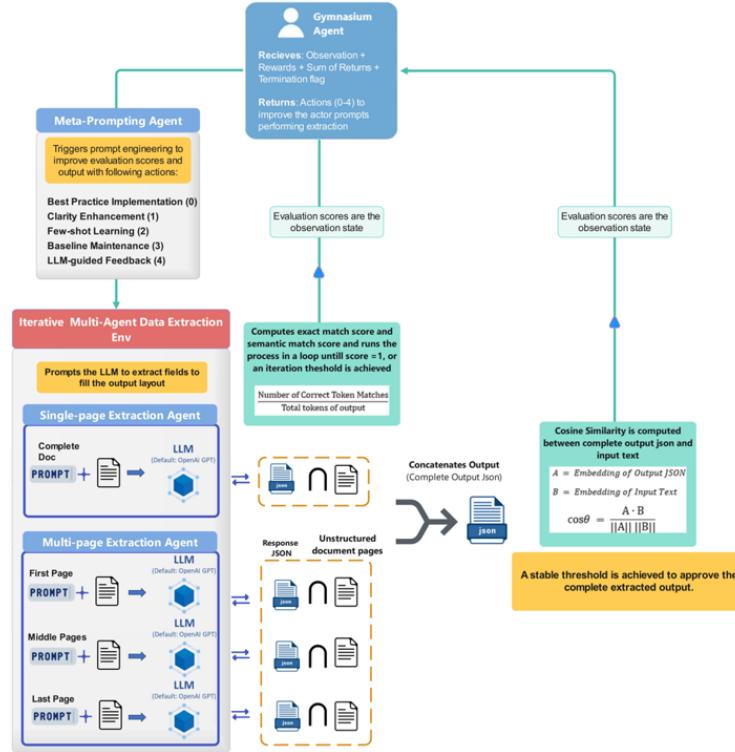


# An agentic system of collaborative AI agents with reinforcement-learned subsystem improvements, applied to parsing form-like documents

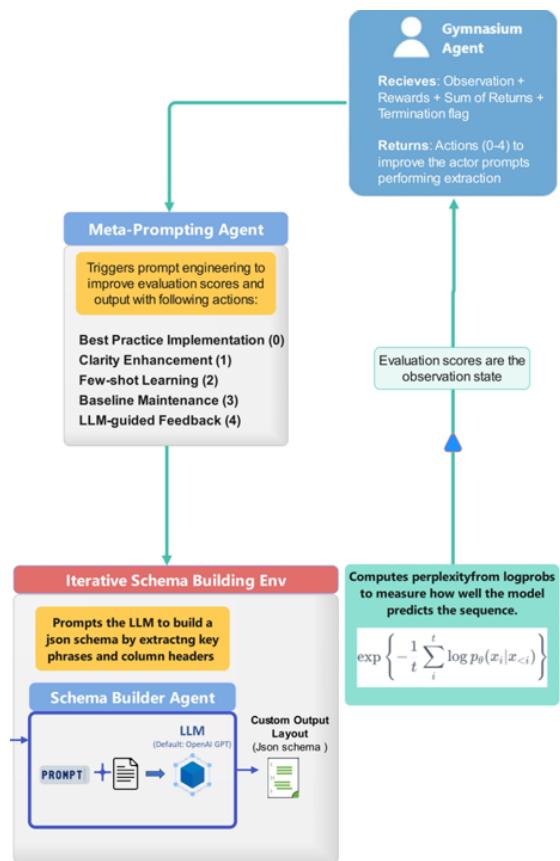
Ayesha Amjad<sup>1</sup>, Saurav Sthapit<sup>2[0000-0002-7421-0479]</sup>, and Tahir Syed<sup>1[0000-0003-0638-9689]</sup>

Institute of Business Administration Karachi [tahirqsyed@gmail.com](mailto:tahirqsyed@gmail.com)  
 Coventry University, UK  
[{Ae0066}@coventry.ac.uk](mailto:{Ae0066}@coventry.ac.uk)

