TASER: Table Agents for Schema-guided Extraction and Recommendation

Nicole Cho **Kirsty Fielding** **William Watson**
Sumitra Ganesh **Manuela Veloso**
J.P. Morgan AI Research
nicole.cho@jpmorgan.com

Abstract

Real-world financial filings report critical information about an entity's investment holdings, essential for assessing that entity's risk, profitability, and relationship profile. Yet, these details are often buried in messy, multi-page, fragmented tables that are difficult to parse, hindering downstream QA and data normalization. Specifically, 99.4% of the tables in our financial table dataset lack bounding boxes, with the largest table spanning 44 pages. To address this, we present **TASER (Table Agents for Schema-guided Extraction and Recommendation)**, a continuously learning, agentic table extraction system that converts highly unstructured, multi-page, heterogeneous tables into normalized, schema-conforming outputs. Guided by an initial portfolio schema, TASER executes table detection, classification, extraction, and recommendations in a single pipeline. Our Recommender Agent reviews unmatched outputs and proposes schema revisions, enabling TASER to outperform vision-based table detection models such as Table Transformer by 10.1%. Within this continuous learning process, larger batch sizes yield a 104.3% increase in useful schema recommendations and a 9.8% increase in total extractions. To train TASER, we manually labeled 22,584 pages and 3,213 tables covering $731.7 billion in holdings, culminating in **TASERTab** to facilitate research on real-world financial tables and structured outputs. Our results highlight the promise of continuously learning agents for robust extractions from complex tabular data.

1 Introduction

Financial documents, particularly annual regulatory filings for funds, house tables that govern $68.9 trillion of investments globally (Investment Company Institute, 2024). By comparison, $68.9 trillion is more than twice the total Gross Domestic Product (GDP) of the United States ($29.1 trillion) (WorldBank, 2025). This critical data is housed in

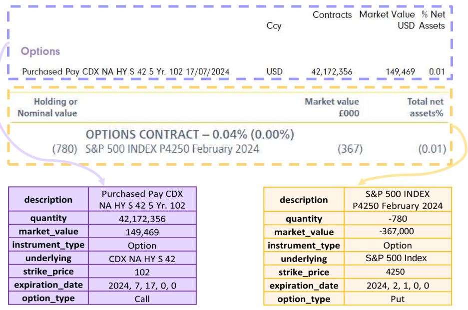


Figure 1: **Complexity of Holdings Table in Regulatory Filings.** In the original format, multiple data attributes are displayed in a single line, with no bounding boxes, rendering the generation of structured outputs highly challenging. TASER enables the generation of structured outputs from highly variable, multi-page financial tables for complex instrument holdings. Negative quantities or market values denote short positions. See Appendix K for additional outputs.

the Financial Holdings Table (Figure 1), which outlines the entirety of an entity's investment holdings (U.S. Congress, 1934; EU Commission, 2019); this table has the highest row count (maximum 426 rows)–more than double the average row count of all other table types (Table 7). These Financial Holdings Tables are long and highly heterogeneous in layout (Figure 2). While generating structured outputs from these tables is critical for many regulatory and financial institutions to undertake basic QA (Question-Answering) tasks using an LLM (Large Language Model) or libraries such as pandas (Cho et al., 2024), there is a relative dearth of studies that focus on continual learning to extract from Financial Holdings Tables, compared to web or SQL tables (Herzig et al., 2020; Pasupat and Liang, 2015; Zhong et al., 2017). Therefore, the following challenges exist in terms of parsing Financial Holdings Tables into structured, machine-readable outputs: **(1) One-to-many relationships between a document and the tables it houses** exacerbate standard model performance for table

Proceedings of the 19th Conference of the European Chapter of the Association for Computational Linguistics (Volume 5: Industry Track), pages 226–252
March 25-27, 2026

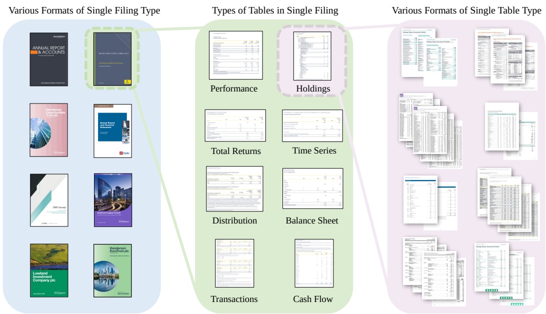


Figure 2: **Variety and complexity of financial tables.** From leftmost column - for a single financial filing type, such as annual reports, there is no consistency among reports. Within each report, there are numerous table types with each type housing very different types of information. Even within a single table type (such as the Financial Holdings Table), there are numerous layout structures, as seen in the rightmost column. Due to the extreme heterogeneity of formatting, document layout, and table structure, traditional table extraction methods fail to perform for financial filings.

detection or structure recognition tasks. **(2) Financial Holdings Tables span across multiple pages**, rendering models that operate at the page level inefficient. **(3) Financial instruments are highly complex** with nested hierarchies. Therefore, details are often clumped in a single cell as seen in Figure 1. **(4) Tabular layouts are heterogeneous with no bounding boxes**, mixing tables, text blocks, footnotes, and images, often without consistent labeling or alignment. 99.4% of tables in our dataset, **TASERTab**, lack bounding boxes to efficiently identify a single cell. These challenges motivate our agentic table extraction methodology capable of goal-driven parsing and self-refinement, continuously learning and reasoning from errors.

Contribution 1: We propose a continuously learning, agentic table extraction methodology, **TASER** (Table Agents for Schema-guided Extraction and Recommendation) that performs detection, classification, extraction, and recommendation in a single pipeline by leveraging the schema invoked as a tool call. TASER is layout-agnostic and can operate for tables of any format. We compare our methodology against predominant methodologies and report TASER's 10.1% improvement over Table Transformer (Yang et al., 2022) for detection.

Contribution 2: We demonstrate the effectiveness of our Recommender Agent to continuously improve the initial schema - reflecting a tunable and continuous self-learning loop. Throughout our training, we found that small batches are optimal for providing diverse and comprehensive recommendations to the original schema–however, at the cost of redundant recommendations. In contrast, large batches drive high precision recommendations at the cost of diversity. Thus, our results establish that self-learning via agents for table extraction is tunable; through adjusting batch size, we can control schema refinement to maximize actionable coverage while minimizing redundancy.

Contribution 3: We have constructed a manually labeled dataset **TASERTab** of ground truth labels for 3,213 real-world Financial Holdings Tables amounting to $731.7B in value. We sourced the filings from fund websites, labeled the total net assets for each fund, and recorded the span of each Financial Holdings Table. We believe that this is the first dataset of its kind to provide access to real-world financial tables side by side with structured outputs.

2 Related Work

Information & Table Extraction: Early information extraction relied on statistical models (HMMs (Borkar et al., 2001), CRFs (Lafferty et al., 2001), heuristics (Press, 2003), and graph-based layouts (Liu et al., 2019; Qian et al., 2019; Meuschke et al., 2023), but still struggle with complex, heterogeneous tables.

Table Representation Learning: Transformer-based table understanding and QA include TaPaS (Herzig et al., 2020), TaBERT (Yin et al., 2020), TaPEX (Liu et al., 2022), TURL (Deng et al., 2020), TUTA (Wang et al., 2021), and TableFormer (Yang et al., 2022). These methods encode text, structure, and layout, but few are benchmarked on long, dense, multi-page financial reports.

LLMs for Structured Data: General LLMs have strong performance for schema-conformant extraction via fine-tuning & prompting (Brown et al., 2020; Liu and Contributors, 2024), while multimodal approaches (LayoutLM (Xu et al., 2020, 2021; Huang et al., 2022), DONUT (Kim et al., 2022), DocFormer (Appalaraju et al., 2021), UniTable (Peng et al., 2024), and Table Transformer (Smock et al., 2021; Carion et al., 2020) improve layout awareness but still lag on long, fragmented tables (Zhao et al., 2024).

Financial Document Parsing: (Watson and Liu, 2020) has focused on table extraction from images while (Cho et al., 2024) has focused on expert agent pipelines. Large-scale benchmarks such as

DocILE (Šimsa et al., 2023), BuDDIE (Wang et al., 2025) have also focused on financial documents.

Agentic and Recursive Extraction: Recent methods cast LLMs as agents capable of iterative extraction and self-correction (Shen et al., 2023; Roucher et al., 2025; Watson et al., 2023; Yuan and Xie, 2025). Prompt-based feedback, introspective refinement, and episodic memory frameworks (Madaan et al., 2023; Shinn et al., 2023; Yao et al., 2023) drive improvements in reasoning for complex extraction.

3 Methodology

3.1 System Architecture

Our methodology is composed of three core Large Language Model (LLM) agents, each with a distinct role. We conduct rigorous ablations to evaluate the importance of each agent.

1. **Detector Agent:** Identifies candidate pages containing Financial Holdings Tables leveraging the initial schema provided. The prompt is tuned to maximize recall to avoid missing any Financial Holdings Tables. We provide our prompts in the Appendix (Figure 12).
2. **Extractor Agent:** Processes detected pages by prompting the LLM with the current `Portfolio` schema embedded in the prompt context. The LLM's output is validated inline against the schema using Pydantic & Instructor, producing a set of structured, type-checked instrument entries (Figure 1).
3. **Recommender Agent:** Reviews unmatched extractions containing both false and true positives. A *false positive* is a spurious extraction (e.g., headers/subtotals/footnotes/OCR noise or cells from non-holdings tables) that fails schema/type/consistency checks; a *true positive* is a valid holdings field from a genuine row that the current schema cannot yet classify but passes those checks. The agent first filters false positives by re-validating each candidate under the current schema; it proposes schema modifications for the remaining true positives and, per class, recommends the minimal change needed.

All agents interact through explicit artifacts: structured outputs, episodic error stacks, and schema definitions. Output validation is integrated into each agent's forward pass via Instructor (Liu and Contributors, 2024). TASER implements a recursive feedback loop, where errors and unmatched holdings identified in the initial extraction are escalated to the Recommender Agent, which provides recommendations to refine the schema and triggers re-extraction. This loop repeats until all entries are matched. A schematic of the full agentic pipeline is shown in Figure 3.

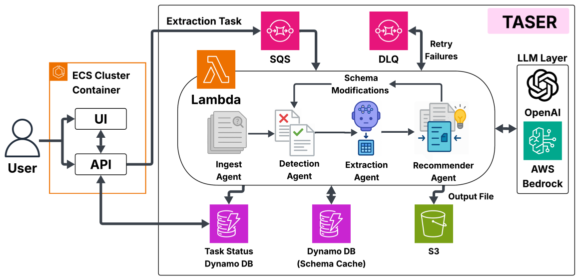


Figure 3: **TASER deployment architecture.** Users submit extraction requests through a UI or API hosted in an ECS container, which enqueues an extraction task to SQS. An AWS Lambda function orchestrates the Ingest, Detection, Extraction, and Recommender Agents, persisting task status and schema cache in DynamoDB and writing intermediary and final outputs to S3; failed tasks are routed to a dead-letter queue (DLQ) for later inspection. The Recommender Agent analyzes the intermediary output file and associated error stack to propose schema enhancement recommendations. Users can accept or reject these recommendations; accepted updates are written back to the schema cache and automatically retrigger the extraction pipeline, enabling TASER to continuously refine its extraction artifacts over time.

3.2 Initial Schema Definition and Application

TASER's extraction process is anchored by an explicit, user-modifiable `Portfolio` schema that defines the target structure for Holdings Tables. We implement this schema using Pydantic models; our initial schema reflects is informed by leveraging external knowledge (U.S. Congress, 1934). Each schema consists of a base `Instrument` model, subclassed for common asset types (e.g., `Equity`, `Bond`, `Option`, `Swap`, `Forward`, `Future`, `Debt`, and an `Other` class for uncategorized rows. Each subclass specifies instrument-specific fields and validation logic (see App. G).

Schema-Guided Extraction: For each candidate page, the Extractor Agent prompts the LLM with the current schema embedded in the prompt context. The LLM is instructed to return a structured output, which is immediately parsed and validated against the schema using Pydantic's type checking and validation logic. Outputs that fail schema validation (e.g., missing fields, type errors, or undeclared instruments) are flagged.

Schema Recommendations for Iterative Refine-

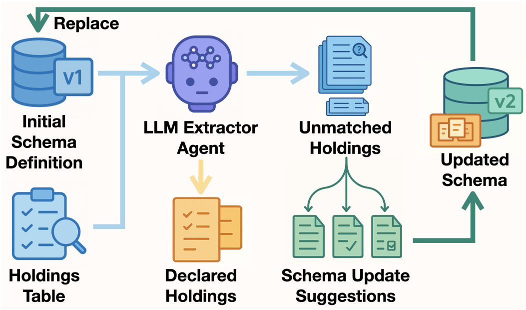


Figure 4: **Schema-Guided Agentic Refinement Loop.** The extraction pipeline begins with an *Initial Schema Definition* (v1), which guides the LLM Extractor Agent as it processes the raw Holdings Table to produce *Declared Holdings*. Holdings that do not match the schema are routed as *Unmatched Holdings*, triggering the generation of *Schema Update Suggestions*. These suggestions are reviewed, clustered, and aggregated by our Recommender Agent before updating the schema (v2), replacing the prior definition and closing the agentic feedback loop. This process enables continuous schema refinement and robust extraction.

Algorithm 1 LLM Iterative Schema Refinement

Require: Unmatched holdings $H = \{h_1, h_2, \ldots, h_N\}$, initial schema $\Sigma^{(0)}$, LLM schema suggestion function g_θ, batch size B, stopping criterion T

Initialize $\ell \leftarrow 0$
$H^{(0)} \leftarrow H$ {Current unmatched holdings}
$\Sigma^{(0)} \leftarrow$ initial schema
while not stopping criterion T met **do**
  Partition $H^{(\ell)}$ into batches $H_j^{(\ell)}$ of size at most B
  $S^{(\ell)} \leftarrow \emptyset$ {Suggested schema modifications}
  for each batch $H_j^{(\ell)}$ **do**
    $S_j^{(\ell)} \leftarrow g_\theta(H_j^{(\ell)}, \Sigma^{(\ell)})$
    $S^{(\ell)} \leftarrow S^{(\ell)} \cup S_j^{(\ell)}$
  end for
  $S_{\text{selected}}^{(\ell)} \leftarrow$ AggregateAndSelect$(S^{(\ell)})$ {Aggregate suggestions}
  $\Sigma^{(\ell+1)} \leftarrow$ UpdateSchema$(\Sigma^{(\ell)}, S_{\text{selected}}^{(\ell)})$
  $H^{(\ell+1)} \leftarrow$ UnmatchedHoldings$(H, \Sigma^{(\ell+1)})$
  if $H^{(\ell+1)} = \emptyset$ **then**
    break
  end if
  $\ell \leftarrow \ell + 1$
end while
return $\Sigma^{(\ell+1)}$

ment: We formalize schema refinement as an iterative, LLM-driven clustering process that updates the schema to accommodate unmatched or novel holdings discovered during extraction (Novikov et al., 2025; Zhang et al., 2023). At each iteration, the agent operates on the episodic error stack to propose schema modifications, and extraction is retried using the updated schema. This process continues until all entries are matched or no further improvements are possible. Let $H = \{h_1, h_2, \ldots, h_N\}$ denote the set of unmatched holdings, and let $\Sigma^{(0)}$ be the initial schema. For each iteration ℓ:

- $H^{(\ell)}$: Unmatched holdings at iteration ℓ.
- $\Sigma^{(\ell)}$: Current schema.
- g_θ: LLM-based schema suggestion function.
- B: Batch size for error grouping.

