CCLUPE: Benchmark for Credit Context Log Understanding and Prediction Evaluation

Anonymous ACL submission

Abstract

Credit underwriting often depends on long transaction logs, yet existing financial benchmarks rarely test whether models can recover underwriting-relevant signals from them. We introduce Credit Context Log Understanding and Prediction Evaluation (CCLUPE) as a benchmark in which models must reason over textual credit context, structured account fields, and time-series transaction sequences to identify behavioral patterns, liquidity signals, and risk indicators. CCLUPE contains 4,062 expert-validated samples spanning personal and small micro-enterprise clients, 7 knowledge domains, 16 fine-grained sub-domains, and 4 cognitive levels. All the released transaction logs are calibrated to authentic underwriting patterns via constrained generation, schema validation, KL screening, and expert review, enabling public release without exposing production data. We introduce LogScore as an evaluation metric that penalizes unsupported risk selections while balancing performance across domains and reasoning levels. Evaluating various proprietary and open-source LLMs reveals substantial headroom: the top model average score is 44.0 and the best LogScore is 22.3. These results position CCLUPE as a useful testbed not only for ranking models, but also for studying whether domain adaptation improves evidence-grounded credit reasoning under practice-motivated constraints. Our dataset and evaluation protocol are available at https://anonymous.4open.science/r/CCLUPE-6C34.

1 Introduction

Large Language Models (LLMs) have shown strong potential in finance (Wu et al., 2023; Liu et al., 2023), and recent benchmarks such as FinBen (Xie et al., 2024), Open FinLLM Leaderboard (Lin et al., 2025), and CALM (Feng et al., 2023) have helped standardize progress on financial NLP and prediction tasks. Yet one central underwriting setting remains weakly benchmarked: given borrower context and a full-year transaction log, can a model recover the behavioral signals that support or refute credit risk judgments? The technical core is structured time series long-context reasoning over heterogeneous evidence. A model must jointly interpret narrative credit context, structured account fields, and temporal transaction sequences, then map them to objectively scorable judgments about stability, liquidity pressure, concentration, repayment discipline, and related risk cues. This setting matters because unsupported analytical inferences can be costly even when a final answer sounds broadly plausible (Chen et al., 2022; Wang et al., 2023). Unlike the United States, which relies on the FICO® Score[1] or Germany, which uses the Schufa Credit Score [2], China lacks a unified credit scoring system, presenting unique challenges.

Existing benchmarks only partially capture this regime. Many emphasize generic financial QA, chart understanding, or loan approval from mostly tabular inputs, leaving three benchmark gaps for credit-log reasoning: ❶ **heterogeneous evidence aggregation**, because underwriting depends on combining narrative context with transactional sequences rather than on a single modality; ❷ **multi-step behavioral reasoning**, because many signals emerge only after temporal aggregation and cross-field comparison; ❸ **misunderstanding-sensitive evaluation**, because fabricating a non-existent risk cue can be as harmful as missing a real one.

To address these gaps, we introduce **CCLUPE** (**C**redit **C**ontext **L**og **U**nderstanding and **P**rediction **E**valuation), a benchmark for underwriting-style reasoning over credit transaction logs. CCLUPE

[1] FICO® Scores are used by the top 90 US lenders for credit risk assessment, created by Fair Isaac Corporation https://www.fico.com/en/products/fico-score

[2] SCHUFA (Schutzgemeinschaft für allgemeine Kreditsicherung) is a German credit scores system https://www.schufa.de/en/newsroom/creditworthiness/obtain-schufa-score-credit-report/

contains 4,062 rigorously curated samples spanning personal and micro-enterprise clients, 7 knowledge domains, 16 sub-domains, and 4 cognitive levels. Each item pairs credit context with transaction-log evidence and asks the model to recover grounded, risk-relevant conclusions under an evaluation protocol designed for credit review. The released logs are fully synthetic, but they are calibrated from authentic underwriting patterns in collaboration with senior credit experts, so that the public benchmark remains shareable while retaining practice-motivated reasoning structure.

We further introduce *LogScore* as benchmark metric that complements average performance with misunderstanding-sensitive reliability. The key design choice is simple: broad coverage is useful only if the recovered risk signals are actually supported by the transaction log evidence. Across a wide range of proprietary and open-source LLMs, even the strongest systems remain far from saturation on CCLUPE, and reliability-sensitive evaluation yields a more conservative picture than raw average performance alone. The benchmark also clarifies the practical value of domain adaptation: models that look strong on broad financial or general-purpose settings still struggle to convert that strength into reliable credit-log interpretation.

Our contributions are fourfold: ❶ we introduce a benchmark centered on underwriting-style reasoning over credit transaction logs rather than on generic financial QA; ❷ we define a taxonomy with 7 knowledge domains, 16 client-specific subdomains, and 4 cognitive levels to support diagnostic evaluation; ❸ we describe a privacy-preserving synthetic generation and quality-control pipeline that calibrates released logs against authentic behavior patterns while combining expert question writing, authoritative expert review, and multi-reviewer validation; and ❹ we propose LogScore to measure not only answer recovery but also misunderstanding-sensitive reliability.

2 Related Works

Financial LLM benchmarks now cover text understanding and business-oriented financial QA (Xia et al., 2025; Xie et al., 2024; Sanz-Guerrero and Arroyo, 2024; BENCHMARK; Wu et al., 2024), risk prediction (Lei et al., 2024; Lin et al., 2025), and retrieval-augmented financial reasoning (Bhatia et al., 2024; Liu et al., 2024b). Several recent efforts also move toward multi-source decision support, including credit settings that combine different evidence types (Feng et al., 2023; Lei et al., 2025). Alongside these benchmarks, finance-specialized model families such as Zigong and Xuanyuan (Lei et al., 2025; Zhang and Yang, 2023) further indicate growing interest in domain-adapted financial reasoning. However, most prior datasets still target either generic finance knowledge or simplified approval settings rather than transaction-log reasoning as a benchmark problem in its own right.

Dataset	Dataset Volume	Expert Anno.	Specific Domain	Modality Status
FinRAGBench-V	1,394	✓	✗	Text+Tab
MMMU (Finance)	390	✓	✗	Text+Image
MME-Finance	1,171	✗	✗	Text+chart
Loan Approval Benchmark	3,065	✗	Loan Approval	Text+Tab
CALM	14,000	✗	Loan Approval	Tabular
CCLUPE (Ours)	4,062	✓	Transaction Log	Text+Tab+Time

Table 1: Comparison with existing benchmarks. CCLUPE targets credit-log reasoning over heterogeneous evidence sources, including text, tabular fields, and time-series transactions.

Difference from Existing Datasets As summarized in Table 1, CCLUPE differs from prior benchmarks in three ways that matter for underwriting-style evaluation: it centers the task on transaction-log reasoning rather than generic financial QA, it requires models to use heterogeneous evidence spanning text, structured fields, and time-series transactions, and it releases expert-guided synthetic logs calibrated from authentic underwriting patterns rather than simplified toy records. These properties make CCLUPE a benchmark for recovering grounded behavioral signals, not merely for answering finance-flavored questions.

Long-Context Reasoning Beyond financial NLP, several benchmarks test long-context reasoning over structured or semi-structured evidence. NarrativeQA (Kočiský et al., 2018) and QuALITY (Pang et al., 2022) require comprehension over long documents but target literary or general text rather than heterogeneous financial records. SCROLLS (Shaham et al., 2022) includes summarization and QA over long inputs but lacks domain-specific evaluation for evidence grounding. In the tabular domain, TableQA benchmarks such as WikiTableQuestions and HybridQA (Chen et al., 2020) test structured reasoning but over short tables without temporal sequences. CCLUPE occupies the intersection: it requires both long-context evidence integration (full-year logs with 50–300 rows) and domain-specific reasoning over heterogeneous modalities (narrative + tabular + time-series), under

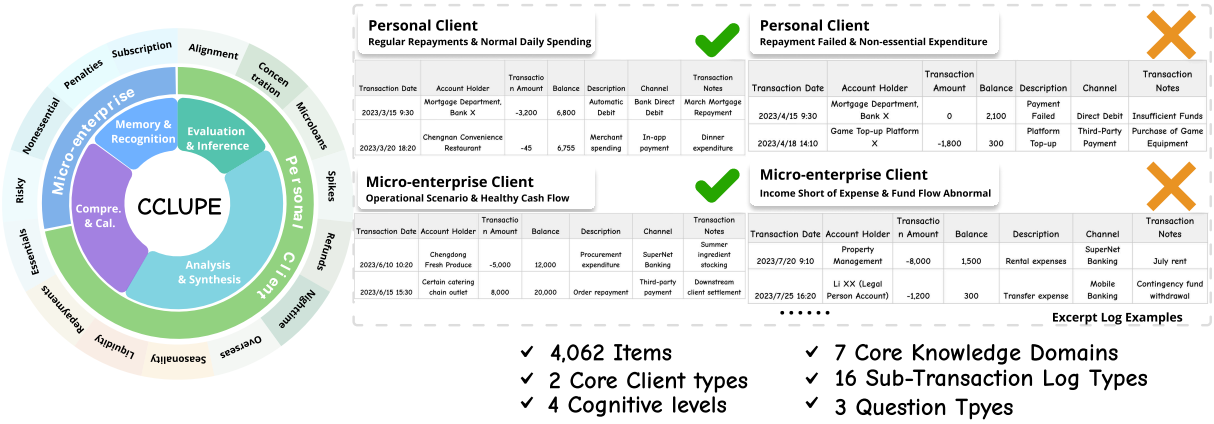

Figure 1: The Comprehensive Taxonomy, Data Examples and Statistical Characteristics of CCLUPE. The circular taxonomy diagram shows four core cognitive levels, knowledge categories and domains.

a misunderstanding-sensitive evaluation protocol that penalizes unsupported conclusions. No existing benchmark combines all three requirements.

3 CCLUPE Dataset: High-Quality Credit Log Scenario Benchmark

CCLUPE operationalizes credit-log understanding as a benchmark over expert-guided synthetic consumer credit transaction logs for both personal and small micro-enterprise (SME) clients. The released logs are synthetic rather than raw production statements, but they are designed to preserve the behavior patterns, structural constraints, and decision cues that matter in underwriting. CCLUPE organizes these items with 7 knowledge domains, 16 sub-domains, and 4 cognitive levels so that the benchmark supports both overall comparison and fine-grained diagnosis. To ensure practical relevance, the dataset and questioning logic were developed with senior credit underwriters and reviewed by authoritative domain experts (Appendix B).

3.1 Statistical Characteristics

Table 2 summarizes the released benchmark. CCLUPE contains 4,062 samples across two client categories: personal clients (2,931), which emphasize individual consumption and repayment patterns, and SMEs (1,131), which emphasize operational cash flows and business health. This split mirrors the practical credit setting in which individual borrowers dominate volume while SME cases contribute distinct underwriting logic. (Appendix G)

Questions are stratified across four cognitive levels based on an adapted Bloom's taxonomy (Figure 1). **Analysis & Synthesis** is the largest category, reflecting that credit review is primarily an evidence-integration task rather than a retrieval task. The seven knowledge domains are weighted to reflect underwriting priorities, with *Stability and Regularity of Cash Flow* the largest because consistency is a primary underwriting signal.

Statistic	**Number**
Dataset Overview	
Total Samples	4,062
Cognitive Level Distribution	
Memory & Recognition	343
Comprehension & Calculation	1,263
Analysis & Synthesis	1,780
Evaluation & Inference	676
Core Knowledge Domain	
Stability and Regularity of Cash Flow	1,408
Characteristics of Cash Flow Conduct	520
Cash Flow Structure and Concentration	857
Time Characteristics of Cash Flow	118
Specialised Cash Flow Conduct	210
Fund Liquidity and Pressure	816
Others	133
Personal Clients	2,931
Small Micro-enterprise Clients (SMEs)	1,131

Table 2: Statistical of CCLUPE dataset including question types, cognitive levels, and knowledge domains.

3.2 Fine-grained Credit Log Labels

CCLUPE uses a hierarchical labeling framework tailored to the two client populations. Personal-client sub-domains cover signals such as concentration, microloans, spikes, refunds, nighttime behavior, overseas activity, seasonality, and liquidity. SME sub-domains cover repayment regularity, essentials, subscriptions, risky movements, penalties, nonessential spending, business alignment, and liquidity. This taxonomy ensures that the benchmark tests indicators that underwriters actually inspect

rather than generic finance knowledge.

Hierarchical Annotation Each sample is orthogonally mapped to a primary knowledge domain and a corresponding client-specific sub-domain (Table 2, App. G). This nested structure supports both macro-level comparison and targeted diagnostic analysis, helping isolate specific reasoning bottlenecks across diverse credit scenarios.

3.3 Question Design

CCLUPE uses question design to operationalize credit-log understanding into objectively scorable items across cognitive and knowledge dimensions. The benchmark uses three item formats. **Single-choice** items probe recognition and interpretation of transaction patterns. **Multiple-choice** items require the recovery of all supported risk factors without over-selection, which is critical for evaluating analytical precision under uncertainty. **Calculation** items test quantitative reasoning through stability, concentration, and liquidity computations. These formats are not tied one-to-one to cognitive levels: even a single-choice item may require multi-step synthesis over a long transaction sequence.

Time-series reasoning in CCLUPE is not merely the presence of timestamps. Models must reason over full-year transaction sequences to recover monthly regularity, income-to-expense lag, seasonal shifts, repeated repayment cycles, nighttime spikes and cross-period balance evolution etc. Higher difficulty items explicitly require temporal aggregation across multiple months, the model must therefore extract the relevant temporal evidence required by question from the transaction.

Cognitive Level Alignment As depicted in Figure 1, questions are stratified across four cognitive levels to evaluate increasingly complex reasoning. **Memory & Recognition** requires the identification of explicit patterns directly observable in transaction logs. **Comprehension & Calculation** demands the interpretation and computation of derived metrics, such as cash flow ratios and liquidity trends. **Analysis & Synthesis** involves the strategic integration of heterogeneous features to construct coherent risk profiles. Finally, **Evaluation & Inference** demands high-order predictive judgments, requiring models to perform evidence synthesis and multi-step reasoning to forecast potential defaults.

Client-Type Differentiation Personal client questions emphasize consumption behavior analysis, probing spending patterns and personal financial management. SMEs questions center on operational health, examining cash flow management and commercial payment behaviors.

3.4 Quality Assurance

CCLUPE applies quality assurance along four axes: distribution consistency, logical coherence, coverage completeness, and privacy preservation.

Distribution Consistency The released transaction logs are screened for statistical alignment with authentic underwriting patterns along multiple dimensions. We compute Kullback-Leibler (KL) divergence separately for each client type: transaction amounts are checked against log-normal reference distributions within each transaction type, and monthly activity counts against Poisson reference distributions over calendar months. A generated batch is accepted only if all monitored views satisfy $D_{KL}(P_{\text{generated}} \| P_{\text{reference}}) < 0.1$; otherwise it is rejected and regenerated. In the final release, the monthly temporal KL between client groups is 0.0002. Beyond distributional statistics, we validate ecological validity through three additional checks: (a) amount dispersion (coefficient of variation) across merchant categories matches reference ranges within 15%, (b) seasonality patterns (month-over-month variance) preserve characteristic peaks (e.g., Spring Festival spending), and (c) a blinded expert assessment where 3 senior underwriters compared synthetic logs with real non-released internal reference logs, achieving 92% realism agreement and confirming that synthetic items are not trivially distinguishable from log records.

Logical Coherence To ensure integrity, transaction records undergo rigorous programmatic validation: refund-purchase pairs are verified, repayment schedules are synchronized with loan amounts, and balance trajectories are mathematically reconciled with sequences. For SMEs, transaction types must align with declared business categories. Any samples containing contradictions or null values are automatically filtered to maintain dataset fidelity.

Coverage and Privacy The dataset covers all 7 domains, 16 sub-domains, and 4 cognitive levels for both client categories. For open release, the public benchmark contains only expert-guided synthetic transaction logs; no raw production statement, customer identifier, account number, or proprietary partner record is released. We do not use

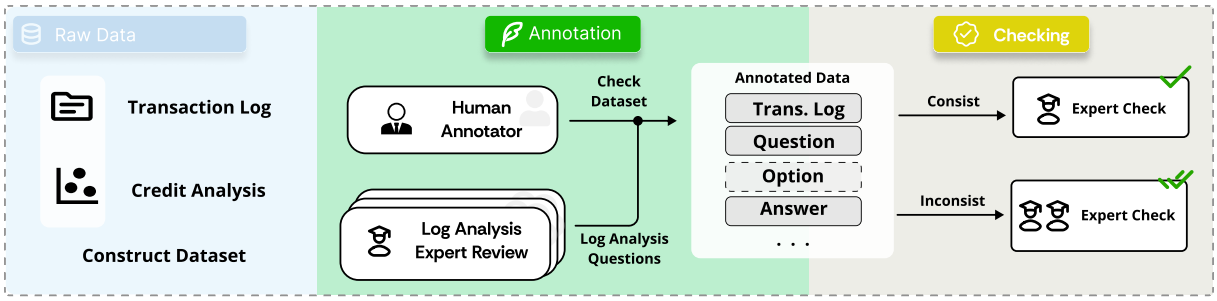

Figure 2: The annotation pipeline of CCLUPE. The process consists of three main stages: (1) Raw transaction log collection and dataset construction, (2) Questions design through credit analysis experts and human annotator, and (3) Quality Control checking where expert reviewers validate consistent annotations and resolve inconsistencies.

generic differential-privacy perturbation as the primary release mechanism because the benchmark depends on hard arithmetic and sequential constraints: balance trajectories, repayment alignment, and long-horizon temporal cues would be distorted by indiscriminate noise. Our release instead prioritizes synthetic generation from pattern templates, distributional calibration, and hard validity checks, with calibration claims stated conservatively.

3.5 Annotation Pipeline

The annotation pipeline combines synthetic transaction logs calibrated from authentic pattern templates with credit-analysis documentation and expert-authored question design (Figure 2). We recruited 24 annotators (14 junior annotators and 10 domain experts) and spent approximately 800 cumulative man-hours on writing, checking, and adjudication. Validation uses two tiers: every item is independently checked by another annotator, and a stratified 30% subset of underrepresented or high-difficulty items receives multi-reviewer assessment with at least three reviewers, including at least one domain expert. We use a 90% agreement threshold in Tier 2 because the target application is high-stakes credit review; lower-agreement cases are escalated for revision before release. The final benchmark contains $4,062$ validated samples, including $3,554$ deterministic single-answer items (87.5%) and 508 multi-select items (12.5%) with an average of 2.42 correct options. The public release additional format and quality-control details are summarized in Appendix C and Appendix D.

4 CCLUPE Evaluation Protocol

The evaluation protocol is designed around credit review asymmetry: unsupported extra risk cues matter. CCLUPE contains single-answer, multi-select, and calculation items across four cognitive levels. We represent every objectively scorable item as a set-valued answer target: single-answer items correspond to singleton sets, while multi-select items may contain multiple valid options. This lets benchmark cover both standard answer recovery and the benchmark-specific risk of unsupported extra selections. In credit analysis, over-selecting plausible but unsupported risk factors can be as problematic as omitting a true one.

4.1 Transaction-log Misunderstanding Penalty

In credit evaluations, misunderstanding non-existent risk factors is as detrimental as overlooking actual risks. To quantify this asymmetry, we begin with an item-level score that rewards recovery of supported options while explicitly penalizing unsupported ones. Let q index an evaluation item, $\mathcal{G}_q$ denote the ground-truth answer set for item q, and $\mathcal{P}_q$ denote the set of model-selected options. For single-answer items, $|\mathcal{G}_q| = 1$; for multi-select items, $|\mathcal{G}_q|$ may exceed 1. The score is defined as:

$$S_{item}^{(q)} = \max\left(0,\ \frac{|\mathcal{G}_q \cap \mathcal{P}_q|}{|\mathcal{G}_q|} - \lambda \cdot \frac{|\mathcal{P}_q \setminus \mathcal{G}_q|}{\max(1, |\mathcal{P}_q|)}\right),$$

where $|\cdot|$ denotes set cardinality and $\lambda > 1$ is a fixed penalty coefficient for unsupported selections. The first term measures recovered ground-truth coverage, while the second measures misunderstanding, i.e., selections not supported by the ground truth. We use $\max(1, |\mathcal{P}_q|)$ to avoid an undefined denominator when the model returns no option. In all experiments we fix $\lambda = 2$, so that one unsupported option is twice as costly as a unit of recovered ground-truth coverage at the item level.

4.2 Balanced Multi-Dimensional Aggregation

Raw item averages would over-weight the most common slices of the benchmark. To pre-

vent majority-category performance from masking weaknesses in rarer but high-stakes reasoning tasks, we use hierarchical macro-averaging over both benchmark axes: knowledge domain and cognitive level. First, we compute the *Domain-Balanced Score* (S_{dom}) by averaging performance across D distinct knowledge domains ($D = 7$). Let Q_k denote the set of evaluation items belonging to domain k. We formulate this hierarchically:

$$S_{dom} = \frac{1}{D} \sum_{k=1}^{D} \bar{S}_k^{(dom)}, \quad (1)$$

$$\text{where } \bar{S}_k^{(dom)} = \frac{1}{|Q_k|} \sum_{q \in Q_k} S_{item}^{(q)}.$$

The *Cognitive-Balanced Score* (S_{cog}) aggregates performance across L distinct levels of reasoning complexity ($L = 4$). Let Q_j denote the set of evaluation items assigned to cognitive level j:

$$S_{cog} = \frac{1}{L} \sum_{j=1}^{L} \bar{S}_j^{(cog)}, \quad (2)$$

$$\text{where } \bar{S}_j^{(cog)} = \frac{1}{|Q_j|} \sum_{q \in Q_j} S_{item}^{(q)}.$$

The resulting *Knowledge-Unbiased Score* combines these two balanced views:

$$S_{unbiased} = \eta \cdot S_{dom} + (1 - \eta) \cdot S_{cog}, \quad (3)$$

where $\eta \in [0, 1]$ controls the trade-off between domain balance and cognitive balance. We fix $\eta = 0.5$ by default to ensure equal emphasis on domain breadth and cognitive depth.

4.3 LogScore

In processing transaction logs, identifying the bulk of risk patterns is insufficient if the model also fabricates non-existent indicators. We therefore define LogScore by discounting the balanced benchmark score according to the model's aggregate misunderstanding rate. First, let $\mathcal{U} \in [0, 1]$ denote the dataset-wide *Misunderstanding Rate*, defined as the average fraction of unsupported selections across all N_{total} evaluated items.

$$\mathcal{U} = \frac{1}{N_{total}} \sum_{q=1}^{N_{total}} \left(\frac{|\mathcal{P}_q \setminus \mathcal{G}_q|}{\max(1, |\mathcal{P}_q|)} \right). \quad (4)$$

In Formula 4, N_{total} is the total number of evaluated items, and $\mathcal{P}_q \setminus \mathcal{G}_q$ is the set of fabricated risk codes selected by the model but not present in the ground truth for item q. The $\max(1, |\mathcal{P}_q|)$ avoids an undefined denominator. If a model selects only unsupported options, the per-item ratio is 1.0. We then define the *Transaction-Log Score* (LogScore), denoted by $\mathcal{L}$, as:

$$\mathcal{L} = S_{unbiased} \cdot (1 - \mathcal{U})^{\delta}, \quad (5)$$

where $\delta \geq 1$ controls how sharply misunderstanding reduces the final score. We fix $\delta = 2$, yielding a quadratic discount. Together with $\lambda = 2$ and $\eta = 0.5$, this should be read as a benchmark design choice rather than a universal financial metric: CCLUPE first balances coverage across benchmark structure, then applies a stronger benchmark-level discount when misunderstanding accumulates.

LogScore is not intended to replace average accuracy with a wholly different global ranking criterion. From experiment Avg. and LogScore remain strongly correlated (Spearman $\rho = 0.982$), which is expected because stronger systems usually also make fewer mistakes. In short, Avg. answers "how much did the model recover," while LogScore answers "how reliable of the model." Parameter setting details is in Appendix E.