## Appendix



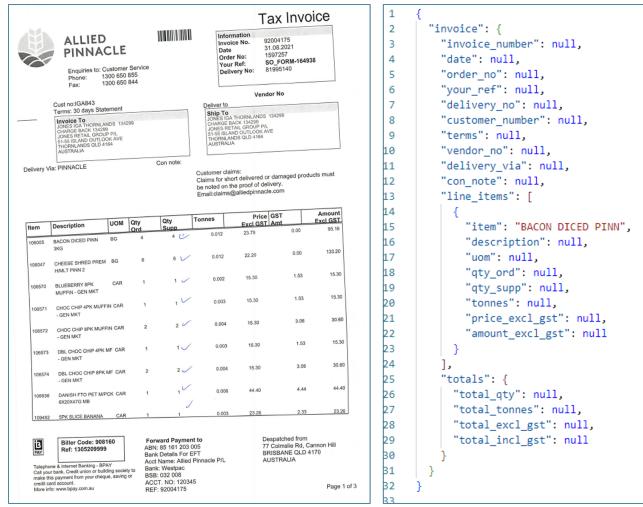
**Fig. 1.** An iterative, reinforcement learning-based framework for extracting form-like data.



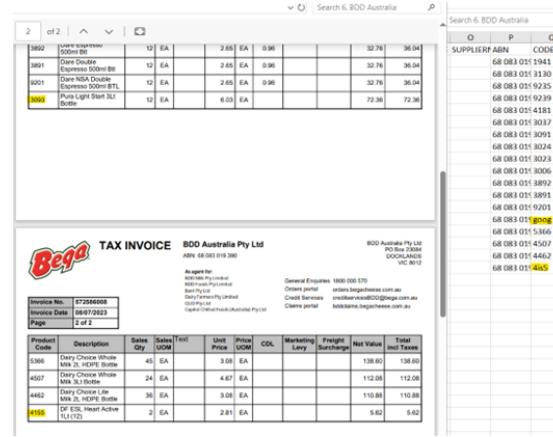
**Fig. 2.** An iterative, reinforcement learning-based framework for building json schema.



**Fig. 3.** a) extracted JSON of a receipt with a resolution of 96 dpi. Despite low quality resolution, its corresponding extracted JSON scored 90% on exact matches, 100% on semantic matches, and 96% on cosine similarity. b) Extracted JSON of a utility bill with a resolution of 66 dpi. Evaluated with 70% on exact matches, 85% on semantic matches, 71% on similarity.



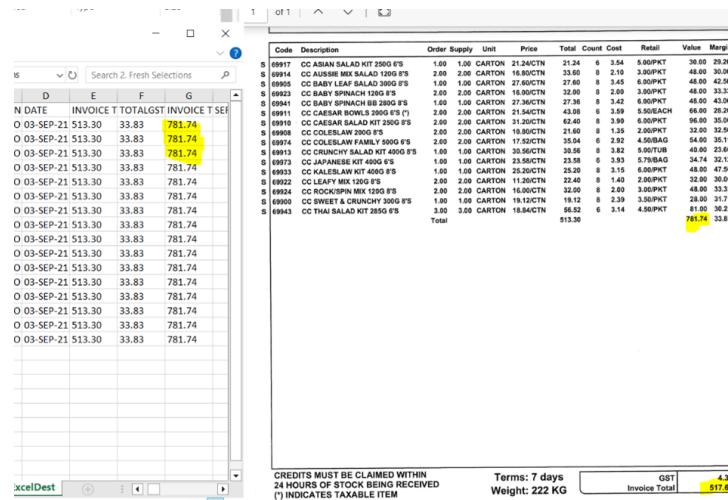
**Fig. 4.** Extracted output of a 3-page scanned document, with a tilt and overwriting, returns an empty JSON. Upon first rerun, it returned all key-value pairs, and some line items. Upon consecutive reruns, output inconsistency is observed.