The refinement loop (Algo. 1) proceeds as follows:

1. Partition $H^{(\ell)}$ into batches of size at most B.
2. For each batch, invoke g_θ with batch errors and $\Sigma^{(\ell)}$ to propose schema modifications.
3. Aggregate, cluster, and select recommendations
4. Update schema to $\Sigma^{(\ell+1)}$ and re-extract.
5. Update error stack and repeat until $H^{(\ell+1)}$ is empty or no new schema changes are suggested.

3.3 Ablation Strategies and Efficiency

We systematically ablate TASER to isolate the impact of schema-guided extraction, prompt engineering, and agentic feedback across four strategies:

1. **Raw Text Prompting:** The LLM is prompted only with the page text; extraction is based solely on a yes/no detection.
2. **Structured Chain-of-Thought (CoT):** Prompts include a minimal schema and few-shot examples, eliciting explicit reasoning traces before a final boolean decision.
3. **Full Schema Prompting:** The full `Portfolio` schema is embedded in the prompt, instructing the LLM to return structured, schema-conformant entries.
4. **Direct Schema Application:** The schema is directly applied to parsed page content without prior detection; extraction succeeds if any schema sub-model instantiates.

Table 2 reports detection and extraction via absolute dollar difference, and Table 10 compares computational efficiency in tokens and latency.

4 Experimental Setup

Detection Metrics: We report *recall*, *precision*, *F1*, and *accuracy* for table detection, prioritizing recall to avoid missing Financial Holdings Tables.

Extraction Metrics: We assess extraction completeness by comparing TASER's outputs to ground truth labels. We manually label a total net asset value for each Holdings Table. We then compare this ground truth with our extractions, dubbed the *total absolute difference (TAD)*.

Schema Refinement Metrics: *Coverage* is the fraction of unmatched holdings aligned with at least one schema suggestion, using RapidFuzz string

Provider	Model	Recall (%)	Precision (%)	F1 Score (%)	Accuracy (%)
Camelot	Hybrid	56.92	23.46	33.23	47.35
Microsoft	Table Transformer	99.76	32.75	49.31	46.27
OpenAI	`gpt-4o-2024-11-20`	**100.00**	43.43	59.44	66.35
OpenAI	`gpt-5-mini-2025-08-07`	94.92	54.15	68.96	80.33
OpenAI	`gpt-4.1-2025-04-14`	95.80	54.32	69.33	80.49
OpenAI	`gpt-5-nano-2025-08-07`	95.97	55.63	70.44	81.46
OpenAI	`gpt-5-2025-08-07`	95.97	68.16	**79.71**	**88.75**
Anthropic	`claude_sonnet-3-7`	88.97	57.34	69.73	82.22
Amazon	`nova_pro-v1-0`	85.90	**69.45**	76.84	88.07

Table 1: **Detector performance across models on TASERTab.** Recall, precision, F1, and accuracy are reported for baselines and the Detector Agent instantiated with different LLMs. `gpt-4o-2024-11-20` attains perfect recall, while `gpt-5-2025-08-07` achieves the best overall F1 and accuracy. `nova_pro-v1-0` delivers the highest precision but at the cost of lower recall, illustrating the trade-off between missing holdings tables and avoiding false positives.

	Method	Recall (%)	Precision (%)	F1 (%)	Accuracy (%)	TAD (USD)	Unaccounted
TASER	(a) Raw Text Prompting	**100.00**	38.62	55.73	58.38	$ 107,066,845	0.015%
	(b) Structured CoT	**100.00**	34.42	51.21	50.10	$ 120,577,458	0.016%
	(c) Full Schema Prompting	**100.00**	**43.43**	**59.44**	**66.35**	$ 102,836,797	0.014%
	(d) Direct Schema Application	**100.00**	41.84	58.30	63.99	$ 118,881,312	0.016%

Table 2: **Detection and Extraction Performance Across Strategies.** While all TASER ablations achieve perfect recall, Full Schema Prompting yields the highest precision (43.43%), F1 score (59.44%), and overall accuracy (66.35%), as well as the lowest total absolute difference (TAD) and unaccounted fraction, underscoring the value of embedding the complete Portfolio schema in the detection prompt. Percentage of unaccounted holdings is out of $731.7 billion (ground truth). Lower TAD and unaccounted percentages indicate higher dollar-value fidelity.

similarity with a lenient (≥ 70) threshold. We also report the number of new matched holdings after re-extraction with the suggested schemas added to `Portfolio`. *Diversity* is the average pairwise Levenshtein distance between suggestion attributes (name and generated schema). *Collision rate* denotes the proportion of duplicate suggestions.

Dataset and Model: We curate a diverse corpus totaling **22,584 pages**, **28M tokens**, and **$731.7B** in holdings. Among **3,213 tables**, **57.53%** exhibit hierarchical structure (via spanning cells). All Holdings Tables (**100%**) are hierarchical. While **39%** of portfolios are single-page, **60.2%** span multiple pages. The average length is **3.24 pages** ($\sigma = 3.41$, max = 19). This variability underscores the need for multi-page detection and consolidation. Unless explicitly stated otherwise, all experiments use `gpt-4o-2024-11-20` as the LLM.

5 Results and Discussion

5.1 Quantitative Evaluation

Detection: Table 2 shows that all TASER ablations achieve perfect recall ($\sim$100%), but precision ranges from 32.8% (Table Transformer) up to 43.4% (Full Schema Prompting), driving F1 scores between 49.3% and 59.4%. Embedding the full Portfolio schema in the prompt boosts precision by over 10% relative to the vision-only baseline and yields the highest F1 (59.4%) and accuracy (66.4%), demonstrating that in-context schema guidance is critical.

Extraction: Table 2 confirms schema-anchored extraction improves dollar-value fidelity. Full Schema Prompting attains the lowest absolute difference ($102.8M) and smallest unaccounted share (0.014%), outperforming Raw Text Prompting ($107.1M, 0.015%) and Structured CoT ($120.6M, 0.016%). Direct Schema Application (skipping detection) incurs a higher error ($118.9M; 0.016%) by parsing spurious non-holding pages.

5.2 Success Highlights

Cross-Document Consistency: TASER classifies and extracts Holdings Tables despite varying titles (e.g., "Portfolio of Investments", "Schedule of Holdings", or "Investment Portfolio") and diverse structural formats. Despite the immense complexity of inputs, TASER consistently extracts and transforms these tables, ensuring that the final output appears as if sourced from a uniform set.

Contextual Understanding: TASER excels in handling contextual nuances, such as interpreting negative values denoted by parentheses (e.g., (140)) in zero-shot settings. Such domain-specific attributes are important for financial tables.

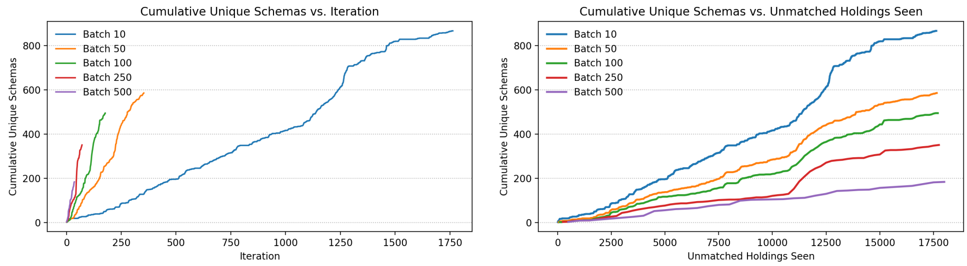


Figure 5: **Left:** Cumulative unique schemas per iteration; larger batches discover schemas rapidly but plateau quickly. **Right:** Cumulative unique schemas per unmatched holding seen; smaller batches ultimately yield more unique schemas but require more suggestions and generate more redundancy.

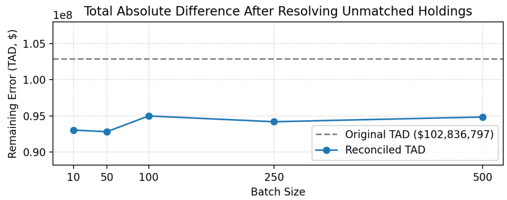


Figure 6: **Reduction in TAD after resolving unmatched holdings.** Remaining TAD is calculated after sequential reconciliation of unmatched holdings. Error reduction is achieved by resolving the most significant unmatched holdings (Figs. 8 & 9).

Extracting Intricate Semantics: TASER demonstrates a strong semantic understanding of financial terminology, which empowers it to extract accurately. For instance, TASER adeptly parsed the table entry "GBP 4,700,000 | UK Treasury 0% 19/02/2024 | 4668 | 1.48," correctly identifying the holding as a bond and extracting its attributes: `quantity`, `market value`, `coupon rate`, `maturity date`, and `issuer`.

5.3 Batch Size Tradeoffs in Refinement

Figure 5 (left) reveals that larger batch sizes (250, 500) rapidly expand the schema. However, this early acceleration comes at the cost of early saturation, after which few new unique schemas are discovered. In contrast, smaller batches require more iterations to reach the same number of unmatched holdings seen, but continue yielding unique schemas, resulting in the highest diversity when normalized by data processed (Figure 5, right). This improvement in coverage, however, is offset by increased redundancy. As shown in Appendix Figure 7, smaller batches incur substantially more overlapping suggestions, reflecting a more granular and exploratory nature. Overall, these results highlight a key tradeoff: larger batches accelerate early discovery but plateau quickly, while smaller batches maximize cumulative schema diversity at the cost of redundancy and computation. Our results establish that schema refinement via agentic feedback is both tractable and tunable. This indicates that an *adaptive batching strategy* may be optimal: using larger batches to quickly identify high-yield schemas, followed by smaller batches for exhaustive diversity.

Schema Diversity and Utilization: Schema diversity, as measured by the average pairwise Levenshtein distance, is maximized for moderate batch sizes (100–250), as shown in Appendix J. While larger batch sizes (500) yield a higher proportion of utilized schemas—up to 59%—smaller, more diverse batches tend to have lower utilization rates (Table 9). Furthermore, the accretive gain in 402 additional unique schemas yielded only marginal improvements in holding coverage (6.1%). Figure 10 illustrates this tradeoff: smaller batch sizes cover more unmatched holdings at the expense of efficiency (96.1% coverage for 29.0% utilization at batch size 10), whereas larger batches achieve higher schema utilization (59.0% at batch size 500).

Improvements in TAD: Resolving the largest unmatched holdings yields a reduction in TAD of approximately 7–10% across batch sizes, with the majority of improvement achieved by reconciling just the top 10–20% of holdings (Table 5).

5.4 Deployment of TASER

We outline the deployment architecture of TASER in Figure 3, with additional system architecture details provided in Appendix A.

6 Conclusion

We present TASER for extracting complex Holdings Tables from documents through continual learning. Our high precision and recall across diverse layouts underscore the potential of agentic continual learning for financial table extraction.

Disclaimer

This paper was prepared for informational purposes by the Artificial Intelligence Research group of JPMorgan Chase & Co. and its affiliates ("JPMorgan") and is not a product of the Research Department of JPMorgan. JPMorgan makes no representation and warranty whatsoever and disclaims all liability, for the completeness, accuracy or reliability of the information contained herein. This document is not intended as investment research or investment advice, or a recommendation, offer or solicitation for the purchase or sale of any security, financial instrument, financial product or service, or to be used in any way for evaluating the merits of participating in any transaction, and shall not constitute a solicitation under any jurisdiction or to any person, if such solicitation under such jurisdiction or to such person would be unlawful.

Limitations

Despite its strong performance, TASER remains susceptible to errors in low-resolution or scanned PDFs, where visual degradation can hinder accurate extraction. Ambiguities in financial documents, such as undefined asset classes or implicit references, pose challenges that cannot always be resolved without external knowledge or manual intervention. While recursive prompting enhances completeness, it introduces added latency and computational overhead. Additionally, TASER relies on prompt-based weak supervision due to the lack of fine-grained, labeled datasets for complex instrument types, which may limit generalization. Finally, TASER does not yet model interactions between table rows or instrument relationships beyond the schema level, which may affect downstream tasks such as portfolio risk analysis or exposure aggregation.

References

Srikar Appalaraju, Bhavan Jasani, Bhargava Urala Kota, Yusheng Xie, and R. Manmatha. 2021. Docformer: End-to-end transformer for document understanding. *Preprint*, arXiv:2106.11539.

Petr Babkin, William Watson, Zhiqiang Ma, Lucas Cecchi, Natraj Raman, Armineh Nourbakhsh, and Sameena Shah. 2023. Bizgraphqa: A dataset for image-based inference over graph-structured diagrams from business domains. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '23, page 2691–2700, New York, NY, USA. Association for Computing Machinery.

Vinayak Borkar, Kaustubh Deshmukh, and Sunita Sarawagi. 2001. Automatic segmentation of text into structured records. In *Proceedings of the 2001 ACM SIGMOD International Conference on Management of Data*, SIGMOD '01, page 175–186, New York, NY, USA. Association for Computing Machinery.

Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, and 12 others. 2020. Language models are few-shot learners. *Preprint*, arXiv:2005.14165.

Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. 2020. End-to-end object detection with transformers. *Preprint*, arXiv:2005.12872.

Zewen Chi, Heyan Huang, Heng-Da Xu, Houjin Yu, Wanxuan Yin, and Xian-Ling Mao. 2019. Complicated table structure recognition. *arXiv preprint arXiv:1908.04729*.

Nicole Cho, Nishan Srishankar, Lucas Cecchi, and William Watson. 2024. Fishnet: Financial intelligence from sub-querying, harmonizing, neural-conditioning, expert swarms, and task planning. In *Proceedings of the 5th ACM International Conference on AI in Finance*, ICAIF '24, page 591–599, New York, NY, USA. Association for Computing Machinery.

Xiang Deng, Huan Sun, Alyssa Lees, You Wu, and Cong Yu. 2020. Turl: Table understanding through representation learning. *Preprint*, arXiv:2006.14806.

EU Commission. 2019. *EU Accounts*. EU Commission.

Jonathan Herzig, Pawel Krzysztof Nowak, Thomas Müller, Francesco Piccinno, and Julian Eisenschlos. 2020. Tapas: Weakly supervised table parsing via pre-training. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics.

Yupan Huang, Tengchao Lv, Lei Cui, Yutong Lu, and Furu Wei. 2022. Layoutlmv3: Pre-training for document ai with unified text and image masking. *Preprint*, arXiv:2204.08387.

Investment Company Institute. 2024. *Factbook*. Investment Company Institute, Washington, DC.

Geewook Kim, Teakgyu Hong, Moonbin Yim, Jeongyeon Nam, Jinyoung Park, Jinyeong Yim, Wonseok Hwang, Sangdoo Yun, Dongyoon Han, and Seunghyun Park. 2022. Ocr-free document understanding transformer. *Preprint*, arXiv:2111.15664.

