

動的知識グラフの状態を測る統一ゲージ・フレームワークの提案

—Gauge what Knowledge Graph needs.



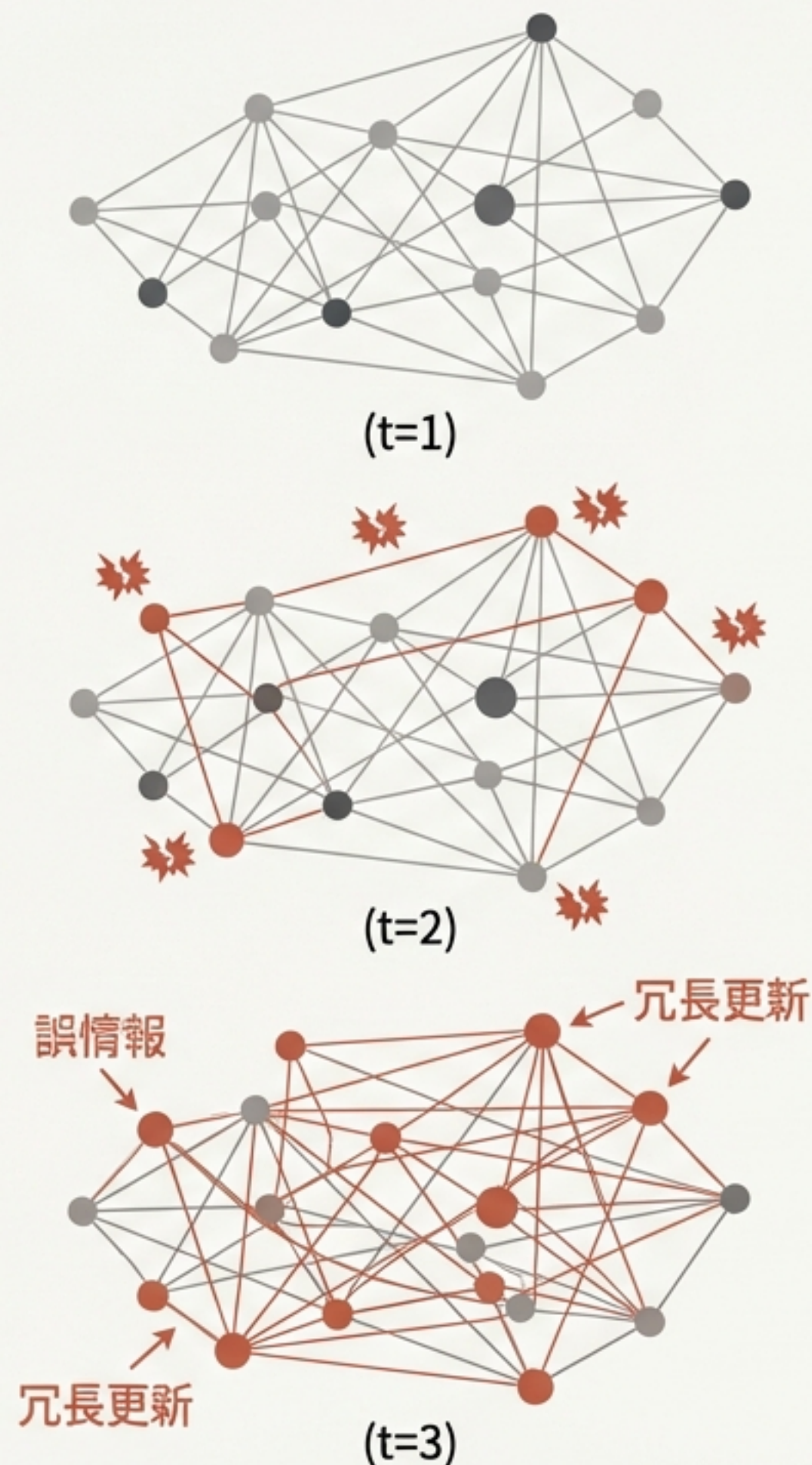
宮内和義, 2025年11月

動的RAGには「いつ知識を受け入れるか(When)」の規範が欠落している

キーマッセージ： 現行のRAGは「何(What)を取るか」の最適化には長けているが、「いつ(When)更新を採用し、いつ見送るか」を決める明示的な基準を持たない。