**Fig. 5.** Incorrect data extracted for a 2-page scanned document. “3093” is extracted as “goog” and “4155” is extracted as “4is5”

Scheduler: DEFAULT Job Monitor SPROC_10_2040011_1602...		Form Editor PDF Tools									
		Invoice Date 13-AUG-24 Account Code BLOOM Page 3 of 3									
Code	Description	Order	Supply	Unit	Price	Total	Count	Cost	Retail	Value	\$
Z7274	POTATO WASHED 2KG P/P MED	15.00	15.00	EACH	\$3.50	1	4.20	6.48	6.48	97.35	
S 22320	RASPBERRIES 125G 12'S INDIVIDUAL	3.00	3.00	TRAY	\$0.80/TRAY	24.00	12	7.33	9.59	355.64	
S 22320	RASPBERRIES 125G 12'S CHOICE	3.00	3.00	TRAY	\$0.80/TRAY	24.00	12	7.33	9.59	355.64	
S 30019	SPROUTS ALFALFA 15G GRM	2.00	2.00	2.00/TUB	4.40	1	2.20	3.40	3.40	6.98	
S 30019	SPROUTS ALFALFA 15G GRM	2.00	2.00	2.00/TUB	4.40	1	2.20	3.40	3.40	6.98	
M 24694	STRAW DIP SGH CHOCOLATE IN TUBE	1.00	0.90	CARTON							
M 24694	STRAW DIP SGH CHOCOLATE IN TUBE	1.00	0.90	CARTON							
S 24675	STRAWBERRY 1L'S SPECIAL	6.00	6.00	TRAY	\$0.80/TRAY	36.00	12	3.61	4.50	486.00	
S 24675	STRAWBERRY 1L'S SPECIAL	6.00	6.00	TRAY	\$0.80/TRAY	36.00	12	3.61	4.50	486.00	
M 67795	TOMATO CHERRY 290G 15'S MB	2.00	2.00	TRAY	\$0.80/TRAY	7.60	15	2.00	4.40	134.70	
M 67795	TOMATO CHERRY 290G 15'S MB	2.00	2.00	TRAY	\$0.80/TRAY	7.60	15	2.00	4.40	134.70	
M 67642	TOMATO GOURMET MILDE (MB)	1.00	1.00	CARTON	\$0.80/CARTON	8.00	10	2.00	4.00	39.80	
M 67642	TOMATO GOURMET MILDE (MB)	1.00	1.00	CARTON	\$0.80/CARTON	8.00	10	2.00	4.00	39.80	
Q 67762	TOMATO MINI ROMA 290G 15'S	1.00	3.00	TRAY	\$0.80/TRAY	12.00	5	4.20	4.50	147.00	
Q 67762	TOMATO MINI ROMA 290G 15'S	1.00	3.00	TRAY	\$0.80/TRAY	12.00	5	4.20	4.50	147.00	
M 54010	TURNIPS WHITE KG	1.00	1.00	KILOGRAM	4.50/KG	4.50	1	4.50	6.98	6.98	
M 54010	TURNIPS WHITE KG	1.00	1.00	KILOGRAM	4.50/KG	4.50	1	4.50	6.98	6.98	
M 69518	ZUCCHINI MEDIUM 1KG (MB)	1.00	1.00	CARTON	\$0.80/CARTON	12.00	15	3.00	3.99	178.55	
M 69518	ZUCCHINI MEDIUM 1KG (MB)	1.00	1.00	CARTON	\$0.80/CARTON	12.00	15	3.00	3.99	178.55	
Total 4854.36											

**Fig. 6.** The extracted output of a 3-page invoice pdf contains records from only page 3, while omitting extracting data from the first two pages.