John D. Lafferty, Andrew McCallum, and Fernando C. N. Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the Eighteenth International Conference on Machine Learning*, ICML '01, page 282–289, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Jason Liu and Contributors. 2024. Instructor: A library for structured outputs from large language models. Accessed: 2025-11-12.

Qian Liu, Bei Chen, Jiaqi Guo, Morteza Ziyadi, Zeqi Lin, Weizhu Chen, and Jian-Guang Lou. 2022. Tapex: Table pre-training via learning a neural sql executor. *Preprint*, arXiv:2107.07653.

Xiaojing Liu, Feiyu Gao, Qiong Zhang, and Huasha Zhao. 2019. Graph convolution for multimodal information extraction from visually rich documents. *Preprint*, arXiv:1903.11279.

Aman Madaan, Niket Tandon, Prakhar Gupta, Skyler Hallinan, Luyu Gao, Sarah Wiegreffe, Uri Alon, Nouha Dziri, Shrimai Prabhumoye, Yiming Yang, Shashank Gupta, Bodhisattwa Prasad Majumder, Katherine Hermann, Sean Welleck, Amir Yazdanbakhsh, and Peter Clark. 2023. Self-refine: Iterative refinement with self-feedback. *arXiv*.

Norman Meuschke, Apurva Jagdale, Timo Spinde, Jelena Mitrović, and Bela Gipp. 2023. *A Benchmark of PDF Information Extraction Tools Using a Multi-task and Multi-domain Evaluation Framework for Academic Documents*, page 383–405. Springer Nature Switzerland.

Alexander Novikov, Ngân Vũ, Marvin Eisenberger, Emilien Dupont, Po-Sen Huang, Adam Zsolt Wagner, Sergey Shirobokov, Borislav Kozlovskii, Francisco J. R. Ruiz, Abbas Mehrabian, M. Pawan Kumar, Abigail See, Swarat Chaudhuri, George Holland, Alex Davies, Sebastian Nowozin, Pushmeet Kohli, and Matej Balog. 2025. Alphaevolve: A coding agent for scientific and algorithmic discovery. *Preprint*, arXiv:2506.13131.

Panupong Pasupat and Percy Liang. 2015. Compositional semantic parsing on semi-structured tables. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1470–1480, Beijing, China. Association for Computational Linguistics.

ShengYun Peng, Aishwarya Chakravarthy, Seongmin Lee, Xiaojing Wang, Rajarajeswari Balasubramaniyan, and Duen Horng Chau. 2024. Unitable: Towards a unified framework for table recognition via self-supervised pretraining. *Preprint*, arXiv:2403.04822.

Columbia University Press. 2003. *The Columbia Guide to Digital Publishing*. Columbia University Press.

Yujie Qian, Enrico Santus, Zhijing Jin, Jiang Guo, and Regina Barzilay. 2019. GraphIE: A graph-based framework for information extraction. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 751–761, Minneapolis, Minnesota. Association for Computational Linguistics.

Aymeric Roucher, Albert Villanova del Moral, Thomas Wolf, Leandro von Werra, and Erik Kaunismäki. 2025. 'smolagents': a smol library to build great agentic systems. https://github.com/huggingface/smolagents.

Yongliang Shen, Kaitao Song, Xu Tan, Dongsheng Li, Weiming Lu, and Yueting Zhuang. 2023. Hugginggpt: Solving ai tasks with chatgpt and its friends in hugging face. *Preprint*, arXiv:2303.17580.

Noah Shinn, Federico Cassano, Edward Berman, Ashwin Gopinath, Karthik Narasimhan, and Shunyu Yao. 2023. Reflexion: Language agents with verbal reinforcement learning. *Preprint*, arXiv:2303.11366.

Štěpán Šimsa, Milan Šulc, Michal Uřičář, Yash Patel, Ahmed Hamdi, Matěj Kocián, Matyáš Skalický, Jiří Matas, Antoine Doucet, Mickaël Coustaty, and Dimosthenis Karatzas. 2023. DocILE benchmark for document information localization and extraction.

Brandon Smock, Rohith Pesala, and Robin Abraham. 2021. Pubtables-1m: Towards comprehensive table extraction from unstructured documents. *Preprint*, arXiv:2110.00061.

U.S. Congress. 1934. Securities exchange act of 1934. https://www.govinfo.gov/content/pkg/COMPS-1884/pdf/COMPS-1884.pdf. Codified at 15 U.S.C. § 78a et seq.

Dongsheng Wang, Ran Zmigrod, Mathieu J. Sibue, Yulong Pei, Petr Babkin, Ivan Brugere, Xiaomo Liu, Nacho Navarro, Antony Papadimitriou, William Watson, Zhiqiang Ma, Armineh Nourbakhsh, and Sameena Shah. 2025. BuDDIE: A business document dataset for multi-task information extraction. In *Proceedings of the Joint Workshop of the 9th Financial Technology and Natural Language Processing (FinNLP), the 6th Financial Narrative Processing (FNP), and the 1st Workshop on Large Language Models for Finance and Legal (LLMFinLegal)*, pages 35–47, Abu Dhabi, UAE. Association for Computational Linguistics.

Zhiruo Wang, Haoyu Dong, Ran Jia, Jia Li, Zhiyi Fu, Shi Han, and Dongmei Zhang. 2021. Tuta: Tree-based transformers for generally structured table pre-training. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, KDD '21. ACM.

William Watson, Nicole Cho, Tucker Balch, and Manuela Veloso. 2023. Hiddentables and pyqtax: A cooperative game and dataset for tableqa to ensure scale and data privacy across a myriad of taxonomies. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, page 7144–7159. Association for Computational Linguistics.

William Watson and Bo Liu. 2020. Financial table extraction in image documents. In *Proceedings of the First ACM International Conference on AI in Finance*, ICAIF '20, page 1–8. ACM.

WorldBank. 2025. Worldbankgdp. `https://data.worldbank.org/indicator/NY.GDP.MKTP.CD?locations=US`. Accessed: 2025-07-22.

Yang Xu, Yiheng Xu, Tengchao Lv, Lei Cui, Furu Wei, Guoxin Wang, Yijuan Lu, Dinei Florencio, Cha Zhang, Wanxiang Che, Min Zhang, and Lidong Zhou. 2021. LayoutLMv2: Multi-modal pre-training for visually-rich document understanding. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 2579–2591, Online. Association for Computational Linguistics.

Yiheng Xu, Minghao Li, Lei Cui, Shaohan Huang, Furu Wei, and Ming Zhou. 2020. Layoutlm: Pre-training of text and layout for document image understanding. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '20, page 1192–1200. ACM.

Jingfeng Yang, Aditya Gupta, Shyam Upadhyay, Luheng He, Rahul Goel, and Shachi Paul. 2022. Tableformer: Robust transformer modeling for table-text encoding. *Preprint*, arXiv:2203.00274.

Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. 2023. React: Synergizing reasoning and acting in language models. *Preprint*, arXiv:2210.03629.

Pengcheng Yin, Graham Neubig, Wen-tau Yih, and Sebastian Riedel. 2020. TaBERT: Pretraining for joint understanding of textual and tabular data. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8413–8426, Online. Association for Computational Linguistics.

Yurun Yuan and Tengyang Xie. 2025. Reinforce llm reasoning through multi-agent reflection. *Preprint*, arXiv:2506.08379.

Yuwei Zhang, Zihan Wang, and Jingbo Shang. 2023. Clusterllm: Large language models as a guide for text clustering. In *EMNLP 2023*. ACL.

Weichao Zhao, Hao Feng, Qi Liu, Jingqun Tang, Shu Wei, Binghong Wu, Lei Liao, Yongjie Ye, Hao Liu, Wengang Zhou, Houqiang Li, and Can Huang. 2024. Tabpedia: Towards comprehensive visual table understanding with concept synergy. *Preprint*, arXiv:2406.01326.

Victor Zhong, Caiming Xiong, and Richard Socher. 2017. Seq2sql: Generating structured queries from natural language using reinforcement learning. *Preprint*, arXiv:1709.00103.

Xu Zhong, Elaheh ShafieiBavani, and Antonio Jimeno Yepes. 2019. Image-based table recognition: data, model, and evaluation. *arXiv preprint arXiv:1911.10683*.

A TASER System Architecture

A.1 Overview

TASER is an event-driven microservice architecture composed of several semi-independent agents, such as text extraction and table detection modules. Each agent operates on a shared cloud infrastructure, which typically consists of an AWS Lambda function that listens to a queue (Amazon SQS), manages file operations in Amazon S3, and utilizes a dead-letter queue (DLQ) for retry handling.

A.2 User Interaction and Deployment

Users may interact with TASER either through a graphical user interface (UI) or directly via the application programming interface (API). Both the UI and API are deployed within an Amazon ECS cluster, with horizontal and vertical scaling enabled to respond to user demand. The UI is updated by periodically polling the API for *task status*, which is maintained in Amazon DynamoDB.

A.3 Task Cache

An additional component is the *task cache*, implemented using DynamoDB. The primary purpose of the cache is to reduce latency and operational costs by retrieving previously computed results. For all agents, the cached value is typically an S3 location indicating where the result of a prior execution is stored. The cache key varies by agent but generally includes model-specific information (e.g., model name, temperature settings) and the prompt used for generation. A cache entry for the table detection agent is structured as follows:

Key:	`{schema_id, prompt_id, document_text_id, model_name, model_configs}`
Value:	S3 location

A.4 Design Rationale: API-Centric Orchestration

A core design decision for TASER was to utilize an API-centric approach for orchestrating workflows. An alternative would have been to configure each agent to automatically trigger an AWS Lambda function upon the arrival of a new file in S3 (either by directly configuring the S3 bucket or using AWS EventBridge). While this method is feasible, it means that any change to workflows would require reprovisioning AWS infrastructure. In contrast, with the current architecture, redesigning a workflow is as simple as refactoring the UI or API client code (for example, changing the order of API calls) and redeploying the application, without modifying the underlying cloud infrastructure.

B Camelot Table Parsing Modes

For completeness, we also compare TASER's detection performance against the four table detection modes in Camelot[1]. The best-performing variant (Hybrid) achieves an F1 score of 0.33, still below TASER's weakest ablation (0.51). Full results are presented in Table 4. Note that our financial tables primarily consist of unruled, whitespace-separated tables with alignment-based structure. Below is a brief summary of each mode:

- **Stream:** Groups text using whitespace and y-axis alignment. Suitable for unruled tables, but yielded low precision on our data ($F1 = 21.6\%$).
- **Lattice:** Uses image-based line detection to extract ruled tables. Less effective for our dataset due to the rarity of bordered layouts ($F1 = 13.3\%$).
- **Network:** Detects tables via text alignment patterns using bounding boxes. Performs better on our format, which lacks explicit ruling ($F1 = 18.6\%$).
- **Hybrid:** Combines Network's structure with Lattice's grid refinement. Achieved the highest F1 score (33.23%) among Camelot modes, confirming the benefit of integrating both visual and alignment cues.

C TASER Annotation Process

We manually sourced each financial document directly from the fund entity's public website, ensuring broad coverage across instrument types. Annotations were performed at the page, table, and holdings level (which may span hundreds of pages). For every filing and fund, we recorded the page-span for the portfolio of investments table and the net asset value across all holdings for that fund.

D TASER Dataset Release

TASER is built on public fund documents. Our release will include labels for the positions of holdings tables, the recorded net asset value, the fund name, multi-page spans, and a URL reference to the public fund document. Each pdf filing is hosted by the fund's advisor, as required by regulation.

[1] https://github.com/camelot-dev/camelot

Model	Modality	Primary Task	Promptable
Camelot	Vision + Spatial	Heuristic Table Detection & Parsing	No
Table Transformer	Vision	Detection & Structure Recognition	No
TaPas	Text	Table-based QA	Partially
TAPEX	Text	Programmatic Extraction (SQL)	Partially
TASER (ours)	Vision + Text	Schema-guided Extraction	Yes

Table 3: **Comparison of representative table extraction and reasoning models.** Our work extends prior methods by introducing a fully agentic, schema-guided extraction framework for highly complex financial tables, leveraging prompt-based self-refinement and continuous schema adaptation.

	Method	Recall (%)	Precision (%)	F1 (%)	Accuracy (%)
Camelot	Stream	28.02	17.56	21.59	53.16
	Lattice	14.01	12.72	13.33	58.08
	Network	42.62	21.50	28.58	50.97
	Hybrid	56.92	23.46	33.23	47.35
	Table Transformer (Smock et al., 2021)	99.76	32.75	49.31	46.27
TASER	(a) Raw Text Prompting	100.0	38.62	55.73	58.38
	(b) Structured CoT	100.0	34.42	51.21	50.10
	(c) Full Schema Prompting	100.0	**43.43**	**59.44**	**66.35**
	(d) Direct Schema Application	100.0	41.84	58.30	63.99

Table 4: **Detection performance across all benchmarked strategies.** Camelot variants underperform across all metrics, with Hybrid achieving the highest F1 score (33.23%) among them. TASER consistently achieves perfect recall and outperforms both Camelot and Table Transformer baselines, with Full Schema Prompting yielding the best precision (43.43%), F1 score (59.44%), and accuracy (66.35%).

Batch Size	Remaining TAD ($)	TAD Reduction (%)	NAV Extracted ($)
500	94,843,638	7.8%	7,993,158
250	94,185,693	8.4%	8,651,103
100	95,985,588	6.7%	7,851,209
50	92,781,421	9.8%	10,025,376
10	93,032,549	9.6%	9,804,248

Table 5: **Remaining Total Absolute Difference** (TAD, $) and Net Asset Value (NAV, $) extracted from reconciled unmatched holdings by batch size.

E Document Preprocessing

For each PDF filing, TASER extracts raw text, layout metadata, and embedded images using a hybrid pipeline based on `pdfplumber`. Each page is parsed into normalized text blocks and layout primitives, preserving spatial relationships and read order. Minimal normalization is applied, including Unicode cleanup and header/footer removal. Each page object includes:

- Raw text blocks (reading order preserved)
- Bounding boxes and font metadata
- Embedded images (if any)

We apply Unicode normalization (NFKC), whitespace collapse, and filter out repeated headers/footers via regex matching. Optionally, OCR is performed if text extraction fails. Code and parameters are available upon request.

F Parallelization and Fund Construction

Extraction: To efficiently process large, multi-page filings, TASER employs parallelization (20 workers) at both the document and page levels. Each agent operates asynchronously across document batches: Detector and Extractor agents process candidate pages in parallel, while the Recommender agent operates downstream on the resulting artifacts.

Merging: For fund-level construction, extracted tables from consecutive pages are merged deterministically. Entity resolution is performed by matching predicted fund names and table headings across pages, while units and currencies are normalized to a consistent reporting standard through a boolean flag `value_in_thousands`. Partial extractions are reconciled using strict types in the response model, whose validation errors re-prompt the LLM on specific extraction errors to ensure a unified, schema-conformant portfolio representation for each fund.