課題：

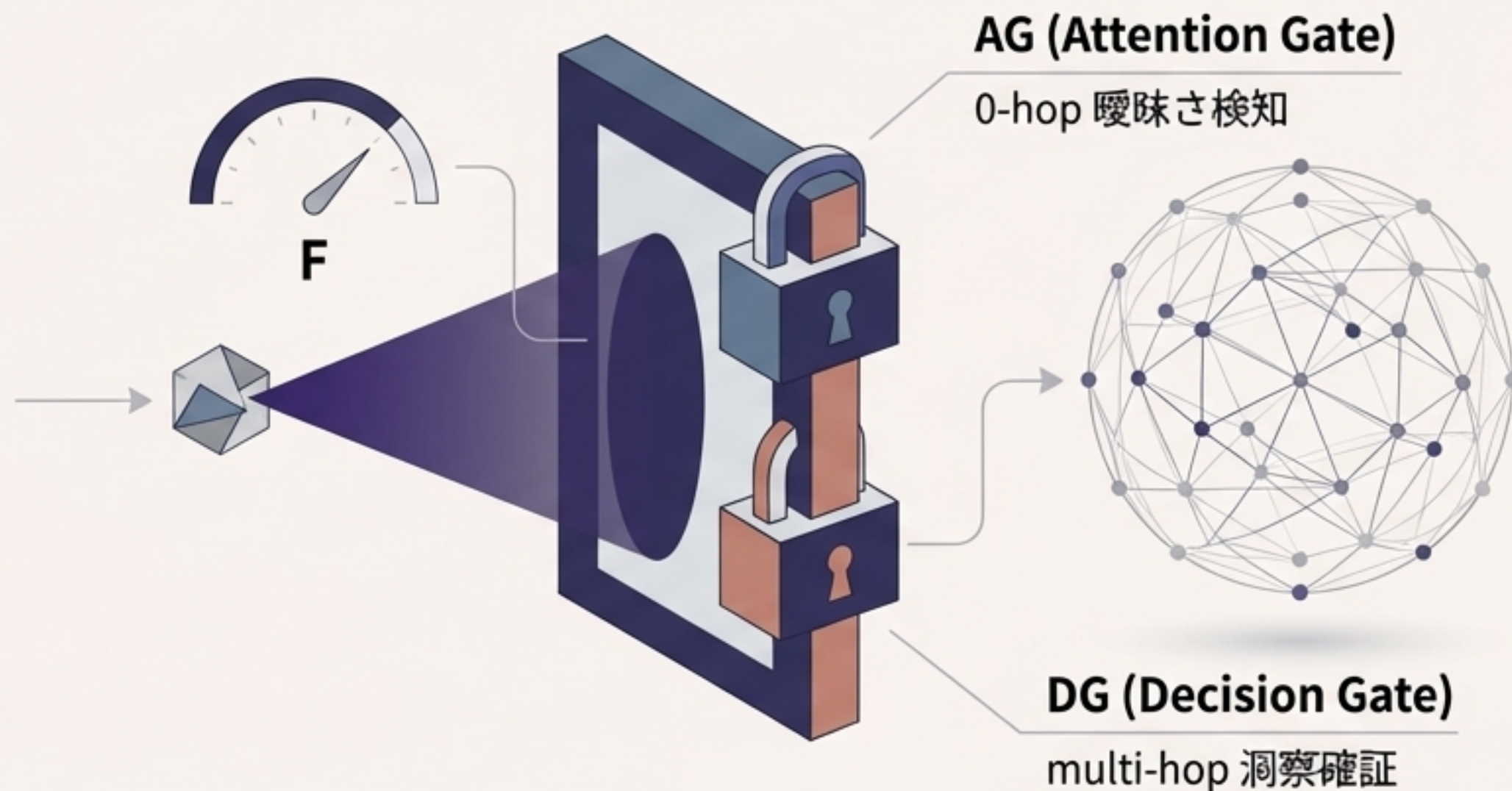
- **知識汚染：** 誤った知識が一度取り込まれると、後続のクエリで繰り返し参照され続ける。
- **冗長化と性能悪化：** 実質的な改善を伴わない高コストな更新が頻発し、システムが劣化する。
- **場当たりのトレードオフ：** 知識の鮮度、汚染リスク、応答遅延のバランスが場当たりのになっている。



統一ゲージ「F」と二段ゲートで「When」を制御するgeDIGフレームワークを提案

コンセプト:

知識グラフの「原理を持つ門番」として機能する統一制御フレームワーク`geDIG` (graph edit Distance and Information Gain)



主要コンポーネント:

1. 単一ゲージ`F`: 知識更新がもたらす「コスト」と「利得」を単一のスカラー値で評価する。
2. 二段ゲート (AG/DG): 2つのステップで更新の是非を判断する。
 - AG (Attention Gate): 0-hopの「曖昧さ」を検知し、不確実なら探索を深化させる。
 - DG (Decision Gate): multi-hopで「洞察」を確証し、真に有益な更新のみを許可する。

ゲージFは「構造コスト」と「情報利得」のバランスを測る



ΔEPC_{norm} (構造コスト)

意味: グラフ編集の「つらさ」。
ノードやエッジの追加に伴うコスト。
方向性: 小さいほど良い。

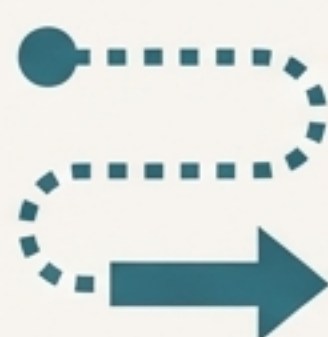
$$F = \Delta EPC_{norm} - \lambda(\Delta H_{norm} + \gamma \Delta SP_{rel})$$