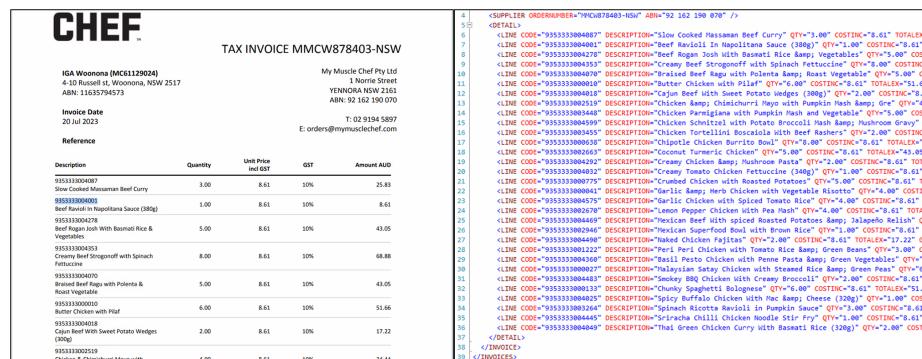
The screenshot shows an Excel spreadsheet with a table of invoice items. The table includes columns for Code, Description, Order Supply, Unit, Price, Total, Count, Cost, Retail, Value, and Margin. The total value for the table is listed as 517.61. A yellow box highlights the value '781.74' in a cell, which is incorrect.

	Code	Description	Order Supply	Unit	Price	Total	Count	Cost	Retail	Value	Margin
S	69917	CC ASIAN SALAD KIT 250G 8'S	1.00	CARTON	21.24CTN	21.24	6	3.54	5.00PCKT	36.00	29.26
S	69914	CC AUSSIE MIX SALAD 120G 8'S	2.00	CARTON	16.80CTN	33.60	8	2.10	3.00PCKT	48.00	30.00
S	69921	CC BABY SPINACH 120G 8'S	2.00	CARTON	16.80CTN	33.60	8	2.00	3.00PCKT	48.00	30.00
S	69923	CC BABY SPINACH 120G 8'S	2.00	CARTON	16.90CTN	32.80	8	2.00	3.00PCKT	48.00	33.33
S	69922	CC BABY SPINACH 120G 8'S	2.00	CARTON	16.90CTN	32.80	8	2.00	3.00PCKT	48.00	33.33
S	69941	CC BABY SPINACH BB 280G 8'S	1.00	CARTON	27.36CTN	27.36	8	3.42	6.00PCKT	48.00	43.06
S	69911	CC CAESAR SALAD KIT 250G 8'S	2.00	CARTON	23.20CTN	46.40	8	3.99	6.00PCKT	96.00	35.00
S	69912	CC CAESAR SALAD KIT 250G FS	2.00	CARTON	23.20CTN	46.40	8	3.99	6.00PCKT	96.00	35.00
S	69974	CC COLESLAW FAMILY 500G 8'S	2.00	CARTON	19.80CTN	39.60	8	1.38	2.00PCKT	32.00	32.00
S	69973	CC COLESLAW FAMILY 500G FS	2.00	CARTON	19.80CTN	39.60	8	1.38	2.00PCKT	32.00	25.51
S	69972	CC COLESLAW KIT 400G 8'S	1.00	CARTON	23.56CTN	23.56	8	3.82	5.00TUB	48.00	23.66
S	69973	CC JAPANESE KIT 400G 8'S	1.00	CARTON	23.56CTN	23.56	8	3.82	5.75TUB	34.72	21.12
S	69933	CC KALESHI KIT 400G 8'S	1.00	CARTON	25.20CTN	25.20	8	3.15	6.00PCKT	32.00	30.00
S	69934	CC KALESHI MIX 120G 8'S	2.00	CARTON	20.20CTN	40.40	8	2.00	3.00PCKT	32.00	30.00
S	69924	CC ROCKSPIN MIX 120G 8'S	2.00	CARTON	16.00CTN	32.00	8	2.00	3.00PCKT	48.00	31.33
S	69900	CC SWEET & CRUNCHY 300G 8'S	1.00	CARTON	19.12CTN	19.12	8	2.39	3.00PCKT	28.00	31.71
S	69943	CC THAI SALAD KIT 280G 8'S	1.00	CARTON	18.94CTN	18.94	8	2.32	3.00PCKT	48.00	30.22
						517.61					

CREDITS MUST BE CLAIMED WITHIN  
24 HOURS OF STOCK BEING RECEIVED  
(\*) INDICATES TAXABLE ITEM

Terms: 7 days G81  
Weight: 222 KG 4.31  
Invoice Total: \$17.61

**Fig. 7.** Incorrect 'Invoice Total' extracted. Correct value is '517.61', but extracted value is '781.74'.



The screenshot shows a PDF receipt from 'CHEF'. The receipt includes a header with the company logo and address, followed by a table of items with descriptions, quantities, unit prices, GST, and amounts. The total amount is listed as 517.61. A yellow box highlights the value '781.74' in a cell, which is incorrect.