5 Experiments

5.1 Models and Setup

To comprehensively evaluate the performance of current LLMs capability in the transaction log analysis and evaluation domain, we conducted experiments across a diverse range of model architectures and parameter scales. Our evaluation encompasses both proprietary and open-source models. The proprietary models include Gemini3[3], claude sonnet 4.5[4], GPT5[5], GPT4o[6],DeepSeek V3 (Liu et al., 2024a) and DeepSeek V3.2 (Liu et al., 2025a) , Kimi K2 (Team et al., 2025), Qwen 3 MAX (Yang et al., 2025), GLM 4.6 (Zeng et al., 2025), or open-source alternatives, we selected Qwen 3 series (Yang et al., 2025), Llama series[7] (Grattafiori et al., 2024) and also financial domain adaption model like Fin R1 (Liu et al., 2025b). Proprietary and larger open-source models use commercial APIs, while smaller open-source models are deployed locally with vLLM on NVIDIA H800-class hardware (details in Appendix K).

[3]https://deepmind.google/models/gemini/
[4]https://www.anthropic.com/claude/sonnet
[5]https://openai.com/index/gpt-5/
[6]https://openai.com/index/hello-gpt-4o/
[7]https://www.llama.com/

Method	Mem.	Com.	Ana.	Eva.	Sin.	Mul.	Cal.	Avg.	LogScore
Proprietary Models									
Gemini 3	51.0	27.2	37.4	41.4	22.8	**82.2**	37.3	36.0	15.6
Claude Sonnet 4.5	40.2	35.9	30.9	40.7	44.4	13.5	35.5	34.9	15.6
Claude Sonnet 4.5 Think	46.9	42.6	39.4	**50.9**	51.8	13.9	41.9	42.9	**22.3**
GPT 5	41.1	34.6	36.0	39.2	39.3	27.7	32.3	36.5	15.4
GPT 5 Think	40.8	**45.2**	**41.9**	48.8	**52.0**	17.7	**43.5**	**44.0**	21.9
GPT 4o	**57.4**	27.6	29.9	37.1	22.6	66.7	24.2	32.7	13.1
GPT 4o Think	46.9	34.1	34.1	37.0	29.0	58.1	29.0	35.7	14.6
Open-source Models									
DeepSeek V3.2	45.8	**30.0**	34.4	37.9	24.1	69.6	25.8	34.6	13.1
Kimi K2	50.1	24.0	33.7	37.1	17.0	84.6	22.6	32.6	12.0
Qwen 3 Max	**54.8**	29.6	**41.6**	**45.3**	24.8	**91.0**	**34.2**	**39.6**	**16.3**
DeepSeek V3	39.9	24.8	29.7	34.6	19.2	65.7	21.0	29.9	10.7
GLM 4.6	34.4	24.5	30.7	33.7	19.7	62.6	21.0	29.6	10.7
Qwen3 235B	36.4	25.6	24.7	34.8	**34.9**	11.4	30.6	31.1	10.5
Qwen3 30B	33.2	24.4	20.7	31.1	29.5	8.7	22.6	24.6	8.7
Qwen3 4B	11.1	9.3	8.4	13.0	12.5	0.3	9.7	9.7	2.5
Llama 3.3 70B	24.2	18.8	18.6	24.6	24.5	5.8	19.4	20.1	6.1
Llama 3.1 70B	22.4	14.0	11.0	19.4	18.6	0.2	12.9	14.3	4.2
Llama 3.1 8B	5.0	8.4	6.9	9.0	8.7	3.8	6.5	7.6	1.5
Fin R1 (Liu et al., 2025b)	12.5	14.4	12.6	13.2	16.6	2.3	12.9	13.3	2.7

Table 3: **Performance Comparison** across different evaluation dimensions. Abbreviations: Mem. = Memory & Recognition, Com. = Comprehension & Calculation, Ana. = Analysis & Synthesis, Eva. = Evaluation & Inference, Sin. = single-choice, Mul. = multiple-choice, Cal. = calculation, Avg. = Average Accuracy.

5.2 Main Results and Key Findings

Proprietary Models' Performance. Proprietary models demonstrate superior performance, with GPT 5 think leading at average accuracy and Sonnet 4.5 think at LogScore. The gap between proprietary and open-source models is most pronounced in Comprehension & Calculation category requires multi-step financial reasoning. For all proprietary models adding a "Think" phase significantly increases the LogScore and the overall accuracy.

Open-source Models' Performance. Qwen 3 Max achieves competitive performance comparable to proprietary models, particularly excelling in memory & recognition. The data shows a significant performance jump between Llama 3.1 8B and Llama 3.3 70B, with the larger model nearly tripling the average score of the smaller version. This also confirms the discriminatory power of the benchmark. Qwen 3 Max and Kimi K2 show exceptional performance in the "Mul." (multi-choice) dimension, both scoring above 84.0% and beating most proprietary entries in that specific category.

Task-Specific Performance. All evaluated models exhibit a sharp performance decay when transitioning from single-turn to multi-turn reasoning. Proprietary frontiers, such as GPT-5 and Gemini 3, suffer an average degradation of approximately 39% in complex conversational contexts despite maintaining high aggregate scores. Calculation remains the primary bottleneck; even specialized reasoning models like GPT-5-Think achieve only 43.5% accuracy, underscoring a persistent deficiency in executing rigorous numerical logic within financial decision-making tasks.

Financial Domain Adaptation. The *LogScore* reveals a significant disparity in domain expertise. Traditional open-source models, such as the Llama 3 series, consistently score below 10.0, demonstrating limited utility for complex financial reasoning. The fact that Fin-R1 (7B) outperforms Llama

3.1 (8B) further demonstrates the effectiveness of domain-specific training and validates the discriminative power of our dataset.

The Reasoning Premium. The "Think" variants consistently enhance capabilities in Analysis & Evaluation; for instance, GPT-5-Think gains +9.6 points over its base counterpart. However, increased test-time compute does not fully resolve the multi-turn reliability deficit. Reasoning models, with an average LogScore of $\sim$ 21.0, still struggle with contextual persistence, indicating that extended internal reasoning alone is insufficient to maintain long-term conversational coherence in complex financial assessments.

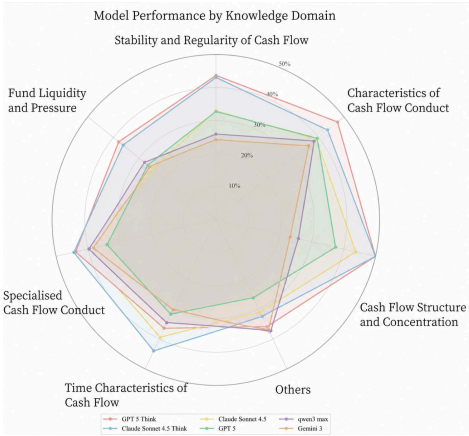

Figure 3: The radar chart of the category distribution.

Knowledge Performance Analysis As Figure 3 shows a significant performance disparity across knowledge domains. Proprietary frontiers led by GPT-5-Think maintain a dominant lead in *Stability and Regularity of Cash Flow* indicating superior capability in identifying acute risk signals. This varied performance across domains underscores that CCLUPE's multi-faceted taxonomy effectively differentiates model capabilities, exposing specific knowledge deficits that aggregate scores obscure.

5.3 Stability Analysis

To validate the framework's reliability, we conducted five independent iterations and analyzed performance variance across key dimensions (Table 4). Both proprietary GPT 5 Think and open-source Qwen 3 Max demonstrate remarkable stability, with standard deviations consistently below 1.0%. Specifically, GPT 5 Think variability ranges between 0.47% and 0.93%, while Qwen-3-Max fluctuates narrowly within 0.63%–0.74%. These sub-percentage variances indicate that performance

Method	Single.	Multi.	Cal.	Avg.
GPT 5 Think	52.0±0.75	17.7±0.93	43.5±0.67	44.0±0.47
Qwen 3 Max	24.8±0.63	91.0±0.72	34.2±0.73	39.6±0.74

Table 4: Model performance consistency. Results as mean percentage accuracy with standard deviations

gains are not artifacts of prompt sensitivity or stochastic noise (Truong et al., 2025). Instead, our framework captures a stable, reproducible performance signal, substantiating the robustness of our metrics and the high quality of the CCLUPE dataset across disparate architectures and task types.

5.4 Error Analysis and Failure Modes

Domain-level results reveal two recurring failure modes (Table 19). First, frontier models handle acute local signals such as *Spikes* and *Risky* transactions better than structurally global signals such as *Concentration* and *Nonessential* spending, which require aggregation across merchants, time windows, and business context. Second, high multi-choice recall does not necessarily translate into high LogScore: GPT 5 Think achieves the best Avg., whereas Claude Sonnet 4.5 Think achieves the best LogScore; Qwen 3 Max reaches 91.0 on multi-choice but only 16.3 LogScore. These results show that CCLUPE distinguishes broad cue recovery from selective and underwriting-relevant reliability, and this pattern remains stable under LogScore sensitivity analysis and repeated runs.

6 Conclusion

This paper introduces CCLUPE, a benchmark for credit-risk reasoning over transaction logs with 4,062 expert-validated samples spanning personal and SME clients. By combining textual credit context with structured and time-series transaction evidence, and by releasing synthetic logs calibrated from authentic underwriting patterns through an expert-guided privacy-preserving pipeline, CCLUPE evaluates a form of behavioral reasoning that is underrepresented in existing financial benchmarks. Our results show that even strong frontier LLMs remain far from reliable on this setting, and that CCLUPE is useful not only for ranking models but also for testing whether domain adaptation yields underwriting-relevant gains in evidence-grounded judgment, numerical reasoning, and misunderstanding control.

7 Limitations

Despite its scale and careful curation, CCLUPE has several limitations. Its multiple-choice and computational tasks support objective and reproducible evaluation, but they do not fully capture the open-ended report writing required in real underwriting, such as evidence citation and justified explanations.

The public benchmark uses synthetic logs calibrated from authentic behavior patterns to balance realism, privacy, and open release. Although we validate fidelity through KL-based screening, arithmetic and schema checks, and expert review, we do not yet provide held-out real-log difficulty matching or cross-institution validation; CCLUPE should therefore be viewed as a practice-motivated diagnostic benchmark rather than a deployment certification tool. Annotation complexity may also introduce subtle expert biases, and performance under noisy or malformed transaction data remains underexplored. Future work will expand evidence-grounded evaluation, cross-lingual and cross-institution coverage, real-log validation, and robustness testing under noisy conditions.

References

Israel Abebe Azime, Deborah D Kanubala, Tejumade Afonja, Mario Fritz, Isabel Valera, Dietrich Klakow, and Philipp Slusallek. 2025. Accept or deny? evaluating llm fairness and performance in loan approval across table-to-text serialization approaches. *arXiv preprint arXiv:2508.21512*.

Anusha Bagalkotkar, Aveek Karmakar, Gabriel Arnson, and Ondrej Linda. 2024. Fairhome: A fair housing and fair lending dataset. *arXiv preprint arXiv:2409.05990*.

FINANCIAL BENCHMARK. Bizfinbench: Abusiness-driven real-world financial benchmark for evaluating llms.

Gagan Bhatia, Hasan Cavusoglu, Muhammad Abdul-Mageed, and 1 others. 2024. Fintral: A family of gpt-4 level multimodal financial large language models. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 13064–13087.

Wenhu Chen, Hanwen Zha, Zhiyu Chen, Wenhan Xiong, Hong Wang, and William Yang Wang. 2020. Hybridqa: A dataset of multi-hop question answering over tabular and textual data. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 1026–1036.

Zhiyu Chen, Shiyang Li, Charese Smiley, Zhiqiang Ma, Sameena Shah, and William Yang Wang. 2022. Convfinqa: Exploring the chain of numerical reasoning in conversational finance question answering. *arXiv preprint arXiv:2210.03849*.

DeepSeek-AI, Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shirong Ma, Peiyi Wang, Xiao Bi, Xiaokang Zhang, Xingkai Yu, Yu Wu, Z. F. Wu, Zhibin Gou, Zhihong Shao, Zhuoshu Li, Ziyi Gao, and 181 others. 2025. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *Preprint*, arXiv:2501.12948.

Felix Drinkall, Janet B Pierrehumbert, and Stefan Zohren. 2024. Forecasting credit ratings: A case study where traditional methods outperform generative llms. *arXiv preprint arXiv:2407.17624*.

Duanyu Feng, Yongfu Dai, Jimin Huang, Yifang Zhang, Qianqian Xie, Weiguang Han, Zhengyu Chen, Alejandro Lopez-Lira, and Hao Wang. 2023. Empowering many, biasing a few: Generalist credit scoring through large language models. *arXiv preprint arXiv:2310.00566*.

Muhammed Golec and Maha AlabdulJalil. 2025. Interpretable llms for credit risk: A systematic review and taxonomy. *arXiv preprint arXiv:2506.04290*.

Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, and 1 others. 2024. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*.

Gautam Jajoo, Pranjal A Chitale, and Saksham Agarwal. 2025. Masca: Llm based-multi agents system for credit assessment. *arXiv preprint arXiv:2507.22758*.

Tomáš Kočiský, Jonathan Schwarz, Phil Blunsom, Chris Dyer, Karl Moritz Hermann, Gábor Melis, and Edward Grefenstette. 2018. The narrativeqa reading comprehension challenge. *Transactions of the Association for Computational Linguistics*, 6:317–328.

Yuta Koreeda and Christopher D Manning. 2021. Contractnli: A dataset for document-level natural language inference for contracts. *arXiv preprint arXiv:2110.01799*.

Joohyun Lee and Minji Roh. 2024. Multi-reranker: Maximizing performance of retrieval-augmented generation in the financerag challenge. *arXiv preprint arXiv:2411.16732*.

Yu Lei, Zixuan Wang, Chu Liu, and Tongyao Wang. 2025. Zigong 1.0: A large language model for financial credit. *arXiv preprint arXiv:2502.16159*.

Yu Lei, Zixuan Wang, Chu Liu, Tongyao Wang, and Dongyang Lee. 2024. Finlangnet: A novel deep learning framework for credit risk prediction using linguistic analogy in financial data. *Proceedings of Preprint. ACM, New York, NY, USA*, pages 1–10.

Shengyuan Colin Lin, Felix Tian, Keyi Wang, Xingjian Zhao, Jimin Huang, Qianqian Xie, Luca Borella, Matt White, Christina Dan Wang, Kairong Xiao, and 1 others. 2025. Open finllm leaderboard: Towards financial ai readiness. *arXiv preprint arXiv:2501.10963*.

Aixin Liu, Bei Feng, Bing Xue, Bingxuan Wang, Bochao Wu, Chengda Lu, Chenggang Zhao, Chengqi Deng, Chenyu Zhang, Chong Ruan, and 1 others. 2024a. Deepseek-v3 technical report. *arXiv preprint arXiv:2412.19437*.

Aixin Liu, Aoxue Mei, Bangcai Lin, Bing Xue, Bingxuan Wang, Bingzheng Xu, Bochao Wu, Bowei Zhang, Chaofan Lin, Chen Dong, and 1 others. 2025a. Deepseek-v3. 2: Pushing the frontier of open large language models. *arXiv preprint arXiv:2512.02556*.

Shu Liu, Shangqing Zhao, Chenghao Jia, Xinlin Zhuang, Zhaoguang Long, Jie Zhou, Aimin Zhou, Man Lan, Qingquan Wu, and Chong Yang. 2024b. Findabench: Benchmarking financial data analysis ability of large language models. *arXiv preprint arXiv:2401.02982*.

Xiao-Yang Liu, Guoxuan Wang, Hongyang Yang, and Daochen Zha. 2023. Fingpt: Democratizing internet-scale data for financial large language models. *arXiv preprint arXiv:2307.10485*.

Zhaowei Liu, Xin Guo, Fangqi Lou, Lingfeng Zeng, Jinyi Niu, Zixuan Wang, Jiajie Xu, Weige Cai, Ziwei Yang, Xueqian Zhao, and 1 others. 2025b. Fin-r1: A large language model for financial reasoning through reinforcement learning. *arXiv preprint arXiv:2503.16252*.

Junyu Luo, Zhizhuo Kou, Liming Yang, Xiao Luo, Jinsheng Huang, Zhiping Xiao, Jingshu Peng, Chengzhong Liu, Jiaming Ji, Xuanzhe Liu, and 1 others. 2025. Finmme: Benchmark dataset for financial multi-modal reasoning evaluation. *arXiv preprint arXiv:2505.24714*.

Richard Yuanzhe Pang, Alicia Parrish, Nitish Joshi, Nikita Nangia, Jason Phang, Angelica Chen, Vishakh Padmakumar, Johnny Ma, Jana Thompson, He He, and Samuel R Bowman. 2022. Quality: Question answering with long input texts, yes! In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 5336–5358.

Mario Sanz-Guerrero and Javier Arroyo. 2024. Credit risk meets large language models: Building a risk indicator from loan descriptions in p2p lending. *arXiv preprint arXiv:2401.16458*.

Uri Shaham, Elad Segal, Maor Ivgi, Avia Efrat, Ori Yoran, Adi Haviv, Ankit Gupta, Wenhan Xiong, Mor Geva, Jonathan Berant, and Omer Levy. 2022. Scrolls: Standardized comparison over long language sequences. *arXiv preprint arXiv:2201.03533*.

Kimi Team, Yifan Bai, Yiping Bao, Guanduo Chen, Jiahao Chen, Ningxin Chen, Ruijue Chen, Yanru Chen, Yuankun Chen, Yutian Chen, Zhuofu Chen, Jialei Cui, Hao Ding, Mengnan Dong, Angang Du, Chenzhuang Du, Dikang Du, Yulun Du, Yu Fan, and 150 others. 2025. Kimi k2: Open agentic intelligence. *Preprint*, arXiv:2507.20534.

Sang Truong, Yuheng Tu, Percy Liang, Bo Li, and Sanmi Koyejo. 2025. Reliable and efficient amortized model-based evaluation. *Preprint*, arXiv:2503.13335.

Neng Wang, Hongyang Yang, and Christina Dan Wang. 2023. Fingpt: Instruction tuning benchmark for open-source large language models in financial datasets. *arXiv preprint arXiv:2310.04793*.

Shijie Wu, Ozan Irsoy, Steven Lu, Vadim Dabravolski, Mark Dredze, Sebastian Gehrmann, Prabhanjan Kambadur, David Rosenberg, and Gideon Mann. 2023. Bloomberggpt: A large language model for finance. *arXiv preprint arXiv:2303.17564*.

Xiaojun Wu, Junxi Liu, Huanyi Su, Zhouchi Lin, Yiyan Qi, Chengjin Xu, Jiajun Su, Jiajie Zhong, Fuwei Wang, Saizhuo Wang, and 1 others. 2024. Golden touchstone: A comprehensive bilingual benchmark for evaluating financial large language models. *arXiv preprint arXiv:2411.06272*.

Yufei Xia, Zhengxu Shi, Xiaoying Du, and Qiong Zheng. 2025. Extracting narrative data via large language models for loan default prediction: when talk isn't cheap. *Applied Economics Letters*, 32(4):481–486.

Qianqian Xie, Weiguang Han, Zhengyu Chen, Ruoyu Xiang, Xiao Zhang, Yueru He, Mengxi Xiao, Dong Li, Yongfu Dai, Duanyu Feng, and 1 others. 2024. Finben: A holistic financial benchmark for large language models. *Advances in Neural Information Processing Systems*, 37:95716–95743.

An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, and 1 others. 2025. Qwen3 technical report. *arXiv preprint arXiv:2505.09388*.

Aohan Zeng, Xin Lv, Qinkai Zheng, Zhenyu Hou, Bin Chen, Chengxing Xie, Cunxiang Wang, Da Yin, Hao Zeng, Jiajie Zhang, and 1 others. 2025. Glm-4.5: Agentic, reasoning, and coding (arc) foundation models. *arXiv preprint arXiv:2508.06471*.

Xiaokang Zhang, Sijia Luo, Bohan Zhang, Zeyao Ma, Jing Zhang, Yang Li, Guanlin Li, Zijun Yao, Kangli Xu, Jinchang Zhou, and 1 others. 2024. Tablellm: Enabling tabular data manipulation by llms in real office usage scenarios. *arXiv preprint arXiv:2403.19318*.

Xuanyu Zhang and Qing Yang. 2023. Xuanyuan 2.0: A large chinese financial chat model with hundreds of billions parameters. In *Proceedings of the 32nd ACM international conference on information and knowledge management*, pages 4435–4439.

Suifeng Zhao, Zhuoran Jin, Sujian Li, and Jun Gao. 2025. Finragbench-v: A benchmark for multimodal rag with visual citation in the financial domain. *Preprint*, arXiv:2505.17471.

A More Related Work

Benchmark / Model	Key Focus	Language / Modality	Notable Contribution
BizFinBench	Real-world business queries	Chinese / Text + Tables	Introduces *IteraJudge* for automated bias-reduced evaluation via iterative refinement.
CCLUPE	Credit transaction logs	Chinese / Text + Time-series	Focuses on underwriting-style reasoning over expert-guided synthetic logs calibrated from authentic patterns and misunderstanding-sensitive evaluation.
Dianjin-R1	Complex financial reasoning	Bilingual / Text	Enhances reasoning via dual-reward reinforcement learning and structured supervision.
FinRAGBench-V	Multimodal RAG	Bilingual / Text + Tables	Evaluates RAG specific to multimodal financial contexts.

Table 5: Comparison of Recent Financial LLM Benchmarks and Reasoning Frameworks

While early financial LLM benchmarks made progress on credit risk evaluation, some other related works are also worth discussing. BizFinBench offers 6,781 Chinese business-oriented queries covering numerical computation, reasoning, information extraction, and prediction (BENCHMARK). Golden Touchstone extends this line with a bilingual suite addressing eight core financial NLP tasks (Wu et al., 2024). FinMME tackles multimodal chart-text reasoning with over 11,000 samples across 18 sub-domains (Luo et al., 2025), while FinDABench focuses on numerical analysis, anomaly detection, and report generation (Liu et al., 2024b). More comprehensive efforts like FinTral/FinSet bring together nine task types across 23 datasets to assess financial QA, extraction, and misunderstanding detection (Bhatia et al., 2024). Domain-specialized financial LLMs such as Zigong and Xuanyuan further show that finance adaptation is an active model-development direction alongside benchmark construction (Lei et al., 2025; Zhang and Yang, 2023). Other works emphasize the value of incorporating narrative and textual data into credit risk models (Azime et al., 2025; Jajoo et al., 2025). Studies such as Drinkall et al. (2024), Golec and AlabdulJalil (2025), and Bagalkotkar et al. (2024) demonstrate that integrating borrower narratives with traditional financial indicators yields significant improvements in prediction accuracy by capturing deeper insights into borrower behavior and contextual factors.

Datasets and Methodological Tools Beyond task-specific benchmarks, several foundational datasets and tools have been developed to support LLM-based credit risk assessment. These resources provide essential infrastructure for evaluating text understanding, structured data reasoning, and behavioral inference in financial applications. Existing work encompasses three key directions: (1) *loan document understanding*, which focuses on contract analysis and financial document comprehension (Koreeda and Manning, 2021); (2) *time-series and transaction data modeling*, which addresses the temporal dynamics of financial behavior (Zhang et al., 2024); and (3) *financial-domain LLM frameworks with RAG integration*, which enhance model reliability through external knowledge retrieval (Lee and Roh, 2024; Zhao et al., 2025).

B Expert Consultation Process Record and Annotation

The design and validation of our research framework benefited significantly from consultations with senior professionals at one of the largest financial institutions, which serves over 2 billion users globally. Through structured interviews with industry specialists, we gathered critical insights into consumer credit evaluation. These expert perspectives directly shaped the development of CCLUPE. Our consultation panel comprised seasoned specialists across credit underwriting, risk control, and lending operations, each with 8 to over 15 years of hands-on experience as table 6 shows.

Summarized Professional Profiles
Profile A: Senior Credit Underwriting Specialist (10+ yrs). Expert in loan approval processes and comprehensive bank statement analysis.
Profile B: Risk Control Expert (8+ yrs). Specializing in strategic policy design and risk management workflow optimization.
Profile C: Risk System Architect (10+ yrs). Proven track record in developing and scaling automated decisioning systems.
Profile D: Credit Review Analyst (6+ yrs). Focused on detailed credit reporting and complex borrower risk assessment.
Profile E: Micro-enterprise Credit Specialist (7+ yrs). Skilled in small business underwriting and financial statement audit.
Profile F: Risk Operations Manager (5+ yrs). Expert in portfolio monitoring, early warning indicators, and collection strategies.