Dataset	Avg. # Tables Per Topology	Avg. Rows Per Table	Avg. Columns Per Table	Avg. Spanning Cells per Table
SciTSR	5.70	9.28	5.19	0.77
PubTabNet	4.13	14.05	5.39	2.24
FinTabNet	11.80	11.93	4.36	1.01
PubTables-1M	3.78	13.41	5.46	3.01
TASERTab	**11.00**	**53.70**	**6.36**	**2.67**

Table 6: **Complexity of table instances across datasets.** TASERTab exhibits almost five times the number of rows compared to other datasets. The maximum row count in TASERTab is 426 rows across 44 pages for a single Financial Holdings Table.

Dataset	# Tables	# Unique Cell Topologies	Avg. # Tables Per Topology	Avg. Rows Per Table	Avg. Columns Per Table	Maximum Page Span
Financial Holdings Table	1933	621	3.11	**53.7**	6.36	44
All Other Tables	1280	331	4.32	**26.9**	3.87	37

Table 7: **Complexity of Financial Holdings Tables**

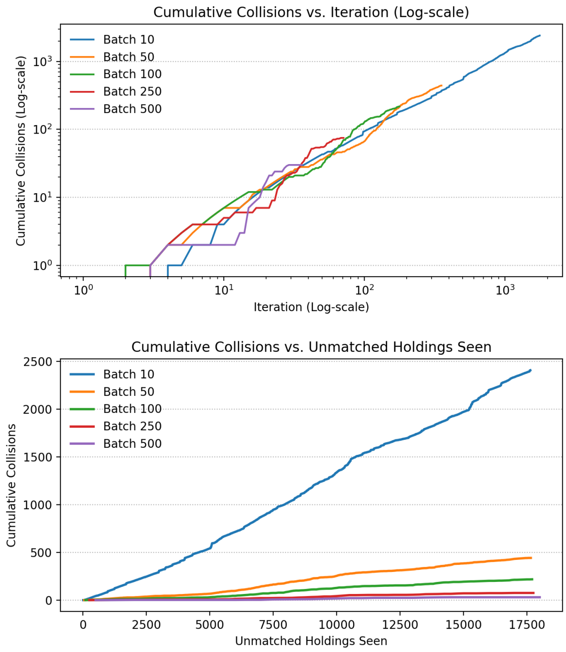


Figure 7: **Cumulative collisions per unmatched holding;** smaller batches incur more collisions, reflecting greater redundancy.

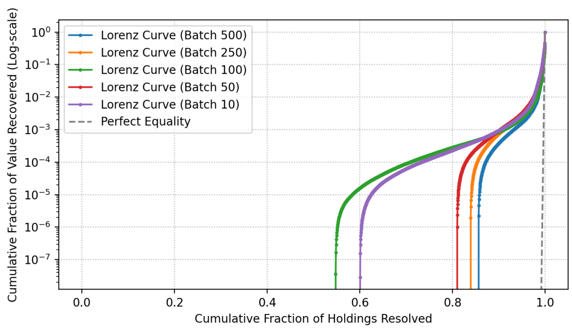


Figure 8: **Heavy-tailed distribution of value recovery from unmatched holdings across batch sizes.** We report the Lorenz curves for the cumulative fraction of value recovered as a function of the fraction of "other" holdings resolved. For all batch sizes, a small number of matches account for the vast majority of recovered net asset value, while most resolved holdings contribute negligibly. The bow of each curve away from the diagonal illustrates the extreme concentration of recoverable value in the "head," characteristic of a heavy-tailed regime.

G Schema Definitions and Portfolio Model

Portfolio Base Model: The core `Instrument` base model in our Pydantic model is subclassed into the following classes (see Figure 13 for the full class diagram). This is our initial schema composed of some of the most well-known financial instruments:

- `Equity`: a share of ownership in a corporation, representing residual claims on earnings and assets.
- `Bond`: a fixed-income security issued by governments or corporations, paying periodic coupons and returning principal at maturity.
- `Future`: an exchange-traded contract obligating the buyer or seller to transact an asset at a predetermined price on a specified future date.
- `Forward`: an over-the-counter agreement to buy or sell an underlying asset at a set price on a future date, customizable but counterparty-risky.
- `Swap`: a bilateral contract to exchange cash flows (e.g., fixed vs. floating interest rates or different currencies), with terms set at initiation.

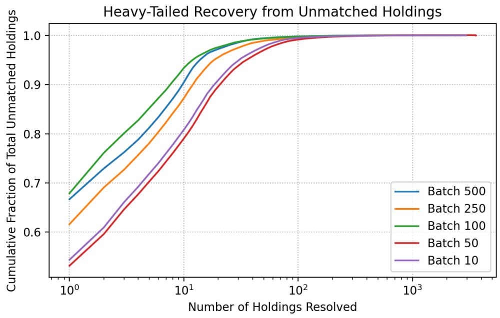


Figure 9: **Cumulative recovery fraction vs. number of holdings resolved.** Cumulative fraction of total value recovered as a function of the number of unmatched holdings resolved (log-log scale). The steep initial rise for each batch size indicates that the largest recoveries are concentrated among the first few resolved holdings; subsequently, improvement plateaus, indicating diminishing returns from resolving additional holdings in the long tail.

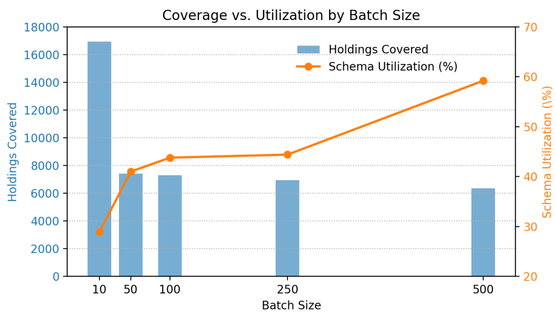


Figure 10: **Coverage vs. Utilization by Batch Size.** The number of unmatched holdings covered (bars, left axis) decreases with increasing batch size, while the fraction of schema suggestions utilized (line, right axis) increases. This highlights a tradeoff: small batches are more exhaustive in coverage, but large batches are more efficient—yielding fewer "wasted" schema suggestions.

- **`Option`**: a derivative granting the right, but not the obligation, to buy (call) or sell (put) an underlying asset at a specified strike price before or at expiry.
- **`Debt`**: a broad class of fixed-income securities including variable return notes, medium-term notes, and government bonds, not otherwise classified as standard bonds.
- **`Equity Linked Note (ELN)`**: a structured product whose returns are linked to the performance of an underlying equity or basket of equities.
- **`Other`**: a catch-all for instrument types not covered by the above classes, enabling schema extension and novelty detection.

Batch Size	Name Diversity Avg	Min	Max	Schema Diversity Avg	Min	Max
10	25.94	0	82	331.80	0	1387
50	22.67	0	78	313.41	0	1305
100	24.21	0	71	350.32	0	1569
250	22.60	0	55	342.97	0	1230
500	20.35	0	54	246.40	0	737

Table 8: Diversity metrics of unique schema suggestions for varying batch sizes. We report the average/minimum/maximum pairwise Levenshtein distance; "schema" metrics are over the entire generated schema, "name" is on the generated holding class name.

H Ablation Strategies

Raw Text Prompting. For the baseline ablation, we prompt the LLM solely with the raw page text, asking whether a portfolio table is present via a simple yes/no detection prompt. Upon affirmative detection, the LLM is instructed to extract a portfolio table from the same text, returning the result as a structured object with a `portfolio` field, but without access to any schema or structural guidance. This strategy measures the LLM's extraction performance in the absence of schema scaffolding or explicit reasoning.

Structured Chain-of-Thought (CoT). To assess the impact of explicit reasoning on table detection, we prompt the LLM with the page text and require a structured Pydantic output containing both a chain-of-thought explanation (`table_chain_of_thought`) and a boolean indicating the presence of a portfolio table (`has_portfolio_table`). This ablation isolates the effect of minimal schema guidance and encourages the model to make its decision transparent through explicit intermediate reasoning. Upon positive detection, extraction is performed identically to the baseline, without additional schema context.

Full Schema Prompting. In this ablation, we inject the complete `Portfolio` Pydantic schema directly into the detection prompt, alongside the page text. The LLM is instructed to reason about the presence of a portfolio table, outputting a chain-of-thought (`chain_of_thought`), a boolean detection (`has_portfolio_table`), and, if present, an extracted `portfolio` object conforming to the provided schema. This strategy evaluates the effect of strong schema supervision on both detection and extraction performance, requiring the model to both reason and map raw text into the structured schema within a single step.

Direct Schema Application. For the final ablation, we bypass explicit table detection and directly apply the `Portfolio` schema extraction to every page. The LLM is prompted to extract a portfolio table from the provided text and return a Pydantic object with a `portfolio` field, irrespective of any prior detection or reasoning. Extraction is considered successful if any portion of the schema can be instantiated from the text. This approach evaluates schema-constrained extraction in the absence of explicit detection or intermediate supervision.

I Aggregation and Conflict Resolution of Schema Suggestions

After the LLM returns a batch of schema suggestions, we aggregate and cluster similar proposals as follows:

1. **Deduplication**: Suggestions with Levenshtein similarity ≥ 0.9 (on class name and field structure) are merged.

2. **Clustering**: All proposals are clustered by semantic similarity of class names and required fields, using LLMs as the decision process.

3. **Selection**: For each cluster, the most frequent or most comprehensive schema suggestion is selected.

4. **Validation**: Each selected schema is validated by re-extracting unmatched holdings; suggestions that do not match any holding are dropped.

5. **Manual review**: If ambiguity remains, a manual review is triggered for final decision. We validated 64 resolved schemas for the second phase of extraction. Listing 15 displays the reconciled JSON schema for Forward Currency Contract, corresponding Pydantic model via `pydantic.create_model`, and several re-extracted holdings.

J Schema Suggestion Diversity

Table 8 summarizes the diversity among schema suggestions across batch sizes. Moderate batch sizes (100–250) achieve the highest average and maximum diversity, while the largest batch size (500) yields the lowest. This indicates that extremely large batches tend to generate more homogeneous or redundant suggestions, while moderate batches foster a broader range of candidate schemas.

K Example Holdings Tables

We show example holdings tables, alongside TASER's extractions in Figures 16 - 34.

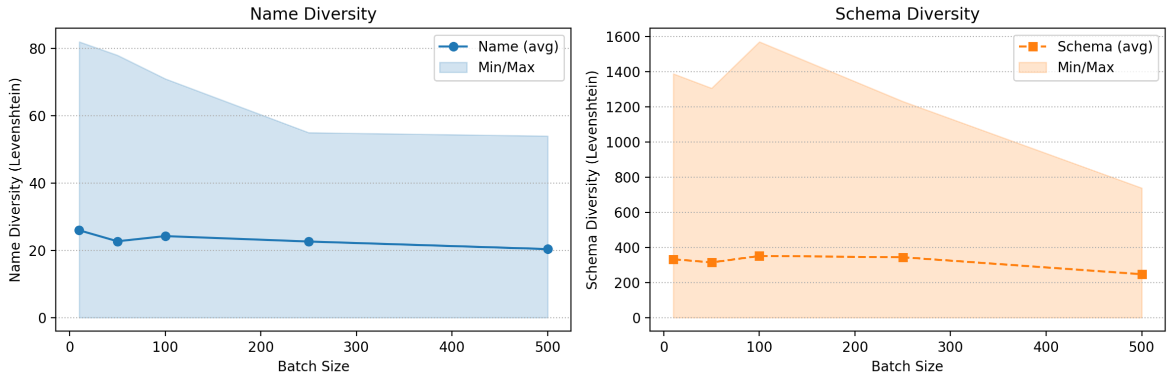


Figure 11: **Left:** Name diversity (average, minimum, and maximum pairwise Levenshtein distance) among schema suggestions for varying batch sizes. **Right:** Schema diversity for the same. Moderate batch sizes (100–250) maximize diversity, while very large batches yield more homogeneous outputs.

Batch	Coverage (Holdings)			Utilization (Schemas)			Reported	Collision
Size	Count	% Covered	NAV (%)	Total #	# Utilized	%	Collisions	Rate (%)
10	16,942	96.1	99.65	867	251	29.0	2,409	73.5
50	7,416	42.1	35.35	586	240	41.0	442	57.0
100	7,311	41.5	35.39	495	217	43.8	218	30.6
250	6,955	39.5	35.46	351	156	44.4	75	17.6
500	6,349	36.0	35.10	184	109	59.2	30	14.0

Table 9: **Schema Utilization Efficiency.** We report the proportion of generated schema suggestions that were utilized (i.e., matched at least one holding), for matching. Larger batch sizes result in a higher fraction of utilized schemas, suggesting that bulkier suggestion rounds are more efficient at targeting actionable schemas, albeit at the expense of overall diversity and coverage.

		Detection		Extraction		End to End	
	Method	Tokens	Latency (s)	Tokens	Latency (s)	Tokens	Latency (s)
TASER	(a) Raw Text Prompting	1,495	0.33	5,414	20.37	6,909	20.69
	(b) Structured CoT	1,514	1.70	5,440	20.20	6,954	21.90
	(c) Full Schema Prompting	5,706	1.58	5,235	20.48	10,941	22.07
	(d) Direct Schema Application	—	—	5,693	21.47	5,693	21.47

Table 10: **Efficiency comparison of each ablation strategy**. We report the token consumption and inference latency for detection, extraction, and end-to-end processing. Raw Text Prompting minimizes detection cost (1,495 tokens, 0.33 s) and achieves a total pipeline latency of 20.69 s; Structured CoT incurs additional reasoning overhead (1.70 s) with similar extraction performance; Full Schema Prompting uses the most detection tokens (5,706) but maintains comparable end-to-end latency (22.07 s); Direct Schema Application skips the detection stage entirely, applying schema validation directly in extraction. Dashes (—) indicate stages not performed by the method.

Instrument Category	Count	Example
Equities	28,737	Taiwan Semiconductor Manufacturing
Debt	17,105	US Treasury 4.69% 09/05/2024
Unmatched Instruments	16,822	EUR
Forwards	8,023	Bought USD Sold KRW at 0.00072513
Options	977	Written Call Unilever 4050
Futures	720	US 5 Year Bond Future
Swaps	776	Pay fixed 3.026% receive float. (1d SOFR)
ELNs	292	BNP (Laobaixing Pharm. Chain (A)) ELN 22/07/2024

Table 11: **Distribution of instrument categories** in the dataset, with an example for each.

```python
# Ablation 1: Raw Text Prompting
detection_prompt = (
    "Is there a table present in the following text? Reply with 'yes' or 'no'.\n\n"
    f"Text:\n{page.text}"
)


# TableDetectionResponse Pydantic Model
class TableDetectionResponse(BaseModel):
    table_chain_of_thought: str = Field(...,
        description="Chain of thoughts on if the page text contains table-like content")
    has_portfolio_table: bool = Field(...,
        description="True if the page has a holdings table, False otherwise")


# Ablation 2: Structured Chain-of-Thought (CoT)
detection_prompt = (
    "Analyze the following text and determine if it contains a portfolio table. "
    "Provide your chain of thought and final decision in a structured output "
    "response model that includes 'chain_of_thought' and 'has_portfolio_table' fields.\n\n"
    f"Text:\n{page.text}"
)


# Ablation 3: Full Schema Prompting
detection_prompt = (
    "Using the provided Portfolio JSON schema, analyze the following text and "
    "if it can be extracted into that schema. Provide your chain of thought. "
    "You will output a response model object including 'chain_of_thought', "
    "'has_portfolio_table', and 'extracted portfolio'.\n\n"
    f"Schema:\n{json.dumps(schema, indent=2)}\n\n"
    f"Text:\n{page.text}"
)


# Ablation 4: Direct Schema Application
detection_prompt = (
    "Extract a portfolio table from the following text following the Portfolio schema. "
    "Return a response object with a 'portfolio' field.\n\n"
    f"Text:\n{page.text}"
)
```

Figure 12: **Detection prompts for all ablation strategies.** Each section is labeled with its corresponding ablation strategy.