TAX INVOICE MMCW878403-NSW					
Description	Quantity	Unit Price	Incl GST	GST	Amount AUD
My Muscle Chef Pty Ltd					
1 Norrie Street					
YEMMELDALE NSW 2517					
ABN: 92 162 190 070					
Invoice Date	T: 02 9194 5897				
20 Jul 2023					
Reference	E: orders@mymusclechef.com				
Item Description	Quantity	Unit Price	Incl GST	GST	Amount AUD
9153113004087	3.00	8.61	10%		25.83
Slow Cooked Basmati Beef Curry					
9153113004001	1.00	8.61	10%		8.61
Beef Revol in Napoletana Sauce (80g)					
9153113004278	5.00	8.61	10%		43.05
Beef Revol With Basmati Rice & Vegetables					
9153113004551	8.00	8.61	10%		68.88
Creamy Beef Stroganoff with Spinach Fettuccine					
9153113004087	5.00	8.61	10%		43.05
Beefed Beef Ragu with Polenta & Root Vegetable					
9153113004001	6.00	8.61	10%		51.66
Butter Chicken with Pilaf					
9153113004278	2.00	8.61	10%		17.22
Cow Beef With Sweet Potato Wedges (30g)					
9153113004018	2.00	8.61	10%		17.22
Chicken & Chorizo Mays with					
9153113002519	4.00	8.61	10%		34.44

```

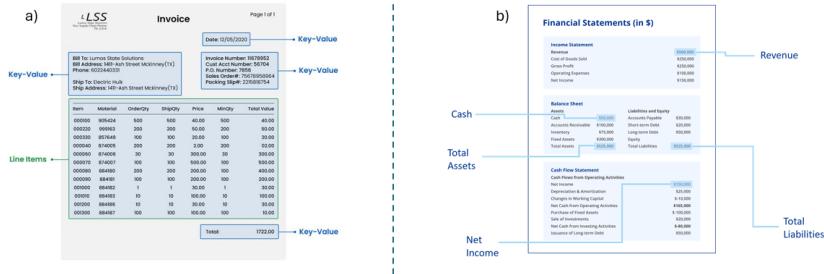
<!--SUPPLIER ORDERNUMBER-->YMCW878403-HSIJ ABN=<--92 162 190 070-->
<!--DETAIL-->
<!--CODE=>9153113004087<!--DESCRIPTION="Slow Cooked Basmati Beef Curry" QTY=>3.00<!--COSTINC=>8.61<!--TOTALEX=>25.83<!-->
<!--LINE CODE=>9153113004001<!--DESCRIPTION="Beef Revol in Napoletana Sauce (30g)" QTY=>1.00<!--COSTINC=>8.61<!--TOTALEX=>8.61<!-->
<!--LINE CODE=>9153113004278<!--DESCRIPTION="Beef Revol with Basmati Rice & Vegetables" QTY=>5.00<!--COSTINC=>43.05<!--TOTALEX=>215.25<!-->
<!--LINE CODE=>9153113004551<!--DESCRIPTION="Creamy Beef Stroganoff with Spinach Fettuccine" QTY=>8.00<!--COSTINC=>68.88<!--TOTALEX=>550.64<!-->
<!--LINE CODE=>9153113004087<!--DESCRIPTION="Beefed Beef Ragu with Polenta & Root Vegetable" QTY=>5.00<!--COSTINC=>43.05<!--TOTALEX=>215.25<!-->
<!--LINE CODE=>9153113004001<!--DESCRIPTION="Butter Chicken with Pilaf" QTY=>6.00<!--COSTINC=>51.66<!--TOTALEX=>310.00<!-->
<!--LINE CODE=>9153113004278<!--DESCRIPTION="Cajun Beef with Sweet Potato Wedges (30g)" QTY=>2.00<!--COSTINC=>17.22<!--TOTALEX=>34.44<!-->
<!--LINE CODE=>9153113004001<!--DESCRIPTION="Chicken Parmigiana with Pumpkin Hash & Vegetables" QTY=>1.00<!--COSTINC=>8.61<!--TOTALEX=>8.61<!-->
<!--LINE CODE=>9153113004551<!--DESCRIPTION="Chicken Schnitzel with Potato Broccoli Hash & Mushroom Gravy" QTY=>2.00<!--COSTINC=>17.22<!--TOTALEX=>34.44<!-->