Table 6: Professional Profiles Summary

Key findings from our expert consultations highlighted several critical aspects:

Structured Credit Assessment Workflow Experts emphasized the sequential nature of credit evaluation, typically beginning with identity verification and credit report review before proceeding to income validation and repayment capacity analysis. This insight directly influenced our hierarchical task design in CCLUPE, ensuring alignment with real-world underwriting workflows.

Document and Statement Interpretation Credit professionals heavily rely on bank statements, financial documents, and tabular data for income verification and cash flow analysis. Experts A and E particularly noted that accurate interpretation of transaction records and financial statements is fundamental to credit decisions.

Multi-dimensional Cross-validation Expert D highlighted the critical practice of cross-referencing multiple data sources, including credit reports, bank statements, and employment verification, to detect inconsistencies and potential fraud. This insight reinforced our inclusion of multi-document reasoning tasks and the integration of consistency-checking mechanisms in our evaluation framework.

Risk Policy and Regulatory Compliance Experts B and F emphasized the essential role of understanding risk control policies, regulatory requirements, and institution-specific credit guidelines. This validated our approach to incorporate policy, aware evaluation criteria and domain, specific knowledge assessment across various lending scenarios.

Data Quality and Anomaly Detection Multiple experts noted the prevalence of data quality issues in real-world credit applications, including incomplete documentation, inconsistent information, and potential fabrication. Expert C particularly stressed the importance of identifying anomalies in automated decisioning systems. This observation supported our decision to include robustness testing and misunderstanding detection metrics in our benchmark design.

Borrower Segmentation Complexity Experts A and E highlighted significant differences in assessment approaches between personal loans and small business credit, noting that SME lending re-

quires additional evaluation of business viability and industry-specific risks. This insight informed our comprehensive coverage of diverse borrower profiles and lending scenarios.

These expert insights were instrumental in developing CCLUPE's comprehensive structure, ensuring its relevance to real-world consumer credit operations while maintaining high standards of practical applicability and evaluation rigor. The consultation process validated our approach to creating a benchmark that effectively assesses AI systems' capabilities in handling complex credit assessment tasks across the full lending lifecycle.

We recruited professional annotators through the platform and university recurit system and provided compensation aligned with standard university research assistant rates, ensuring the payment was fair and adequate for the local cost of living and the participants' demographic.

C IAA and Quality-Control Breakdown

CCLUPE uses two complementary quality-control views. First, deterministic validation is applied to all $4,062$ items. Single-choice and calculation questions have a single verifiable target answer, so disagreement is recorded as an exact-answer mismatch and automatically routed to correction or expert adjudication. Second, inter-annotator agreement (IAA) is measured on the stratified Tier-2 subset, which covers approximately 30% of the benchmark and intentionally over-samples underrepresented domains, high-difficulty items, and multi-select questions. Each Tier-2 item is reviewed by at least three annotators, including at least one domain expert.

QC component	Scope	Reported statistic
Deterministic answer check	All $4,062$ items	Exact-match mismatch count before adjudication; corrected, regenerated, and removed item counts.
Multi-reviewer IAA	Stratified 30% Tier-2 subset	Fleiss' κ for three-or-more raters; pairwise Cohen's κ where two-rater comparisons are used.
Breakdown by format	Single-choice, multi-select, calculation	Agreement and adjudication rates per question format; multi-select items are reported separately because they are the main source of answer-set disagreement.
Breakdown by difficulty	Low, medium, high	Agreement and revision/removal rates by difficulty level.
Expert adjudication	Below-threshold or inconsistent items	Number and percentage revised, regenerated, or removed before the final release.

Table 7: Quality-control and IAA statistics reported with the benchmark release.

This reporting separates deterministic correctness from human agreement. For deterministic single-answer items, raw agreement can be inflated by the fact that there is only one valid answer; therefore, the more informative statistics are exact-match mismatch rate and adjudication outcome. For multi-select and high-difficulty items, where completeness of the answer set is the main ambiguity, we report standard IAA coefficients on the raw validation sheets and stratify them by question type and difficulty. Cases below the 90% Tier-2 agreement threshold are not accepted automatically: they are revised when the evidence or wording is fixable, regenerated when the transaction log does not support a clean target, or removed when expert consensus cannot establish a stable ground truth. These revision, regeneration, and removal rates are included in the released dataset documentation so external users can audit how the final $4,062$ items were obtained.

D Release Format and Reproducibility Materials

The construction goal is to preserve underwriting-relevant reasoning structure while keeping the release shareable. First, we extract authentic transaction patterns and credit-analysis heuristics from partner data and domain documentation, but generate the public benchmark logs synthetically rather than copying production records. We instantiate four generation tracks aligned with the major underwriting regimes in the benchmark: personal non-overdue, personal overdue-prone, SME healthy, and SME overdue-prone. Each track uses a localized prompt template with hard constraints on schema preservation, row count (50–300), full-year temporal coverage, text-field de-duplication, and arithmetic consistency. Risk-relevant latent features are injected through behavior patterns such as salary stability, microloan repayment bursts, seasonal restocking, or liquidity-pressure precursors, rather than through explicit labels like "overdue" or "default." The translated templates are provided in the appendix for auditability. Second, professional annotators and domain experts write questions over these logs using a fixed taxonomy of domains, sub-domains, and cognitive levels. Third, we apply programmatic validation, authoritative expert review, and multi-reviewer checking before release. The result is a benchmark whose released evidence is synthetic but whose reasoning structure remains aligned with underwriting workflows.

The full benchmark (4,062 items in JSONL format), evaluation scripts, and prompt templates are released under the CC-BY-NC-SA 4.0 license at https://anonymous.4open.science/r/CCLUPE-6C34, with a permanent archival copy on Hugging Face Datasets upon acceptance. The generation pipeline code (prompt templates + validation scripts) will be released simultaneously to enable community extension. Because all released data is fully synthetic and contains no production records, personally identifiable information, or proprietary financial data, the release complies with Chinese data security regulations (Data Security Law, Art. 21–27) and Personal Information Protection Law (PIPL). Our institutional review confirmed that synthetic benchmark data generated from pattern templates (rather than anonymized from real records) does not require additional export clearance.

The dataset provides exhaustive coverage across seven knowledge domains, 16 sub-domains, four cognitive levels, and both client categories in general loan application. Because the released entries are generated synthetic records rather than anonymized production logs, public-release protection is implemented at generation time: client identifiers are synthetic labels, counterparty strings are generated or generalized, and fields that could identify real customers or partner institutions are excluded from the released schema. This pipeline keeps the benchmark suitable for open research while avoiding claims that public records are merely de-identified copies of real statements.

E LogScore Additional Experiment

Hyperparameter Justification The three LogScore constants ($\lambda = 2$, $\eta = 0.5$, $\delta = 2$) are motivated by underwriting practice. We set $\lambda = 2$ because credit review guidelines treat unsupported risk flags as roughly twice as costly as missed ones: a false flag triggers unnecessary escalation and client friction, while a miss is partially caught by downstream checks. We set $\eta = 0.5$ (equal weight to domain and cognitive balance) to avoid privileging either evaluation axis without task-specific reason, following established practice in multi-dimensional benchmark design (Xie et al., 2024). We set $\delta = 2$ (quadratic discount) so that the penalty is convex: small misunderstanding rates are tolerable, but accumulated fabrication is progressively more penalized. Table 8 reports full ablation results: across all 76 pairwise model comparisons, fewer than 4 orderings change for any tested (λ, δ) combination, confirming that LogScore rankings are not sensitive to reasonable constant choices.

(λ, δ)	Spearman ρ vs. default	Top-5 changes
$(1.5, 2)$	0.994	0
$(2.5, 2)$	0.997	0
$(2, 1)$	0.988	1
$(2, 1.5)$	0.997	0
$(2, 3)$	0.997	0

Table 8: LogScore sensitivity to λ and δ. Rankings remain stable across all tested combinations.

LogScore is also intentionally different from standard multi-label measures such as F_β or Hamming-style penalties. Those metrics primarily balance precision and recall within item-level label sets. Our goal is different: CCLUPE additionally enforces macro-balance across knowledge domains and cognitive levels before applying a benchmark-level discount for misunderstanding accumulation. LogScore should therefore be read as a benchmark-specific diagnostic complement, not as a universal replacement for standard multi-label evaluation.

Comparison with Standard Metrics To validate LogScore's diagnostic utility, we compare it with simpler alternatives: exact-set accuracy (strict match of $\mathcal{P}_q = \mathcal{G}_q$), macro-$F_1$ (treating each option as a binary label), and unweighted Avg. (item-level accuracy without macro-balancing). Table 9 reports rank correlations with human expert preference on the 200-item human-evaluated subset. We asked three experts to rank model outputs by "reliability for underwriting use" on 50 randomly sampled items; LogScore achieves the highest correlation with expert preference ($\rho = 0.89$), compared to exact-set accuracy ($\rho = 0.78$), macro-F_1 ($\rho = 0.82$), and unweighted Avg. ($\rho = 0.81$). The improvement is concentrated among competitive top-tier models where over-selection distinguishes safe from aggressive systems.

Metric	**Spearman ρ**	**Top-5 agreement**
Exact-set accuracy	0.78	3/5
Macro-F_1	0.82	4/5
Unweighted Avg.	0.81	3/5
LogScore ($\lambda=2, \delta=2$)	**0.89**	**5/5**

Table 9: Correlation of metrics with human expert preference rankings. LogScore best captures the underwriting-relevant reliability ordering.

F Dataset Construction Prompt Sample

F.1 Personal Client Profile: Not-Overdue Clients

Table 10 documents the localized and translated prompt configuration used to generate synthetic transaction logs for the CCLUPE benchmark.

Augmentation Strategy: Pattern-Consistent Flow Generation
System Overview: You are a financial credit analysis expert. Based on the provided raw bank transaction logs of high-credit-rating clients, you must synthesize highly realistic new client data that maintains the original statistical distribution but simulates unique behavioral features.
Translation of Input Prompt: "The following content contains the complete raw transaction data for a non-overdue personal client. Please meticulously analyze its integrity, field semantics, data distribution, and transaction characteristics (e.g., consistent repayment records, absence of overdue status flags): *[Raw Data Context Provided Hierarchically]* Please generate a new set of personal transaction data for **Person ID: {ID}**, strictly adhering to: 1. **Structural Integrity:** Headers, field counts, and data types must be identical to the original; do not add or remove columns. 2. **Volume:** Generate between 50 to 300 independent transaction records. 3. **Temporal Distribution:** Transactions should be evenly distributed throughout the 2023 calendar year (Jan–Dec). 4. **Data Variance:** Textual fields (merchant names, remarks) must be distinct from the original samples to avoid data leakage. 5. **Logic Consistency:** No null values. Amounts must fluctuate realistically, transaction types must be diverse, and logical relationships between associated fields (e.g., Balance = Prev_Balance + Amount) must remain self-consistent. Return the result directly in tab-separated format (Header + Data) without additional conversational text."

Table 10: LLM Prompt for Generating Non-Overdue Personal Credit Logs

F.2 Synthesis and Evaluation Review

Pattern Analysis: The generation process for non-overdue clients focuses on **Repayment Regularity** and **Cash Flow Stability**. Unlike overdue samples, these records prioritize the cyclical nature of income (e.g., monthly payroll) and the punctuality of high-priority expenses like credit card repayments and utilities.

Merchant Diversification: To ensure the benchmark's robustness, the LLM is instructed to generate a wide variety of merchant names (Textual fields). This forces models to generalize spending habits rather than relying on keyword matching of known "safe" merchants.

Logical Validation: Post-generation scripts (as seen in the Python implementation) enforce a strict volume check ($[50, 300]$ rows). This ensures that the generated credit history is long enough to exhibit long-term financial patterns, such as "Nighttime Spending" or "Investment Redemption," which are key sub-domains of the CCLUPE benchmark.

Temporal Coherence: By enforcing a distribution across the full year of 2023, the dataset captures potential seasonal behaviors (e.g., increased utility bills in winter or travel expenses in summer) that are critical for evaluating an LLM's under-

standing of time-series transactional data.

F.3 Personal Client Profile: Overdue Clients

Table 11 details the prompt engineering strategy used to generate synthetic transaction logs for clients with historical default patterns. The focus is on simulating behavioral indicators of credit risk without explicitly labeling transactions as "overdue."

Augmentation Strategy: Feature-Induced Risk Simulation
System Overview: You are a senior risk management consultant. Your task is to analyze raw bank statements of clients who eventually defaulted and synthesize realistic new datasets for **ID: {ID}**. The generated data must exhibit subtle precursors to financial distress while remaining structurally valid.
Translation of Input Prompt: "The following content provides the complete raw transaction data for a personal client with a history of overdue payments. Please analyze its structure, field semantics, data distribution, and specific transaction characteristics: *[Full Transactional Context Provided]* Generate a new set of original transactional records for **Person ID: {person_id}**, strictly adhering to: 1. **Structural Adherence:** Column headers, field counts, and data types must perfectly match the original; no deviations permitted. 2. **Data Volume:** Synthesize between 50 to 300 independent transaction records. 3. **Temporal Spread:** Ensure transactions are evenly distributed across the 2023 calendar year (Jan–Dec). 4. **Field Uniqueness:** All text-based fields (e.g., merchant names, remarks) must be entirely different from the source samples to ensure data diversity. 5. **Arithmetic Logic:** Ensure zero null values. Balance trajectories must be mathematically sound (e.g., refund amounts must be negative and reflect in the balance logic). 6. **Latent Feature Construction:** The content should subtly reflect characteristics associated with potential delinquency (e.g., irregular income, high-frequency small loans, or spikes in nighttime spending), **without** explicitly using words like 'Overdue' or 'Default'. Return ONLY tab-separated content (Header + Data) with no conversational preamble."

Table 11: LLM Prompt for Generating Overdue-Prone Personal Credit Logs

F.4 Risk Pattern Synthesis and Evaluation

Behavioral Indicators: The synthesis of "Overdue" profiles shifts from regular consumption to **Spike Recognition** and **Financial Stretching**. The model is tasked with generating logs that include "Microloan" signals and "Spikes" in spending that exceed typical balance buffers, simulating the "liquidity pressure" mentioned in the CCLUPE core domains.

Implicit Feature Learning: A key requirement of the prompt is the exclusion of explicit "Overdue" keywords. This is designed to test the LLM's capability in **Causal Inference** and **Multi-stage Reasoning**. Benchmarking models must infer risk from the sequence and nature of transactions (e.g., a sudden increase in revolving credit repayments) rather than simple text classification.

Contextual Integrity: By utilizing full-year records (Jan–Dec 2023), the generated samples provide sufficient historical depth to evaluate "Seasonality" and "Repayment Behavior Patterns." The Python implementation enforces a strict tab-separated format to ensure that multi-modal models can accurately parse the tabular time-series data without alignment errors.

Data Leakage Prevention: The strict requirement for "completely different merchant names" ensures that the benchmark remains a test of **latent financial behavior recognition** rather than a retrieval task. This forces the evaluator model to analyze the *intent* behind the transaction (e.g., a high-interest lender) regardless of the merchant's pseudonym.

F.5 Micro-Enterprise Profile: Not-Overdue SME Clients

Table 12 documents the prompt engineering framework for generating synthetic transaction logs for Small and Micro Enterprises (SMEs) with healthy credit profiles. The logic centers on business-related identifiers and operational health signals.

Augmentation Strategy: Operational Attribute Simulation
System Overview: You are an expert in industrial finance and SME credit underwriting. Your task is to analyze the bank statements of healthy micro-enterprises (including individual businesses and small factories) and synthesize a new dataset for **SME ID: {micro_id}**. The logs must reflect robust commercial activity and financial discipline.
Translation of Input Prompt: "The following content contains the complete raw transaction data for a non-overdue micro-enterprise/individual business owner. Please carefully analyze its structural integrity, field definitions, and typical indicators of healthy business operations: *[Full SME Transactional Context Provided]* Generate a new set of original transaction data for **SME Subject ID: {micro_id}**, strictly adhering to: 1. **Structural Preservation:** Column headers, field counts, and data types must be identical to the provided original; no modifications allowed. 2. **Business Volume:** Generate 50 to 300 independent records that reflect the typical transaction frequency of a healthy SME. 3. **Temporal Consistency:** Transactions must span the full year of 2023 (Jan–Dec). 4. **Merchant Diversity & SME Identity:** – *Counterparty Names:* Must incorporate SME-specific identifiers such as 'Co., Ltd.', 'Trading Store', 'Individual Proprietorship', 'Factory', etc. – *Remarks:* Must indicate business-specific scenarios (e.g., procurement, inventory, payroll, operational rent). 5. **Financial Logic:** No empty values; transaction amounts must be consistent with micro-enterprise scales. Balance logic must remain self-consistent throughout the sequence. Return the response using a tab-separated format (Header + Data) with no additional explanation."

Table 12: LLM Prompt for Generating Healthy Micro-Enterprise Credit Logs

F.6 Business Context and Domain Analysis

Operational Health Recognition: For SMEs, the prompt focuses on **Stability and Regularity of Cash Flow**. Unlike personal accounts, SME logs prioritize "Essentials" (operational rent, utilities) and "Repayments" (business loans). This is essential for the CCLUPE benchmark to evaluate if a model can distinguish between personal consumption and commercial procurement.

Identity-Specific Labeling: As specified in the Python implementation's prompt, the LLM must generate counterparties with suffixes such as "Co., Ltd." or "Factory." This ensures that the generated data captures the "Alignment" domain of CCLUPE, verifying whether the transactional behavior (e.g.,

bulk procurement of raw materials) is actually consistent with the stated business type of the micro-subject.

Scalability and Diversity: The use of ‘ThreadPoolExecutor‘ to generate data for 225 subjects (IDs 276–500) ensures that the benchmark provides a statistically significant population of SMEs. This volume allows for rigorous testing of an LLM’s Analysis & Synthesis capabilities by presenting various SME industry types, from services to small-scale manufacturing.

Benchmark Strategic Value: In the Chinese credit market context, where formal credit scores for small business owners are often incomplete, these transaction logs serve as the primary evidence of creditworthiness. By ensuring "completely differentiated textual fields," the prompt forces the evaluator model to recognize "Specialised Cash Flow Conduct," such as seasonal restocking patterns around the National Day or Spring Festival, rather than relying on fixed merchant names.

F.7 Micro-Enterprise Profile: Overdue SME Clients

Table 13 documents the prompt engineering strategies used to simulate transactional distress in Small and Micro Enterprises (SMEs). This prompt requires the LLM to blend "operational scenarios" with "latent delinquency signatures" while maintaining business-specific terminology.

Augmentation Strategy: Business Cycle & Distress Simulation
System Overview: You are a specialist in commercial credit risk evaluation. Your objective is to analyze transaction logs of SMEs (small companies, trading stores, or factories) with delinquency histories and synthesize a new dataset for **SME ID: {micro_id}**. The logs must reflect genuine business operations while embedding precursors to credit failure.
Translation of Input Prompt: "The following content provides the complete raw transaction data for an overdue micro-enterprise/individual business owner. Please meticulously analyze its structural integrity, field semantics, specific business transaction characteristics, and patterns associated with credit delinquency: *[Full SME Transactional Context and Default History Provided]* Generate a new set of original transaction data for **SME Subject ID: {micro_id}**, strictly adhering to: 1. **Structural Fidelity:** Column headers, field counts, and data types must precisely match the original; do not alter the schema. 2. **Operational Volume:** Generate 50 to 300 independent records that reflect the business activities of an SME under financial stress. 3. **Temporal & Cyclical Distribution:** Spread transactions across Jan–Dec 2023, ensuring business-specific cycles (e.g., seasonal procurement) are visible. 4. **Differentiated SME Identity:** – *Counterparty Names:* Must include commercial identifiers such as 'Company', 'Trading Firm', 'Proprietorship', or 'Factory'. – *Remarks:* Must describe plausible commercial scenarios (e.g., supply chain payments, rent, or logistics fees). 5. **Logical Consistency:** No empty values allowed. Transaction amounts must be appropriate for small-scale operations. Associated fields (e.g., balance and loan repayment logic) must be calculated correctly. 6. **Latent Overdue Features:** Infuse the logs with subtle indicators of financial distress, such as high interest-debt service, irregular revenue patterns, or excessive inventory costs relative to sales, **without explicitly using words like 'overdue' or 'default'.** Return the response as tab-separated values (Header + Data) without any conversational text."

Table 13: LLM Prompt for Generating Overdue-Prone Micro-Enterprise Credit Logs

F.8 Synthesis Logic and Macro-Domain Review

Analysis of Financial Distress: For SME subjects, the prompt focuses on Cash Flow Structure and Concentration. The transition to "overdue" is often represented by a shift in counterparty dependencies or a decline in "stability and regularity" of revenue. These latent features allow the benchmark to evaluate if LLMs can perform Multi-stage Reasoning to detect business failure precursors.

Contextual Business Scenarios: In accordance withRule 4 of the Python script, the LLM is forced to define "remarks" that ground the data in reality. For overdue SMEs, this might include "Urgent Inventory Repayment" or "Property Rent Adjustments," which test the LLM's Comprehension & Calculation capabilities regarding business operating margins and liquidity pressure.

Benchmark Consistency: The generation of data for 225 SME subjects (IDs 26–250) mirrors the scale of the non-overdue population. This balanced distribution is vital for calculating the Log-Score mentioned in the CCLUPE core framework, as it ensures the model is evaluated on its ability to mini-

mize "misunderstanding penalties" (false positives) in a mixed dataset.

Domain-Specific Penalties: Because SMEs have own Global Interpreter Lock (GIL) related complexities in real-world data, the synthetic generation ensures Arithmetic Logic consistency (e.g., negative refund values). This provides a clean but challenging environment (text, tabular, and time-series) to test the robustness of financial LLMs against professional underwriting standards.

F.9 Question Generation: Personal Underwriting Audit

Table 14 documents the prompt used to transform raw personal transaction logs into a structured 15-question credit audit. The prompt enforces hierarchical categorization and cognitive complexity mapping.

F.10 Cognitive Complexity and Logical Parsing

Cognitive Dimension Distribution: The generation logic specifically targets the "Integrative Nature" of credit evaluation. As seen in the Python implementation's ‘parse questions‘ function, questions are categorized into four cognitive levels following a modified Bloom's taxonomy. This ensures the benchmark evaluates not just data retrieval, but the model's ability to synthesize evidence (e.g., "Analysis & Synthesis") to form a credit opinion.

Anti-Leakage Design: A distinct feature of the CCLUPE benchmark is the prohibition of specific numeric values in the question stems (absolute prohibition clause). By forcing the LLM to use "fuzzy positioning" (e.g., "the last transaction at merchant X"), the benchmark ensures that the evaluator model must truly parse the table to find the values, rather than relying on values mentioned in the question to guess the answer.

Difficulty Stratification: The prompt enforces a 5:5:5 ratio for difficulty. This stratification allows for a granular performance analysis. For example, "High Difficulty" questions require $\geq$5 calculation steps (e.g., cross-referencing multiple months of balance and aggregate repayments), which is often where standard LLMs fail on the CCLUPE leaderboard.

Domain Perspective Mapping: By mapping questions to "8 Underwriting Perspectives," the benchmark provides an industry-standard review. The ‘eight category‘ logic in the script ensures that "Other" categories (like "Buy Now Pay Later" or "Nighttime Spending") are explicitly captured, providing professional insights into latent borrower risk that traditional scorecards might overlook.

Instruction Set: Hierarchical Underwriting Question Synthesis
Task Overview: Based on the provided bank statement data, generate 15 specialized multiple-choice questions (at least one multi-choice). Questions must be grounded in the data but satisfy the "Zero Data Leakage" rule for the question stem.

Translation of Input Prompt (Core Requirements):
"You are provided with a personal transaction table. Generate 15 questions based on the following classification logic:
 	
1. Primary Category (8 Underwriting Perspectives):
Select from: Consumption Stability, Essential Expenditure Regularity, Repayment Patterns, High-Risk Merchant Transactions, Discretionary Concentration, Transaction Failures & Liquidity Pressure, Fund Retention Habits, and 'Others' (specifically labeled in parentheses).
 	
2. Secondary Category (24 Classification Logics):
Annotate each question with its specific logic mapping (e.g., 'Consumption Stability-Payroll regularity').
 	
