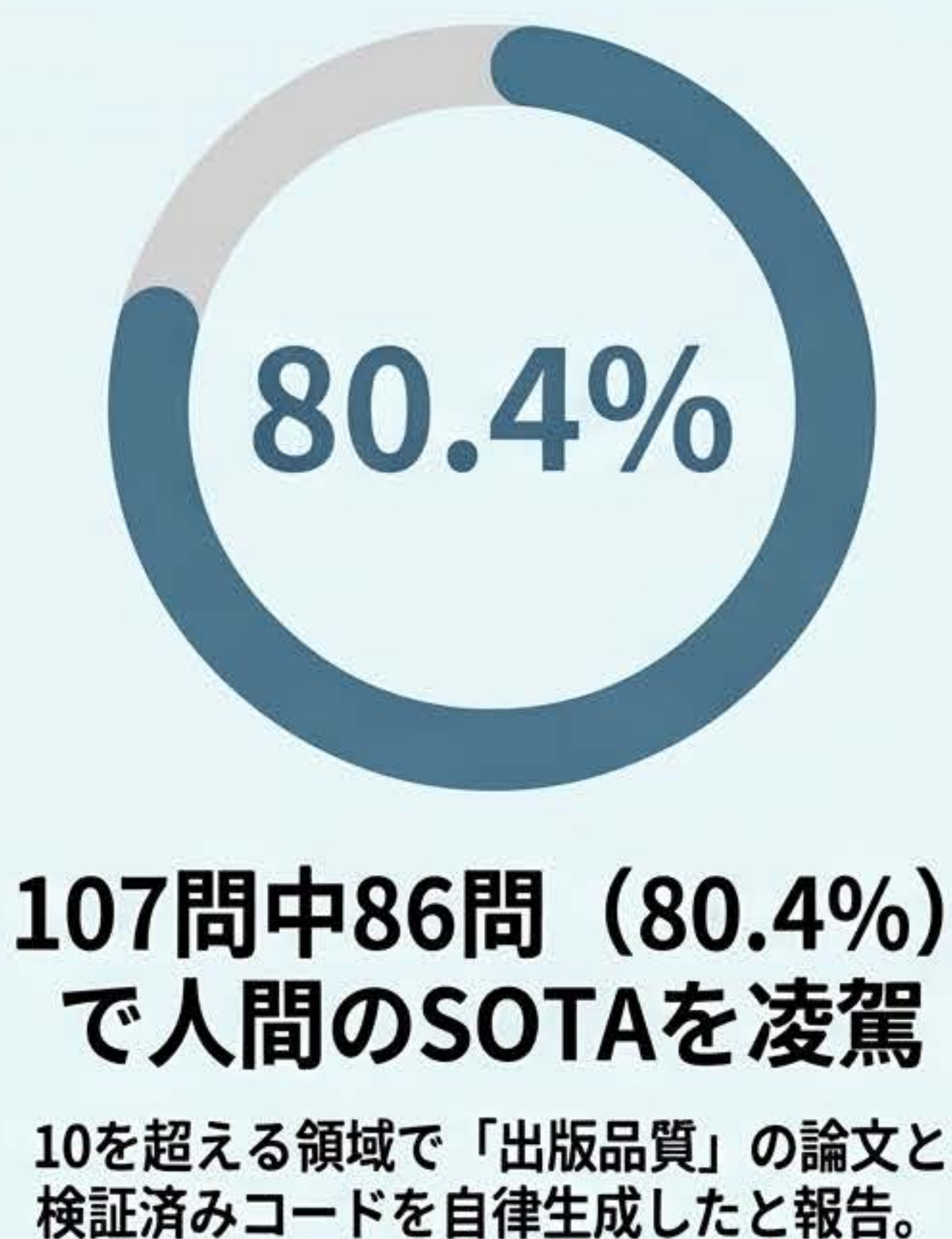


ScientistTwoの実像：自律型AI研究の進展と「フロンティア」への距離

2026年9月発表の「ScientistTwo」は自律型AIとして人間を凌駕する性能を主張するが、その評価は内部AI査読に依存しており、人間による検証や独立した再現は未だ「ゼロ」とであるという主張と実態の乖離が指摘されています。

主張される驚異的な成果とアーキテクチャ



6つのエージェントによる
完全自律の閉ループ



検証の「壁」

参照ハルシネーション「ゼロ」の整合性監査



0

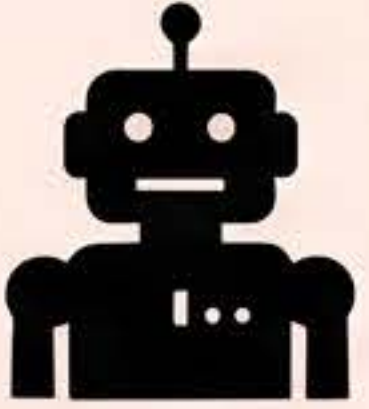
49本の論文、1,814件の参照において
誤情報がないことを内部監査で確認。

分析が示す「実証の乖離」と課題

評価の循環依存（グッドハルトの法則）



人間による検証・再現は現時点で「ゼロ」

ScientistTwo	他社システム例 (Zochi, The AI Scientist v2)
 0 	 
翫合（Sakana AI等）が人間査読を通過したのと対照的に、客観的裏付けが欠如。	ACL 2025 本会議 Nature誌掲載 人間による査読通過
	人間による検証状況：あり

主張の過大評価：改善の中央値は7.7%



平均25.2%の改善を謳うが、実態は既存論文の繊細なインクリメンタル改良が中心。