Python×LLM×人間による特許AI読解の品質管理（QA）アーキテクチャ

5層ハイブリッド処理と「2つの関門」

データ成形と事前カウント

段落ID付与とPython事前集計

第一関門：機械による実在性照合

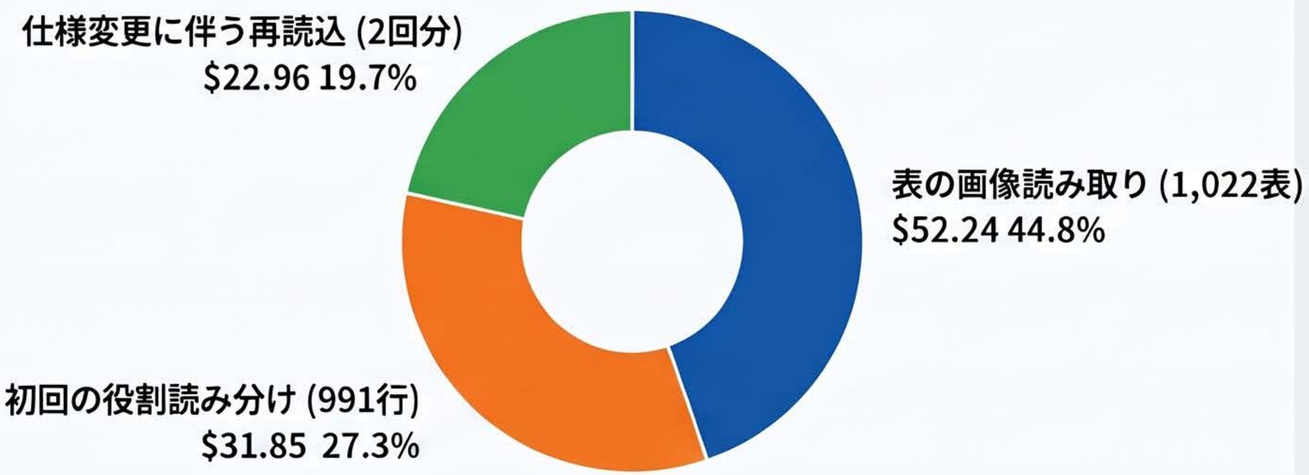
引用文の実在性を決定論的照合

第二関門：人間による妥当性検証

抽出結果の妥当性を目視監査

実証データ・コスト構造と課題

194件の公報処理費用割合（総額$116.54）



LLM文脈読解による誤分類回避

文脈理解で40行の誤分類を回避

表読み取りと再処理のコスト罠

表OCRと再処理が費用の約6割

隠れた設計コストと辞書依存

設計人件費と辞書の補完が必要

Gemini Notebook