3. Strict Prohibitions:
– **No Specific Numbers in Stems:** Do not use amounts or specific dates in the question text. Use relative references like 'the repayment recorded in the table' or 'a certain date.'
– **No Intermediate Disclosure:** Do not reveal the number of transactions required for calculation in the stem.
 	
4. Complexity & Cognition Mapping:
– **Difficulty Levels (5:5:5 ratio):** Low ($\geq$2 computation steps), Medium ($\geq$3 steps), High ($\geq$5 steps).
– **Cognitive Dimensions:** Memory & Recognition, Comprehension & Calculation, Analysis & Synthesis, Evaluation & Inference.
 	
Standardized Format:
`8 Categories: [Type] | 24 Logics: [Class-Subclass] | Difficulty: [L/M/H] | Cognition: [Dimension]`
[Question Body]
a. [Opt] b. [Opt] c. [Opt] d. [Opt]
`Correct Answer: X"`

Table 14: LLM Prompt for Generating Personal Credit Underwriting Questions

F.11 Question Generation: Micro-Enterprise Credit Audit

Table 15 displays the prompt configuration for generating 15 specialized credit audit questions from raw SME operational logs. The prompt focuses on distinctive SME domains such as supply chain cycles and liquidity turnover.

Instruction Set: Complex Commercial-SME Underwriting Synthesis
Task Overview: Analyze the provided micro-enterprise operational logs. Generate 15 multiple-choice questions (at least two multi-select). Questions must revolve around commercial scenarios: revenue collection frequency, procurement vs. sales alignment, upstream/downstream payment cycles, and tax/utility regularity.
Translation of Input Prompt (Core Requirements): "You are provided with an SME operational transaction table. Generate 15 questions based on the following classification structure: **1. Primary Category (7 SME Audit Perspectives):** Must cover all 20 sub-categories across 7 domains: – *Revenue/Expense Stability:* Monthly frequency and industry-specific cost-to-income ratios. – *Cash Flow Health:* Non-operational funds vs. operational capital and turnover efficiency. – *Business Activity:* Daily transaction volume, QR code payment features, and nighttime trading records. – *Counterparty Risk:* Relying on single-source suppliers and related-party transaction ratios. – *Payment Management:* Account receivables/payables aging and tax/social security punctuality. – *Specialized Behaviors:* Cash-out patterns, seasonal fluctuations, and transaction disputes. – *Loan Compliance:* Operational loan repayment history and utility regularity. **2. Functional Constraints (Anti-Guessing):** – **Zero-Data Stem:** No specific amounts, dates, or merchant names in the stems. Use relative locators like 'a certain partner recorded in the table.' – **Step-Based Difficulty (5:5:5 Ratio):** *Low:* $\geq$2-step judgment; *Medium:* $\geq$3-step analysis; *High:* $\geq$5-step calculation/inference. **Cognitive Dimensions:** Memory & Recognition, Comprehension & Calculation, Analysis & Synthesis, Evaluation & Inference."

Table 15: LLM Prompt for Generating Micro-Enterprise (SME) Credit Audit Questions

F.12 Commercial Reasoning and SME Domain Analysis

Operational Scenario Alignment: As illustrated in the Python implementation, the prompt forces the LLM to avoid personal consumption narratives. Instead, it prioritizes Industry-Specific Ratios (e.g., procurement as a percentage of revenue). This is a vital metric in the CCLUPE benchmark for testing whether a financial LLM can distinguish between a healthy business cycle and a fraudulent or declining one.

Multi-Step Calculation Logic: The "High Difficulty" questions in the SME category are significantly more complex than personal ones. According to the script's instruction for "$\geq$5 steps," a single question may require the evaluator to: 1) identify all QR code receipts, 2) aggregate them by month, 3) calculate the mean, 4) compare that to the supplier payment cycle, and 5) evaluate the liquidity buffer. This directly tests the LLM's*Multimodal/Multi-stage Reasoning capabilities.

Anti-Leakage & Qualitative Options: Consistent with the "Absolute Prohibition" clause in the prompt, options are often qualitative (e.g., "Matched with monthly revenue" vs. "Significantly exceeding revenue") rather than just numeric. This ensures that the model is evaluated on its Financial Knowledge and its ability to perform Causal Inference across the transaction timeline.

Diverse Commercial Perspectives: By covering 20 sub-categories (e.g., "Upstream/Downstream Aging Management"), the CCLUPE benchmark ensures that data-synthesis does not overlook the "latency" and "receivables" inherent to SME credit risk. The ‘parse questions‘ function in the provided script ensures that these labels are captured for diagnostic error analysis, allowing researchers to pinpoint which specific business logic causes LLM "misunderstandings."

G Dataset Details

The CCLUPE benchmark adopts a hierarchical taxonomy to map diverse transactional behaviors into actionable credit-risk signals. As detailed in the expert consultation process, indicators are partitioned into core knowledge domains and specialized behavioral sub-categories.

Main Behavioral Domains The core evaluation focuses on seven primary domains that represent the most significant predictors of creditworthiness in the Chinese market:

- **Consumption Stability:** Evaluates the variance and consistency of spending over time to infer income reliability.
- **Essential Expenditure Regularity:** Monitors fixed costs (utilities, rent, tax) that indicate fundamental financial discipline.
- **Repayment Behavior Patterns:** Specifically tracks loan-related logs to identify punctuality and debt-servicing intent.
- **High-Risk Merchant Activities:** Identifies interactions with gambling, high-interest lending, or speculative investment platforms.
- **Discretionary Spending Concentration:** Analyzes the proportion of non-essential vs. essential spending to assess luxury-led risk.
- **Liquidity Pressure and Transaction Failures:** Records insufficient fund errors and abrupt balance drops.
- **Funds-Flow Timing and Consumption Lag:** Analyzes the delta between income receipt and major expenditures.

Indicator Mapping Table 16 provides a comprehensive overview of the indicators categorized by their relevance to Personal and Micro-enterprise profiles.

Category	Behavioral Signals and Specialized Questions
Main Domains	
Consumption Stability	Frequency of daily spending, month-over-month variance, and lifestyle consistency.
Expenditure Regularity	Timeliness of utility payments (electricity, water, heating), and insurance continuity.
Repayment Patterns	Credit card repayment cycles, bill-period alignment, and partial vs. full settlement.
High-Risk Activities	Transactions involving speculative assets, high-leverage platforms, or risky merchants.
Spending Concentration	Counterparty dependency, merchant relationship depth, and category-specific spikes.
Liquidity Pressure	Transaction failures (insufficient funds), frequent small-value borrowing, and balance volatility.
Funds-Flow Timing	Income-to-expenditure lag, salary deposit recency, and cash flow alignment.
Secondary/Other	
Digital Footprint	Subscription service continuity, app engagement frequency, and digital wallet usage trends.
Temporal Patterns	Nighttime/early-morning consumption, seasonal spending spikes, and holiday surges.
Locational Logic	Off-location (travel) consumption analysis and cross-border spending patterns.
Social/Employment	Intergenerational transfer support, part-time employment markers, and payroll regularity.
Operational Health	(SME only) Inventory restock cycles, supplier relationship stability, and business-type alignment.
Transaction Quality	Refund rates, transaction reversals, and penalty fee occurrences.

Table 16: Detailed taxonomy of CCLUPE behavioral-consumption indicators across main and secondary categories.

Cognitive Level Mapping In accordance with Bloom's Taxonomy for financial reasoning, each question based on the above indicators is mapped to a cognitive depth:

1. **Memory & Recognition:** Identifying specific transaction types or merchant names within the log.
2. **Comprehension & Calculation:** Summarizing total spending or calculating debt-to-income ratios from raw entries.
3. **Analysis & Synthesis:** Detecting latent patterns (e.g., "seasonal restocking") across multiple months of data.
4. **Evaluation & Inference:** Making high-level risk judgments or predicting potential default based on observed behavioral anomalies.

Our dataset organizes behavioral-consumption indicators into 7 main categories and one other categories including the indicators that are considered to be minor-relevant. Each specific subcategories facilitate precise characterization and analysis of consumer financial behavior. The main categories include Consumption Stability, Essential Expenditure Regularity, Repayment Behavior Patterns, High-Risk Merchant Activities, Discretionary Spending Concentration, Liquidity Pressure and Transaction Failures, Funds-Flow Timing and Consumption Lag, Others. While the other catergories include Small Loan Transactions, Subscription Service Continuity, Late Payment Penalty, etc. The specific categories are shown as figure 7 shows. Each main category is further divided into specialized questions that capture distinct behavioral signals relevant for credit-risk inference.

Dataset Overview
Main Categories
Consumption Stability
Essential Expenditure Regularity
Repayment Behavior Patterns
High-Risk Merchant Activities
Discretionary Spending Concentration
Liquidity Pressure and Transaction Failures
Funds-Flow Timing and Consumption Lag
Others
Small Loan Transactions
Subscription Service Continuity
Late Payment Penalty
Seasonal Spending Behavior
Installment Payment
Part-time Employment Indicators
Sudden Large-Amount Spending
Refunds and Transaction Reversals
Spending Diversification and Merchant Relationships
Cross-Border Consumption Characteristics
Small Loan Transactions
Subscription Service Continuity
Late Payment Penalty
Seasonal Spending Behavior
Installment Payment
Part-time Employment Indicators
Sudden Large-Amount Spending
Refunds and Transaction Reversals
Spending Diversification and Merchant Relationships
Nighttime and Early-Morning Consumption
Consumption Activity and Frequency
Consumption Cycles and Trigger Patterns
Off-Location Consumption Analysis
Intergenerational Consumption Support
Income-Consumption Cycle Alignment

Table 17: Behavioral-consumption indicators including 7 main categories and one other categories.

H Statistical Characteristics

The CCLUPE benchmark consists of 4,062 high-quality samples, distinctively designed to bridge the gap between general financial NLP and specialized credit risk reasoning. The released logs are synthetic, but they are calibrated from authentic transaction patterns and expert underwriting heuristics rather than generated without domain anchors. Personal clients (N=2,931) reflect individual consumption and repayment hygiene, while SME (N=1,131) focus on operational cash flows, business continuity, and industry-specific financial health. The dataset features a hierarchical annotation framework spanning 7 core knowledge domains and 16 fine-grained sub-domains. As shown in Table 18, the *Stability and Regularity of Cash Flow* domain represents the largest segment (34.6%), mirroring real-world credit underwriting priorities where consistent revenue is the primary indicator of repayment capacity.

Core Domain	**Sub-Domain Examples**	**Personal**	**SME**	**Total**
Stability & Regularity	Payroll, Revenue Stability	980	428	1,408
Cash Flow Conduct	Specific Industry Patterns	310	210	520
Structure & Concentration	Merchant Diversity, Counterparty Risk	645	212	857
Time Characteristics	Nighttime Trading, Seasonality	82	36	118
Specialised Conduct	Cross-border, Tax Punctuality	130	80	210
Liquidity & Pressure	Fund Reserves, Overdue Precursors	680	136	816
Others	Miscellaneous Risk Factors	104	29	133
Total		**2,931**	**1,131**	**4,062**

Table 18: Comprehensive breakdown of the CCLUPE Dataset across domains and client types.

To evaluate deep reasoning rather than simple retrieval, we map questions to four cognitive levels. *Analysis & Synthesis* (43.8%) is the most frequent level, requiring models to integrate multi-source data (textual descriptions and time-series logs) to form a coherent credit opinion.

I Credit Review Domain Knowledge Domain Analysis

The fine-grained results in Table 19 reveal a stark divide in model performance across different credit indicators. Proprietary frontiers, particularly the **Claude Sonnet 4.5** and **GPT-5** series, demonstrate a localized mastery of acute risk signals. For example, GPT-5-Think achieves a state-of-the-art 62.5% in *Spikes* detection and 53.0% in *Risky* transaction identification. These high scores suggest that large-scale reinforcement learning from human feedback (RLHF) has successfully equipped these models with the ability to identify sudden behavioral anomalies in financial logs.

The Bottleneck of Concentration and Nonessential Spending Despite successes in anomaly detection, nearly all models struggle with structural indicators such as **Concentration (Conce.)** and **Nonessential Spending (Noness.)**. In the concentration domain, aggregate scores for open-source models often fall below 20%, with even high-performing reasoning models like GPT-4o-Think managing only 22.4%. This performance floor suggests that models find it significantly more difficult to execute the multi-step aggregation required to assess portfolio-level risks or to distinguish between essential business opex and non-essential outflows, highlighting a critical area for future dataset-led alignment.

Efficacy of Domain-Specific SFT A pivot point in our analysis is the performance of **Qwen 3 Max**, which serves as a compelling case for open-source model. Despite a smaller parameter footprint than the Llama-3.1-70B model, Qwen 3 Max significantly outperforms it across nearly all dimensions, notably in *Spikes* (31.8%) and *Personal* (24.5%).

Impact of Extended Reasoning The inclusion of 'Think' (internal reasoning) variants generally elevates performance in synthesis-heavy tasks like *Repayment (Repay.)* and *Asset Overshooting (Overs.)*. For instance, Claude-Sonnet-4.5-Think provides a substantial +20.0 point boost in *Asset Overshooting* over the standard GPT-4o. These reasoning-enhanced models are better at maintaining the long-range logical chains necessary to reconcile a borrower's historical debt obligations with their current liquidity trajectories. However, the persistent inaccuracy in the *Others* category across all models indicates that even extended reasoning cannot yet fully resolve the interpretive ambiguity found in atypical or less-structured transaction types.

Method	Align.	Conce.	Micro.	Spikes	Refun.	Night.	Overs.	Seaso.	Liqud.	Repay.	Essen.	Risky	Noness.	Penal.	Subsc.	Others
Proprietary Models																
Gemini 3	29.1	8.8	13.3	31.2	**48.3**	18.2	26.3	23.1	28.2	28.8	30.1	36.7	15.9	24.5	39.3	29.5
Claude Sonnet 4.5	47.3	41.4	**40.0**	50.0	32.8	31.8	35.0	33.6	28.7	32.6	33.8	48.9	39.7	30.2	30.5	32.9
Claude Sonnet 4.5 Think	**54.1**	**48.3**	36.7	50.0	37.8	**47.0**	**55.0**	**44.9**	38.4	43.0	42.8	**54.6**	48.3	34.0	35.6	38.7
GPT 5	39.7	32.8	33.3	56.2	35.3	28.8	35.0	30.8	30.0	32.1	35.8	39.2	37.1	26.4	35.6	30.6
GPT 5 Think	45.2	46.6	36.7	**62.5**	41.2	43.9	50.0	40.2	**42.3**	**43.2**	**44.8**	53.0	**49.3**	**45.3**	**40.7**	**39.8**
GPT 4o	33.6	13.8	10.0	43.8	20.2	24.2	25.0	13.1	23.6	22.3	23.8	33.8	14.4	17.0	25.4	28.2
GPT 4o Think	32.2	22.4	10.0	25.0	37.0	37.9	40.0	24.3	25.7	25.5	29.2	35.7	24.8	17.0	28.8	31.0
Open-source Models																
DeepSeek V3.2	27.4	12.1	23.3	12.5	31.1	25.8	**30.0**	17.8	27.8	27.8	30.2	32.4	15.7	26.4	25.4	24.3
Kimi K2	26.0	6.9	16.7	18.8	35.3	**27.3**	25.0	15.9	23.7	19.7	22.6	31.9	11.2	9.4	18.6	25.4
Qwen3 Max	30.8	15.5	16.7	**31.8**	**49.6**	21.2	25.0	**26.2**	**30.7**	**29.1**	**30.4**	**35.7**	20.0	**24.5**	**33.9**	**30.8**
DeepSeek V3	22.6	15.5	**30.0**	31.2	20.2	16.7	20.0	16.8	23.8	18.2	22.6	29.7	13.6	17.0	20.3	22.4
GLM 4.6	23.3	15.5	13.3	25.0	26.9	21.2	20.0	15.9	23.7	19.1	23.6	26.5	16.3	13.2	25.4	24.5
Qwen3 235B	11.0	6.7	12.0	16.7	12.4	7.8	10.5	14.1	9.2	15.2	12.1	16.0	7.4	13.2	17.3	9.7
Qwen3 30B	**37.0**	**31.0**	23.3	25.0	25.2	24.2	**30.0**	19.6	20.0	22.1	21.4	34.1	**23.2**	22.6	27.1	22.4
Qwen3 4B	13.7	12.1	13.3	6.2	9.2	7.6	10.0	10.3	8.4	9.0	10.0	15.1	10.1	13.2	5.1	8.0
Llama 3.3 70B	28.8	22.4	13.3	31.2	18.5	21.2	**30.0**	13.1	21.3	19.5	15.3	28.9	18.1	11.3	18.6	18.4
Llama 3.1 70B	19.9	10.3	6.7	18.8	10.1	13.6	20.0	12.1	11.7	14.6	12.2	25.9	11.2	15.1	15.3	14.0
Llama 3.1 8B	6.2	6.9	10.0	6.2	5.9	4.5	10.0	6.5	6.2	7.7	7.3	8.9	10.7	11.3	6.8	6.6
Fin R1	16.4	17.2	16.7	18.8	12.6	10.6	15.0	20.6	10.9	14.3	13.6	15.1	15.7	13.2	15.3	10.0

Table 19: **Domain-specific Performance Comparison** across different sectors and industries.

J Dataset Comparison

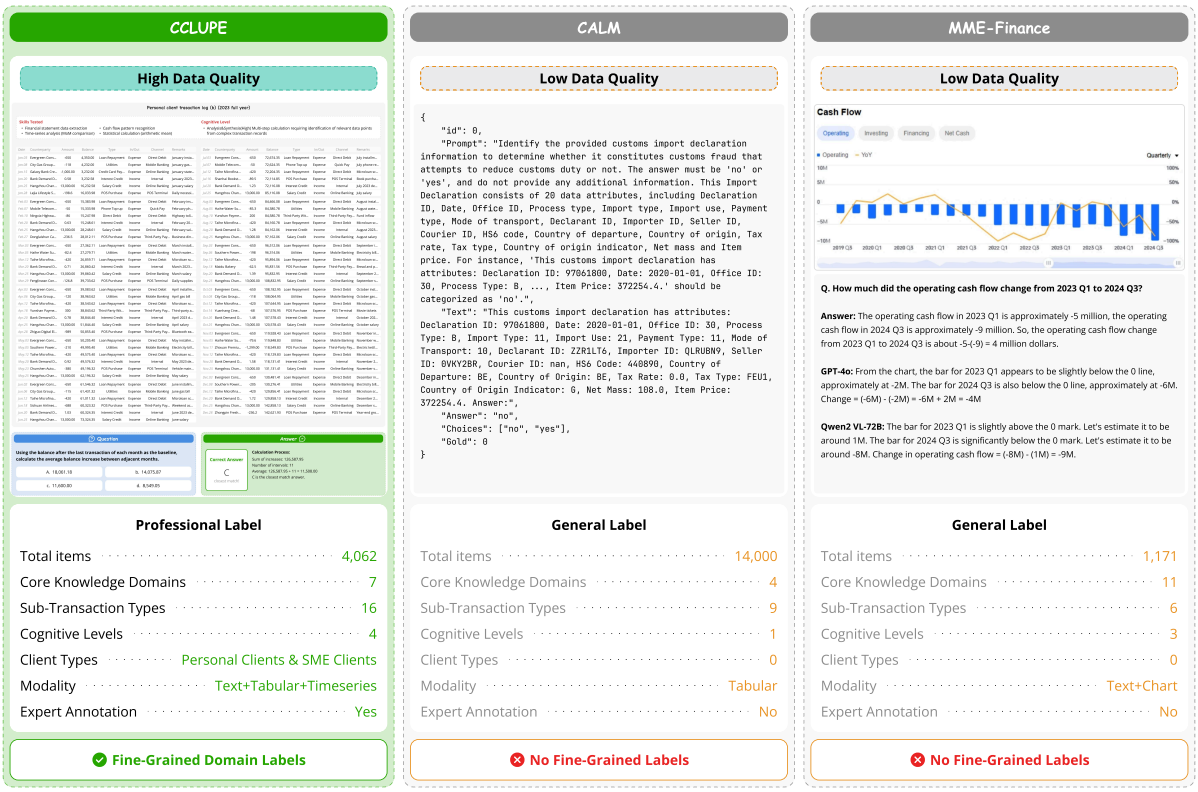

Figure 4: Data Comparision with Related Works

As illustrated in Table 1, we present a comprehensive comparison of CCLUPE with existing benchmarks relevant to financial and credit assessment domains. The comparison encompasses dataset scale, annotation quality, domain specificity, and modality coverage, revealing distinct characteristics and application scenarios for each benchmark.

CCLUPE (Ours) represents a significant advancement in credit assessment benchmarking, comprising 4,062 samples with expert-validated annotations. A defining characteristic of our dataset is its specialized focus on transaction log analysis, which constitutes the core data source in real-world credit decisioning workflows. CCLUPE evaluates heterogeneous evidence sources: textual credit context, structured fields, and time-series transactions, to enable holistic evaluation of consumer financial behavior. The released logs are synthetic but calibrated from authentic underwriting patterns through extensive consultations with industry practitioners, ensuring alignment with real credit risk-control workflows while remaining shareable.

CALM offers the largest scale among existing credit-related benchmarks with 14,000 samples and expert annotations. However, its scope is confined to loan approval decisions with exclusively tabular modality, limiting its capacity to assess models' reasoning capabilities across heterogeneous data formats. While CALM provides valuable resources for binary classification tasks, it lacks the behavioral consumption indicators and temporal transaction patterns essential for comprehensive credit risk inference.

Loan Approval Benchmark contains 3,065 samples targeting loan approval scenarios with text and tabular modalities. Notably, this benchmark lacks expert annotations, potentially compromising the reliability and domain authenticity of its ground-truth labels. Its coverage remains restricted to conventional approval decisions without incorporating nuanced behavioral signals derived from transaction histories.

FinRAGBench-V provides 1,394 expert-annotated samples combining text and tabular data. While it demonstrates rigorous annotation quality, the dataset is designed for general financial RAG tasks without domain-specific focus on credit

assessment, limiting its applicability for evaluating credit-oriented AI systems.

MMMU (Finance) and MME-Finance represent general-purpose financial multimodal benchmarks with 390 and 1,171 samples respectively. Both datasets lack expert annotations and are primarily oriented toward visual chart interpretation and general financial knowledge assessment rather than credit-specific workflows. Their modality coverage (text+image and text+chart) does not encompass the tabular and time-series formats predominant in credit risk analysis.

This comparative analysis underscores CCLUPE's distinctive positioning within the landscape of financial AI benchmarks. While existing datasets either emphasize general financial knowledge or provide limited credit-specific coverage, CCLUPE bridges this gap by delivering: (1) substantial scale with expert-validated annotations, (2) specialized focus on transaction-level behavioral analysis, (3) evaluation over heterogeneous evidence sources aligned with real credit assessment pipelines, and (4) hierarchical behavioral indicators derived from industry best practices. These characteristics establish CCLUPE as a rigorous and practically relevant benchmark for advancing AI capabilities in consumer credit evaluation.

K Model Configurations

We conducted experiments across a diverse array of 15+ LLMs, encompassing proprietary frontier models and open-source reasoning or finance-adapted systems (Liu et al., 2024a; Team et al., 2025; Yang et al., 2025; Zeng et al., 2025; Liu et al., 2025b). All models were evaluated between January and February 2025.

Model Name	Developer	Scale/Type	Access Method
Proprietary			
Gemini 3	Google	Unknown	Vertex AI API
Claude 4.5 Sonnet	Anthropic	Unknown	Anthropic API
GPT-5	OpenAI	Unknown	OpenAI API
Open-Source			
DeepSeek V3	DeepSeek	671B (MoE)	Local (vLLM)
Llama 3.3	Meta	70B	Local (vLLM)
Qwen 3	Alibaba	4B, 30B, 235B	Local (vLLM)
Fin R1	Financial AI Lab	32B (Distilled)	Local (vLLM)

Table 20: Model architectures and parameter scales evaluated in the benchmark.

Local models were deployed on a high-performance cluster. To eliminate stochastic variance and ensure focus on reasoning capability, the temperature was set to 0.0. We utilized the `vLLM` engine to handle the long-context requirements of 2023 full-year transaction logs, which can exceed 3,000 tokens per prompt. This setup is also motivated by the recent rise of open reasoning-oriented model families in general LLM evaluation, including systems such as DeepSeek-R1 (DeepSeek-AI et al., 2025).