Fが小さいほど「少ない編集で、よく整理され、近道も増えた良い統合」と評価される。



ΔH_{norm} (情報利得 - 秩序化)

意味: 情報整理の「すっきり度」。
エントロピーの低下度合い。
方向性: 正に大きいほど良い (秩序化が進む)。



ΔSP_{rel} (情報利得 - 効率化)

意味: 経路の「回り道の減り具合」。
平均最短路長の短縮率。
方向性: 正に大きいほど良い (経路が短縮される)。

二段ゲートが「曖昧なら探索、確信できれば統合」をイベント駆動で実行する

0-hop: 曖昧検知 (Attention Gate - AG)

- 仮接続直後の局所的な悪化 (g_0) を評価する。
- $g_0 > \theta_{AG}$ の場合、AGが発火し、さらなる探索を促す (FEP的誤差検出)。

multi-hop: 洞察確証 (Decision Gate - DG)

- 探索により経路短縮などの構造的改善 (g_{min}) が生じたかを評価する。
- $\min\{g_0, g_{min}\} \leq \theta_{DG}$ を満たす場合、DGが発火し、更新を確定する (MDL的圧縮検出)。

制御ロジック

状態	0-hop (g_0)	Multi-hop (g_h)	制御
明確な統合	$g_0 < \theta_{AG}$	-	即座に受容
曖昧な局面	$g_0 > \theta_{AG}$	未評価	AG 発火→探索深化
真の洞察	$g_0 > \theta_{AG}$	$g_h < \theta_{DG}$	AG→統合確定
擬似洞察	$g_0 > \theta_{AG}$	$g_h > \theta_{DG}$	探索深化／棄却

「2つの世界、1つの原理」：geDIGの普遍性を2つの環境で検証する

同じ「F」ゲージとAG/DGゲートが、抽象レベルの異なる2つの世界で有効に機能するかを検証し、フレームワークの汎用性を示す。



1. 迷路 PoC (Proof of Concept)

- 世界: 言語要素を排した、制御可能な最小環境。
- 目的: 学習（エピソード統合）と推論（探索/バックトラック）の同時“制御という基本原理が機能するかを確認する。

2. RAG 実証実験

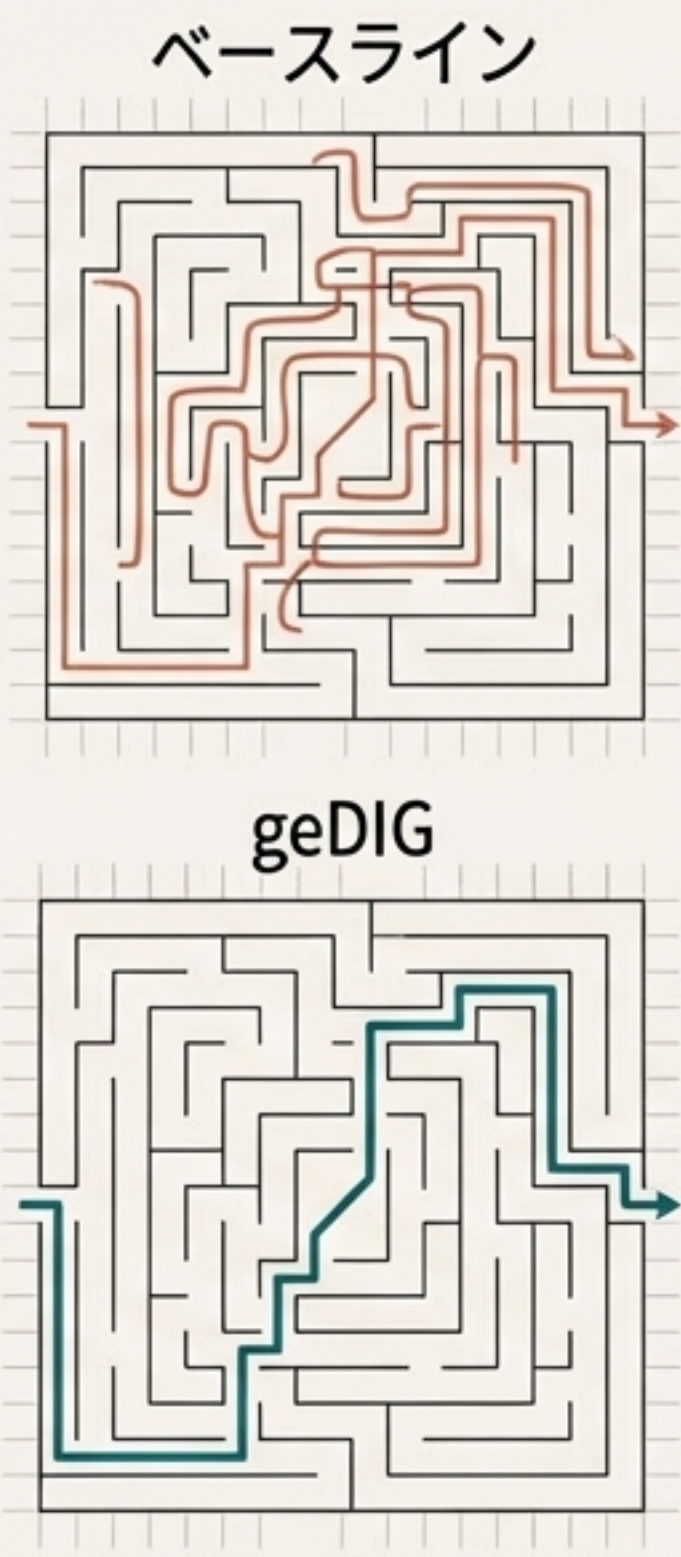
- 世界: 50ドメイン・500クエリ規模の、ノイズが多く複雑な実運用環境。
- 目的: 実世界のRAGパイプラインにおいて、知識汚染の抑制や運用健全性の改善に貢献できるかを実証する。

