

AI-native Memory 2.0: Second Me

～人間の記憶を拡張する新しいパラダイム～

Mindverse.ai

Jiale Wei, Xiang Ying, Tao Gao, Fangyi Bao, Felix Tao, Jingbo Shang

2025/04/16

1. はじめに
2. SECOND MEの概要
3. アーキテクチャと設計
4. トレーニングとモデル構造
5. 実験結果
6. アプリケーションと可能性
7. まとめと今後の展望

- 人間と外部世界のインタラクションは**個人の記憶**に大きく依存している
 - 会話での情報想起
 - デジタルプラットフォームでの個人情報への繰り返し入力
- 既存のソリューション（自動入力、統合認証システムなど）は限定的
 - 静的なデータ保存のみ
 - コンテキスト理解や適応性に欠ける
 - 断片的で最適とは言えない体験
- 大規模言語モデル（LLM）の登場により**記憶管理の再定義**が可能に

- **SECOND ME**：インテリジェントかつ永続的な記憶オフロードシステム
 - ユーザーとマシンの相互作用における動的な仲介者として機能
 - ユーザー固有の知識を自律的に取得、整理、適用
 - LLMベースのメモリパラメータ化を活用
- 従来の記憶ストレージソリューションとの違い：
 - 静的なデータ保持を超えた機能
 - 構造化された組織化、コンテキスト推論、適応的な知識検索
 - 記憶管理へのより体系的かつインテリジェントなアプローチ

SECOND MEの3層アーキテクチャ

- **L0: Raw Data Layer（生データ層）**
 - 非構造化データの全体
 - RALM（Ram et al., 2023）またはRAG（Lewis et al., 2020）を直接適用
- **L1: Natural Language Memory Layer（自然言語記憶層）**
 - 自然言語で要約可能な記憶
 - ユーザーの短いバイオグラフィー、重要な文やフレーズのリスト、プリファレンスタグなど
- **L2: AI-Native Memory Layer（AI-ネイティブ記憶層）**
 - 必ずしも自然言語での説明を必要としない記憶
 - モデルパラメータを通じて学習・整理
 - 各LPMがニューラルネットワークとして機能



center

Figure 1: SECOND MEのハイブリッドアーキテクチャ

トレーニング目標

SECOND MEの主要な機能：

1. **Memory QA**：知識検索、概念理解、行動予測、アイテム推奨
 - Memory (Self)：ユーザーに直接サービス
 - Memory (Third-party)：外部インタラクションでユーザーを代表
2. **Context Enhancement**：ユーザーの専門モデルへのクエリに関連する詳細を追加
3. **Context Critic**：外部エージェントとのインタラクションを、ユーザーのコンテキストとフィードバックを取り入れて調整

自動化されたトレーニングパイプライン

 center

Figure 2: LLMをジャッジおよびデータ合成器として使用する自動Personal Modelパイプライン

Chain-of-Thought (COT) の活用

3つの戦略でCOTデータを生成：

- **Weak:** フォーマット強制や内容制約なしでCOTパターンで応答
- **Multi-step:** 最初の推論ステップで推論プロセスのみを生成、2番目のステップでクエリ、コンテキスト、推論に基づいて最終回答を生成
- **Strong:** Deepseek-R1 (DeepSeek-AI et al., 2025) を専門モデルとして使用し、詳細なCOT推論と回答を厳格なフォーマット制約と長さ制限で生成

パフォーマンス評価

COT	DPO	Memory (Self)	Memory (Third-Party)	Context Enhance	Context Critic
Strong	Yes	0.96	0.76	0.85	0.86
Strong	No	0.91	0.71	0.75	0.85
Weak	Yes	0.90	0.60	0.83	0.70
Weak	No	0.86	0.58	0.87	0.64

Table 2: COTとDPOの使用に関する実験結果（フルスコアに対する比率として表示）

結果の考察

- **Strong COT**がモデルのパフォーマンスを大幅に向上
 - 記憶関連の質問に回答する能力を強化
 - 専門家とのコミュニケーションを促進
- **DPO** (Direct Preference Optimization) は大幅な改善をもたらす
 - COTの反復的な改良とDPOの使用により、すべてのタスクで一貫したパフォーマンス向上
- 人間による評価では、Strong COT (DPOなし) は平均スコア0.95、Strong COT (DPOあり) はほぼ1.0を達成

SECOND MEの応用分野

- **情報管理の効率化**
 - 関連情報の予測提供
 - フォームの自動入力
 - 過去のやり取りの記憶
 - アプリケーション間でのコンテキスト維持
- **内部的サポート**
 - 思考の整理
 - 決断の振り返り
 - 感情調整
- **ネットワークインテリジェンス**

SECOND MEの目指す未来

- 記憶管理をAI-ネイティブなパラダイムで再定義
 - 個人の思考記録から始まり、データ合成、微調整、強化学習を統合する自動化パイプラインへ
- 課題:
 - 単一ターントレーニングへの依存
 - モデルアライメントの高度化
 - 実世界のユーザーフィードバックの制限
- 将来のビジョン:
 - より良いAI応答を超えて、ユーザーと共に考え、進化し、リアルタイムで認知状態を理解するAIの創造
 - マルチモーダル個人データの統合による人間の認知の完全な把握

- プロジェクトページ: <https://github.com/Mindverse/Second-Me>
- 論文: AI-native Memory 2.0: Second Me
- 参考文献:
 - Shang et al. (2024). AI-native memory: A pathway from LLMs towards AGI.
 - Lewis et al. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks.
 - Ram et al. (2023). In-context retrieval-augmented language models.
 - Wei et al. (2023). Chain-of-thought prompting elicits reasoning in large language models.
 - Rafailov et al. (2024). Direct preference optimization: Your language model is secretly a reward model.