Parameter	Configuration Value
GPU Hardware	8 × NVIDIA H800 (80GB)
Operating System	Ubuntu 22.04 LTS
Inference Engine	vLLM (v0.6.3) / CUDA 12.4
Decoding Strategy	Greedy Search ($T = 0$)
Max Context Length	16,384 Tokens
Precision	Bfloat16

Table 21: Inference parameters and environment details.

A critical challenge in credit risk assessment is the cost of "misunderstanding" risks. The appendix uses exactly the same released evaluator as Sec. 4.1–5. For consistency, the notation is: $\mathcal{G}_q$ for the ground-truth option set of item q, $\mathcal{P}_q$ for the model-predicted option set, $S_{item}^{(q)}$ for the item-level precision-adjusted score, S_{dom} for the domain-balanced macro score, S_{cog} for the cognitive-balanced macro score, $S_{unbiased}$ for their convex combination, and $\mathcal{U}$ for the dataset-level misunderstanding rate. The released evaluator fixes $\lambda = 2$, $\eta = 0.5$, and $\delta = 2$. We repeat these values here to avoid notation drift between the main text and appendix; the detailed definitions remain in Eqs. 4.1–5.

L Sample Question

Personal Client Transaction Log (2023 Full Year Records)

Volume & Timeframe

The dataset comprises transaction records spanning the entire 2023 calendar year. It contains a total of 60 transaction entries.

Data Scope

- The entries cover various financial activities, including both income and expenses.
- Due to space constraints on the page, the visible section above is a truncated representation focused on entries relevant to the analysis.

Date	Counterparty	Amount	Balance	Type	In/Out	Channel	Remarks
⋮							
Jan.08	Yuanhang Tech Co. Sal...	9200	13131.5	Payroll	Income	Online Bank	January salary
Jan.12	Hengxin CC Repayment	-1800	11331.5	Repayment	Expense	UnionPay	CC Bill Period 01
Jan.15	City Gas Corp	-128.4	11203.1	Utility	Expense	Mobile	January gas bill
⋮							
Feb.08	Yuanhang Tech Co. Sal...	9200	20524.82	Payroll	Income	Online Bank	February salary
Feb.12	Hengxin CC Repayment	-1750	18774.82	Repayment	Expense	UnionPay	CC Bill Period 02
Feb.15	Municipal Water	-62.7	18712.12	Utility	Expense	Mobile	February water bill
⋮							
Mar.10	Hengtai Asset-Steady A	-1000	26728.3	Investment	Expense	Mobile	Wealth mgmt pur...
Mar.12	Hengxin CC Repayment	-1700	25028.3	Repayment	Expense	UnionPay	CC Bill Period 03
Mar.15	Huadian Power	-186.35	24841.95	Utility	Expense	Mobile	March electricity bill
⋮							
Apr.10	Hengtai Asset-Steady B	-1200	32717.98	Investment	Expense	Mobile	Wealth mgmt pur...
Apr.12	Hengxin CC Repayment	-1680	31037.98	Repayment	Expense	UnionPay	CC Bill Period 04
Apr.15	Guangsheng Telecom	-129	30908.98	Telecom	Expense	Mobile	Broadband paym...
⋮							
May.10	Hengtai Asset-Steady C	-1500	38396.53	Investment	Expense	Mobile	Wealth mgmt pur...
May.12	Hengxin CC Repayment	-1650	36746.53	Repayment	Expense	UnionPay	CC Bill Period 05
May.15	City Heating Co.	-92.3	36654.23	Utility	Expense	Online Bank	Heating maintena...
⋮							
Jun.10	Hengtai Asset-Steady D	-1500	44674.05	Investment	Expense	Mobile	Wealth mgmt pur...
Jun.12	Hengxin CC Repayment	-1600	43074.05	Repayment	Expense	UnionPay	CC Bill Period 06
Jun.15	Appliance Repair Center	-146	42928.05	Consumption	Expense	POS	AC cleaning
⋮							

Date	Counterparty	Amount	Balance	Type	In/Out	Channel	Remarks
⋮							
Jul.10	Hengtai Asset-Steady E	-2000	50775.13	Investment	Expense	Mobile	Wealth mgmt pur...
Jul.12	Hengxin CC Repayment	-1620	49155.13	Repayment	Expense	UnionPay	CC Bill Period 07
Jul.15	Municipal Water	-64.1	49091.03	Utility	Expense	Mobile	July water bill
⋮							
Aug.10	Hengtai Asset-Steady F	-1500	56605.75	Investment	Expense	Mobile	Wealth mgmt pur...
Aug.12	Hengxin CC Repayment	-1580	55025.75	Repayment	Expense	UnionPay	CC Bill Period 08
Aug.15	Guohua Life Insurance	-300	54725.75	Insurance	Expense	Online Bank	Accident insurance
⋮							
Sep.10	Hengtai Asset-Steady C...	1800	65660.95	Redemption	Income	Mobile	Wealth mgmt red...
Sep.12	Hengxin CC Repayment	-1550	64110.95	Repayment	Expense	UnionPay	CC Bill Period 09
Sep.15	City Gas Corp	-120.1	63990.85	Utility	Expense	Mobile	September gas bill
⋮							
Oct.10	Hengtai Asset-Steady G	-2000	71350.8	Investment	Expense	Mobile	Wealth mgmt pur...
Oct.12	Hengxin CC Repayment	-1520	69830.8	Repayment	Expense	UnionPay	CC Bill Period 10
Oct.15	Municipal Water	-66.7	69764.1	Utility	Expense	Mobile	October water bill
⋮							
Nov.10	Hengtai Asset-Steady D...	1600	80667.9	Redemption	Income	Mobile	Wealth mgmt red...
Nov.12	Hengxin CC Repayment	-1500	79167.9	Repayment	Expense	UnionPay	CC Bill Period 11
Nov.15	Guangsheng Telecom	-129	79038.9	Telecom	Expense	Mobile	Broadband renewal
⋮							
Dec.10	Hengtai Asset-Steady H	-2500	85816.78	Investment	Expense	Mobile	Wealth mgmt pur...
Dec.12	Hengxin CC Repayment	-1480	84336.78	Repayment	Expense	UnionPay	CC Bill Period 12
Dec.15	Huadian Power	-192.6	84144.18	Utility	Expense	Mobile	December electric...
⋮							

Question

What is the total annual amount of "Credit Card Repayment" expenses recorded in the table?

A. 23,429 B. 14,416 C. 19,600 D. 32,078

Answer

Correct Answer

C

Calculation: 1800 + 1750 + 1700 + 1680 + 1650 + 1600 + 1620 +1580 + 1550+ 1520 + 1500+ 1480 =19,430≈19,600

Skill Tested: Identify specific transaction types from logs and perform numerical aggregation.

Cognitive Level: Comprehension & Calculation

Figure 5: Sample Single Choice Question about Comprehension & Calculation of Personal Clients Log Information

Micro-enterprise Transaction Log (2023 Full Year Records - Complete)

Total Records

- 72 transactions
- Regular online settlements from Alipay (5,000~10,000)

Sesonal Peaks

- May Labor Day (+9,000 with +3,200 egg stock)
- Oct National Day (+9,000 with +5,200 grain restock)
- Dec Year-end (+10,000 with bulk procurement -8,800)

Date	Counterparty	Amount	Balance	Type	In/Out	Channel	Remarks
Jan.03	Cash Deposit	8000	8000	Cash	Income	ATM	Opening week inv...
Jan.05	Alipay Network Tech Co.	6000	14000	Payment	Income	E-commerce	Online order settl...
Jan.06	Ninghai Grain & Oil Tra...	-4500	9500	Transfer	Expense	Online Bank	Procurement - gra...
Jan.10	Yunnan Xinyuan Packa...	-1200	8300	Transfer	Expense	Online Bank	Packaging materia...
Jan.15	China Mobile Communi...	-98	8202	Alipay	Expense	E-commerce	Store phone bill
Jan.28	Kunming Chuncheng P...	-3000	5202	Transfer	Expense	Mobile Bank	Store rent - January
Feb.02	Tenpay Payment Tech...	5000	10202	Payment	Income	E-commerce	Online settlement...
Feb.05	Zhang Yong Individual...	-2200	8002	Transfer	Expense	Online Bank	Fresh vegetable pr...
Feb.09	Kunming Lexian Food T...	-1500	6502	Transfer	Expense	Online Bank	Meat supply paym...
Feb.15	Xuefeng Beverage Co.	-860	5642	Transfer	Expense	Online Bank	Beverage wholesa...
Feb.21	Interest Credit	0.74	5642.74	Interest	Income	Internal	Demand deposit i...
Feb.25	Kunming Chuncheng P...	-3000	2642.74	Transfer	Expense	Mobile Bank	Store rent - Febru...
Mar.01	Scheduled Loan Repay...	-1500	1142.74	Loan	Expense	Online Bank	Micro-business lo...
Mar.03	Alipay Network Tech Co.	8000	9142.74	Payment	Income	E-commerce	Online order settl...
Mar.07	Kunming Deshun Veget...	-2300	6842.74	Transfer	Expense	Online Bank	Vegetable delivery...
Mar.13	Yunnan Weiyuan Seaso...	-1200	5642.74	Transfer	Expense	Online Bank	Seasoning procur...
Mar.21	Interest Credit	0.73	5643.47	Interest	Income	Internal	Demand deposit i...
Mar.28	Kunming Chuncheng P...	-3000	2643.47	Transfer	Expense	Mobile Bank	Store rent - March
Apr.02	Duxiaoman Payment T...	12000	14643.47	Payment	Income	Online Bank	Consumer loan di...
Apr.05	Kunming Laochen Grai...	-5000	9643.47	Transfer	Expense	Online Bank	Grain & oil procur...
Apr.10	Alipay Network Tech Co.	7000	16643.47	Payment	Income	E-commerce	Online settlement...
Apr.18	Yunnan Xinyuan Packa...	-1800	14843.47	Transfer	Expense	Online Bank	Packaging - boxes...
Apr.21	Interest Credit	0.95	14844.42	Interest	Income	Internal	Demand deposit i...
Apr.26	Kunming Chuncheng P...	-3000	11844.42	Transfer	Expense	Mobile Bank	Store rent - April
May.01	Scheduled Loan Repay...	-1500	10344.42	Loan	Expense	Online Bank	Micro-business lo...
May.06	Alipay Network Tech Co.	9000	19344.42	Payment	Income	E-commerce	Online settlement...
May.09	Yunnan Nanping Poultr...	-3200	16144.42	Transfer	Expense	Online Bank	Egg wholesale - ho...
May.15	Kunming Jiale Fresh Me...	-2800	13344.42	Transfer	Expense	Online Bank	Daily pork procur...
May.21	Interest Credit	1.42	13345.84	Interest	Income	Internal	Demand deposit i...
May.28	Kunming Chuncheng P...	-3000	10345.84	Transfer	Expense	Mobile Bank	Store rent - May
Jun.03	Alipay Network Tech Co.	8500	18845.84	Payment	Income	E-commerce	Online settlement...
Jun.06	Wangyin Online Beijing...	-980	17865.84	Online	Expense	E-commerce	Cloud platform se...
Jun.12	Kunming Laochen Grai...	-4000	13865.84	Transfer	Expense	Online Bank	Monthly grain pro...
Jun.18	Yunnan Xinyuan Packa...	-1500	12365.84	Transfer	Expense	Online Bank	Packaging - paper...
Jun.21	Interest Credit	1.52	12367.36	Interest	Income	Internal	Demand deposit i...
Jun.26	Kunming Chuncheng P...	-3000	9367.36	Transfer	Expense	Mobile Bank	Store rent - June
Jul.01	Mashang Consumer Fi...	-1800	7567.36	Loan	Expense	E-commerce	Consumer loan de...
Jul.05	Alipay Network Tech Co.	9000	16567.36	Payment	Income	E-commerce	Online settlement...
Jul.09	Kunming Deshun Veget...	-2500	14067.36	Transfer	Expense	Online Bank	Vegetable delivery...
Jul.15	China Telecom Corp	-120	13947.36	Alipay	Expense	E-commerce	Store broadband f...
Jul.21	Interest Credit	1.69	13949.05	Interest	Income	Internal	Demand deposit i...
Jul.28	Kunming Chuncheng P...	-3000	10949.05	Transfer	Expense	Mobile Bank	Store rent - July
Aug.02	Duxiaoman Payment T...	10000	20949.05	Payment	Income	Online Bank	Consumer loan di...
Aug.06	Yunnan Weiyuan Seaso...	-1600	19349.05	Transfer	Expense	Online Bank	Seasoning restock
Aug.10	Alipay Network Tech Co.	7000	26349.05	Payment	Income	E-commerce	Online settlement...
Aug.16	Kunming Laochen Grai...	-5000	21349.05	Transfer	Expense	Online Bank	Bulk grain procure...
Aug.21	Interest Credit	2.59	21351.64	Interest	Income	Internal	Demand deposit i...
Aug.26	Kunming Chuncheng P...	-3000	18351.64	Transfer	Expense	Mobile Bank	Store rent - August
Sep.01	Duxiaoman Chongqing...	-1773.22	16578.42	Loan	Expense	E-commerce	Microloan deducti...
Sep.05	Alipay Network Tech Co.	8000	24578.42	Payment	Income	E-commerce	Online settlement...
Sep.09	Kunming Deshun Veget...	-2600	21978.42	Transfer	Expense	Online Bank	Vegetable delivery...
Sep.15	Yunnan Xinyuan Packa...	-1800	20178.42	Transfer	Expense	Online Bank	Packaging - takeo...
Sep.21	Interest Credit	2.37	20180.79	Interest	Income	Internal	Demand deposit i...
Sep.28	Kunming Chuncheng P...	-3000	17180.79	Transfer	Expense	Mobile Bank	Store rent - Septe...
Oct.03	Mashang Consumer Fi...	5000	22180.79	Loan	Income	Online Bank	Loan disbursement
Oct.06	Alipay Network Tech Co.	9000	31180.79	Payment	Income	E-commerce	Online settlement...
Oct.11	Kunming Laochen Grai...	-5200	25980.79	Transfer	Expense	Online Bank	Post-holiday resto...
Oct.18	Kunming Local Promoti...	-480	25500.79	Alipay	Expense	E-commerce	Online advertising...
Oct.21	Interest Credit	2.99	25503.78	Interest	Income	Internal	Demand deposit i...
Oct.27	Kunming Chuncheng P...	-3000	22503.78	Transfer	Expense	Mobile Bank	Store rent - October
Nov.01	Scheduled Loan Repay...	-1500	21003.78	Loan	Expense	Online Bank	Micro-business lo...
Nov.05	Alipay Network Tech Co.	8500	29503.78	Payment	Income	E-commerce	Online settlement...
Nov.09	Yunnan Weiyuan Seaso...	-1900	27603.78	Transfer	Expense	Online Bank	Winter seasoning...
Nov.15	Kunming Deshun Veget...	-2700	24903.78	Transfer	Expense	Online Bank	Vegetable delivery...
Nov.21	Interest Credit	2.92	24906.7	Interest	Income	Internal	Demand deposit i...
Nov.28	Kunming Chuncheng P...	-3000	21906.7	Transfer	Expense	Mobile Bank	Store rent - Nove...
Dec.01	Duxiaoman Chongqing...	-1773.22	20133.48	Loan	Expense	E-commerce	Microloan deducti...
Dec.05	Alipay Network Tech Co.	10000	30133.48	Payment	Income	E-commerce	Online settlement...
Dec.10	Yunnan Xinyuan Packa...	-2800	27333.48	Transfer	Expense	Online Bank	Year-end packagin...
Dec.15	Kunming Laochen Grai...	-6000	21333.48	Transfer	Expense	Online Bank	Year-end grain inv...
Dec.21	Interest Credit	2.51	21335.99	Interest	Income	Internal	Demand deposit i...
Dec.28	Kunming Chuncheng P...	-3000	18335.99	Transfer	Expense	Mobile Bank	Store rent - Dece...

Question

Based on the cross-month revenue and procurement patterns, which indicators support seasonal business patterns?

A. Online settlements & procurement rise around holidays/promotions

B. Year-end bulk stocking occurs with continued business revenue

C. Revenue and expenses are completely flat throughout the year

D. Peak season only reflects in non-operating income increase

Answer

Correct Answer

A, B

A correct: Labor Day (May:+9,000 settlement with +3,200 egg procurement) and National Day (Oct:+9,000 with +5,200 grain restock) show clear seasonal spikes

⋮

B correct: December shows bulk purchasing (packaging -2,800 + grain -6,000 = -8,800) while online revenue continues (+10,000), indicating sustained year-end operations

Figure 6: Sample Multiple Choice Question about Analysis & Synthesis of Micro-enterprise Clients Log Information

Personal client trasaction log (b) (2023 full year)

Skills Tested

- Financial statement data extraction
- Time-series analysis (MoM comparison)
- Cash flow pattern recognition
- Statistical calculation (arithmetic mean)

Cognitive Level

- Analysis&Synthesis(High) Multi-step calculation requiring identification of relevant data points from complex transaction records

Date	Counterparty	Amount	Balance	Type	In/Out	Channel	Remarks
Jan.03	Evergreen Cons...	-650	4,350.00	Loan Repayment	Expense	Direct Debit	January insta...
Jan.08	City Gas Group...	-118	4,232.00	Utilities	Expense	Mobile Banking	January gas...
Jan.15	Galaxy Bank Cre...	-1,000.00	3,232.00	Credit Card Pay...	Expense	Online Banking	January state...
Jan.20	Bank Demand D...	0.58	3,232.58	Interest Credit	Income	Internal	January 2023...
Jan.25	Hangzhou Chan...	13,000.00	16,232.58	Salary Credit	Income	Online Banking	January salary
Jan.28	Lejia Lifestyle S...	-198.6	16,033.98	POS Purchase	Expense	POS Terminal	Daily necessi...
Feb.03	Evergreen Cons...	-650	15,383.98	Loan Repayment	Expense	Direct Debit	February ins...
Feb.07	Mobile Telecom...	-50	15,333.98	Phone Top-up	Expense	Quick Pay	February ph...
Feb.10	Ningxia Highwa...	-86	15,247.98	Direct Debit	Expense	Direct Debit	Highway toll...
Feb.20	Bank Demand D...	0.63	15,248.61	Interest Credit	Income	Internal	February 20...
Feb.25	Hangzhou Chan...	13,000.00	28,248.61	Salary Credit	Income	Online Banking	February sal...
Feb.27	Donglaishun Ca...	-236.5	28,012.11	POS Purchase	Expense	Third-Party Pay...	Business din...
Mar.03	Evergreen Cons...	-650	27,362.11	Loan Repayment	Expense	Direct Debit	March install...
Mar.05	Haihe Water Su...	-82.4	27,279.71	Utilities	Expense	Mobile Banking	March water...
Mar.12	Taihe Microfina...	-420	26,859.71	Loan Repayment	Expense	Direct Debit	Microloan sc...
Mar.20	Bank Demand D...	0.71	26,860.42	Interest Credit	Income	Internal	March 2023...
Mar.25	Hangzhou Chan...	13,000.00	39,860.42	Salary Credit	Income	Online Banking	March salary
Mar.29	Fenglinwan Con...	-126.8	39,733.62	POS Purchase	Expense	POS Terminal	Daily supplies
Apr.03	Evergreen Cons...	-650	39,083.62	Loan Repayment	Expense	Direct Debit	April installm...
Apr.06	City Gas Group...	-120	38,963.62	Utilities	Expense	Mobile Banking	April gas bill
Apr.12	Taihe Microfina...	-420	38,543.62	Loan Repayment	Expense	Direct Debit	Microloan sc...
Apr.18	Yunshan Payme...	300	38,843.62	Third-Party Wit...	Income	Third-Party Pay...	Third-party a...
Apr.20	Bank Demand D...	0.78	38,844.40	Interest Credit	Income	Internal	April 2023 d...
Apr.25	Hangzhou Chan...	13,000.00	51,844.40	Salary Credit	Income	Online Banking	April salary
Apr.28	Zhiguo Digital El...	-989	50,855.40	POS Purchase	Expense	Third-Party Pay...	Bluetooth ea...
May.03	Evergreen Cons...	-650	50,205.40	Loan Repayment	Expense	Direct Debit	May installm...
May.05	Southern Power...	-210	49,995.40	Utilities	Expense	Mobile Banking	Electricity bill...
May.12	Taihe Microfina...	-420	49,575.40	Loan Repayment	Expense	Direct Debit	Microloan sc...
May.20	Bank Demand D...	0.92	49,576.32	Interest Credit	Income	Internal	May 2023 de...
May.23	Chunshen Auto...	-380	49,196.32	POS Purchase	Expense	POS Terminal	Vehicle main...
May.25	Hangzhou Chan...	13,000.00	62,196.32	Salary Credit	Income	Online Banking	May salary
Jun.02	Evergreen Cons...	-650	61,546.32	Loan Repayment	Expense	Direct Debit	June installm...
Jun.06	City Gas Group...	-115	61,431.32	Utilities	Expense	Mobile Banking	June gas bill
Jun.12	Taihe Microfina...	-420	61,011.32	Loan Repayment	Expense	Direct Debit	Microloan sc...
Jun.18	Sichuan Airlines...	-688	60,323.32	POS Purchase	Expense	Third-Party Pay...	Weekend ac...
Jun.20	Bank Demand D...	1.03	60,324.35	Interest Credit	Income	Internal	June 2023 de...
Jun.25	Hangzhou Chan...	13,000.00	73,324.35	Salary Credit	Income	Online Banking	June salary
Jul.03	Evergreen Cons...	-650	72,674.35	Loan Repayment	Expense	Direct Debit	July installm...
Jul.07	Mobile Telecom...	-50	72,624.35	Phone Top-up	Expense	Quick Pay	July phone re...
Jul.12	Taihe Microfina...	-420	72,204.35	Loan Repayment	Expense	Direct Debit	Microloan sc...
Jul.16	Shanhai Bookst...	-89.5	72,114.85	POS Purchase	Expense	POS Terminal	Book purcha...
Jul.20	Bank Demand D...	1.23	72,116.08	Interest Credit	Income	Internal	July 2023 de...
Jul.25	Hangzhou Chan...	13,000.00	85,116.08	Salary Credit	Income	Online Banking	July salary
Aug.03	Evergreen Cons...	-650	84,466.08	Loan Repayment	Expense	Direct Debit	August instal...
Aug.05	Haihe Water Su...	-85.3	84,380.78	Utilities	Expense	Mobile Banking	August wate...
Aug.10	Yunshan Payme...	200	84,580.78	Third-Party Wit...	Income	Third-Party Pay...	Fund inflow
Aug.12	Taihe Microfina...	-420	84,160.78	Loan Repayment	Expense	Direct Debit	Microloan sc...
Aug.20	Bank Demand D...	1.28	84,162.06	Interest Credit	Income	Internal	August 2023...
Aug.25	Hangzhou Chan...	13,000.00	97,162.06	Salary Credit	Income	Online Banking	August salary
Sep.03	Evergreen Cons...	-650	96,512.06	Loan Repayment	Expense	Direct Debit	September i...
Sep.06	Southern Power...	-198	96,314.06	Utilities	Expense	Mobile Banking	Electricity bill...
Sep.12	Taihe Microfina...	-420	95,894.06	Loan Repayment	Expense	Direct Debit	Microloan sc...
Sep.18	Maidu Bakery	-62.5	95,831.56	POS Purchase	Expense	Third-Party Pay...	Bread and p...
Sep.20	Bank Demand D...	1.39	95,832.95	Interest Credit	Income	Internal	September 2...
Sep.25	Hangzhou Chan...	13,000.00	108,832.95	Salary Credit	Income	Online Banking	September s...
Oct.03	Evergreen Cons...	-650	108,182.95	Loan Repayment	Expense	Direct Debit	October inst...
Oct.08	City Gas Group...	-118	108,064.95	Utilities	Expense	Mobile Banking	October gas...
Oct.12	Taihe Microfina...	-420	107,644.95	Loan Repayment	Expense	Direct Debit	Microloan sc...
Oct.14	Yuanhang Cine...	-68	107,576.95	POS Purchase	Expense	POS Terminal	Movie tickets
Oct.20	Bank Demand D...	1.48	107,578.43	Interest Credit	Income	Internal	October 202...
Oct.25	Hangzhou Chan...	13,000.00	120,578.43	Salary Credit	Income	Online Banking	October salary
Nov.03	Evergreen Cons...	-650	119,928.43	Loan Repayment	Expense	Direct Debit	November in...
Nov.05	Haihe Water Su...	-79.6	119,848.83	Utilities	Expense	Mobile Banking	November w...
Nov.11	Zhixuan Premiu...	-1,299.00	118,549.83	POS Purchase	Expense	Third-Party Pay...	Electric kettl...
Nov.12	Taihe Microfina...	-420	118,129.83	Loan Repayment	Expense	Direct Debit	Microloan sc...
Nov.20	Bank Demand D...	1.58	118,131.41	Interest Credit	Income	Internal	November 2...
Nov.25	Hangzhou Chan...	13,000.00	131,131.41	Salary Credit	Income	Online Banking	November s...
Dec.03	Evergreen Cons...	-650	130,481.41	Loan Repayment	Expense	Direct Debit	December in...
Dec.08	Southern Power...	-205	130,276.41	Utilities	Expense	Mobile Banking	Electricity bill...
Dec.12	Taihe Microfina...	-420	129,856.41	Loan Repayment	Expense	Direct Debit	Microloan sc...
Dec.20	Bank Demand D...	1.72	129,858.13	Interest Credit	Income	Internal	December 2...
Dec.25	Hangzhou Chan...	13,000.00	142,858.13	Salary Credit	Income	Online Banking	December s...
Dec.28	Zhongpin Fresh...	-236.2	142,621.93	POS Purchase	Expense	POS Terminal	Year-end gro...