Listing 1: Portfolio schema with all matched instrument types.

```python
from enum import Enum
from typing import Optional, List, Literal
from pydantic import BaseModel, Field
from datetime import datetime

class BaseInstrument(BaseModel):
    cusip: Optional[str] = Field(None, description="CUSIP identifier")
    isin: Optional[str] = Field(None,
        description="International Securities Identification Number")
    ticker: Optional[str] = Field(None, description="Ticker Symbol")
    description: Optional[str] = Field(None,
        description="Description or name of the instrument")
    quantity: Optional[float] = Field(None, description="Number of units held")
    market_value: Optional[float] = Field(None, description="Market value of the holding")

class Equity(BaseInstrument):
    instrument_type: Literal["Equity"] = "Equity"
    exchange: Optional[str] = Field(None, description="Trading exchange for the equity")

class Option(BaseInstrument):
    instrument_type: Literal["Option"] = "Option"
    underlying: Optional[str] = Field(None, description="Identifier for the underlying asset")
    strike_price: Optional[float] = Field(None, description="Strike price of the option")
    expiration_date: Optional[datetime] = Field(None,
        description="Expiration date of the option")
    option_type: Optional[str] = Field(None, description="Call or Put option")

class Swap(BaseInstrument):
    instrument_type: Literal["Swap"] = "Swap"
    notional_amount: Optional[float] = Field(None, description="Notional amount of the swap")
    fixed_rate: Optional[float] = Field(None,
        description="Fixed rate component (if applicable)")
    floating_rate_index: Optional[str] = Field(None,
        description="Index used for floating rate leg")
    maturity_date: Optional[datetime] = Field(None, description="Maturity date of the swap")
    counterparty: Optional[str] = Field(None, description="The name of the counterparty")

class Forward(BaseInstrument):
    instrument_type: Literal["Forward"] = "Forward"
    forward_price: Optional[float] = Field(None, description="Agreed forward price")
    settlement_date: Optional[datetime] = Field(None,
        description="Settlement date for the forward")

class Future(BaseInstrument):
    instrument_type: Literal["Future"] = "Future"
    contract_size: Optional[int] = Field(None, description="Size of the contract")
    expiration_date: Optional[datetime] = Field(None,
        description="Expiration date of the future")

class Debt(BaseInstrument):
    instrument_type: Literal["Debt"] = "Debt"
    coupon_rate: Optional[float] = Field(None,
        description="Annual coupon rate of the debt/bond")
    maturity_date: Optional[datetime] = Field(None,
        description="Maturity date of the debt/bond")
    issuer: Optional[str] = Field(None, description="Issuer of the debt/bond")

class EquityLinkedNote(BaseInstrument):
    instrument_type: Literal["Equity Linked Note"] = "Equity Linked Note"
    issuer: Optional[str] = Field(None, description="Issuer of the ELN")
    product: Optional[str] = Field(None, description="Underlying product of the ELN")
    maturity_date: Optional[datetime] = Field(None, description="Maturity date of the ELN")
```

Listing 2: Main Portfolio Model with Unmatched (Other) Holdings class

```python
class Other(BaseModel):
    description: str = Field(...,
        description="Text of the unknown instrument.")
    name: str = Field(...,
        description="Suggested classification of the description or type")
    market_value: Optional[float] = Field(None,
        description="Market value associated with the instrument"
    )

class Portfolio(BaseModel):
    fund_name: Optional[str] = Field(None,
        description="Name of the fund that the portfolio belongs to")
    value_in_thousands: bool = Field(False,
        description="True if the market value is based on thousands")
    equities: Optional[List[Equity]] = Field(default_factory=list,
        description="List of equities")
    options: Optional[List[Option]] = Field(default_factory=list,
        description="List of options")
    swaps: Optional[List[Swap]] = Field(default_factory=list,
        description="List of swaps")
    forwards: Optional[List[Forward]] = Field(default_factory=list,
        description="List of forwards")
    futures: Optional[List[Future]] = Field(default_factory=list,
        description="List of futures")
    debt: Optional[List[Debt]] = Field(default_factory=list,
        description="List of debt instruments")
    elns: Optional[List[EquityLinkedNote]] = Field(default_factory=list,
        description="List of equity linked notes")
    other_instruments: Optional[List[Other]] = Field(default_factory=list,
        description="The list of instruments that do not match any other type")
```

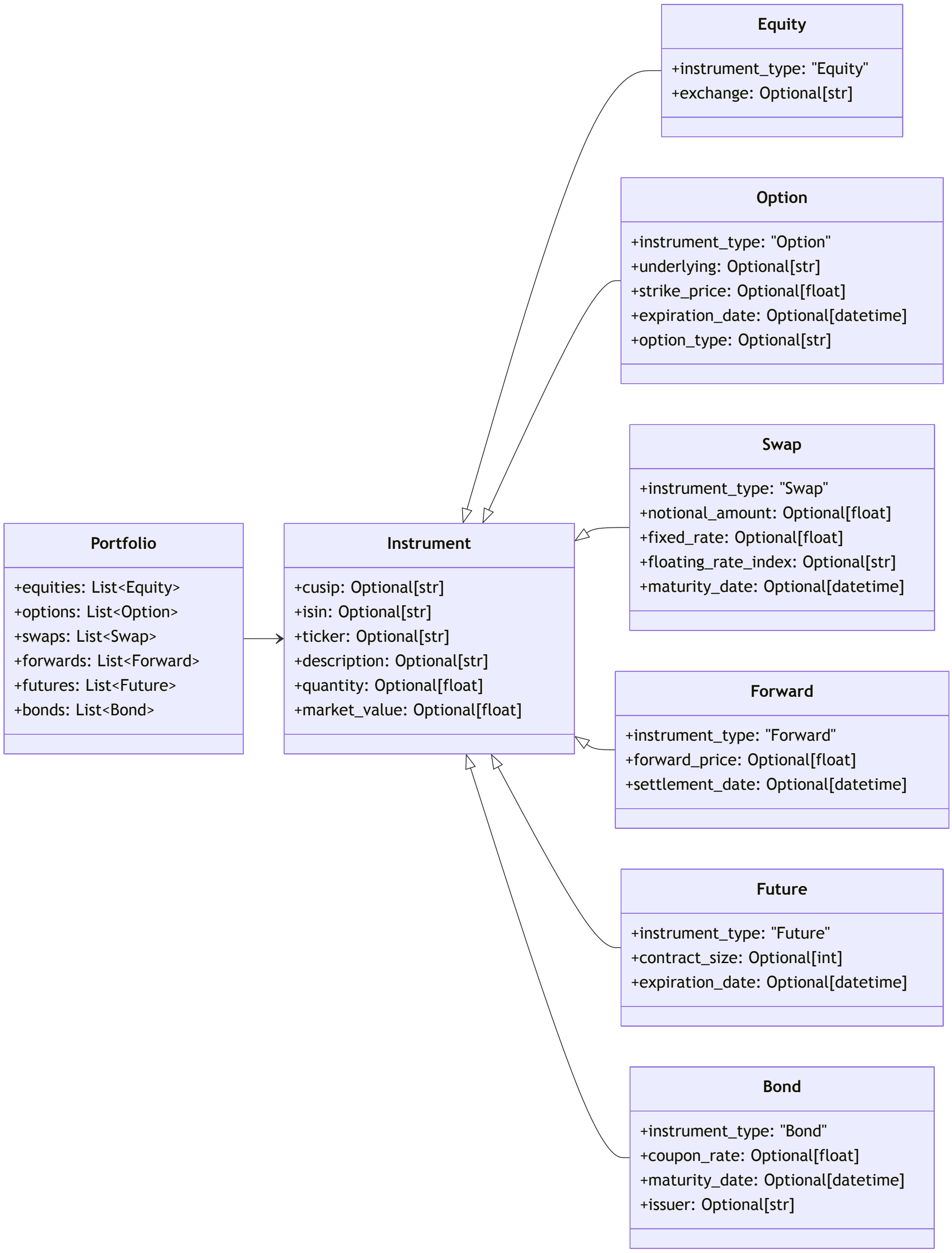


Figure 13: **Class diagram of the initial `Portfolio` schema,** showing the top-level `Portfolio` containing a collection of `Instrument` objects, each subclassed into specific security types (`Equity`, `Bond`, `Future`, `Forward`, `Swap`, `Option`) to capture their unique attributes.

```python
# Prompt template for Recommender Agent, using batch size parameterization

def recommender_agent_prompt(
    portfolio_schema: dict,
    unmatched_holdings: list,
    batch_size: int,
    start: int = 0,
    previous_suggestions: list = None
):
    return """
You are a schema refinement assistant for financial tables. Your task is:
- Review a batch of {batch_size} unmatched financial holdings.
- Given the current schema (JSON below), propose new classes or modifications so each holding
    can be classified.
- If a holding matches a previously suggested class, propose new optional fields if needed.
- Return your schema suggestions as a list of Pydantic SchemaSuggestion model objects.

Current Portfolio Schema:
{Portfolio.model_json_schema()}

Batch of unmatched holdings:
{unmatched_holdings[start : start + batch_size]}

Previously seen suggestions (optional, from prior batches):
{previous_suggestions if previous_suggestions else None}

For each unique holding, propose:
- A new schema class, or a modification to an existing class (add or refine fields).
- Specify all required and optional fields with Python type hints.
- If similar to an earlier suggestion, mark only new fields as optional.
- Provide a sample match (the original holding string).
- Output format: a Python list of SchemaSuggestion objects, as defined below.
"""


class SchemaSuggestion(BaseModel):
    name: str                 # Name of new or modified schema class
    suggested_schema: str   # JSON schema for the instrument.
    example: str              # Example instrument seen in unmatched holdings
```

Figure 14: **Recommender Agent schema suggestion prompt, output model, and example LLM response.** The agent sees a batched portion of unmatched holdings to recommend new alterations to the Portfolio schema. This prompt is batch-specific and may include `previous_suggestions` for cross-batch refinement and de-duplication.

Listing 3: Currency Forward Generated JSON Schema

```json
{
  "title": "Currency Forward",
  "type": "object",
  "properties": {
    "description": {
      "type": "string",
      "title": "Description",
      "description": "Description or name of the currency forward"
    },
    "market_value": {
      "anyOf": [
        { "type": "number" },
        { "type": "null" }
      ],
      "title": "Market Value",
      "description": "Market value of the currency forward",
      "default": null
    },
    "instrument_type": {
      "type": "string",
      "title": "Instrument Type",
      "const": "Currency Forward",
      "default": "Currency Forward"
    },
    "currency_pair": {
      "anyOf": [
        { "type": "string" },
        { "type": "null" }
      ],
      "title": "Currency Pair",
      "description": "Currency pair involved in the forward contract",
      "default": null
    },
    "forward_rate": {
      "anyOf": [
        { "type": "number" },
        { "type": "null" }
      ],
      "title": "Forward Rate",
      "description": "Agreed forward rate",
      "default": null
    },
    "settlement_date": {
      "anyOf": [
        { "type": "string", "format": "date-time" },
        { "type": "null" }
      ],
      "title": "Settlement Date",
      "description": "Settlement date for the currency forward",
      "default": null
    }
  }
}
```

Listing 4: Currency Forward Pydantic Model

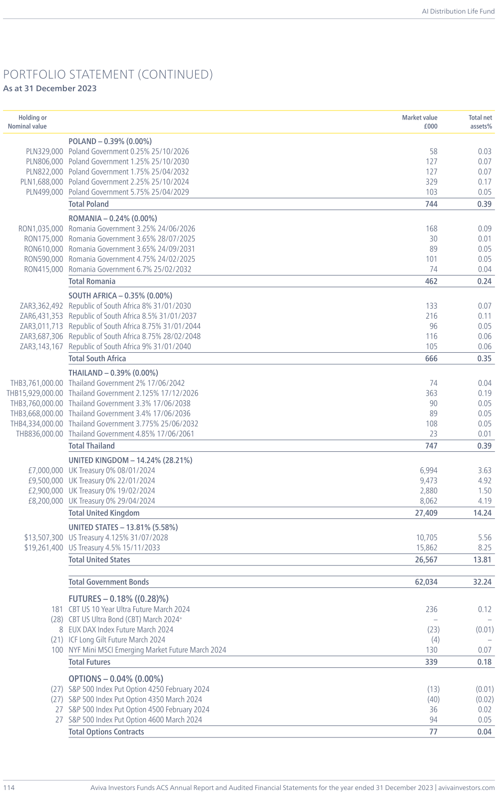
```python
class CurrencyForward(BaseModel):
    description: str
    market_value: Optional[float]
    instrument_type: str = "Currency Forward"
    currency_pair: Optional[str]
    forward_rate: Optional[float]
    settlement_date: Optional[datetime]
```

Listing 5: Refined Extraction

```
# Raw inputs
"Bought EUR Sold USD at 0.93035372 11/06/2024"
"Bought USD Sold GBP at 1.25473636 31/05/2024"
"Bought GBP Sold USD at 0.79368122 16/05/2024"

# Extracted as fields
{
  "description":"Bought EUR Sold USD at 0.93035372 11/06/2024",
  "market_value": -282515.0,
  "instrument_type": "Currency Forward",
  "currency_pair": "EUR/USD",
  "forward_rate": 0.93035372,
  "settlement_date": "2024-06-11T00:00:00"
},
{
  "description":"Bought USD Sold GBP at 1.25473636 31/05/2024",
  "market_value": 20651.0,
  "instrument_type": "Currency Forward",
  "currency_pair": "USD/GBP",
  "forward_rate": 1.25473636,
  "settlement_date": "2024-05-31T00:00:00"
},
{
  "description":"Bought GBP Sold USD at 0.79368122 16/05/2024",
  "market_value": 1429313.0,
  "instrument_type": "Currency Forward",
  "currency_pair": "GBP/USD",
  "forward_rate": 0.79368122,
  "settlement_date": "2024-05-16T00:00:00"
}
```

Figure 15: **Left:** Final `Currency Forward` JSON schema. **Top right:** Equivalent Pydantic model. **Bottom right:** Example input string and its extraction into schema fields. This demonstrates schema-driven parsing of text into structured portfolio data. A currency forward contract is a financial instrument in the foreign exchange market that locks in the price at which an entity can buy or sell a currency at a future date.