迷路PoC：geDIGは冗長な探索を抑制し、ゴールまでのステップ数を20-30%削減

- **成功率:** 95–100%の高い成功率を維持。
- **探索効率:** ベースライン比でステップ数を約20–30%削減。
- **枝刈り性能:** AG/DGゲートが不要な候補エッジを約95%剪定。
- **メカニズム:** AGによる行き止まりの即時検知と、DGによる未探索分岐への効率的なバックトラックが、無駄な探索の最小化に直結。

手法	成功率 (%)	平均ステップ	エッジ削減 (%)
ベースライン (Curiosity)	82.3	85.4	-
geDIG (提案)	100.0	69.0	94.8

*15×15迷路, N=100エピソードでの結果



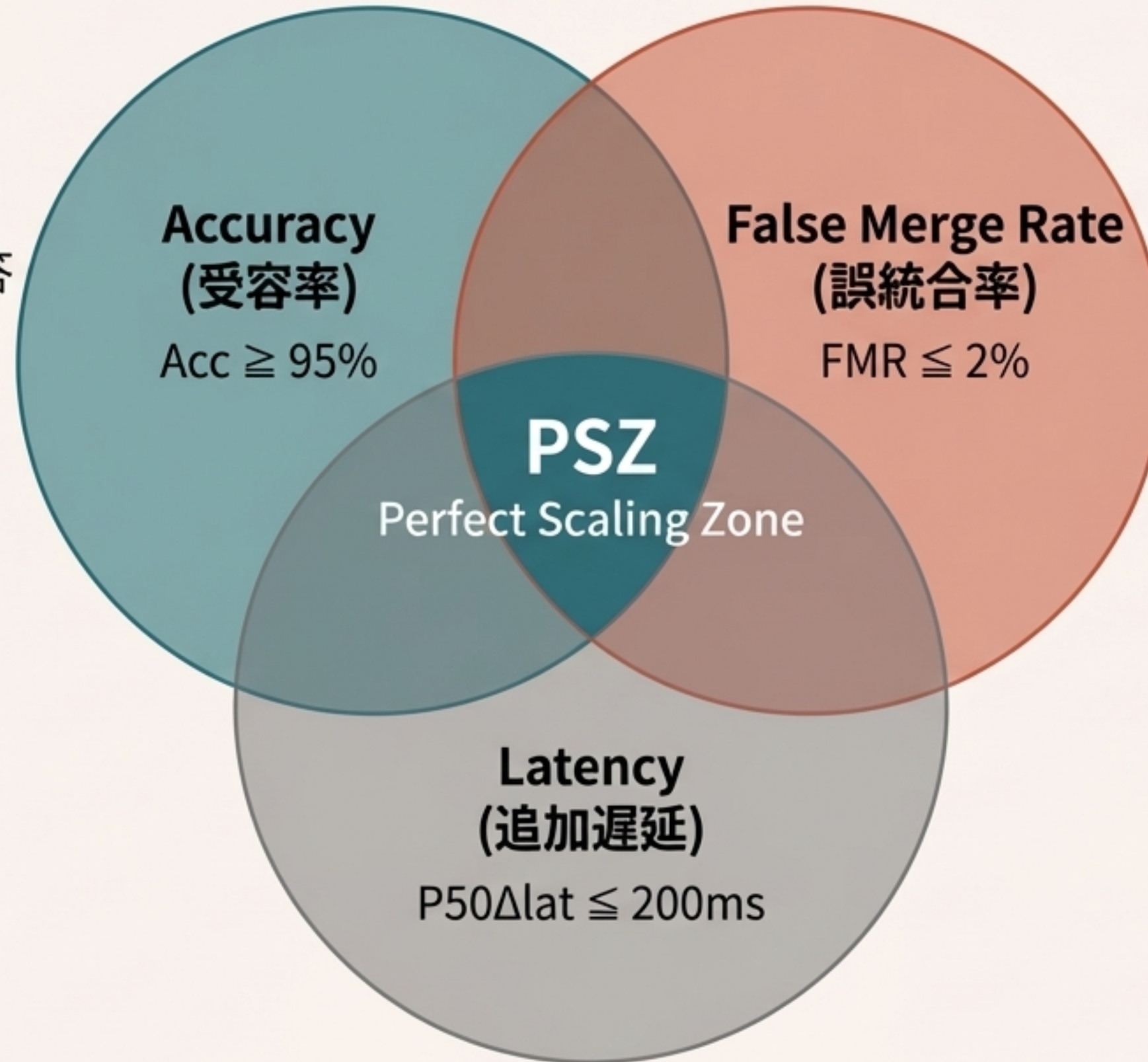
RAG評価の新たな指標：理想的な運用領域「PSZ」を定義する

動的RAG評価の課題：

従来の正答率(EM/F1)だけでは、知識汚染(FMR)や応答遅延(Latency)といった運用上の健全性を測れない。

提案指標：PSZ (Perfect Scaling Zone)

定義：精度、誤統合率、追加遅延の3つの目標を同時に満たす、運用上の「理想状態」。



評価メトリクス：

sPSZ (PSZからの短欠)：

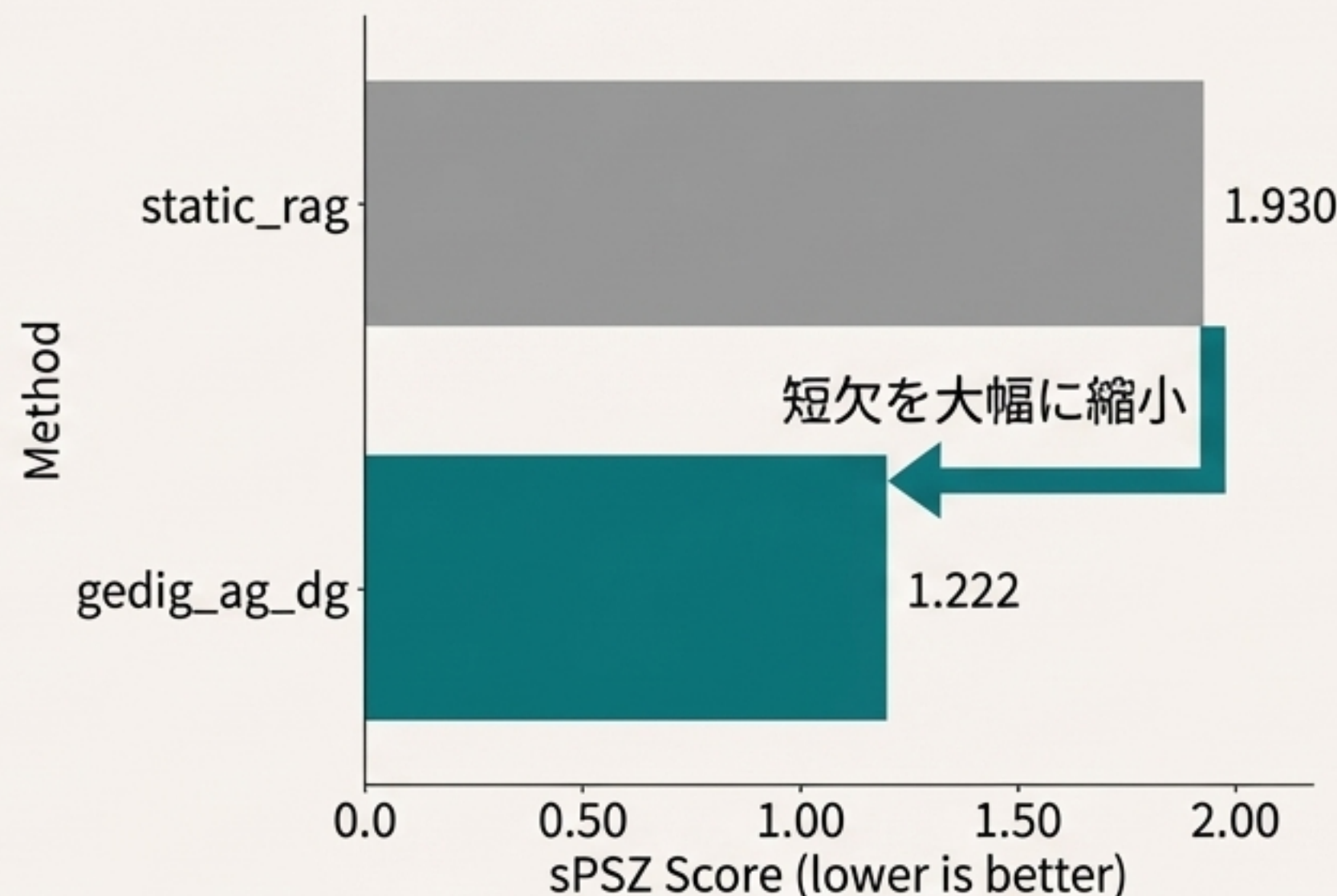
PSZの目標値からどれだけ離れているかを測る統合スコア。`sPSZ=0`が理想状態。動的実験の目標は、この`sPSZ`を一貫して`#spSZ`して縮小させること。