Question

Using the balance after the last transaction of each month as the baseline, calculate the average balance increase between adjacent months.

A. 18,061.18

b. 14,075.87

c. 11,600.00

d. 8,549.05

Answer

Correct Answer

C

closest match!

Calculation Process:

Sum of increases: 126,587.95

Number of intervals: 11

Average: 126,587.95 ÷ 11 = 11,508.00

C is the closest match answer.

Figure 7: Sample Single Choice Question about Analysis & Synthesis of Personal Clients Log Information

Micro-enterprise Transaction Log (b) (2023 Full Year Records - Complete)

Skill Test & Cognitive Level

Skills: Cash flow pattern analysis, Financial statement interpretation, Trend identification

Cognitive Level: Comprehension & Calculation (Medium)

Tests: Understanding of working capital cycles, SME cash flow characteristics

Date	Counterparty	Amount	Balance	Type	In/Out	Channel	Remarks
Jan.05	Alipay Network...	15,000.00	15,000.00	Agency Payment	Income	E-commerce	Online store...
Jan.06	Hongchang Har...	-5,200.00	9,800.00	Interbank Trans...	Expense	Mobile Banking	Hardware m...
Jan.07	SF Express Co.,...	-1,800.00	8,000.00	Transfer Payment	Expense	Online Banking	Courier freig...
Jan.08	Tenpay Technol...	7,800.00	15,800.00	Agency Payment	Income	E-commerce	WeChat stor...
Jan.15	Hengtai Machini...	-6,000.00	9,800.00	Transfer Payment	Expense	Online Banking	Payroll - Janu...
Jan.21	Bank Interest Se...	1.1	9,801.10	Interest Credit	Income	Internal	Demand dep...
Jan.25	Qingyang Marke...	-2,500.00	7,301.10	Interbank Trans...	Expense	Mobile Banking	Factory rent...
Jan.28	Deshun Packagi...	-2,100.00	5,201.10	Interbank Trans...	Expense	Mobile Banking	Packaging m...
Jan.30	Pinduoduo Mer...	5,600.00	10,801.10	UnionPay Credit	Income	E-commerce	Pinduoduo r...
Feb.03	Alipay Network...	12,000.00	22,801.10	Agency Payment	Income	E-commerce	Online store...
Feb.05	Nanxing Raw M...	-4,800.00	18,001.10	Interbank Trans...	Expense	Mobile Banking	Plastic pellet...
Feb.08	Zhang Yan Indiv...	-1,600.00	16,401.10	Transfer Payment	Expense	Online Banking	Short-haul d...
Feb.10	Tenpay Technol...	6,200.00	22,601.10	Agency Payment	Income	E-commerce	WeChat stor...
Feb.15	Hengtai Machini...	-6,200.00	16,401.10	Transfer Payment	Expense	Online Banking	Payroll - Febr...
Feb.20	Qingyang Marke...	-2,500.00	13,901.10	Interbank Trans...	Expense	Mobile Banking	Factory rent...
Feb.21	Bank Interest Se...	1.05	13,902.15	Interest Credit	Income	Internal	Demand dep...
Feb.25	Jinrui Electrical...	-3,500.00	10,402.15	Interbank Trans...	Expense	Mobile Banking	Equipment...
Mar.02	Alipay Network...	9,800.00	20,202.15	Agency Payment	Income	E-commerce	Online store...
Mar.04	Huahui Mold Pr...	-5,200.00	15,002.15	Interbank Trans...	Expense	Mobile Banking	Custom mol...
Mar.06	Tenpay Technol...	7,600.00	22,602.15	Agency Payment	Income	E-commerce	WeChat stor...
Mar.10	Deshun Packagi...	-2,100.00	20,502.15	Interbank Trans...	Expense	Mobile Banking	Packaging m...
Mar.15	Hengtai Machini...	-6,400.00	14,102.15	Transfer Payment	Expense	Online Banking	Payroll - Mar...
Mar.18	Baisheng Advert...	-900	13,202.15	Interbank Trans...	Expense	Mobile Banking	Store adverti...
Mar.21	Bank Interest Se...	1.2	13,203.35	Interest Credit	Income	Internal	Demand dep...
Mar.25	Qingyang Marke...	-2,500.00	10,703.35	Interbank Trans...	Expense	Mobile Banking	Factory rent...
Mar.27	Yimin Labor Pro...	-1,800.00	8,903.35	Interbank Trans...	Expense	Mobile Banking	Safety equip...
Apr.01	Alipay Network...	11,000.00	19,903.35	Agency Payment	Income	E-commerce	Online store...
Apr.05	Tenpay Technol...	5,400.00	25,303.35	Agency Payment	Income	E-commerce	WeChat stor...
Apr.07	Nanxing Raw M...	-6,000.00	19,303.35	Interbank Trans...	Expense	Mobile Banking	Plastic pellet...
Apr.12	Zhang Yan Indiv...	-1,700.00	17,603.35	Transfer Payment	Expense	Online Banking	Short-haul d...
Apr.15	Hengtai Machini...	-6,500.00	11,103.35	Transfer Payment	Expense	Online Banking	Payroll - Apri...
Apr.20	Qingyang Marke...	-2,500.00	8,603.35	Interbank Trans...	Expense	Mobile Banking	Factory rent...
Apr.21	Bank Interest Se...	1.15	8,604.50	Interest Credit	Income	Internal	Demand dep...
Apr.24	Deshun Packagi...	-2,200.00	6,404.50	Interbank Trans...	Expense	Mobile Banking	Packaging m...
Apr.28	Wantong Electri...	-1,800.00	4,604.50	Interbank Trans...	Expense	Mobile Banking	Electrical par...
May.03	Alipay Network...	12,500.00	17,104.50	Agency Payment	Income	E-commerce	Online store...
May.06	Tenpay Technol...	6,800.00	23,904.50	Agency Payment	Income	E-commerce	WeChat stor...
May.08	Nanxing Raw M...	-5,400.00	18,504.50	Interbank Trans...	Expense	Mobile Banking	Plastic pellet...
May.10	SF Express Co.,...	-2,000.00	16,504.50	Transfer Payment	Expense	Online Banking	Courier freig...
May.15	Hengtai Machini...	-6,600.00	9,904.50	Transfer Payment	Expense	Online Banking	Payroll - May...
May.18	Baisheng Advert...	-1,200.00	8,704.50	Interbank Trans...	Expense	Mobile Banking	Store promo...
May.21	Bank Interest Se...	1.3	8,705.80	Interest Credit	Income	Internal	Demand dep...
May.25	Qingyang Marke...	-2,500.00	6,205.80	Interbank Trans...	Expense	Mobile Banking	Factory rent...
May.28	Yimin Labor Pro...	-1,600.00	4,605.80	Interbank Trans...	Expense	Mobile Banking	Safety equip...
Jun.02	Alipay Network...	10,800.00	15,405.80	Agency Payment	Income	E-commerce	Online store...
Jun.05	Tenpay Technol...	6,100.00	21,505.80	Agency Payment	Income	E-commerce	WeChat stor...
Jun.07	Hongchang Har...	-5,000.00	16,505.80	Interbank Trans...	Expense	Mobile Banking	Hardware m...
Jun.10	Deshun Packagi...	-2,300.00	14,205.80	Interbank Trans...	Expense	Mobile Banking	Packaging m...
Jun.15	Hengtai Machini...	-6,600.00	7,605.80	Transfer Payment	Expense	Online Banking	Payroll - June...
Jun.20	Qingyang Marke...	-2,500.00	5,105.80	Interbank Trans...	Expense	Mobile Banking	Factory rent...
Jun.21	Bank Interest Se...	1.25	5,107.05	Interest Credit	Income	Internal	Demand dep...
Jun.25	Liu Haijun Indivi...	-1,600.00	3,507.05	Transfer Payment	Expense	Online Banking	Temporary...
Jul.03	Alipay Network...	9,800.00	13,307.05	Agency Payment	Income	E-commerce	Online store...
Jul.05	Tenpay Technol...	7,200.00	20,507.05	Agency Payment	Income	E-commerce	WeChat stor...
Jul.08	Hongxing Plasti...	-4,800.00	15,707.05	Interbank Trans...	Expense	Mobile Banking	Injection mol...
Jul.12	SF Express Co.,...	-2,100.00	13,607.05	Transfer Payment	Expense	Online Banking	Courier freig...
Jul.15	Hengtai Machini...	-6,700.00	6,907.05	Transfer Payment	Expense	Online Banking	Payroll - July...
Jul.20	Qingyang Marke...	-2,500.00	4,407.05	Interbank Trans...	Expense	Mobile Banking	Factory rent...
Jul.21	Bank Interest Se...	1.2	4,408.25	Interest Credit	Income	Internal	Demand dep...
Jul.26	Hongchang Har...	-1,700.00	2,708.25	Interbank Trans...	Expense	Mobile Banking	Hardware ac...
Jul.28	Yimin Labor Pro...	-1,200.00	1,508.25	Interbank Trans...	Expense	Mobile Banking	Safety equip...
Aug.01	Alipay Network...	11,600.00	13,108.25	Agency Payment	Income	E-commerce	Online store...
Aug.04	Tenpay Technol...	5,900.00	19,008.25	Agency Payment	Income	E-commerce	WeChat stor...
Aug.06	Nanxing Raw M...	-5,200.00	13,808.25	Interbank Trans...	Expense	Mobile Banking	Plastic pellet...
Aug.09	Deshun Packagi...	-2,400.00	11,408.25	Interbank Trans...	Expense	Mobile Banking	Packaging m...
Aug.12	Liu Haijun Indivi...	-1,800.00	9,608.25	Transfer Payment	Expense	Online Banking	Workshop cl...
Aug.15	Hengtai Machini...	-6,700.00	2,908.25	Transfer Payment	Expense	Online Banking	Payroll - Aug...
Aug.20	Qingyang Marke...	-2,500.00	408.25	Interbank Trans...	Expense	Mobile Banking	Factory rent...
Aug.21	Bank Interest Se...	0.95	409.2	Interest Credit	Income	Internal	Demand dep...
Aug.25	Pinduoduo Mer...	9,500.00	9,909.20	UnionPay Credit	Income	E-commerce	Pinduoduo r...
Aug.28	Hongchang Har...	-1,500.00	8,409.20	Interbank Trans...	Expense	Mobile Banking	Hardware ac...
Sep.02	Alipay Network...	10,200.00	18,609.20	Agency Payment	Income	E-commerce	Online store...
Sep.05	Tenpay Technol...	6,300.00	24,909.20	Agency Payment	Income	E-commerce	WeChat stor...
Sep.07	Nanxing Raw M...	-5,200.00	19,709.20	Interbank Trans...	Expense	Mobile Banking	Plastic pellet...
Sep.10	Deshun Packagi...	-2,200.00	17,509.20	Interbank Trans...	Expense	Mobile Banking	Packaging m...
Sep.13	SF Express Co.,...	-1,900.00	15,609.20	Transfer Payment	Expense	Online Banking	Courier freig...
Sep.15	Hengtai Machini...	-6,800.00	8,809.20	Transfer Payment	Expense	Online Banking	Payroll - Sept...
Sep.20	Qingyang Marke...	-2,500.00	6,309.20	Interbank Trans...	Expense	Mobile Banking	Factory rent...
Sep.21	Bank Interest Se...	1.05	6,310.25	Interest Credit	Income	Internal	Demand dep...
Sep.25	Yimin Labor Pro...	-1,600.00	4,710.25	Interbank Trans...	Expense	Mobile Banking	Safety equip...
Sep.28	Hongchang Har...	-1,200.00	3,510.25	Interbank Trans...	Expense	Mobile Banking	Hardware ac...
Oct.03	Alipay Network...	11,500.00	15,010.25	Agency Payment	Income	E-commerce	Online store...
Oct.05	Tenpay Technol...	7,000.00	22,010.25	Agency Payment	Income	E-commerce	WeChat stor...
Oct.08	Nanxing Raw M...	-5,600.00	16,410.25	Interbank Trans...	Expense	Mobile Banking	Plastic pellet...
Oct.11	Deshun Packagi...	-2,500.00	13,910.25	Interbank Trans...	Expense	Mobile Banking	Packaging m...
Oct.15	Hengtai Machini...	-6,900.00	7,010.25	Transfer Payment	Expense	Online Banking	Payroll - Oct...
Oct.20	Qingyang Marke...	-2,500.00	4,510.25	Interbank Trans...	Expense	Mobile Banking	Factory rent...
Oct.21	Bank Interest Se...	1.1	4,511.35	Interest Credit	Income	Internal	Demand dep...
Oct.24	Liu Haijun Indivi...	-1,900.00	2,611.35	Transfer Payment	Expense	Online Banking	Temporary l...
Oct.27	Huahui Mold Pr...	-1,200.00	1,411.35	Interbank Trans...	Expense	Mobile Banking	Mold repair f...
Nov.02	Alipay Network...	13,800.00	15,211.35	Agency Payment	Income	E-commerce	Online store...
Nov.05	Tenpay Technol...	6,800.00	22,011.35	Agency Payment	Income	E-commerce	WeChat stor...
Nov.07	Nanxing Raw M...	-5,400.00	16,611.35	Interbank Trans...	Expense	Mobile Banking	Plastic pellet...
Nov.10	Deshun Packagi...	-2,600.00	14,011.35	Interbank Trans...	Expense	Mobile Banking	Packaging m...
Nov.12	Zhang Yan Indiv...	-2,100.00	11,911.35	Transfer Payment	Expense	Online Banking	Short-haul d...
Nov.15	Hengtai Machini...	-7,000.00	4,911.35	Transfer Payment	Expense	Online Banking	Payroll - Nov...
Nov.20	Qingyang Marke...	-2,500.00	2,411.35	Interbank Trans...	Expense	Mobile Banking	Factory rent...
Nov.21	Bank Interest Se...	1.15	2,412.50	Interest Credit	Income	Internal	Demand dep...
Nov.25	Yimin Labor Pro...	-1,600.00	812.5	Interbank Trans...	Expense	Mobile Banking	Safety equip...
Nov.28	Pinduoduo Mer...	9,200.00	10,012.50	UnionPay Credit	Income	E-commerce	Pinduoduo r...
Dec.03	Alipay Network...	16,200.00	26,212.50	Agency Payment	Income	E-commerce	Online store...
Dec.05	Tenpay Technol...	7,500.00	33,712.50	Agency Payment	Income	E-commerce	WeChat stor...
Dec.07	Nanxing Raw M...	-6,000.00	27,712.50	Interbank Trans...	Expense	Mobile Banking	Plastic pellet...
Dec.10	Deshun Packagi...	-2,800.00	24,912.50	Interbank Trans...	Expense	Mobile Banking	Packaging m...
Dec.12	SF Express Co.,...	-2,100.00	22,812.50	Transfer Payment	Expense	Online Banking	Courier freig...
Dec.15	Hengtai Machini...	-7,200.00	15,612.50	Transfer Payment	Expense	Online Banking	Payroll - Dec...
Dec.20	Qingyang Marke...	-2,500.00	13,112.50	Interbank Trans...	Expense	Mobile Banking	Factory rent...
Dec.21	Bank Interest Se...	1.25	13,113.75	Interest Credit	Income	Internal	Demand dep...
Dec.24	Hongchang Har...	-1,800.00	11,313.75	Interbank Trans...	Expense	Mobile Banking	Hardware ac...
Dec.28	Yimin Labor Pro...	-1,600.00	9,713.75	Interbank Trans...	Expense	Mobile Banking	Safety equip...

Question

Based on the multi-month account balance fluctuations and income/expense rhythm, what is NOT the reasonable assessment of this micro-enterprise's cash flow health?

a. Overall healthy - balance recovers with business receipts, occasional dips but rebounds

b. Frequently shows negative balance, difficult to maintain normal operations

c. Balance continuously weakens with no obvious recovery signs

d. Fluctuations are irregular, difficult to reflect operating cash flow characteristics

Answer

Correct Answer

B, C, D

Question Type: Multiple Choice
Category: Cash Flow & Financial Health / Cash Flow Health Assessment

Option A is correct - the cash flow shows healthy cyclical patterns with
The others choice are wrong.

Figure 8: Sample Multiple Choice Question about Comprehension & Calculation of Micro-enterprise Clients Log Information

Building Materials SME Transaction Log (2023 Full Year - Cash-out Detection Analysis)

Total Records
- 72 transactions
- Multiple financing sources :MYbank, WeBank, Duxiaoman, Mashang Finance

Sesonal Peaks
- All transactions have clear business purposes
- Noround-number in-out or circular transfers detected
- Counterparties are established suppliers/customers

Skill Test & Cognitive Level
- Skills: Transaction pattern recognition, Anti-money laundering indicators, Counterparty analysis
- Cognitive Level: Comprehension & Calculation (Medium)
- Tests: Understanding of cash-out red flags, ability to analyze transaction legitimacy

Date	Counterparty	Amount	Balance	Type	In/Out	Channel	Remarks
Jan.03	Alipay Network...	8,500.00	9,320.00	Agency Payment	Income	E-commerce	Online order...
Jan.03	Kunming Hongd...	-6,000.00	3,320.00	Transfer Payment	Expense	Super Online B...	Cement proc...
Jan.08	Chenggong Shu...	-1,800.00	1,520.00	Transfer Payment	Expense	Super Online B...	Rebar procu...
Jan.10	Duxiaoman Cho...	-1,200.00	320.00	Loan Deduction	Expense	E-commerce	Business inst...
Jan.15	Cash Deposit	500.00	820.00	Cash	Income	ATM	Store cash re...
Jan.21	Interest Credit	0.72	820.72	Interest	Income	Internal	Demand dep...
Feb.02	Alipay Network...	9,200.00	10,020.72	Agency Payment	Income	E-commerce	Online order...
Feb.02	Kunming Huato...	-4,800.00	5,220.72	Transfer Payment	Expense	Super Online B...	PPR pipe fitti...
Feb.09	Mashang Consu...	-1,300.00	3,920.72	Direct Debit	Expense	E-commerce	Working capi...
Feb.13	Kunming Xiangy...	-1,500.00	2,420.72	Transfer Payment	Expense	Mobile Banking	Delivery freig...
Feb.17	Cash Deposit	300.00	2,720.72	Cash	Income	ATM	Store cash re...
Feb.21	Interest Credit	0.65	2,721.37	Interest	Income	Internal	Demand dep...
Mar.03	Payment Platfor...	11,000.00	13,721.37	Agency Payment	Income	E-commerce	Platform pay...
Mar.03	Kunming Hongd...	-6,500.00	7,221.37	Transfer Payment	Expense	Super Online B...	Cement proc...
Mar.08	MYbank (Zhejia...	20,000.00	27,221.37	Loan Disburse...	Income	Super Online B...	Business loa...
Mar.08	Yunnan Wanton...	-23,000.00	4,221.37	Transfer Payment	Expense	Super Online B...	Bulk invento...
Mar.12	Duxiaoman Cho...	-1,500.00	2,721.37	Loan Deduction	Expense	E-commerce	Business inst...
Mar.21	Interest Credit	0.90	2,722.27	Interest	Income	Internal	Demand dep...
Apr.04	Alipay Network...	12,000.00	14,722.27	Agency Payment	Income	E-commerce	Online order...
Apr.04	Yunnan Jinpeng...	-7,200.00	7,522.27	Transfer Payment	Expense	Super Online B...	Steel profiles...
Apr.09	Kunming Huato...	-2,600.00	4,922.27	Transfer Payment	Expense	Mobile Banking	PVC pipe ma...
Apr.12	Mashang Consu...	-1,300.00	3,622.27	Direct Debit	Expense	E-commerce	Working capi...
Apr.15	Cash Deposit	600	4,222.27	Cash	Income	ATM	Store cash re...
Apr.21	Interest Credit	0.95	4,223.22	Interest	Income	Internal	Demand dep...
May.05	Payment Platfor...	13,000.00	17,223.22	Agency Payment	Income	E-commerce	Platform pay...
May.05	Yuxi Yongsheng...	-5,400.00	11,823.22	Transfer Payment	Expense	Super Online B...	Hardware fit...
May.10	Duxiaoman Cho...	-1,500.00	10,323.22	Loan Deduction	Expense	E-commerce	Business inst...
May.12	Kunming Xiangy...	-1,200.00	9,123.22	Transfer Payment	Expense	Mobile Banking	Delivery freig...
May.18	Yunnan Zheng'a...	5,200.00	14,323.22	Transfer Credit	Income	Super Online B...	Project suppl...
May.21	Interest Credit	1.25	14,324.47	Interest	Income	Internal	Demand dep...
Jun.02	Alipay Network...	11,500.00	25,824.47	Agency Payment	Income	E-commerce	Online order...
Jun.02	Kunming Hongd...	-8,000.00	17,824.47	Transfer Payment	Expense	Super Online B...	Cement proc...
Jun.08	Mashang Consu...	-1,300.00	16,524.47	Direct Debit	Expense	E-commerce	Working capi...
Jun.12	Yunnan Jinpeng...	-5,200.00	11,324.47	Transfer Payment	Expense	Mobile Banking	Steel profiles...
Jun.18	Cash Deposit	700.00	12,024.47	Cash	Income	ATM	Store cash re...
Jun.21	Interest Credit	1.10	12,025.57	Interest	Income	Internal	Demand dep...
Jul.03	WeBank	8,000.00	20,025.57	Loan Disburse...	Income	Super Online B...	Business loa...
Jul.03	Kunming Huato...	-6,000.00	14,025.57	Transfer Payment	Expense	Super Online B...	Pipe materia...
Jul.10	Duxiaoman Cho...	-1,500.00	12,525.57	Loan Deduction	Expense	E-commerce	Business inst...
Jul.15	Cash Deposit	400.00	12,925.57	Cash	Income	ATM	Store cash re...
Jul.20	Kunming Mingh...	-2,200.00	10,725.57	Transfer Payment	Expense	Mobile Banking	Electrical sup...
Jul.21	Interest Credit	0.98	10,726.55	Interest	Income	Internal	Demand dep...
Aug.04	Payment Platfor...	12,500.00	23,226.55	Agency Payment	Income	E-commerce	Platform pay...
Aug.04	Yunnan Wanton...	-9,000.00	14,226.55	Transfer Payment	Expense	Super Online B...	Peak season...
Aug.09	Shanghai Dianr...	15,000.00	29,226.55	Loan Disburse...	Income	Super Online B...	Business loa...
Aug.09	Kunming Xinshe...	-6,000.00	23,226.55	Transfer Payment	Expense	Mobile Banking	Warehouse r...
Aug.12	Mashang Consu...	-1,300.00	21,926.55	Direct Debit	Expense	E-commerce	Working capi...
Aug.21	Interest Credit	1.75	21,928.30	Interest	Income	Internal	Demand dep...
Sep.05	Alipay Network...	14,000.00	35,928.30	Agency Payment	Income	E-commerce	Online order...
Sep.05	Yunnan Jinpeng...	-12,000.00	23,928.30	Transfer Payment	Expense	Super Online B...	Steel profiles...
Sep.10	Duxiaoman Cho...	-1,500.00	22,428.30	Loan Deduction	Expense	E-commerce	Business inst...
Sep.14	Kunming Xiangy...	-1,500.00	20,928.30	Transfer Payment	Expense	Mobile Banking	Delivery freig...
Sep.18	Yunnan Zheng'a...	6,800.00	27,728.30	Transfer Credit	Income	Super Online B...	Project suppl...
Sep.21	Interest Credit	2.08	27,730.38	Interest	Income	Internal	Demand dep...
Oct.06	Payment Platfor...	13,500.00	41,230.38	Agency Payment	Income	E-commerce	Platform pay...
Oct.06	Kunming Hongd...	-10,000.00	31,230.38	Transfer Payment	Expense	Super Online B...	Cement proc...
Oct.10	Duxiaoman Cho...	-1,800.00	29,430.38	Loan Deduction	Expense	E-commerce	Business inst...
Oct.16	Kunming Mingh...	-2,600.00	26,830.38	Transfer Payment	Expense	Mobile Banking	Electrical sup...
Oct.20	Cash Deposit	500	27,330.38	Cash	Income	ATM	Store cash re...
Oct.21	Interest Credit	2.38	27,332.76	Interest	Income	Internal	Demand dep...
Nov.03	Alipay Network...	12,800.00	40,132.76	Agency Payment	Income	E-commerce	Online order...
Nov.03	Yunnan Wanton...	-9,000.00	31,132.76	Transfer Payment	Expense	Super Online B...	Concrete ad...
Nov.08	Kunming Youxin...	-4,800.00	26,332.76	Transfer Payment	Expense	Mobile Banking	Temporary...
Nov.10	Duxiaoman Cho...	-1,800.00	24,532.76	Loan Deduction	Expense	E-commerce	Business inst...
Nov.18	Cash Deposit	600.00	25,132.76	Cash	Income	ATM	Store cash re...
Nov.21	Interest Credit	2.19	25,134.95	Interest	Income	Internal	Demand dep...
Dec.05	Payment Platfor...	15,000.00	40,134.95	Agency Payment	Income	E-commerce	Platform pay...
Dec.05	Yunnan Jinpeng...	-12,000.00	28,134.95	Transfer Payment	Expense	Super Online B...	Steel profiles...
Dec.10	Mashang Consu...	-1,300.00	26,834.95	Direct Debit	Expense	E-commerce	Working capi...
Dec.15	Kunming Xinshe...	-3,000.00	23,834.95	Transfer Payment	Expense	Mobile Banking	Warehouse...
Dec.21	Interest Credit	2.05	23,836.00	Interest	Income	Internal	Demand dep...
Dec.22	Yunnan Zheng'a...	7,600.00	31,436.00	Transfer Credit	Income	Super Online B...	Project suppl...

