20AI%20(Published)_75dccafd-9c43-46f5-9f56-06a11da7a7cf.pdf
4. Echoes in AI: Quantifying lack of plot diversity in LLM outputs - PNAS, 11月 26, 2025にアクセス、https://www.pnas.org/doi/10.1073/pnas.2504966122
5. Verbalized Sampling: How to Mitigate Mode Collapse and Unlock LLM Diversity - arXiv, 11月 26, 2025にアクセス、https://arxiv.org/abs/2510.01171
6. Verbalized Sampling: How to Mitigate Mode Collapse and Unlock LLM Diversity - arXiv, 11月 26, 2025にアクセス、https://arxiv.org/html/2510.01171v1
7. mode_collapse_explanation.md - GitHub Gist, 11月 26, 2025にアクセス、https://gist.github.com/jimmc414/0f89daaa6269b82a55ae9466ec859378
8. Verbalized Sampling: Turning Mode Collapse into Multiplicity | by Mark Craddock | Oct, 2025, 11月 26, 2025にアクセス、https://medium.com/@mcraddock/verbalized-sampling-turning-mode-collapse-into-multiplicity-78c5fea11570
9. 'AI mode collapse' directory - Gwern.net, 11月 26, 2025にアクセス、https://gwern.net/doc/reinforcement-learning/preference-learning/mode-collapse/index
10. Reinforcement Learning from Human Feedback (RLHF) Explained - IntuitionLabs, 11月 26, 2025にアクセス、https://intuitionlabs.ai/articles/reinforcement-learning-human-feedback
11. An Empirical Study of the Anchoring Effect in LLMs: Existence, Mechanism, and Potential Mitigations - ResearchGate, 11月 26, 2025にアクセス、https://www.researchgate.net/publication/391954185_An_Empirical_Study_of_the_Anchoring_Effect_in_LLMs_Existence_Mechanism_and_Potential_Mitigations
12. An Empirical Study of the Anchoring Effect in LLMs: Existence, Mechanism, and Potential Mitigations - arXiv, 11月 26, 2025にアクセス、https://arxiv.org/pdf/2505.15392?
13. Understanding Design Fixation in Generative AI - arXiv, 11月 26, 2025にアクセス、https://arxiv.org/html/2502.05870v1
14. AI Alignment in CAD Design: Teaching Machines to Understand Design Intent in AutoConstrain - Autodesk Research, 11月 26, 2025にアクセス、

https://www.research.autodesk.com/blog/ai-alignment-in-cad-design-teaching-machines-to-understand-design-intent-in-autoconstrain/
15. ShapeCraft: LLM Agents for Structured, Textured and Interactive 3D Modeling - arXiv, 11月 26, 2025にアクセス、https://arxiv.org/html/2510.17603v1
16. The Jungle of Generative Drug Discovery: Traps, Treasures, and Ways Out - arXiv, 11月 26, 2025にアクセス、https://arxiv.org/html/2501.05457v1
17. Deep Generative AI for Multi-Target Therapeutic Design: Toward Self-Improving Drug Discovery Framework - MDPI, 11月 26, 2025にアクセス、https://www.mdpi.com/1422-0067/26/23/11443
18. Generative Deep Learning for de Novo Drug Design A Chemical Space Odyssey, 11月 26, 2025にアクセス、https://pubs.acs.org/doi/10.1021/acs.jcim.5c00641
19. 1 Introduction - arXiv, 11月 26, 2025にアクセス、https://arxiv.org/html/2511.02824v1
20. The AI Scientist: Towards Fully Automated Open-Ended Scientific Discovery - Sakana AI, 11月 26, 2025にアクセス、https://sakana.ai/ai-scientist/
21. OPTIMIZING DIVERSITY AND QUALITY THROUGH BASE-ALIGNED MODEL COLLABORATION - OpenReview, 11月 26, 2025にアクセス、https://openreview.net/pdf/5b91c47163ee9ee0a19a99c5e0cd79aabda40ff0.pdf
22. The AI Scientist Generates its First Peer-Reviewed Scientific Publication, 11月 26, 2025にアクセス、https://sakana.ai/ai-scientist-first-publication/
23. ShinkaEvolve: Evolving New Algorithms with LLMs, Orders of ..., 11月 26, 2025にアクセス、https://sakana.ai/shinka-evolve/
24. Verbalized Sampling: How to Mitigate Mode Collapse and Unlock LLM Diversity - arXiv, 11月 26, 2025にアクセス、https://arxiv.org/html/2510.01171v2
25. Multiagent Finetuning: Self Improvement with Diverse Reasoning Chains, 11月 26, 2025にアクセス、https://llm-multiagent-ft.github.io/
26. Debate-to-Write: A Persona-Driven Multi-Agent Framework for Diverse Argument Generation - ACL Anthology, 11月 26, 2025にアクセス、https://aclanthology.org/2025.coling-main.314.pdf
27. Multi-Agent Debate Strategies to Enhance Requirements Engineering with Large Language Models - arXiv, 11月 26, 2025にアクセス、https://arxiv.org/html/2507.05981v1
28. LLM Output Homogenization is Task Dependent - arXiv, 11月 26, 2025にアクセス、https://arxiv.org/html/2509.21267v1
29. The Ultimate Guide to Fine-Tuning LLMs from Basics to Breakthroughs: An Exhaustive Review of Technologies, Research, Best Practices, Applied Research Challenges and Opportunities (Version 1.0) - arXiv, 11月 26, 2025にアクセス、https://arxiv.org/html/2408.13296v1
30. How Teams Can Avoid Groupthink - Aperian Global, 11月 26, 2025にアクセス、https://aperian.com/blog/how-teams-can-avoid-groupthink/
31. [2312.00506] Generative artificial intelligence enhances creativity but reduces the diversity of novel content - arXiv, 11月 26, 2025にアクセス、https://arxiv.org/abs/2312.00506
32. Homogenization Effects of Large Language Models on Human Creative Ideation - Max Kreminski, 11月 26, 2025にアクセス、https://mkremins.github.io/publications/Homogenization_C&C2024.pdf

33. Anchoring-Guidance Fine-Tuning (AnGFT): Elevating Professional Response Quality in Role-Playing Conversational Agents - ACL Anthology, 11月 26, 2025にアクセス、https://aclanthology.org/2025.emnlp-main.172.pdf