geDIGはPSZからの短欠(sPSZ)を一貫して縮小し、システムの健全性を改善

実験結果のハイライト: equal-resources条件下で、geDIGはベースライン手法と比較して、受容率(Acc)を向上させつつ、誤統合率(FMR)を抑制し、PSZからの短欠(sPSZ)を有意に改善した。

主要指標の比較データ (exp2to4_lite)

Method	PER	Acc (%)	FMR (%)	P50 (ms)	sPSZ
static_rag	0.172	0.0	100.0	160.0	1.930
gedig_ag_dg	0.421	37.4	62.6	240.0	1.222



PSZへの完全到達は未達だが、短欠sPSZはベースラインの1.930から1.222へ大幅に縮小

geDIGの各構成要素は、性能に対して有意に寄与している

分析手法：迷路PoC環境において、geDIGの主要な構成要素（ Δ EPC, Δ IG, Multi-hop, AG/DGゲート, Δ SP）を一つずつ無効化（アブレーション）し、性能（成功率、平均ステップ）がどの程度劣化するかを測定。

アブレーション分析結果（迷路 15×15）

構成	成功率（%）	平均ステップ
Full geDIG	100.0	69.0
w/o Δ EPC（構造コストなし）	▼ 85.2	▼ 82.6
w/o Δ IG（情報利得なし）	▼ 88.5	▼ 78.2
w/o Multi-hop（洞察確証なし）	▼ 92.3	▼ 74.8
w/o Δ SP（経路短縮なし）	▼ 96.7	▼ 71.5

結論：構造コスト、情報利得、multi-hopでの洞察評価など、すべての要素が性能向上に不可欠である。

geDIGは、自由エネルギー原理(FEP)と最小記述長(MDL)に操作的に対応する

理論的ブリッジ	操作的対応 (Operational Correspondence)	
$F \propto \Delta MDL + O(1/N)$ geDIGのゲージ F は、特定の仮定下で、 記述長の総変化量 ΔMDL に比例する。	0-hop評価 (AG) ↔ FEP的「誤差・驚き」の検出 予測誤差が大きい（不確実性が高い） 状態を検知し、探索を促す。	multi-hop評価 (DG) ↔ MDL的「複雑さ・記述長」の圧縮 グラフ構造をよりコンパクトに記述できる（=経路が短縮される）洞察を発見し、 統合を確定する。

対応表

枠組み	構造的コスト成分	情報/精度成分
最小記述長(MDL)	モデル記述長 $L(M)$	データ記述長 $L(D M)$
自由エネルギー(FEP)	複雑性	精度/驚き低減
本稿指標 F	$\Delta EPC_{\text{norm}} - \lambda\gamma\Delta SP_{\text{rel}}$	ΔH_{norm}