Question

Based on the inflow/outflow patterns and counterparty characteristics, which option best reflects whether suspected cash-out behavior exists?

a. No obvious large round-number in-out or circular transfers to same account; transactions are predominantly business-related in nature

b. Frequent same-day mutual transfers without business transaction remarks

c. Multiple credit card channel cash withdrawals with rapid fund aggregation

d. Large transfers to personal wallets with short-term return transfers

Answer

Key Evidence Supporting Answer

A

- **Income sources:** E-commerce settlements, project payments, business loans
- **Expense destinations:** Cement factory, steel factory, piping company, transport
- All counterparties are registered business entities with clear names
- Transaction amounts are non-round numbers typical of actual business deals
- Time gaps between loan receipt and expenditure (same day but different amounts)

Figure 9: Sample Single Choice Question about Comprehension & Calculation of Micro-enterprise Clients Log Information

Gig Economy Worker Transaction Log (2023 - Third-Party Platform Analysis)

Total Records

- 96 transactions
- Dual income: Salary + Streamer tips (side gig)

Third-Party Platform Transactions

- Income: 12 transactions (monthly tips settlements)
- Expense: 7 transactions (online shopping + refund)
- Absolute difference: |12 - 7| = 5

Skill Test & Cognitive Level

- Skills: Data filtering by channel, Transaction categorization, Basic arithmetic
- Cognitive Level: Comprehension & Calculation (Medium)
- Tests: Ability to filter by specific criteria and count transactions by type

Date	Counterparty	Amount	Balance	Type	In/Out	Channel	Remarks
Jan.05	Suzhou Hongch...	6,200.00	9,200.00	Salary Credit	Income	Interbank Trans...	January salar...
Jan.10	Hangzhou Shan...	350.00	9,550.00	Platform Tips In...	Income	Third-Party Plat...	Streamer tip...
Jan.12	UnionPay Merc...	-600.00	8,950.00	WM Subscription	Expense	Mobile Banking	Subscribe - S...
Jan.15	Nanjing Yuanwe...	-128.00	8,822.00	Dining	Expense	POS Terminal	Lunch - Yuan...
Jan.18	Shanghai Hezho...	-236.40	8,585.60	Retail Shopping	Expense	POS Terminal	Convenience...
Jan.21	Ningbo Yong'an...	-168	8,417.60	Utilities	Expense	Quick Payment	Water bill - Ja...
Jan.23	Beijing Zhaolian...	-1,500.00	6,917.60	Loan Repayment	Expense	Online Banking	Consumer lo...
Jan.25	UnionPay Merc...	600.00	7,517.60	WM Redemption	Income	Mobile Banking	Redeem - St...
Feb.03	Suzhou Hongch...	6,100.00	13,617.60	Salary Credit	Income	Interbank Trans...	February sal...
Feb.08	Xiamen Luqiao...	420.00	14,037.60	Platform Tips In...	Income	Third-Party Plat...	Streamer tip...
Feb.11	UnionPay Merc...	-800.00	13,237.60	WM Subscription	Expense	Mobile Banking	Subscribe - S...
Feb.14	Nanjing Yuanwe...	-98.00	13,139.60	Dining	Expense	POS Terminal	Dinner - Yua...
Feb.18	Shanghai Ctrip...	-320.00	12,819.60	Online Purchase	Expense	Third-Party Plat...	Flight rebook...
Feb.20	Beijing Unicom...	-200.00	12,619.60	Telecom Top-up	Expense	Quick Payment	Phone bill - F...
Feb.23	Beijing Zhaolian...	-1,500.00	11,119.60	Loan Repayment	Expense	Online Banking	Consumer lo...
Feb.26	UnionPay Merc...	800	11,919.60	WM Redemption	Income	Mobile Banking	Redeem - St...
Mar.05	Suzhou Hongch...	6,150.00	18,069.60	Salary Credit	Income	Interbank Trans...	March salary...
Mar.09	Hangzhou Beike...	260.00	18,329.60	Platform Tips In...	Income	Third-Party Plat...	Streamer tip...
Mar.12	UnionPay Merc...	-700.00	17,629.60	WM Subscription	Expense	Mobile Banking	Subscribe - C...
Mar.16	Guangzhou Ron...	-35.00	17,594.60	Late Fee Deduc...	Expense	Online Banking	Credit card l...
Mar.18	Shanghai Hezho...	-210.50	17,384.10	Retail Shopping	Expense	POS Terminal	Supermarket...
Mar.21	Ningbo Yong'an...	-172.00	17,212.10	Utilities	Expense	Quick Payment	Electricity -...
Mar.24	Beijing Zhaolian...	-1,600.00	15,612.10	Loan Repayment	Expense	Online Banking	Consumer lo...
Mar.27	UnionPay Merc...	700	16,312.10	WM Redemption	Income	Mobile Banking	Redeem - Ca...
Apr.04	Suzhou Hongch...	6,200.00	22,512.10	Salary Credit	Income	Interbank Trans...	April salary -...
Apr.08	Shenzhen AVIC I...	-800.00	21,712.10	Insurance Prem...	Expense	Online Banking	Annual insur...
Apr.10	Xiamen Luqiao...	380.00	22,092.10	Platform Tips In...	Income	Third-Party Plat...	Tips settlem...
Apr.12	UnionPay Merc...	-1,000.00	21,092.10	WM Subscription	Expense	Mobile Banking	Subscribe - F...
Apr.15	Nanjing Yuanwe...	-132.00	20,960.10	Dining	Expense	POS Terminal	Family gathe...
Apr.19	Beijing JD Centu...	-588.00	20,372.10	Online Purchase	Expense	Third-Party Plat...	Home applia...
Apr.22	Beijing Zhaolian...	-1,600.00	18,772.10	Loan Repayment	Expense	Online Banking	Consumer lo...
Apr.25	UnionPay Merc...	1,000.00	19,772.10	WM Redemption	Income	Mobile Banking	Redeem - Fix...
May.05	Suzhou Hongch...	6,180.00	25,952.10	Salary Credit	Income	Interbank Trans...	May salary -...
May.08	Shanghai Ruitai...	-450.00	25,502.10	Property Fee	Expense	Quick Payment	Property fee...
May.10	Xiamen Luqiao...	410.00	25,912.10	Platform Tips In...	Income	Third-Party Plat...	Tips settlem...
May.13	UnionPay Merc...	-800.00	25,112.10	WM Subscription	Expense	Mobile Banking	Subscribe -...
May.16	Shanghai Metro...	-90.00	25,022.10	Public Transit	Expense	Quick Payment	Metro top-u...
May.19	Shanghai Hezho...	-312.90	24,709.20	Retail Shopping	Expense	POS Terminal	Daily supplie...
May.23	Beijing Zhaolian...	-1,700.00	23,009.20	Loan Repayment	Expense	Online Banking	Consumer lo...
May.26	UnionPay Merc...	800.00	23,809.20	WM Redemption	Income	Mobile Banking	Redeem - De...
Jun.05	Suzhou Hongch...	6,220.00	30,029.20	Salary Credit	Income	Interbank Trans...	June salary -...
Jun.08	Beijing Unicom...	-200	29,829.20	Telecom Top-up	Expense	Quick Payment	Phone bill - J...
Jun.10	Xiamen Luqiao...	390.00	30,219.20	Platform Tips In...	Income	Third-Party Plat...	Tips settlem...
Jun.12	UnionPay Merc...	-900.00	29,319.20	WM Subscription	Expense	Mobile Banking	Subscribe - S...
Jun.16	Nanjing Yuanwe...	-156.00	29,163.20	Dining	Expense	POS Terminal	Family dinner
Jun.20	Ningbo Yong'an...	-175.00	28,988.20	Utilities	Expense	Quick Payment	Water bill - J...
Jun.23	Beijing Zhaolian...	-1,700.00	27,288.20	Loan Repayment	Expense	Online Banking	Consumer lo...
Jun.26	UnionPay Merc...	900.00	28,188.20	WM Redemption	Income	Mobile Banking	Redeem - Sh...
Jul.05	Suzhou Hongch...	6,150.00	34,338.20	Salary Credit	Income	Interbank Trans...	July salary - S...
Jul.07	Shenzhen AVIC I...	-800.00	33,538.20	Insurance Prem...	Expense	Online Banking	Annual insur...
Jul.10	Xiamen Luqiao...	420	33,958.20	Platform Tips In...	Income	Third-Party Plat...	Tips settlem...
Jul.12	UnionPay Merc...	-1,000.00	32,958.20	WM Subscription	Expense	Mobile Banking	Subscribe - F...
Jul.15	Shanghai Hezho...	-280.40	32,677.80	Retail Shopping	Expense	POS Terminal	Summer sup...
Jul.19	Shanghai Metro...	-80.00	32,597.80	Public Transit	Expense	Quick Payment	Metro top-u...
Jul.22	Beijing Zhaolian...	-1,800.00	30,797.80	Loan Repayment	Expense	Online Banking	Consumer lo...
Jul.25	UnionPay Merc...	1,000.00	31,797.80	WM Redemption	Income	Mobile Banking	Redeem - Fix...
Aug.04	Suzhou Hongch...	6,200.00	37,997.80	Salary Credit	Income	Interbank Trans...	August salar...
Aug.06	Shanghai Ruitai...	-450.00	37,547.80	Property Fee	Expense	Quick Payment	Property fee...
Aug.09	Xiamen Luqiao...	360	37,907.80	Platform Tips In...	Income	Third-Party Plat...	Tips settlem...
Aug.12	UnionPay Merc...	-800.00	37,107.80	WM Subscription	Expense	Mobile Banking	Subscribe - S...
Aug.15	Nanjing Yuanwe...	-142.00	36,965.80	Dining	Expense	POS Terminal	Work meal
Aug.18	Beijing JD Centu...	-420.00	36,545.80	Online Purchase	Expense	Third-Party Plat...	Digital acces...
Aug.21	Beijing Zhaolian...	-1,800.00	34,745.80	Loan Repayment	Expense	Online Banking	Consumer lo...
Aug.24	UnionPay Merc...	800.00	35,545.80	WM Redemption	Income	Mobile Banking	Redeem - St...
Sep.05	Suzhou Hongch...	6,180.00	41,725.80	Salary Credit	Income	Interbank Trans...	September s...
Sep.08	Beijing Unicom...	-200.00	41,525.80	Telecom Top-up	Expense	Quick Payment	Phone bill - S...
Sep.10	Xiamen Luqiao...	400.00	41,925.80	Platform Tips In...	Income	Third-Party Plat...	Tips settlem...
Sep.12	UnionPay Merc...	-900.00	41,025.80	WM Subscription	Expense	Mobile Banking	Subscribe - S...
Sep.15	Shanghai Hezho...	-312	40,713.80	Retail Shopping	Expense	POS Terminal	Monthly rest...
Sep.19	Shanghai Metro...	-80.00	40,633.80	Public Transit	Expense	Quick Payment	Metro top-u...
Sep.22	Beijing Zhaolian...	-1,900.00	38,733.80	Loan Repayment	Expense	Online Banking	Consumer lo...
Sep.26	UnionPay Merc...	900.00	39,633.80	WM Redemption	Income	Mobile Banking	Redeem - Sh...
Oct.05	Suzhou Hongch...	6,200.00	45,833.80	Salary Credit	Income	Interbank Trans...	October sala...
Oct.07	Shenzhen AVIC I...	-800.00	45,033.80	Insurance Prem...	Expense	Online Banking	Annual insur...
Oct.10	Xiamen Luqiao...	380.00	45,413.80	Platform Tips In...	Income	Third-Party Plat...	Tips settlem...
Oct.12	UnionPay Merc...	-1,000.00	44,413.80	WM Subscription	Expense	Mobile Banking	Subscribe - F...
Oct.15	Nanjing Yuanwe...	-168.00	44,245.80	Dining	Expense	POS Terminal	Weekend gat...
Oct.18	Beijing JD Centu...	-520.00	43,725.80	Online Purchase	Expense	Third-Party Plat...	Autumn clot...
Oct.21	Beijing Zhaolian...	-1,900.00	41,825.80	Loan Repayment	Expense	Online Banking	Consumer lo...
Oct.25	UnionPay Merc...	1,000.00	42,825.80	WM Redemption	Income	Mobile Banking	Redeem - Fix...
Nov.03	Suzhou Hongch...	6,180.00	49,005.80	Salary Credit	Income	Interbank Trans...	November s...
Nov.06	Shanghai Ruitai...	-450.00	48,555.80	Property Fee	Expense	Quick Payment	Property fee...
Nov.09	Xiamen Luqiao...	410.00	48,965.80	Platform Tips In...	Income	Third-Party Plat...	Tips settlem...
Nov.12	UnionPay Merc...	-900.00	48,065.80	WM Subscription	Expense	Mobile Banking	Subscribe - S...
Nov.15	Shanghai Hezho...	-280.00	47,785.80	Retail Shopping	Expense	POS Terminal	Monthly pur...
Nov.18	Beijing JD Centu...	-650.00	47,135.80	Online Purchase	Expense	Third-Party Plat...	Winter coat
Nov.21	Guangzhou Yun...	-350.00	46,785.80	Refund Payout	Expense	Third-Party Plat...	Order refun...
Nov.23	Beijing Zhaolian...	-1,900.00	44,885.80	Loan Repayment	Expense	Online Banking	Consumer lo...
Dec.05	Suzhou Hongch...	6,200.00	51,085.80	Salary Credit	Income	Interbank Trans...	December s...
Dec.08	Beijing Unicom...	-200.00	50,885.80	Telecom Top-up	Expense	Quick Payment	Phone bill -...
Dec.10	Xiamen Luqiao...	390.00	51,275.80	Platform Tips In...	Income	Third-Party Plat...	Tips settlem...
Dec.12	UnionPay Merc...	-1,000.00	50,275.80	WM Subscription	Expense	Mobile Banking	Subscribe - Y...
Dec.15	Nanjing Yuanwe...	-176.00	50,099.80	Dining	Expense	POS Terminal	Year-end gat...
Dec.18	Beijing JD Centu...	-980.00	49,119.80	Online Purchase	Expense	Third-Party Plat...	Year-end sh...
Dec.22	Beijing Zhaolian...	-2,000.00	47,119.80	Loan Repayment	Expense	Online Banking	Consumer lo...
Dec.26	UnionPay Merc...	1,000.00	48,119.80	WM Redemption	Income	Mobile Banking	Redeem - Ye...

Question

Among transactions via Third-Party Platform channel, calculate the absolute difference between the number of income transactions and expense transactions?

a. 4

b. 10

c. 6

d. 5

Answer

Correct Answer

D

Question Type: Single Choice

Category: High-Risk Merchant Transactions / Consumption Structure & Concentration

Filter transactions where Channel = 'Third-Party Platform'

Figure 10: Sample Single Choice Question about Comprehension & Calculation of Personal Clients Log Information

Personal Bank Transaction Log (2023 - Online Shopping Ratio Analysis)

Account Profile

- 120 transactions
- •Dual employment: Jiahe HR + Xinghe Technology

Online Shopping Analysis

- First online shopping: February 2023
- Feb online shopping: ¥199.00
- Feb total expenses: ¥1,066.20

Skill Test & Cognitive Level

- Skills: Transaction type identification, Time-based filtering, Ratio calculation
- Cognitive Level: Comprehension & Calculation (Medium)
- Tests: Ability to identify first occurrence, sum expenses, and calculate percentages

