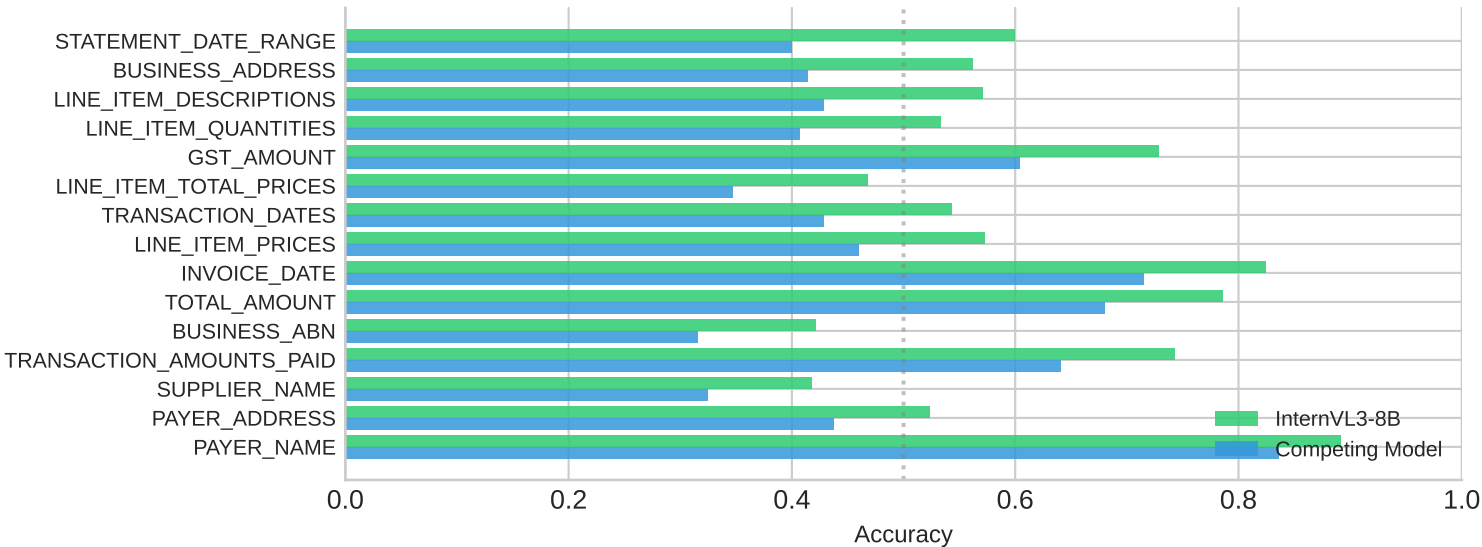


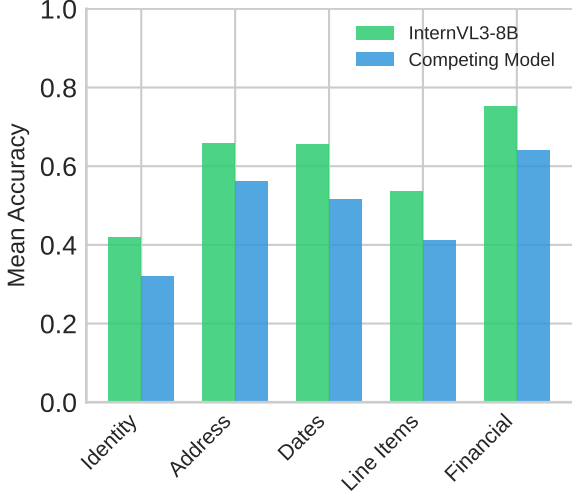
Model Accuracy Comparison Dashboard

InternVL3-8B vs Competing Model | 15 fields | Mean improvement: +11.7%

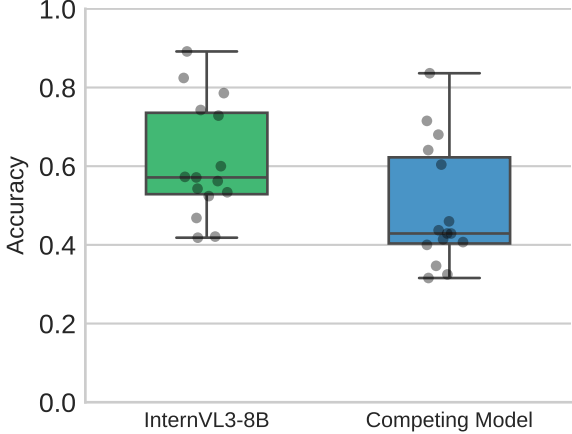
A. Field-Level Accuracy



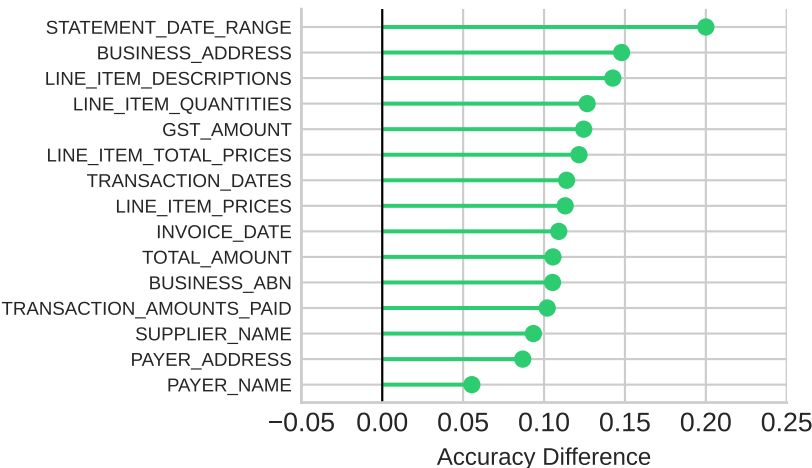
B. By Category



C. Distribution



D. Improvement per Field



E. Summary Statistics

Metric	InternVL3-8B	Competing Model
Mean	61.3%	49.6%
Median	57.1%	42.9%
Std Dev	14.7%	15.9%
Critical Fields	58.8%	48.1%
Statistical Test	Value	Result
p-value (t-test)	0.0000	Significant
Cohen's d	0.79	medium
95% CI	[-10.2%, +13.3%]	