AI Distribution Life Fund

PORTFOLIO STATEMENT (CONTINUED)
As at 31 December 2023

Holding or Nominal value		Market value £000	Total net assets%
	POLAND – 0.39% (0.00%)		
PLN329,000	Poland Government 0.25% 25/10/2026	58	0.03
PLN806,000	Poland Government 1.25% 25/10/2030	127	0.07
PLN822,000	Poland Government 1.75% 25/04/2032	127	0.07
PLN1,688,000	Poland Government 2.25% 25/10/2024	329	0.17
PLN499,000	Poland Government 5.75% 25/04/2029	103	0.05
	Total Poland	**744**	**0.39**
	ROMANIA – 0.24% (0.00%)		
RON1,035,000	Romania Government 3.25% 24/06/2026	168	0.09
RON175,000	Romania Government 3.65% 28/07/2025	30	0.01
RON610,000	Romania Government 3.65% 24/09/2031	89	0.05
RON590,000	Romania Government 4.75% 24/02/2025	101	0.05
RON415,000	Romania Government 6.7% 25/02/2032	74	0.04
	Total Romania	**462**	**0.24**
	SOUTH AFRICA – 0.35% (0.00%)		
ZAR3,362,492	Republic of South Africa 8% 31/01/2030	133	0.07
ZAR6,431,353	Republic of South Africa 8.5% 31/01/2037	216	0.11
ZAR3,011,713	Republic of South Africa 8.75% 31/01/2044	96	0.05
ZAR3,687,306	Republic of South Africa 8.75% 28/02/2048	116	0.06
ZAR3,143,167	Republic of South Africa 9% 31/01/2040	105	0.06
	Total South Africa	**666**	**0.35**
	THAILAND – 0.39% (0.00%)		
THB3,761,000.00	Thailand Government 2% 17/06/2042	74	0.04
THB15,929,000.00	Thailand Government 2.125% 17/12/2026	363	0.19
THB3,760,000.00	Thailand Government 3.3% 17/06/2038	90	0.05
THB3,668,000.00	Thailand Government 3.4% 17/06/2036	89	0.05
THB4,334,000.00	Thailand Government 3.775% 25/06/2032	108	0.05
THB836,000.00	Thailand Government 4.85% 17/06/2061	23	0.01
	Total Thailand	**747**	**0.39**
	UNITED KINGDOM – 14.24% (28.21%)		
£7,000,000	UK Treasury 0% 08/01/2024	6,994	3.63
£9,500,000	UK Treasury 0% 22/01/2024	9,473	4.92
£2,900,000	UK Treasury 0% 19/02/2024	2,880	1.50
£8,200,000	UK Treasury 0% 29/04/2024	8,062	4.19
	Total United Kingdom	**27,409**	**14.24**
	UNITED STATES – 13.81% (5.58%)		
$13,507,300	US Treasury 4.125% 31/07/2028	10,705	5.56
$19,261,400	US Treasury 4.5% 15/11/2033	15,862	8.25
	Total United States	**26,567**	**13.81**
	Total Government Bonds	**62,034**	**32.24**
	FUTURES – 0.18% ((0.28)%)		
181	CBT US 10 Year Ultra Future March 2024	236	0.12
(28)	CBT US Ultra Bond (CBT) March 2024+	–	–
8	EUX DAX Index Future March 2024	(23)	(0.01)
(21)	ICF Long Gilt Future March 2024	(4)	–
100	NYF Mini MSCI Emerging Market Future March 2024	130	0.07
	Total Futures	**339**	**0.18**
	OPTIONS – 0.04% (0.00%)		
(27)	S&P 500 Index Put Option 4250 February 2024	(13)	(0.01)
(27)	S&P 500 Index Put Option 4350 March 2024	(40)	(0.02)
27	S&P 500 Index Put Option 4500 February 2024	36	0.02
27	S&P 500 Index Put Option 4600 March 2024	94	0.05
	Total Options Contracts	**77**	**0.04**

114 Aviva Investors Funds ACS Annual Report and Audited Financial Statements for the year ended 31 December 2023 | avivainvestors.com

Figure 16: Holdings Table Example 1

Description	Quantity	Market Value	Type	Coupon Rate	Maturity Date	Issuer	Debt Type
Poland Government 0.25% 25/10/2026	329000	58000	Debt	0.25	10/25/2026	Poland Government	Government Bond
Poland Government 1.25% 25/10/2030	806000	127000	Debt	1.25	10/25/2030	Poland Government	Government Bond
Poland Government 1.75% 25/04/2032	822000	127000	Debt	1.75	04/25/2032	Poland Government	Government Bond
Poland Government 2.25% 25/10/2024	1688000	329000	Debt	2.25	10/25/2024	Poland Government	Government Bond
Poland Government 5.75% 25/04/2029	499000	103000	Debt	5.75	04/25/2029	Poland Government	Government Bond
Romania Government 3.25% 24/06/2026	1035000	168000	Debt	3.25	06/24/2026	Romania Government	Government Bond
Romania Government 3.65% 28/07/2025	175000	30000	Debt	3.65	07/28/2025	Romania Government	Government Bond
Romania Government 3.65% 24/09/2031	610000	89000	Debt	3.65	09/24/2031	Romania Government	Government Bond
Romania Government 4.75% 24/02/2025	590000	101000	Debt	4.75	02/24/2025	Romania Government	Government Bond
Romania Government 6.7% 25/02/2032	415000	74000	Debt	6.7	02/25/2032	Romania Government	Government Bond
Republic of South Africa 8% 31/01/2030	3362492	133000	Debt	8	01/31/2030	Republic of South Africa	Government Bond
Republic of South Africa 8.5% 31/01/2037	6431353	216000	Debt	8.5	01/31/2037	Republic of South Africa	Government Bond
Republic of South Africa 8.75% 31/01/2044	3011713	96000	Debt	8.75	01/31/2044	Republic of South Africa	Government Bond
Republic of South Africa 8.75% 28/02/2048	3687306	116000	Debt	8.75	02/28/2048	Republic of South Africa	Government Bond
Republic of South Africa 9% 31/01/2040	3143167	105000	Debt	9	01/31/2040	Republic of South Africa	Government Bond
Thailand Government 2% 17/06/2042	3761000	74000	Debt	2	06/17/2042	Thailand Government	Government Bond
Thailand Government 2.125% 17/12/2026	15929000	363000	Debt	2.125	12/17/2026	Thailand Government	Government Bond
Thailand Government 3.3% 17/06/2038	3760000	90000	Debt	3.3	06/17/2038	Thailand Government	Government Bond
Thailand Government 3.4% 17/06/2036	3668000	89000	Debt	3.4	06/17/2036	Thailand Government	Government Bond
Thailand Government 3.775% 25/06/2032	4334000	108000	Debt	3.775	06/25/2032	Thailand Government	Government Bond
Thailand Government 4.85% 17/06/2061	836000	23000	Debt	4.85	06/17/2061	Thailand Government	Government Bond
UK Treasury 0% 08/01/2024	7000000	6994000	Debt	0	01/08/2024	UK Treasury	Government Bond
UK Treasury 0% 22/01/2024	9500000	9473000	Debt	0	01/22/2024	UK Treasury	Government Bond
UK Treasury 0% 19/02/2024	2900000	2880000	Debt	0	02/19/2024	UK Treasury	Government Bond
UK Treasury 0% 29/04/2024	8200000	8062000	Debt	0	04/29/2024	UK Treasury	Government Bond
US Treasury 4.125% 31/07/2028	13507300	10705000	Debt	4.125	07/31/2028	US Treasury	Government Bond
US Treasury 4.5% 15/11/2033	19261400	15862000	Debt	4.5	11/15/2033	US Treasury	Government Bond

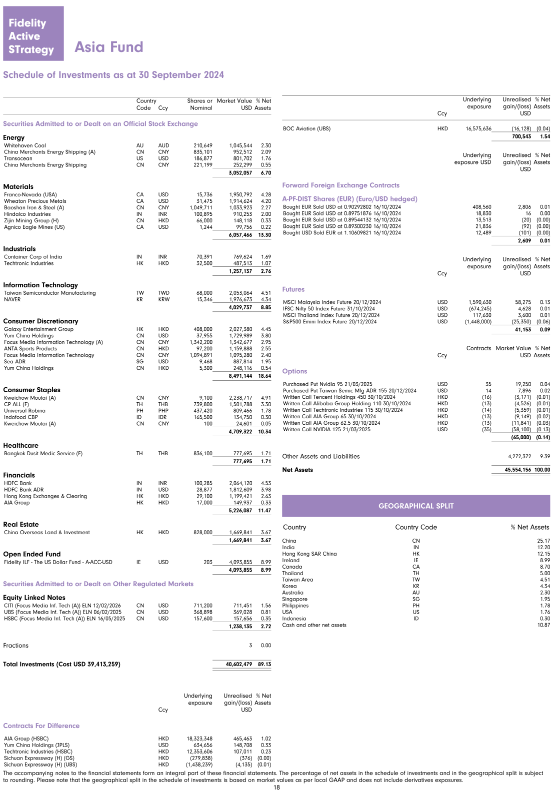
Figure 17: Debt Extracted

Description	Quantity	Market Value	Type	Expiration Date
CBT US 10 Year Ultra Future March 2024	181	236000	Future	03/01/2024
CBT US Ultra Bond (CBT) March 2024+	-28	0	Future	03/01/2024
EUX DAX Index Future March 2024	8	-23000	Future	03/01/2024
ICF Long Gilt Future March 2024	-21	-4000	Future	03/01/2024
NYF Mini MSCI Emerging Market Future March 2024	100	130000	Future	03/01/2024

Figure 18: Futures Extracted

Description	Quantity	Market Value	Type	Strike Price	Expiration Date	Option Type
S&P 500 Index Put Option 4250 February 2024	-27	-13000	Option	4250	02/01/2024	Put
S&P 500 Index Put Option 4350 March 2024	-27	-40000	Option	4350	03/01/2024	Put
S&P 500 Index Put Option 4500 February 2024	27	36000	Option	4500	02/01/2024	Put
S&P 500 Index Put Option 4600 March 2024	27	94000	Option	4600	03/01/2024	Put

Figure 19: Options Extracted



Fidelity Active STrategy

Asia Fund

Schedule of Investments as at 30 September 2024

	Country Code	Ccy	Shares or Nominal	Market Value USD	% Net Assets
Securities Admitted to or Dealt on an Official Stock Exchange					
Energy					
Whitehaven Coal	AU	AUD	210,649	1,045,544	2.30
China Merchants Energy Shipping (A)	CN	CNY	835,101	952,512	2.09
Transocean	US	USD	186,877	801,702	1.76
China Merchants Energy Shipping	CN	CNY	221,199	252,299	0.55
				3,052,057	**6.70**
Materials					
Franco-Nevada (USA)	CA	USD	15,736	1,950,792	4.28
Wheaton Precious Metals	CA	USD	31,475	1,914,624	4.20
Baoshan Iron & Steel (A)	CN	CNY	1,049,711	1,033,923	2.27
Hindalco Industries	IN	INR	100,895	910,253	2.00
Zijin Mining Group (H)	CN	HKD	66,000	148,118	0.33
Agnico Eagle Mines (US)	CA	USD	1,244	99,756	0.22
				6,057,466	**13.30**
Industrials					
Container Corp of India	IN	INR	70,391	769,624	1.69
Techtronic Industries	HK	HKD	32,500	487,513	1.07
				1,257,137	**2.76**
Information Technology					
Taiwan Semiconductor Manufacturing	TW	TWD	68,000	2,053,064	4.51
NAVER	KR	KRW	15,346	1,976,673	4.34
				4,029,737	**8.85**
Consumer Discretionary					
Galaxy Entertainment Group	HK	HKD	408,000	2,027,380	4.45
Yum China Holdings	CN	USD	37,955	1,729,989	3.80
Focus Media Information Technology (A)	CN	CNY	1,342,200	1,342,677	2.95
ANTA Sports Products	CN	HKD	97,200	1,159,888	2.55
Focus Media Information Technology	CN	CNY	1,094,891	1,095,280	2.40
Sea ADR	SG	USD	9,468	887,814	1.95
Yum China Holdings	CN	HKD	5,300	248,116	0.54
				8,491,144	**18.64**
Consumer Staples					
Kweichow Moutai (A)	CN	CNY	9,100	2,238,717	4.91
CP ALL (F)	TH	THB	739,800	1,501,788	3.30
Universal Robina	PH	PHP	437,420	809,466	1.78
Indofood CBP	ID	IDR	165,500	134,750	0.30
Kweichow Moutai (A)	CN	CNY	100	24,601	0.05
				4,709,322	**10.34**
Healthcare					
Bangkok Dusit Medic Service (F)	TH	THB	836,100	777,695	1.71
				777,695	**1.71**
Financials					
HDFC Bank	IN	INR	100,285	2,064,120	4.53
HDFC Bank ADR	IN	USD	28,877	1,812,609	3.98
Hong Kong Exchanges & Clearing	HK	HKD	29,100	1,199,421	2.63
AIA Group	HK	HKD	17,000	149,937	0.33
				5,226,087	**11.47**
Real Estate					
China Overseas Land & Investment	HK	HKD	828,000	1,669,841	3.67
				1,669,841	**3.67**
Open Ended Fund					
Fidelity ILF - The US Dollar Fund - A-ACC-USD	IE	USD	203	4,093,855	8.99
				4,093,855	**8.99**
Securities Admitted to or Dealt on Other Regulated Markets					
Equity Linked Notes					
CITI (Focus Media Inf. Tech (A)) ELN 12/02/2026	CN	USD	711,200	711,451	1.56
UBS (Focus Media Inf. Tech (A)) ELN 06/02/2025	CN	USD	368,898	369,028	0.81
HSBC (Focus Media Inf. Tech (A)) ELN 16/05/2025	CN	USD	157,600	157,656	0.35
				1,238,135	**2.72**
Fractions				3	0.00
Total Investments (Cost USD 39,413,259)				**40,602,479**	**89.13**

	Ccy	Underlying exposure	Unrealised gain/(loss) USD	% Net Assets
Contracts For Difference				
AIA Group (HSBC)	HKD	18,323,348	465,463	1.02
Yum China Holdings (JPLS)	USD	634,656	148,708	0.33
Techtronic Industries (HSBC)	HKD	12,353,606	107,011	0.23
Sichuan Expressway (H) (GS)	HKD	(279,838)	(376)	(0.00)
Sichuan Expressway (H) (UBS)	HKD	(1,438,239)	(4,135)	(0.01)
BOC Aviation (UBS)	HKD	16,575,636	(16,128)	(0.04)
			700,543	**1.54**

	Underlying exposure USD	Unrealised gain/(loss) USD	% Net Assets
Forward Foreign Exchange Contracts			
A-PF-DIST Shares (EUR) (Euro/USD hedged)			
Bought EUR Sold USD at 0.90292802 16/10/2024	408,560	2,806	0.01
Bought EUR Sold USD at 0.89751876 16/10/2024	18,830	16	0.00
Bought EUR Sold USD at 0.89544132 16/10/2024	13,513	(20)	(0.00)
Bought EUR Sold USD at 0.89300230 16/10/2024	21,836	(92)	(0.00)
Bought USD Sold EUR at 1.10609821 16/10/2024	12,489	(101)	(0.00)
		2,609	**0.01**

	Ccy	Underlying exposure	Unrealised gain/(loss) USD	% Net Assets
Futures				
MSCI Malaysia Index Future 20/12/2024	USD	1,590,630	58,275	0.13
IFSC Nifty 50 Index Future 31/10/2024	USD	(674,245)	4,628	0.01
MSCI Thailand Index Future 20/12/2024	USD	117,630	3,600	0.01
S&P500 Emini Index Future 20/12/2024	USD	(1,448,000)	(25,350)	(0.06)
			41,153	**0.09**

	Ccy	Contracts	Market Value USD	% Net Assets
Options				
Purchased Put Nvidia 95 21/03/2025	USD	35	19,250	0.04
Purchased Put Taiwan Semic Mfg ADR 155 20/12/2024	USD	14	7,896	0.02
Written Call Tencent Holdings 450 30/10/2024	HKD	(16)	(3,171)	(0.01)
Written Call Alibaba Group Holding 110 30/10/2024	HKD	(13)	(4,526)	(0.01)
Written Call Techtronic Industries 115 30/10/2024	HKD	(14)	(5,359)	(0.01)
Written Call AIA Group 65 30/10/2024	HKD	(13)	(9,149)	(0.02)
Written Call AIA Group 62.5 30/10/2024	HKD	(13)	(11,841)	(0.03)
Written Call NVIDIA 125 21/03/2025	USD	(35)	(58,100)	(0.13)
			(65,000)	**(0.14)**
Other Assets and Liabilities			4,272,372	9.39
Net Assets			**45,554,156**	**100.00**

GEOGRAPHICAL SPLIT

Country	Country Code	% Net Assets
China	CN	25.17
India	IN	12.20
Hong Kong SAR China	HK	12.15
Ireland	IE	8.99
Canada	CA	8.70
Thailand	TH	5.00
Taiwan Area	TW	4.51
Korea	KR	4.34
Australia	AU	2.30
Singapore	SG	1.95
Philippines	PH	1.78
USA	US	1.76
Indonesia	ID	0.30
Cash and other net assets		10.87

The accompanying notes to the financial statements form an integral part of these financial statements. The percentage of net assets in the schedule of investments and in the geographical split is subject to rounding. Please note that the geographical split in the schedule of investments is based on market values as per local GAAP and does not include derivatives exposures.

