

AI研究自動化の現在地：工学能力と研究判断の「非対称性」

高度な実行力と欠如する判断力

未来への提言：人間とAIのハイブリッド構成

実装能力と戦略判断の大きな乖離



実行：実験構築や論文執筆は自律実行可能



判断：採択水準の判断ができない



ケース1: Personas



Reject

ケース2: TabPFN



Strong Reject

シャドー評価の結果、主要な2課題でRejectおよびStrong Reject判定

トップ会議レベルの評価では「不採択」

AI研究を阻む5つの失敗モード



資源感覚の欠如



戦略的後戻りの不能



指示の漂流 (Drift)



意味付けの弱さ



飛躍した結論

AIは「高スループット研究エンジニア」



AI

文献調査

コード実装

実験の回旋

図表作成

人間は「PI (主宰研究者)」として舵を取る

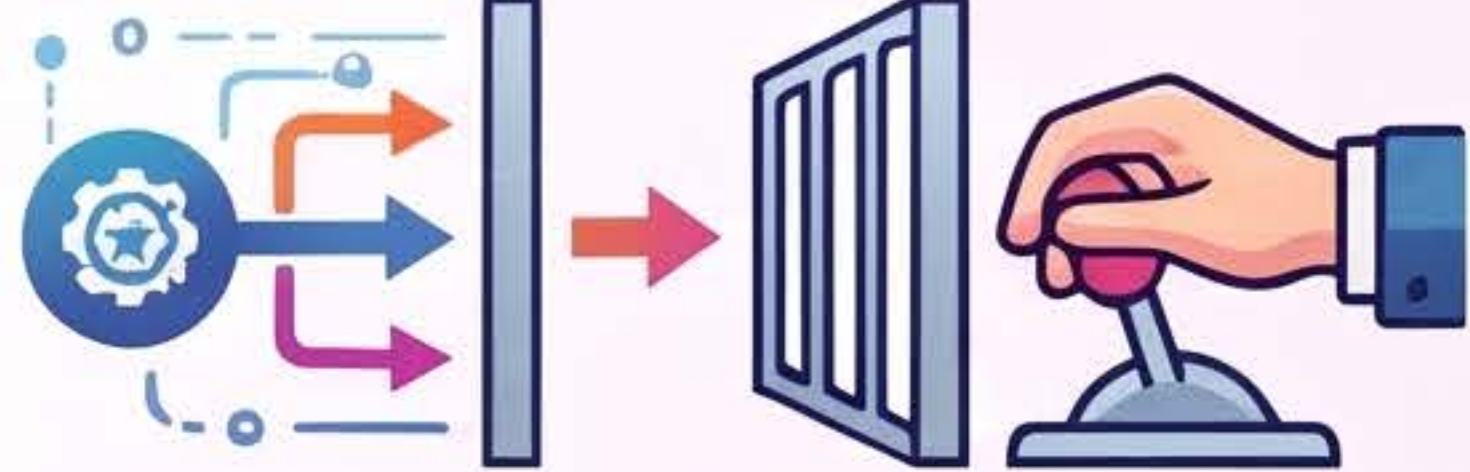


問いの価値

証拠の十分性

大きな戦略変更
研究倫理に責任を持つ

システム側での「介入ゲート」の設置



AI任せにせず、予算管理や戦略転換の判断ポイントを明示的に設計する

	ケース1 (Personas)	ケース2 (TabPFN)
成功点	- 頑韌性確認 - 負の結果の報告	- 多数の内部信号の試験 - 誠実な報告
失敗要因	- 探索の早期収束 - 意味付けの弱さ	- 飛躍した結論 - 根本的再出発の欠如
総合判定	- 採択拒否 (Reject)	- 強い採択拒否 (Strong Reject)