<!--LINE CODE=>9153113004087<!--DESCRIPTION="Chipotle Chicken Burrito Bowl" QTY=>7.00<!--COSTINC=>61.01<!--TOTALEX=>427.07<!-->
<!--LINE CODE=>9153113004001<!--DESCRIPTION="Cocconi Turmeric Chicken" QTY=>3.00<!--COSTINC=>8.61<!--TOTALEX=>25.83<!-->
<!--LINE CODE=>9153113004278<!--DESCRIPTION="Crab & Avocado Risotto" QTY=>1.00<!--COSTINC=>8.61<!--TOTALEX=>8.61<!-->
<!--LINE CODE=>9153113004087<!--DESCRIPTION="Creamy Tomato Chicken Fettuccine (340g)" QTY=>1.00<!--COSTINC=>8.61<!--TOTALEX=>8.61<!-->
<!--LINE CODE=>9153113004001<!--DESCRIPTION="Crumbed Chicken with Roasted Potatoes" QTY=>5.00<!--COSTINC=>8.61<!--TOTALEX=>43.05<!-->
<!--LINE CODE=>9153113004278<!--DESCRIPTION="Deli Delight with Spiced Tomato Aioli" QTY=>1.00<!--COSTINC=>8.61<!--TOTALEX=>8.61<!-->
<!--LINE CODE=>9153113004087<!--DESCRIPTION="Garlic Butter Beef with Spiced Tomato Aioli" QTY=>1.00<!--COSTINC=>8.61<!--TOTALEX=>8.61<!-->
<!--LINE CODE=>9153113004001<!--DESCRIPTION="Lemon Pepper Chicken with Pea Mash" QTY=>4.00<!--COSTINC=>8.61<!--TOTALEX=>34.44<!-->
<!--LINE CODE=>9153113004278<!--DESCRIPTION="Mexican Beef with spiced Roasted Potatoes & Jalapeño Relish" QTY=>1.00<!--COSTINC=>8.61<!--TOTALEX=>8.61<!-->
<!--LINE CODE=>9153113004087<!--DESCRIPTION="Naked Chicken Fajitas" QTY=>2.00<!--COSTINC=>8.61<!--TOTALEX=>17.22<!--GST=17.22<!-->
<!--LINE CODE=>9153113004278<!--DESCRIPTION="Peri Peri Chicken with Tomato Rice & Green Beans" QTY=>3.00<!--COSTINC=>8.61<!--TOTALEX=>25.83<!-->
<!--LINE CODE=>9153113004001<!--DESCRIPTION="Pork Chop with Tomato & Basil" QTY=>1.00<!--COSTINC=>8.61<!--TOTALEX=>8.61<!-->
<!--LINE CODE=>9153113004087<!--DESCRIPTION="Ramen Noodle with Beef" QTY=>1.00<!--COSTINC=>8.61<!--TOTALEX=>8.61<!-->
<!--LINE CODE=>9153113004001<!--DESCRIPTION="Spicy Buffalo Chicken with Mac & Cheese (320g)" QTY=>1.00<!--COSTINC=>8.61<!--TOTALEX=>8.61<!-->
<!--LINE CODE=>9153113003264<!--DESCRIPTION="Spinach Ricotta Revolli in Pumpkin Sauce" QTY=>3.00<!--COSTINC=>8.61<!--TOTALEX=>25.83<!-->
<!--LINE CODE=>9153113004445<!--DESCRIPTION="Spinach Ricotta Revolli in Pumpkin Sauce" QTY=>3.00<!--COSTINC=>8.61<!--TOTALEX=>25.83<!-->
<!--LINE CODE=>9153113004001<!--DESCRIPTION="Thai Green Chicken Curry with Basmati Rice (340g)" QTY=>2.00<!--COSTINC=>8.61<!--TOTALEX=>17.22<!--GST=17.22<!-->
<!--/DETAILS-->
<!--/INVOICES-->