18

Figure 20: Holdings Table Example 2

Description	Quantity	Market Value	Type	Exch.
Whitehaven Coal	210649	1045544	Equity	AU
China Merchants Energy Shipping (A)	835101	952512	Equity	CN
Transocean	186877	801702	Equity	US
China Merchants Energy Shipping	221199	252299	Equity	CN
Franco-Nevada (USA)	15736	1950792	Equity	CA
Wheaton Precious Metals	31475	1914624	Equity	CA
Baoshan Iron & Steel (A)	1049711	1033923	Equity	CN
Hindalco Industries	100895	910253	Equity	IN
Zijin Mining Group (H)	66000	148118	Equity	CN
Agnico Eagle Mines (US)	1244	99756	Equity	CA
Container Corp of India	70391	769624	Equity	IN
Techtronic Industries	32500	487513	Equity	HK
Taiwan Semiconductor Manufacturing	68000	2053064	Equity	TW
NAVER	15346	1976673	Equity	KR
Galaxy Entertainment Group	408000	2027380	Equity	HK
Yum China Holdings	37955	1729989	Equity	CN
Focus Media Information Technology (A)	1342200	1342677	Equity	CN
ANTA Sports Products	97200	1159888	Equity	CN
Focus Media Information Technology	1094891	1095280	Equity	CN
Sea ADR	9468	887814	Equity	SG
Yum China Holdings	5300	248116	Equity	CN
Kweichow Moutai (A)	9100	2238717	Equity	CN
CP ALL (F)	739800	1501788	Equity	TH
Universal Robina	437420	809466	Equity	PH
Indofood CBP	165500	134750	Equity	ID
Kweichow Moutai (A)	100	24601	Equity	CN
Bangkok Dusit Medic Service (F)	836100	777695	Equity	TH
HDFC Bank	100285	2064120	Equity	IN
HDFC Bank ADR	28877	1812609	Equity	IN
Hong Kong Exchanges & Clearing	29100	1199421	Equity	HK
AIA Group	17000	149937	Equity	HK
China Overseas Land & Investment	828000	1669841	Equity	HK

Figure 21: Equities Extracted

Description	Quantity	Market Value	Type	Issuer	Product	Maturity Date
CITI (Focus Media Inf. Tech (A)) ELN 12/02/2026	711200	711451	Equity Linked Note	CITI	Focus Media Inf. Tech (A)	12/02/2026
UBS (Focus Media Inf. Tech (A)) ELN 06/02/2025	368898	369028	Equity Linked Note	UBS	Focus Media Inf. Tech (A)	02/06/2025
HSBC (Focus Media Inf. Tech (A)) ELN 16/05/2025	157600	157656	Equity Linked Note	HSBC	Focus Media Inf. Tech (A)	05/16/2026

Figure 22: ELNs Extracted

Description	Quantity	Market Value	Type	Forward Price	Settlement Date
Bought EUR Sold USD at 0.90292802 16/10/2024	408560	2806	Forward	0.90292802	10/16/2024
Bought EUR Sold USD at 0.89751876 16/10/2024	18830	16	Forward	0.89751876	10/16/2024
Bought EUR Sold USD at 0.89544132 16/10/2024	13513	-20	Forward	0.89544132	10/16/2024
Bought EUR Sold USD at 0.89300230 16/10/2024	21836	-92	Forward	0.89300230	10/16/2024
Bought USD Sold EUR at 1.10609821 16/10/2024	12489	-101	Forward	1.10609821	10/16/2024

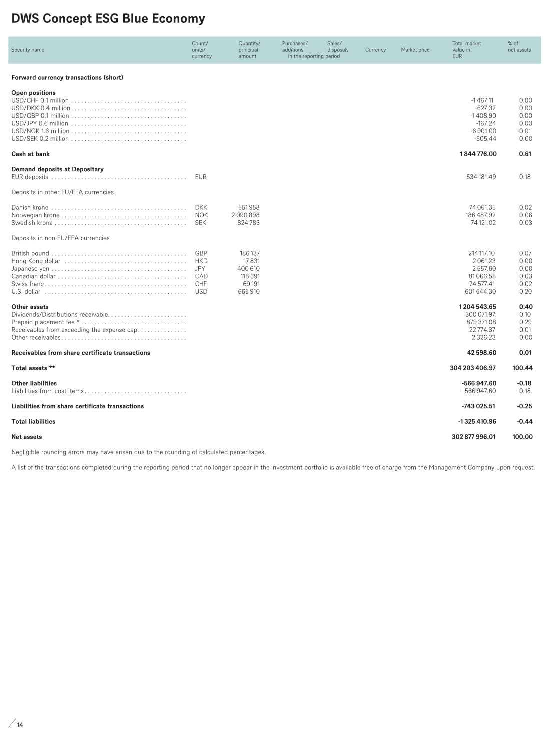
Figure 23: Forwards Extracted

Description	Quantity	Market Value	Type	Contract Size	Expiration Date
MSCI Malaysia Index Future 20/12/2024	1590630	58275	Future	1590630	12/20/2024
IFSC Nifty 50 Index Future 31/10/2024	-674245	4628	Future	674245	10/31/2024
MSCI Thailand Index Future 20/12/2024	117630	3600	Future	117630	12/20/2024
S&P500 Emini Index Future 20/12/2024	-1448000	-25350	Future	1448000	12/20/2024

Figure 24: Futures Extracted

Description	Quantity	Market Value	Type	Underlying	Strike Price	Expiration Date	Option Type	Ticker
Purchased Put Nvidia 95 21/03/2025	35	19250	Option	Nvidia	95	03/21/2025	Put	Nvidia
Purchased Put Taiwan Semic Mfg ADR 155 20/12/2024	14	7896	Option	Taiwan Semic Mfg ADR	155	12/20/2024	Put	Taiwan Semic Mfg ADR
Written Call Tencent Holdings 450 30/10/2024	-16	-3171	Option	Tencent Holdings	450	10/30/2024	Call	Tencent Holdings
Written Call Alibaba Group Holding 110 30/10/2024	-13	-4526	Option	Alibaba Group Holding	110	10/30/2024	Call	Alibaba Group Holding
Written Call Techtronic Industries 115 30/10/2024	-14	-5359	Option	Techtronic Industries	115	10/30/2024	Call	Techtronic Industries
Written Call AIA Group 65 30/10/2024	-13	-9149	Option	AIA Group	65	10/30/2024	Call	AIA Group
Written Call AIA Group 62.5 30/10/2024	-13	-11841	Option	AIA Group	62.5	10/30/2024	Call	AIA Group
Written Call NVIDIA 125 21/03/2025	-35	-58100	Option	NVIDIA	125	03/21/2025	Call	NVIDIA

Figure 25: Options Extracted



DWS Concept ESG Blue Economy

Security name	Count/ units/ currency	Quantity/ principal amount	Purchases/ additions (in the reporting period)	Sales/ disposals (in the reporting period)	Currency	Market price	Total market value in EUR	% of net assets
Forward currency transactions (short)								
Open positions								
USD/CHF 0.1 million							-1 467.11	0.00
USD/DKK 0.4 million							-627.32	0.00
USD/GBP 0.1 million							-1 408.90	0.00
USD/JPY 0.6 million							-167.24	0.00
USD/NOK 1.6 million							-6 901.00	-0.01
USD/SEK 0.2 million							-505.44	0.00
Cash at bank							**1 844 776.00**	**0.61**
Demand deposits at Depositary								
EUR deposits	EUR						534 181.49	0.18
Deposits in other EU/EEA currencies								
Danish krone	DKK	551 958					74 061.35	0.02
Norwegian krone	NOK	2 090 898					186 487.92	0.06
Swedish krona	SEK	824 783					74 121.02	0.03
Deposits in non-EU/EEA currencies								
British pound	GBP	186 137					214 117.10	0.07
Hong Kong dollar	HKD	17 831					2 061.23	0.00
Japanese yen	JPY	400 610					2 557.60	0.00
Canadian dollar	CAD	118 691					81 066.58	0.03
Swiss franc	CHF	69 191					74 577.41	0.02
U.S. dollar	USD	665 910					601 544.30	0.20
Other assets							**1 204 543.65**	**0.40**
Dividends/Distributions receivable							300 071.97	0.10
Prepaid placement fee *							879 371.08	0.29
Receivables from exceeding the expense cap							22 774.37	0.01
Other receivables							2 326.23	0.00
Receivables from share certificate transactions							**42 598.60**	**0.01**
Total assets **							**304 203 406.97**	**100.44**
Other liabilities							**-566 947.60**	**-0.18**
Liabilities from cost items							-566 947.60	-0.18
Liabilities from share certificate transactions							**-743 025.51**	**-0.25**
Total liabilities							**-1 325 410.96**	**-0.44**
Net assets							**302 877 996.01**	**100.00**

Negligible rounding errors may have arisen due to the rounding of calculated percentages.

A list of the transactions completed during the reporting period that no longer appear in the investment portfolio is available free of charge from the Management Company upon request.

/ 14

Figure 26: Holdings Table Example 3

Description	Market Value	Type
USD/CHF 0.1 million	-1467.11	Forward
USD/DKK 0.4 million	-627.32	Forward
USD/GBP 0.1 million	-1408.9	Forward
USD/JPY 0.6 million	-167.24	Forward
USD/NOK 1.6 million	-6901	Forward
USD/SEK 0.2 million	-505.44	Forward

Figure 27: Forwards Extracted

Description	Type	Market Value
Cash at bank	Other	1844776
Demand deposits at Depositary - EUR deposits	Other	534181.49
Deposits in other EU/EEA currencies - Danish krone	Other	74061.35
Deposits in other EU/EEA currencies - Norwegian krone	Other	186487.92
Deposits in other EU/EEA currencies - Swedish krona	Other	74121.02
Deposits in non-EU/EEA currencies - British pound	Other	214117.1
Deposits in non-EU/EEA currencies - Hong Kong dollar	Other	2061.23
Deposits in non-EU/EEA currencies - Japanese yen	Other	2557.6
Deposits in non-EU/EEA currencies - Canadian dollar	Other	81066.58
Deposits in non-EU/EEA currencies - Swiss franc	Other	74577.41
Deposits in non-EU/EEA currencies - U.S. dollar	Other	601544.3
Dividends/Distributions receivable	Other	300071.97
Prepaid placement fee	Other	879371.08
Receivables from exceeding the expense cap	Other	22774.37
Other receivables	Other	2326.23
Receivables from share certificate transactions	Other	42598.6
Liabilities from cost items	Other	-566947.6
Liabilities from share certificate transactions	Other	-743025.51

Figure 28: Other Instruments Extracted

Strategic report | Governance | Financial statements | Additional information

A snapshot of our portfolio continued

Investment portfolio as at 30 September 2024

Ranking 2024	2023	Company	Sector	Country of listing	Valuation 2024 £'000	% of portfolio
1	2	Novo Nordisk	Pharmaceuticals and Biotechnology	Denmark	41,930	6.07
2	–	ASML	Technology Hardware and Equipment	Netherlands	34,798	5.03
3	17	SAP	Software and Computer Services	Germany	31,184	4.51
4	4	TotalEnergies	Oil, Gas and Coal	France	24,622	3.56
5	19	Siemens	General Industrials	Germany	23,147	3.35
6	–	UniCredit	Banks	Italy	23,012	3.33
7	25	Deutsche Boerse	Investment Banking and Brokerage Services	Germany	19,999	2.89
8	–	Munich Re	Non-life Insurance	Germany	19,512	2.82
9	24	Anheuser-Busch InBev	Beverages	Belgium	18,885	2.73
10	–	CRH	Construction and Materials	Ireland	18,877	2.73
11	–	Sanofi	Pharmaceuticals and Biotechnology	France	18,857	2.73
12	–	BNP Paribas	Banks	France	17,815	2.58
13	10	Schneider Electric	Electronic and Electrical Equipment	France	17,192	2.49
14	–	Novartis	Pharmaceuticals and Biotechnology	Switzerland	17,036	2.46
15	–	Alcon	Medical Equipment and Services	Switzerland	16,833	2.43
16	6	Safran	Aerospace and Defence	France	16,448	2.38
17	–	National Grid	Gas, Water and Multi-utilities	United Kingdom	16,195	2.34
18	7	Airbus	Aerospace and Defence	France	15,578	2.25
19	–	Compass	Travel and Leisure	United Kingdom	15,493	2.24
20	8	LVMH Moët Hennessy Louis Vuitton	Personal Goods	France	15,141	2.19
21	34	Infineon	Technology Hardware and Equipment	Germany	14,905	2.15
22	–	SGS	Industrial Support Services	Switzerland	14,327	2.07
23	35	ASM International	Technology Hardware and Equipment	Netherlands	13,782	1.99
24	–	Cellnex Telecom	Telecommunications Service Providers	Spain	12,982	1.88
25	–	Aena	Industrial Transportation	Spain	12,179	1.76
26	–	KONE	Industrial Engineering	Finland	12,163	1.76
27	–	Ryanair	Travel and Leisure	Ireland	11,974	1.73
28	16	Danone	Food Producers	France	11,697	1.69
29	–	Roche	Pharmaceuticals and Biotechnology	Switzerland	11,592	1.68
30	9	Holcim	Construction and Materials	Switzerland	11,387	1.65
31	–	British Land	Real Estate Investment Trusts	United Kingdom	11,252	1.63
32	–	Smurfit Westrock	General Industrials	Ireland	11,087	1.60
33	–	BAWAG	Banks	Austria	10,607	1.54
34	13	Adidas	Personal Goods	Germany	10,392	1.50
35	–	Beiersdorf	Personal Care, Drug and Grocery Stores	Germany	10,390	1.50
36	–	VAT Group	Electronic and Electrical Equipment	Switzerland	10,143	1.47
37	5	Saint-Gobain	Construction and Materials	France	9,829	1.42
38	–	DSV	Industrial Transportation	Denmark	9,267	1.34
39	–	Syensqo	Chemicals	Belgium	8,991	1.30
40	–	Anglo American	Industrial Metals and Mining	United Kingdom	8,085	1.17
41	–	Hermès	Personal Goods	France	7,529	1.09
42	–	Rheinmetall	Aerospace and Defence	Germany	7,456	1.08
43	–	Nestlé	Food Producers	Switzerland	7,222	1.04
44	–	Galderma	Pharmaceuticals and Biotechnology	Switzerland	7,028	1.02
45	–	Bayer	Pharmaceuticals and Biotechnology	Germany	6,988	1.01
46	–	Stellantis	Automobiles and Parts	Netherlands	5,689	0.82
		Total listed equity investments at fair value			**691,497**	**100.00**

