

生成AI基盤による非定型帳票の革新と自律促進・事業成果概要

実施者

AI inside株式会社

事業概要

非定型帳票の誤差部分において30%の精度改善を目指す。同等計算リソースにおける非定型帳票のスループットを10倍にすることを旨とする。

非定型帳票読み取りの精度改善を実施。
さらに、非定型帳票読み取り処理の効率化に向け、LLM-SLMの二段階推論処理と、自律蒸留学習の開発を実施。



また、これらを事業として運用するためにスケーラブルなインフラや分散処理基盤の研究開発も実施。

社会実装イメージ



現在弊社が提供している「DX Suite」の機能として、6月上旬にリリース予定。
また、ベンチマークやTechブログについても、同時期に公開予定。

事業成果

今回、弊社では、非定型帳票の誤差部分の30%の精度を改善するという目標を掲げた。

結果として、非定型帳票の読み取り精度については、通常項目では誤差部分の約68%を改善し、明細項目では誤差部分の約33%を改善した。

さらに、弊社で開発した新モデル(ps3)が主要な他社モデルと比較しても、競争力の高い結果となった。

通常項目			明細項目（テーブル項目）		
model_name	accuracy	std	model_name	accuracy	std
PS3-LLM	0.951	0.002	Gemini-2.5-pro-preview-03-25	0.959	0.007
Gemini-2.5-pro-preview-03-25	0.899	0.008	PS3-VLM	0.939	0.001
PS3-SLM	0.863	0.006	Unsloth-Qwen2.5-VL-32B-Instruct	0.917	0.001
PS2-LLM	0.844	0.007	PS2-LLM	0.908	0.003
Unsloth-Qwen2.5-VL-32B-Instruct	0.809	0.003	Chatgpt-4o-latest	0.871	0.035
Chatgpt-4o-latest	0.805	0.040	Mistral OCR	0.868	0.015
Mistral OCR	0.724	0.033	Azure AI Document Intelligence	0.78	0.000
Azure AI Document Intelligence	0.606	0.000			

また、今回はユーザーが利用するほど、SLMの精度が向上するというアーキテクチャの性質を踏まえ、スループットを10倍にするという目標も掲げた。
こちらについても、既存のLLMと比べ、レイテンシーを約3分の1、GPUメモリ使用量を約4分の1に抑えたSLMを開発したことで、全文読み取り後の項目抽出機能に対して、実運用時のコストを10分の1以下に削減できた。
また、データセンターとクラウドサービス間で連携してスケールする仕組みの研究開発も完了し、従来データセンターでしか扱えていなかった生成AI基盤モデルについて、クラウド側でも分散してスケールできることを検証した。