```

**Fig. 8.** Adding CODE as a required field in the prompt helps extract the product code as a separate field, even though there is no 'CODE' column header found in the document.

**Fig. 9.** The invoice number is not extracted in any of the files as it is highlighted in a dark background.

**Fig. 10.** Examples of diverse set of form-like documents with varying layouts, formats, and complexity, generated by several industry verticals. Typically comprises of key-value pairs and tabular structures (single or multiple)

**Fig. 11.** Key-value pairs and line-items found in a typical form-like document

**INVOICE**

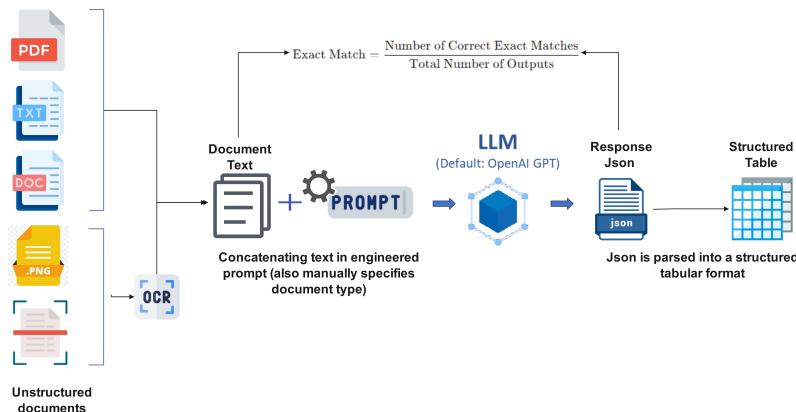
ACCOUNT #:	C1006
STATUS:	Pending
INVOICE #:	I1083
DATE:	08-14-2008
DUE DATE:	08-14-2008
TERMS:	None

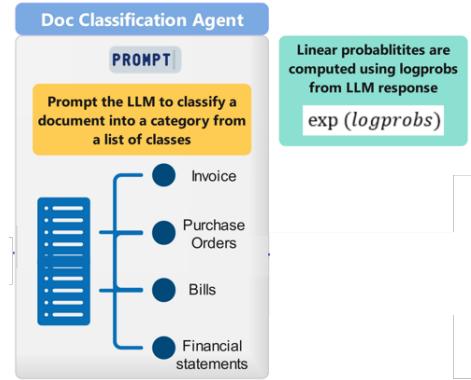
**Purchase Order**

order number:	STU4500279621
order date:	2018/11/06
quote number:	

**Medical Test Report**

Patient: Doe, John
DOB: 12/22/1970 Age: 50 Y Gender: Male
Phone: 111-222-3333 Address: 5678 Summer Avenue, Dallas, TX-75231
Provider: Sandra Bloom MD
Date: 11/11/2021

**Fig. 12.** Two types of Key-value pair variations across different form-like documents; semantic variation and positional variation**Fig. 13.** Baseline unstructured data extraction logic depends on a single static one-shot prompt with up to 3 maximum retries. Type of document is manually defined in each prompt, and the retries are dependent on predefined exact match threshold.



**Fig. 14.** A static prompt is used to categorize documents into predefined classes.