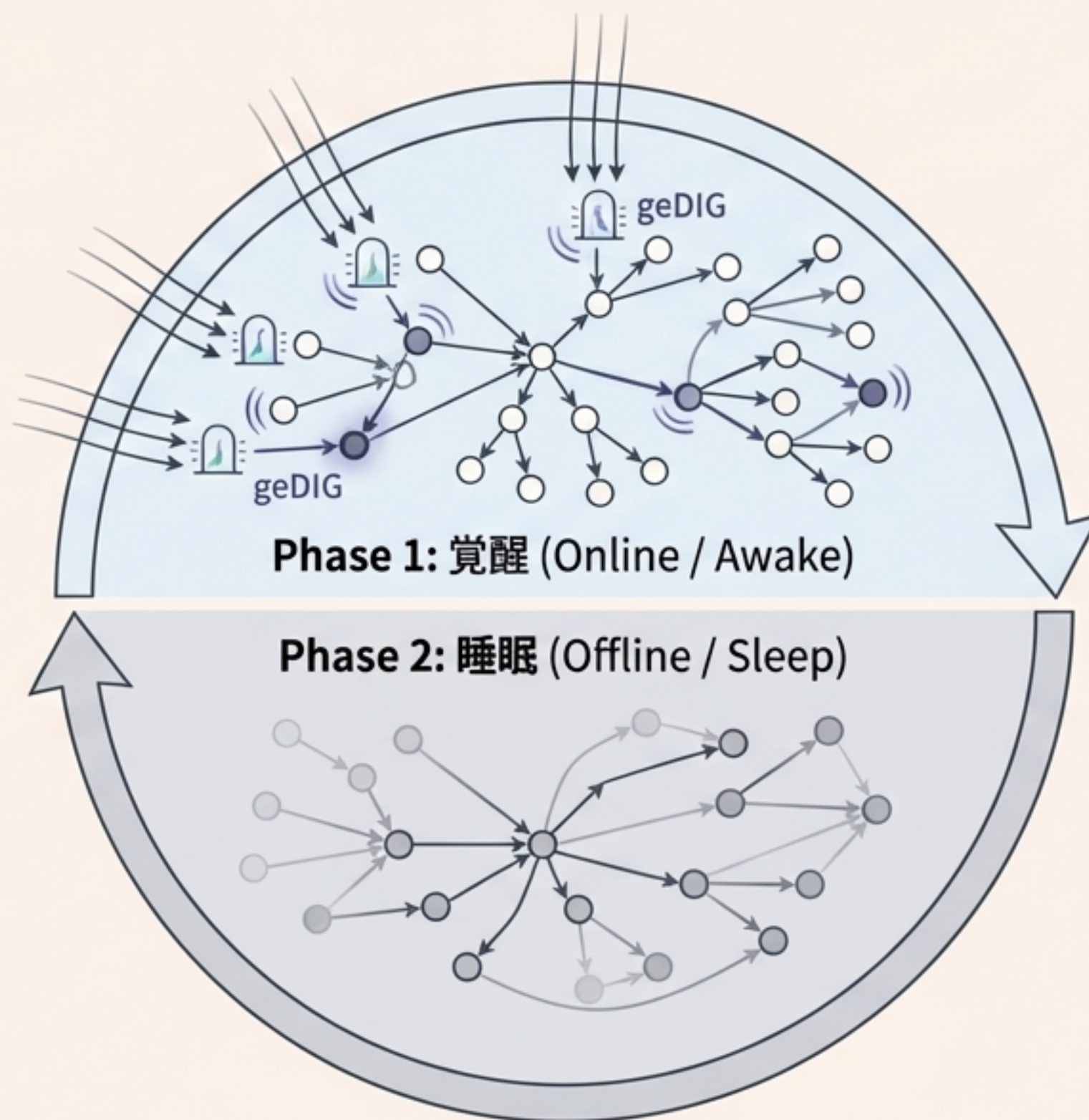
二相アーキテクチャが、オンラインの即時性とオフラインの全体最適化を両立

課題: グラフ全体の最適編集距離(GEDmin)の計算はNP困難であり、リアルタイム運用には工夫が必要。

解決策: 二相アーキテクチャ

Phase 1: 覚醒 (Online / Awake)

- 役割: 入力ストリームに対し、リアルタイムで判断・制御を行う。
- 処理: クエリ中心の局所評価。単一ゲージ F とAG/DGゲートに基づき、受容/棄却/探索/バックトラックを即時決定する。



Phase 2: 睡眠 (Offline / Sleep)

- 役割: 入力を遮断し、グラフ全体の整合性を回復・最適化する。
- 処理: 冗長性の削減、遠隔クラスタ間の再配線、グラフ全体の圧縮などを、時間をかけて実行する。

DGが選別した「洞察サブグラフ」は、LLMの応答方向と統計的に整合する

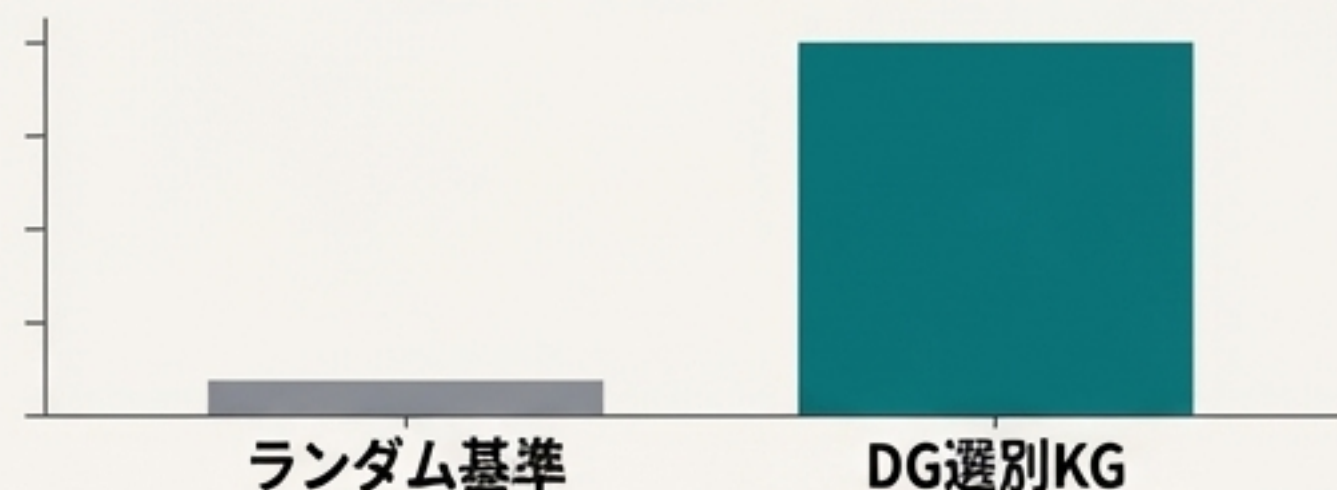
研究仮説:

DGゲートによって選別された高品質な知識サブグラフから導出される「洞察ベクトル」は、LLMの最終的な回答の方向性と一致するはずである。

検証プロトコル:

1. DGで選別されたサブグラフから、統合ベクトル(h_{agg})を生成。
2. LLMの応答埋め込み(r_{LLM})とのコサイン類似度を測定。
3. ランダム選択などのベースラインと比較し、方向一致度の差分(Δs)を評価。

方向一致度 (Δs)



結果 (N=124):

- DGで選別されたKGに基づく洞察ベクトルは、ランダム基準より応答との整合度が有意に高かった。
- コサイン類似度差分: $\Delta s \approx +0.021$ ($p \approx 5.2e-9$, Cohen's $d \approx 1.0$)

示唆:

geDIGによる外部知識の構造化が、LLMの内部推論を補助する可能性。

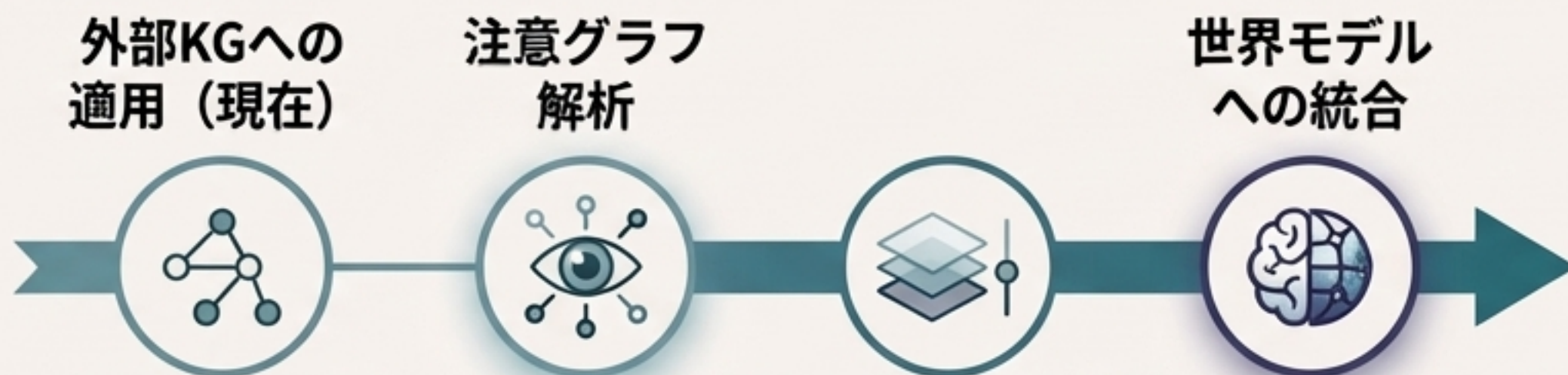


今後の課題と展望：Transformerスケール推論への拡張

現在の限界

- 評価: PSZへの完全到達は未達。より大規模なベンチマークでの検証が必要。
- 実装: 大規模グラフ（10万ノード以上）へのスケーラビリティに課題。
- 理論: FEP-MDLブリッジの数学的厳密化は今後の課題。

将来のロードマップ



外部KGへの適用
（現在）

適応的層制御

- 注意グラフ解析: Transformerの注意機構をグラフとして解析し、推論プロセスを可視化・制御する。
- 適応的層制御: 計算コストと精度のバランスを考慮し、動的に層の深さや複雑さを調整する。
- 世界モデルへの統合: 外部知識と内部推論を統合し、より汎用的でスケーラブルな世界モデルを構築する。

結論：geDIGは、動的知識グラフの「When」を制御する統一原理を提示した

本研究の貢献 (Phase 1の範囲):

1. **One-Gauge統合制御**: 構造コスト(ΔEPC)と情報利得(ΔIG)を単一ゲージFに統合し、「いつ更新すべきか」という問いに明確な基準を与えた。
2. **イベント駆動の意思決定**: AG/DGの二段ゲートにより、曖昧さの検知と洞察の確証をイベント駆動で行い、効率的な知識運用を実現した。
3. **運用可能な評価指標**: 動的RAGの健全性を測るPSZを提案し、sPSZの縮小という形で改善傾向を実証した。
4. **理論と実践の橋渡し**: FEP-MDLブリッジにより、設計の背後にある理論的直観を操作的命題として整理した。

最終メッセージ: 本研究は、動的知識の運用における「When問題」の明示化と、Fという単一原理に基づく解決策の有効性を示した。