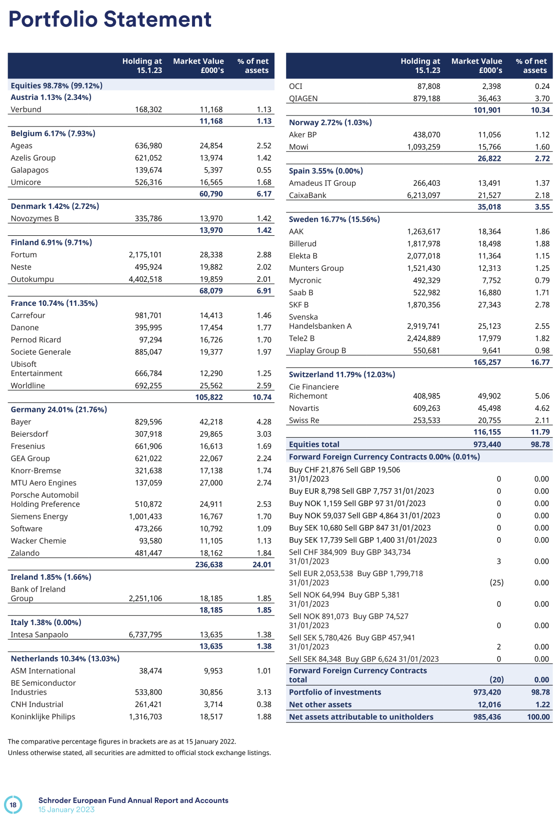
The number of stocks held may increase above 45 for a limited time period if necessary to enable operational settlement of sales and purchases in the portfolio. For this reason the portfolio temporarily held 46 stocks as at 30 September 2024.

Henderson European Trust plc Annual Report 2024 13

Figure 29: Holdings Table Example 4

Description	Market Value	Type
Novo Nordisk	41930000	Equity
ASML	34798000	Equity
SAP	31184000	Equity
TotalEnergies	24622000	Equity
Siemens	23147000	Equity
UniCredit	23012000	Equity
Deutsche Boerse	19999000	Equity
Munich Re	19512000	Equity
Anheuser-Busch InBev	18885000	Equity
CRH	18877000	Equity
Sanofi	18857000	Equity
BNP Paribas	17815000	Equity
Schneider Electric	17192000	Equity
Novartis	17036000	Equity
Alcon	16833000	Equity
Safran	16448000	Equity
National Grid	16195000	Equity
Airbus	15578000	Equity
Compass	15493000	Equity
LVMH Moët Hennessy Louis Vuitton	15141000	Equity
Infineon	14905000	Equity
SGS	14327000	Equity
ASM International	13782000	Equity
Cellnex Telecom	12982000	Equity
Aena	12179000	Equity
KONE	12163000	Equity
Ryanair	11974000	Equity
Danone	11697000	Equity
Roche	11592000	Equity
Holcim	11387000	Equity
British Land	11252000	Equity
Smurfit Westrock	11087000	Equity
BAWAG	10607000	Equity
Adidas	10392000	Equity
Beiersdorf	10390000	Equity
VAT Group	10143000	Equity
Saint-Gobain	9829000	Equity
DSV	9267000	Equity
Syensqo	8991000	Equity
Anglo American	8085000	Equity
Hermès	7529000	Equity
Rheinmetall	7456000	Equity
Nestlé	7222000	Equity
Galderma	7028000	Equity
Bayer	6988000	Equity
Stellantis	5689000	Equity

Figure 30: Equities Extracted



Portfolio Statement

	Holding at 15.1.23	Market Value £000's	% of net assets
Equities 98.78% (99.12%)			
Austria 1.13% (2.34%)			
Verbund	168,302	11,168	1.13
		11,168	**1.13**
Belgium 6.17% (7.93%)			
Ageas	636,980	24,854	2.52
Azelis Group	621,052	13,974	1.42
Galapagos	139,674	5,397	0.55
Umicore	526,316	16,565	1.68
		60,790	**6.17**
Denmark 1.42% (2.72%)			
Novozymes B	335,786	13,970	1.42
		13,970	**1.42**
Finland 6.91% (9.71%)			
Fortum	2,175,101	28,338	2.88
Neste	495,924	19,882	2.02
Outokumpu	4,402,518	19,859	2.01
		68,079	**6.91**
France 10.74% (11.35%)			
Carrefour	981,701	14,413	1.46
Danone	395,995	17,454	1.77
Pernod Ricard	97,294	16,726	1.70
Societe Generale	885,047	19,377	1.97
Ubisoft Entertainment	666,784	12,290	1.25
Worldline	692,255	25,562	2.59
		105,822	**10.74**
Germany 24.01% (21.76%)			
Bayer	829,596	42,218	4.28
Beiersdorf	307,918	29,865	3.03
Fresenius	661,906	16,613	1.69
GEA Group	621,022	22,067	2.24
Knorr-Bremse	321,638	17,138	1.74
MTU Aero Engines	137,059	27,000	2.74
Porsche Automobil Holding Preference	510,872	24,911	2.53
Siemens Energy	1,001,433	16,767	1.70
Software	473,266	10,792	1.09
Wacker Chemie	93,580	11,105	1.13
Zalando	481,447	18,162	1.84
		236,638	**24.01**
Ireland 1.85% (1.66%)			
Bank of Ireland Group	2,251,106	18,185	1.85
		18,185	**1.85**
Italy 1.38% (0.00%)			
Intesa Sanpaolo	6,737,795	13,635	1.38
		13,635	**1.38**
Netherlands 10.34% (13.03%)			
ASM International	38,474	9,953	1.01
BE Semiconductor Industries	533,800	30,856	3.13
CNH Industrial	261,421	3,714	0.38
Koninklijke Philips	1,316,703	18,517	1.88
OCI	87,808	2,398	0.24
QIAGEN	879,188	36,463	3.70
		101,901	**10.34**
Norway 2.72% (1.03%)			
Aker BP	438,070	11,056	1.12
Mowi	1,093,259	15,766	1.60
		26,822	**2.72**
Spain 3.55% (0.00%)			
Amadeus IT Group	266,403	13,491	1.37
CaixaBank	6,213,097	21,527	2.18
		35,018	**3.55**
Sweden 16.77% (15.56%)			
AAK	1,263,617	18,364	1.86
Billerud	1,817,978	18,498	1.88
Elekta B	2,077,018	11,364	1.15
Munters Group	1,521,430	12,313	1.25
Mycronic	492,329	7,752	0.79
Saab B	522,982	16,880	1.71
SKF B	1,870,356	27,343	2.78
Svenska Handelsbanken A	2,919,741	25,123	2.55
Tele2 B	2,424,889	17,979	1.82
Viaplay Group B	550,681	9,641	0.98
		165,257	**16.77**
Switzerland 11.79% (12.03%)			
Cie Financiere Richemont	408,985	49,902	5.06
Novartis	609,263	45,498	4.62
Swiss Re	253,533	20,755	2.11
		116,155	**11.79**
Equities total		**973,440**	**98.78**
Forward Foreign Currency Contracts 0.00% (0.01%)			
Buy CHF 21,876 Sell GBP 19,506 31/01/2023		0	0.00
Buy EUR 8,798 Sell GBP 7,757 31/01/2023		0	0.00
Buy NOK 1,159 Sell GBP 97 31/01/2023		0	0.00
Buy NOK 59,037 Sell GBP 4,864 31/01/2023		0	0.00
Buy SEK 10,680 Sell GBP 847 31/01/2023		0	0.00
Buy SEK 17,739 Sell GBP 1,400 31/01/2023		0	0.00
Sell CHF 384,909 Buy GBP 343,734 31/01/2023		3	0.00
Sell EUR 2,053,538 Buy GBP 1,799,718 31/01/2023		(25)	0.00
Sell NOK 64,994 Buy GBP 5,381 31/01/2023		0	0.00
Sell NOK 891,073 Buy GBP 74,527 31/01/2023		0	0.00
Sell SEK 5,780,426 Buy GBP 457,941 31/01/2023		2	0.00
Sell SEK 84,348 Buy GBP 6,624 31/01/2023		0	0.00
Forward Foreign Currency Contracts total		**(20)**	**0.00**
Portfolio of investments		**973,420**	**98.78**
Net other assets		**12,016**	**1.22**
Net assets attributable to unitholders		**985,436**	**100.00**

The comparative percentage figures in brackets are as at 15 January 2022.
Unless otherwise stated, all securities are admitted to official stock exchange listings.

18 Schroder European Fund Annual Report and Accounts
15 January 2023

Figure 31: Holdings Table Example 5

Description	Quantity	Market Value	Type
Verbund	168302	11168000	Equity
Ageas	636980	24854000	Equity
Azelis Group	621052	13974000	Equity
Galapagos	139674	5397000	Equity
Umicore	526316	16565000	Equity
Novozymes B	335786	13970000	Equity
Fortum	2175101	28338000	Equity
Neste	495924	19882000	Equity
Outokumpu	4402518	19859000	Equity
Carrefour	981701	14413000	Equity
Danone	395995	17454000	Equity
Pernod Ricard	97294	16726000	Equity
Societe Generale	885047	19377000	Equity
Ubisoft Entertainment	666784	12290000	Equity
Worldline	692255	25562000	Equity
Bayer	829596	42218000	Equity
Beiersdorf	307918	29865000	Equity
Fresenius	661906	16613000	Equity
GEA Group	621022	22067000	Equity
Knorr-Bremse	321638	17138000	Equity
MTU Aero Engines	137059	27000000	Equity
Porsche Automobil Holding Preference	510872	24911000	Equity
Siemens Energy	1001433	16767000	Equity
Software	473266	10792000	Equity
Wacker Chemie	93580	11105000	Equity
Zalando	481447	18162000	Equity
Bank of Ireland Group	2251106	18185000	Equity
Intesa Sanpaolo	6737795	13635000	Equity
ASM International	38474	9953000	Equity
BE Semiconductor Industries	533800	30856000	Equity
CNH Industrial	261421	3714000	Equity
Koninklijke Philips	1316703	18517000	Equity
OCI	87808	2398000	Equity
QIAGEN	879188	36463000	Equity
Aker BP	438070	11056000	Equity
Mowi	1093259	15766000	Equity
Amadeus IT Group	266403	13491000	Equity
CaixaBank	6213097	21527000	Equity
AAK	1263617	18364000	Equity
Billerud	1817978	18498000	Equity
Elekta B	2077018	11364000	Equity
Munters Group	1521430	12313000	Equity
Mycronic	492329	7752000	Equity
Saab B	522982	16880000	Equity
SKF B	1870356	27343000	Equity
Svenska Handelsbanken A	2919741	25123000	Equity
Tele2 B	2424889	17979000	Equity
Viaplay Group B	550681	9641000	Equity
Cie Financiere Richemont	408985	49902000	Equity
Novartis	609263	45498000	Equity
Swiss Re	253533	20755000	Equity

Figure 32: Equities Extracted

Description	Type	Market Value
Net other assets	Other	12016000

Figure 33: Other Instruments Extracted

DWS Invest (IE) ICAV

DWS Customised Global Investment Grade Bond Fund*

PORTFOLIO OF INVESTMENTS (Unaudited)(continued)

As at 31 December 2024

No. of Shares	Security	Fair Value USD	Net Assets %
Transferable securities (continued)			
Corporate Bonds (continued)			
United States (continued)			
700,000	3.908% Wells Fargo & Co. 25/04/2026	697,829	0.88
600,000	5.350% Zimmer Biomet Holdings, Inc. 01/12/2028	609,607	0.77
		30,602,469	**38.52**
Total corporate bonds		**75,998,420**	**95.65**
Total transferable securities		**75,998,420**	**95.65**

Financial derivative instruments

Futures contracts

	Broker	Notional	Unrealised Gain USD	Net Assets %
(35) of US 10 Years Note Short futures contracts Expiring 20 March 2025	Deutsche Bank AG	(3,869,414)	55,508	0.07
(17) of US 5 Years Note Short futures contracts Expiring 31 March 2025	Deutsche Bank AG	(1,820,727)	11,953	0.02
Unrealised gain on futures contracts			**67,461**	**0.09**
Total futures contracts			**67,461**	**0.09**

Forward Foreign Exchange Contracts

Currency buy	Buy amount	Currency sell	Sell amount	Counterparty	Contract date	Unrealised Gain USD	Net Assets %
USD	2,190,379	CAD	3,070,762	Royal Bank of Canada	31/01/2025	53,468	0.07
USD	24,869,939	EUR	23,466,061	Deutsche Bank AG	31/01/2025	482,874	0.61
USD	3,559,616	GBP	2,811,466	Deutsche Bank AG	31/01/2025	43,213	0.05
Unrealised gain on forwards contracts						**579,555**	**0.73**
Total forward foreign exchange contracts						**579,555**	**0.73**
Total financial derivative instruments						**647,016**	**0.82**

133

Figure 34: Holdings Table Example 6

Description	Quantity	Market Value	Type	Coupon Rate	Maturity Date	Issuer
CBT US 10 Year Ultra Future March 2024	181	236000	Future	4250	02/01/2024	Poland Government
CBT US Ultra Bond (CBT) March 2024+	-28	0	Future	4250	03/01/2024	Poland Government
EUX DAX Index Future March 2024	8	-23000	Future	4350	03/01/2024	
ICF Long Gilt Future March 2024	-21	-4000	Future	4500	02/01/2024	
NYF Mini MSCI Emerging Market Future March 2024	100	130000	Future	4600	03/01/2024	

Figure 35: Debt Extracted

Description	Quantity	Market Value	Type	Expiration Date
US 10 Years Note Short futures contracts	-35	55508	Future	02/03/2025
US 5 Years Note Short futures contracts	-17	11953	Future	31/03/2025

Figure 36: Futures Extracted

Description	Quantity	Market Value	Type Type	Settlement Date
USD/CAD Forward Contract	1	53468	Forward	31/01/2025
USD/EUR Forward Contract	1	482874	Forward	31/01/2025
USD/GBP Forward Contract	1	43213	Forward	31/01/2025

Figure 37: Forwards Extracted