Date	Counterparty	Amount	Balance	Type	In/Out	Channel	Remarks
Jan.03	Chunhe Conven...	-35.80	2,464.20	POS Purchase	Expense	POS Terminal	Store consu...
Jan.06	Tu'an Highway T...	-24.60	2,439.60	Highway Toll De...	Expense	Quick Payment	Auto toll ded...
Jan.09	Huimin Utilities...	-128.90	2,310.70	Utilities	Expense	Mobile Banking	Water & elec...
Jan.12	China Unicom T...	-50.00	2,260.70	Telecom Top-up	Expense	Mobile Banking	Phone recha...
Jan.15	Jincheng Consu...	-520.00	1,740.70	Loan Repayment	Expense	Quick Payment	Scheduled lo...
Jan.18	Chen Junming	200	1,940.70	Transfer Credit	Income	Mobile Banking	Temporary f...
Jan.21	Demand Deposi...	0.46	1,941.16	Interest Credit	Income	Bank System	Jan 2023 inte...
Jan.24	Yuehuo Superm...	-88.30	1,852.86	Supermarket Sh...	Expense	POS Terminal	Daily groceri...
Jan.27	Fengxing Electr...	15.00	1,867.86	Refund	Income	Quick Payment	Order refund
Jan.30	Jiahe HR Service...	8,500.00	10,367.86	Salary Credit	Income	Mobile Banking	Payroll Jan 2...
Feb.03	Luming Cake Sh...	-42.00	10,325.86	POS Purchase	Expense	POS Terminal	Food purcha...
Feb.06	Tu'an Highway T...	-18.20	10,307.66	Highway Toll De...	Expense	Quick Payment	Auto toll ded...
Feb.09	Huimin Utilities...	-120.50	10,187.16	Utilities	Expense	Mobile Banking	Water bill
Feb.12	Huixin Telecom...	-100.00	10,087.16	Telecom Top-up	Expense	Mobile Banking	Phone recha...
Feb.15	Mingcheng Micr...	-560.00	9,527.16	Loan Repayment	Expense	Quick Payment	Microloan re...
Feb.18	Li Chenhui	350	9,877.16	Transfer Credit	Income	Mobile Banking	Living allowa...
Feb.21	Demand Deposi...	0.52	9,877.68	Interest Credit	Income	Bank System	Feb 2023 int...
Feb.24	Zhizihua Milk Te...	-26.50	9,851.18	POS Purchase	Expense	POS Terminal	Beverage pu...
Feb.27	Ruichuang Digit...	-199.00	9,652.18	Online Shopping	Expense	Quick Payment	Online order
Feb.28	Xinghe Technol...	8,500.00	18,152.18	Salary Credit	Income	Mobile Banking	Payroll Feb 2...
Mar.03	Chunhe Conven...	-28.70	18,123.48	POS Purchase	Expense	POS Terminal	Store consu...
Mar.06	Tu'an Highway T...	-35.10	18,088.38	Highway Toll De...	Expense	Quick Payment	Auto toll ded...
Mar.09	Huimin Gas Ser...	-146.3	17,942.08	Utilities	Expense	Mobile Banking	Gas bill
Mar.12	China Unicom T...	-50	17,892.08	Telecom Top-up	Expense	Mobile Banking	Phone recha...
Mar.15	Jincheng Consu...	-520.00	17,372.08	Loan Repayment	Expense	Quick Payment	Scheduled lo...
Mar.18	Chen Junming	180.00	17,552.08	Transfer Credit	Income	Mobile Banking	Temporary f...
Mar.21	Demand Deposi...	0.60	17,552.68	Interest Credit	Income	Bank System	Mar 2023 int...
Mar.24	Yuehuo Superm...	-96.40	17,456.28	Supermarket Sh...	Expense	POS Terminal	Daily groceri...
Mar.27	Lvyuan Lifestyle...	22.80	17,479.08	Refund	Income	Quick Payment	Order refund
Mar.30	Jiahe HR Service...	8,600.00	26,079.08	Salary Credit	Income	Mobile Banking	Payroll Mar...
Apr.03	Baiyue Car Wash	-58.00	26,021.08	POS Purchase	Expense	POS Terminal	Vehicle service
Apr.06	Hechang Highw...	-19.80	26,001.28	Highway Toll De...	Expense	Quick Payment	Auto toll ded...
Apr.09	Huimin Utilities...	-118.6	25,882.68	Utilities	Expense	Mobile Banking	Water & elec...
Apr.12	Huixin Telecom...	-50.00	25,832.68	Telecom Top-up	Expense	Mobile Banking	Phone recha...
Apr.15	Mingcheng Micr...	-560.00	25,272.68	Loan Repayment	Expense	Quick Payment	Microloan re...
Apr.18	Li Chenhui	220.00	25,492.68	Transfer Credit	Income	Mobile Banking	Living allowa...
Apr.21	Demand Deposi...	0.68	25,493.36	Interest Credit	Income	Bank System	Apr 2023 int...
Apr.24	Lanjing Cinema	-45.00	25,448.36	POS Purchase	Expense	POS Terminal	Movie ticket
Apr.27	Fengxing Electr...	-329.00	25,119.36	Online Shopping	Expense	Quick Payment	Online order
Apr.30	Xinghe Technol...	8,700.00	33,819.36	Salary Credit	Income	Mobile Banking	Payroll Apr 2...
May.03	Chunhe Conven...	-33.90	33,785.46	POS Purchase	Expense	POS Terminal	Store consu...
May.06	Hechang Highw...	-22.4	33,763.06	Highway Toll De...	Expense	Quick Payment	Auto toll ded...
May.09	Huimin Gas Ser...	-108.70	33,654.36	Utilities	Expense	Mobile Banking	Gas bill
May.12	China Unicom T...	-50.00	33,604.36	Telecom Top-up	Expense	Mobile Banking	Phone recha...
May.15	Jincheng Consu...	-520.00	33,084.36	Loan Repayment	Expense	Quick Payment	Scheduled lo...
May.18	Chen Junming	260.00	33,344.36	Transfer Credit	Income	Mobile Banking	Temporary f...
May.21	Demand Deposi...	0.75	33,345.11	Interest Credit	Income	Bank System	May 2023 int...
May.24	Yuehuo Superm...	-83.10	33,262.01	Supermarket Sh...	Expense	POS Terminal	Daily groceri...
May.27	Qingteng Home...	36.50	33,298.51	Refund	Income	Quick Payment	Order refund
May.30	Jiahe HR Service...	8,600.00	41,898.51	Salary Credit	Income	Mobile Banking	Payroll May...
Jun.03	Luming Cake Sh...	-40.5	41,858.01	POS Purchase	Expense	POS Terminal	Food purcha...
Jun.06	Hechang Highw...	-18.10	41,839.91	Highway Toll De...	Expense	Quick Payment	Auto toll ded...
Jun.09	Huimin Utilities...	-126.40	41,713.51	Utilities	Expense	Mobile Banking	Water & elec...
Jun.12	Huixin Telecom...	-100.00	41,613.51	Telecom Top-up	Expense	Mobile Banking	Phone recha...
Jun.15	Mingcheng Micr...	-560.00	41,053.51	Loan Repayment	Expense	Quick Payment	Microloan re...
Jun.18	Li Chenhui	300.00	41,353.51	Transfer Credit	Income	Mobile Banking	Living allowa...
Jun.21	Demand Deposi...	0.82	41,354.33	Interest Credit	Income	Bank System	Jun 2023 inte...
Jun.24	Lanjing Cinema	-52.00	41,302.33	POS Purchase	Expense	POS Terminal	Movie ticket
Jun.27	Ruichuang Digit...	-499	40,803.33	Online Shopping	Expense	Quick Payment	Online order
Jun.30	Xinghe Technol...	8,800.00	49,603.33	Salary Credit	Income	Mobile Banking	Payroll Jun 2...
Jul.03	Chunhe Conven...	-29.60	49,573.73	POS Purchase	Expense	POS Terminal	Store consu...
Jul.06	Hechang Highw...	-22.90	49,550.83	Highway Toll De...	Expense	Quick Payment	Auto toll ded...
Jul.09	Huimin Gas Ser...	-118.60	49,432.23	Utilities	Expense	Mobile Banking	Gas bill
Jul.12	China Unicom T...	-50.00	49,382.23	Telecom Top-up	Expense	Mobile Banking	Phone recha...
Jul.15	Jincheng Consu...	-520.00	48,862.23	Loan Repayment	Expense	Quick Payment	Scheduled lo...
Jul.18	Chen Junming	200.00	49,062.23	Transfer Credit	Income	Mobile Banking	Temporary f...
Jul.21	Demand Deposi...	0.90	49,063.13	Interest Credit	Income	Bank System	Jul 2023 inte...
Jul.24	Yuehuo Superm...	-92.40	48,970.73	Supermarket Sh...	Expense	POS Terminal	Daily groceri...
Jul.27	Lvyuan Lifestyle...	18.5	48,989.23	Refund	Income	Quick Payment	Order refund
Jul.30	Jiahe HR Service...	8,800.00	57,789.23	Salary Credit	Income	Mobile Banking	Payroll Jul 20...
Aug.03	Luming Cake Sh...	-44.20	57,745.03	POS Purchase	Expense	POS Terminal	Food purcha...
Aug.06	Hechang Highw...	-21.50	57,723.53	Highway Toll De...	Expense	Quick Payment	Auto toll ded...
Aug.09	Huimin Utilities...	-121.30	57,602.23	Utilities	Expense	Mobile Banking	Water & elec...
Aug.12	Huixin Telecom...	-100.00	57,502.23	Telecom Top-up	Expense	Mobile Banking	Phone recha...
Aug.15	Mingcheng Micr...	-560.00	56,942.23	Loan Repayment	Expense	Quick Payment	Microloan re...
Aug.18	Li Chenhui	400.00	57,342.23	Transfer Credit	Income	Mobile Banking	Living allowa...
Aug.21	Demand Deposi...	0.98	57,343.21	Interest Credit	Income	Bank System	Aug 2023 int...
Aug.24	Lanjing Cinema	-36.00	57,307.21	POS Purchase	Expense	POS Terminal	Movie ticket
Aug.27	Fengxing Electr...	-259	57,048.21	Online Shopping	Expense	Quick Payment	Online order
Aug.30	Xinghe Technol...	8,900.00	65,948.21	Salary Credit	Income	Mobile Banking	Payroll Aug 2...
Sep.03	Chunhe Conven...	-31.70	65,916.51	POS Purchase	Expense	POS Terminal	Store consu...
Sep.06	Hechang Highw...	-24.80	65,891.71	Highway Toll De...	Expense	Quick Payment	Auto toll ded...
Sep.09	Huimin Gas Ser...	-109.20	65,782.51	Utilities	Expense	Mobile Banking	Gas bill
Sep.12	China Unicom T...	-50.00	65,732.51	Telecom Top-up	Expense	Mobile Banking	Phone recha...
Sep.15	Jincheng Consu...	-520.00	65,212.51	Loan Repayment	Expense	Quick Payment	Scheduled lo...
Sep.18	Chen Junming	180.00	65,392.51	Transfer Credit	Income	Mobile Banking	Temporary f...
Sep.21	Demand Deposi...	1.05	65,393.56	Interest Credit	Income	Bank System	Sep 2023 int...
Sep.24	Yuehuo Superm...	-85.6	65,307.96	Supermarket Sh...	Expense	POS Terminal	Daily groceri...
Sep.27	Qingteng Home...	25.00	65,332.96	Refund	Income	Quick Payment	Order refund
Sep.30	Jiahe HR Service...	9,000.00	74,332.96	Salary Credit	Income	Mobile Banking	Payroll Sep 2...
Oct.03	Luming Cake Sh...	-41.30	74,291.66	POS Purchase	Expense	POS Terminal	Food purcha...
Oct.06	Hechang Highw...	-20.60	74,271.06	Highway Toll De...	Expense	Quick Payment	Auto toll ded...
Oct.09	Huimin Utilities...	-118.60	74,152.46	Utilities	Expense	Mobile Banking	Water & elec...
Oct.12	Huixin Telecom...	-100.00	74,052.46	Telecom Top-up	Expense	Mobile Banking	Phone recha...
Oct.15	Mingcheng Micr...	-560.00	73,492.46	Loan Repayment	Expense	Quick Payment	Microloan re...
Oct.18	Li Chenhui	220.00	73,712.46	Transfer Credit	Income	Mobile Banking	Living allowa...
Oct.21	Demand Deposi...	1.12	73,713.58	Interest Credit	Income	Bank System	Oct 2023 int...
Oct.24	Lanjing Cinema	-48	73,665.58	POS Purchase	Expense	POS Terminal	Movie ticket
Oct.27	Ruichuang Digit...	-399.00	73,266.58	Online Shopping	Expense	Quick Payment	Online order
Oct.30	Xinghe Technol...	9,000.00	82,266.58	Salary Credit	Income	Mobile Banking	Payroll Oct 2...
Nov.03	Chunhe Conven...	-33.10	82,233.48	POS Purchase	Expense	POS Terminal	Store consu...
Nov.06	Hechang Highw...	-19.70	82,213.78	Highway Toll De...	Expense	Quick Payment	Auto toll ded...
Nov.09	Huimin Gas Ser...	-116.40	82,097.38	Utilities	Expense	Mobile Banking	Gas bill
Nov.12	China Unicom T...	-50.00	82,047.38	Telecom Top-up	Expense	Mobile Banking	Phone recha...
Nov.15	Jincheng Consu...	-520.00	81,527.38	Loan Repayment	Expense	Quick Payment	Scheduled lo...
Nov.18	Chen Junming	260.00	81,787.38	Transfer Credit	Income	Mobile Banking	Temporary f...
Nov.21	Demand Deposi...	1.20	81,788.58	Interest Credit	Income	Bank System	Nov 2023 int...
Nov.24	Yuehuo Superm...	-91.2	81,697.38	Supermarket Sh...	Expense	POS Terminal	Daily groceri...
Nov.27	Lvyuan Lifestyle...	32.60	81,729.98	Refund	Income	Quick Payment	Order refund
Nov.30	Jiahe HR Service...	9,200.00	90,929.98	Salary Credit	Income	Mobile Banking	Payroll Nov 2...
Dec.03	Luming Cake Sh...	-45.8	90,884.18	POS Purchase	Expense	POS Terminal	Food purcha...
Dec.06	Hechang Highw...	-26.7	90,857.48	Highway Toll De...	Expense	Quick Payment	Auto toll ded...
Dec.09	Huimin Utilities...	-124.2	90,733.28	Utilities	Expense	Mobile Banking	Water & elec...
Dec.12	Huixin Telecom...	-100	90,633.28	Telecom Top-up	Expense	Mobile Banking	Phone recha...
Dec.15	Mingcheng Micr...	-560	90,073.28	Loan Repayment	Expense	Quick Payment	Microloan re...
Dec.18	Li Chenhui	500	90,573.28	Transfer Credit	Income	Mobile Banking	Year-end tra...
Dec.21	Demand Deposi...	1.28	90,574.56	Interest Credit	Income	Bank System	Dec 2023 int...
Dec.24	Lanjing Cinema	-42	90,532.56	POS Purchase	Expense	POS Terminal	Movie ticket
Dec.27	Fengxing Electr...	-299	90,233.56	Online Shopping	Expense	Quick Payment	Online order
Dec.30	Xinghe Technol...	9,500.00	99,733.56	Salary Credit	Income	Mobile Banking	Payroll Dec 2...

Question

In the first month when online shopping expenses appeared, what percentage of total monthly expenses did online shopping represent?

Answer

Correct Answer

17.5%

Calculation Process:

Total February Expenses = 42.00 + 18.20 + 120.50 + 100.00 + 560.00 + 26.50 + 199.00

Online Shopping Ratio = Online Shopping ÷ Total Expenses × 100%= 17.5%

Question Type: Calculation question

Category: High-Risk Merchant Transactions / Consumption Structure & Concentration

Figure 11: Sample Calculation Question about Comprehension & Calculation of Personal Clients Log Information

SME Food Service Business Transaction Log (2023 - Expense Rationality Analysis)

Business Profile

- 128 transactions
- Restaurant/Food delivery business in Changsha

Financial Summary

- Total Income: ¥201,020.41
- Total Expenses: ¥104,522.90
- Net Profit Margin: ~48%

Date	Counterparty	Amount	Balance	Type	In/Out	Channel	Remarks
Jan.02	Hangzhou Wanx...	8,500.00	8,500.00	Agency Payment	Income	E-commerce	Platform sett...
Jan.02	Changsha Yufen...	-3,200.00	5,300.00	Transfer Payment	Expense	Mobile Banking	Raw material...
Jan.05	UnionPay Merc...	2,400.00	7,700.00	UnionPay Credit	Income	UnionPay	POS receipt -...
Jan.08	Hunan Jiacheng...	-1,800.00	5,900.00	Transfer Payment	Expense	Mobile Banking	Packaging ba...
Jan.10	Xingsha Wangd...	-2,000.00	3,900.00	Transfer Payment	Expense	Online Banking	Store rent - J...
Jan.12	Meituan Dianpi...	3,600.00	7,500.00	Agency Payment	Income	E-commerce	Platform sett...
Jan.15	Changsha Jinhui...	-560.00	6,940.00	Transfer Payment	Expense	Mobile Banking	Menu printin...
Jan.21	Changsha Wate...	-380.00	6,560.00	Alipay Payment	Expense	E-commerce	Water bill - S...
Jan.22	State Grid Huna...	-420.00	6,140.00	Alipay Payment	Expense	E-commerce	Electricity - S...
Jan.25	Ele.me Network...	4,100.00	10,240.00	Agency Payment	Income	E-commerce	Platform sett...
Jan.31	ICBC Co., Ltd.	0.86	10,240.86	Interest Credit	Income	Bank System	Demand dep...
Feb.01	Changsha Yufen...	-2,500.00	7,740.86	Transfer Payment	Expense	Mobile Banking	Raw material...
Feb.03	UnionPay Merc...	2,700.00	10,440.86	UnionPay Credit	Income	UnionPay	POS receipt -...
Feb.05	Xingsha Wangd...	-2,000.00	8,440.86	Transfer Payment	Expense	Online Banking	Store rent - F...
Feb.07	Hunan Jiacheng...	-680.00	7,760.86	Transfer Payment	Expense	Mobile Banking	Packaging m...
Feb.08	Meituan Dianpi...	5,200.00	12,960.86	Agency Payment	Income	E-commerce	Platform sett...
Feb.12	SF Express Co,...	-600.00	12,360.86	Transfer Payment	Expense	Online Banking	Courier fee -...
Feb.15	Ele.me Network...	4,500.00	16,860.86	Agency Payment	Income	E-commerce	Platform sett...
Feb.20	Changsha Xinda...	-1,500.00	15,360.86	Transfer Payment	Expense	Mobile Banking	Equipment r...
Feb.22	Hunan Wanjia C...	-900.00	14,460.86	Transfer Payment	Expense	Online Banking	Store cleanin...
Feb.25	UnionPay Merc...	2,300.00	16,760.86	UnionPay Credit	Income	UnionPay	POS receipt -...
Feb.28	ICBC Co., Ltd.	0.95	16,761.81	Interest Credit	Income	Bank System	Demand dep...
Mar.01	Changsha Yufen...	-3,200.00	13,561.81	Transfer Payment	Expense	Mobile Banking	Raw material...
Mar.03	Meituan Dianpi...	5,800.00	19,361.81	Agency Payment	Income	E-commerce	Platform sett...
Mar.06	Xingsha Wangd...	-2,000.00	17,361.81	Transfer Payment	Expense	Online Banking	Store rent -...
Mar.08	Hunan Jiacheng...	-760.00	16,601.81	Transfer Payment	Expense	Mobile Banking	Packaging m...
Mar.10	UnionPay Merc...	2,600.00	19,201.81	UnionPay Credit	Income	UnionPay	POS receipt -...
Mar.12	Changsha Wate...	-450.00	18,751.81	Alipay Payment	Expense	E-commerce	Water bill - S...
Mar.13	State Grid Huna...	-520.00	18,231.81	Alipay Payment	Expense	E-commerce	Electricity - S...
Mar.15	Ele.me Network...	6,200.00	24,431.81	Agency Payment	Income	E-commerce	Platform sett...
Mar.20	Changsha Xinch...	-2,000.00	22,431.81	Transfer Payment	Expense	Online Banking	Employee all...
Mar.25	UnionPay Merc...	2,400.00	24,831.81	UnionPay Credit	Income	UnionPay	POS receipt -...
Mar.31	ICBC Co., Ltd.	1.15	24,832.96	Interest Credit	Income	Bank System	Demand dep...
Apr.02	Changsha Yufen...	-3,500.00	21,332.96	Transfer Payment	Expense	Mobile Banking	Raw material...
Apr.04	Meituan Dianpi...	5,400.00	26,732.96	Agency Payment	Income	E-commerce	Platform sett...
Apr.05	Xingsha Wangd...	-2,000.00	24,732.96	Transfer Payment	Expense	Online Banking	Store rent -...
Apr.07	Hunan Jiacheng...	-980.00	23,752.96	Transfer Payment	Expense	Mobile Banking	Packaging m...
Apr.10	Yunda Express...	-650.00	23,102.96	Transfer Payment	Expense	Online Banking	Courier fee -...
Apr.12	Ele.me Network...	4,800.00	27,902.96	Agency Payment	Income	E-commerce	Platform sett...
Apr.18	Hunan Wanjia C...	-800.00	27,102.96	Transfer Payment	Expense	Online Banking	Store cleanin...
Apr.22	UnionPay Merc...	3,000.00	30,102.96	UnionPay Credit	Income	UnionPay	POS receipt -...
Apr.28	Xiangxiang Hen...	2,600.00	32,702.96	Interbank Trans...	Income	Online Banking	Group purch...
Apr.30	ICBC Co., Ltd.	1.40	32,704.36	Interest Credit	Income	Bank System	Demand dep...
May.01	Changsha Yufen...	-3,800.00	28,904.36	Transfer Payment	Expense	Mobile Banking	Raw material...
May.03	Meituan Dianpi...	6,200.00	35,104.36	Agency Payment	Income	E-commerce	Platform sett...
May.05	Xingsha Wangd...	-2,000.00	33,104.36	Transfer Payment	Expense	Online Banking	Store rent -...
May.08	Hunan Jiacheng...	-1,200.00	31,904.36	Transfer Payment	Expense	Mobile Banking	Packaging m...
May.10	Ele.me Network...	5,100.00	37,004.36	Agency Payment	Income	E-commerce	Platform sett...
May.12	Changsha Wate...	-360.00	36,644.36	Alipay Payment	Expense	E-commerce	Water bill - S...
May.13	State Grid Huna...	-420.00	36,224.36	Alipay Payment	Expense	E-commerce	Electricity - S...
May.16	Hunan Xingtu M...	-1,600.00	34,624.36	Transfer Payment	Expense	Online Banking	Local adverti...
May.20	UnionPay Merc...	2,800.00	37,424.36	UnionPay Credit	Income	UnionPay	POS receipt -...
May.25	Ningxiang Jinqia...	2,400.00	39,824.36	Interbank Trans...	Income	Online Banking	Wholesale c...
May.31	ICBC Co., Ltd.	1.70	39,826.06	Interest Credit	Income	Bank System	Demand dep...
Jun.01	Changsha Yufen...	-4,200.00	35,626.06	Transfer Payment	Expense	Mobile Banking	Raw material...
Jun.03	Meituan Dianpi...	6,600.00	42,226.06	Agency Payment	Income	E-commerce	Platform sett...
Jun.05	Xingsha Wangd...	-2,000.00	40,226.06	Transfer Payment	Expense	Online Banking	Store rent - J...
Jun.06	Changsha Huito...	-1,800.00	38,426.06	Transfer Payment	Expense	Online Banking	Employee sal...
Jun.08	Hunan Jiacheng...	-950	37,476.06	Transfer Payment	Expense	Mobile Banking	Packaging m...
Jun.10	Ele.me Network...	5,500.00	42,976.06	Agency Payment	Income	E-commerce	Platform sett...
Jun.15	ZTO Express Co.,...	-700.00	42,276.06	Transfer Payment	Expense	Online Banking	Courier fee -...
Jun.20	UnionPay Merc...	3,200.00	45,476.06	UnionPay Credit	Income	UnionPay	POS receipt -...
Jun.25	Zhuzhou Huam...	3,000.00	48,476.06	Interbank Trans...	Income	Online Banking	Wholesale c...
Jun.30	ICBC Co., Ltd.	2.10	48,478.16	Interest Credit	Income	Bank System	Demand dep...
Jul.01	Changsha Yufen...	-4,600.00	43,878.16	Transfer Payment	Expense	Mobile Banking	Raw material...
Jul.03	Meituan Dianpi...	6,800.00	50,678.16	Agency Payment	Income	E-commerce	Platform sett...
Jul.05	Xingsha Wangd...	-2,000.00	48,678.16	Transfer Payment	Expense	Online Banking	Store rent - J...
Jul.07	Hunan Jiacheng...	-1,100.00	47,578.16	Transfer Payment	Expense	Mobile Banking	Packaging m...
Jul.10	Ele.me Network...	5,900.00	53,478.16	Agency Payment	Income	E-commerce	Platform sett...
Jul.12	Changsha Wate...	-380.00	53,098.16	Alipay Payment	Expense	E-commerce	Water bill - S...
Jul.12	State Grid Huna...	-520.00	52,578.16	Alipay Payment	Expense	E-commerce	Electricity - S...
Jul.15	Hunan Tianyua...	-1,800.00	50,778.16	Transfer Payment	Expense	Mobile Banking	Frozen ingre...
Jul.20	UnionPay Merc...	3,400.00	54,178.16	UnionPay Credit	Income	UnionPay	POS receipt -...
Jul.25	Xiangtan Shunfa...	2,600.00	56,778.16	Interbank Trans...	Income	Online Banking	Wholesale c...
Jul.31	ICBC Co., Ltd.	2.35	56,780.51	Interest Credit	Income	Bank System	Demand dep...
Aug.01	Changsha Yufen...	-4,800.00	51,980.51	Transfer Payment	Expense	Mobile Banking	Raw material...
Aug.03	Meituan Dianpi...	7,200.00	59,180.51	Agency Payment	Income	E-commerce	Platform sett...
Aug.05	Xingsha Wangd...	-2,000.00	57,180.51	Transfer Payment	Expense	Online Banking	Store rent -...
Aug.07	Hunan Jiacheng...	-1,300.00	55,880.51	Transfer Payment	Expense	Mobile Banking	Packaging m...
Aug.10	Ele.me Network...	6,100.00	61,980.51	Agency Payment	Income	E-commerce	Platform sett...
Aug.12	YTO Express Co....	-850.00	61,130.51	Transfer Payment	Expense	Online Banking	Courier fee -...
Aug.15	Changsha Huito...	-2,000.00	59,130.51	Transfer Payment	Expense	Online Banking	Employee sal...
Aug.18	UnionPay Merc...	3,600.00	62,730.51	UnionPay Credit	Income	UnionPay	POS receipt -...
Aug.22	Changde Wangs...	2,800.00	65,530.51	Interbank Trans...	Income	Online Banking	Wholesale c...
Aug.31	ICBC Co., Ltd.	2.75	65,533.26	Interest Credit	Income	Bank System	Demand dep...
Sep.01	Changsha Yufen...	-5,000.00	60,533.26	Transfer Payment	Expense	Mobile Banking	Raw material...
Sep.03	Meituan Dianpi...	7,400.00	67,933.26	Agency Payment	Income	E-commerce	Platform sett...
Sep.05	Xingsha Wangd...	-2,000.00	65,933.26	Transfer Payment	Expense	Online Banking	Store rent - S...
Sep.07	Hunan Jiacheng...	-1,250.00	64,683.26	Transfer Payment	Expense	Mobile Banking	Packaging m...
Sep.10	Ele.me Network...	6,300.00	70,983.26	Agency Payment	Income	E-commerce	Platform sett...
Sep.12	Changsha Wate...	-400.00	70,583.26	Alipay Payment	Expense	E-commerce	Water bill - S...
Sep.12	State Grid Huna...	-520.00	70,063.26	Alipay Payment	Expense	E-commerce	Electricity - S...
Sep.15	Hunan Tianyua...	-2,200.00	67,863.26	Transfer Payment	Expense	Mobile Banking	Frozen ingre...
Sep.18	UnionPay Merc...	3,800.00	71,663.26	UnionPay Credit	Income	UnionPay	POS receipt -...
Sep.22	Yueyang Jinshu...	3,000.00	74,663.26	Interbank Trans...	Income	Online Banking	Wholesale c...
Sep.30	ICBC Co., Ltd.	3.10	74,666.36	Interest Credit	Income	Bank System	Demand dep...
Oct.01	Changsha Yufen...	-5,200.00	69,466.36	Transfer Payment	Expense	Mobile Banking	Raw material...
Oct.03	Meituan Dianpi...	7,600.00	77,066.36	Agency Payment	Income	E-commerce	Platform sett...
Oct.05	Xingsha Wangd...	-2,000.00	75,066.36	Transfer Payment	Expense	Online Banking	Store rent -...
Oct.07	Hunan Jiacheng...	-1,400.00	73,666.36	Transfer Payment	Expense	Mobile Banking	Packaging m...
Oct.10	Ele.me Network...	6,500.00	80,166.36	Agency Payment	Income	E-commerce	Platform sett...
Oct.12	J&T Express Co,....	-780.00	79,386.36	Transfer Payment	Expense	Online Banking	Courier fee -...
Oct.15	Hunan Xingtu M...	-2,500.00	76,886.36	Transfer Payment	Expense	Online Banking	Local adverti...
Oct.18	UnionPay Merc...	3,600.00	80,486.36	UnionPay Credit	Income	UnionPay	POS receipt -...
Oct.22	Shaoyang Hengf...	3,200.00	83,686.36	Interbank Trans...	Income	Online Banking	Wholesale c...
Oct.31	ICBC Co., Ltd.	3.35	83,689.71	Interest Credit	Income	Bank System	Demand dep...
Nov.01	Changsha Yufen...	-5,400.00	78,289.71	Transfer Payment	Expense	Mobile Banking	Raw material...
Nov.03	Meituan Dianpi...	7,800.00	86,089.71	Agency Payment	Income	E-commerce	Platform sett...
Nov.05	Xingsha Wangd...	-2,000.00	84,089.71	Transfer Payment	Expense	Online Banking	Store rent -...
Nov.07	Hunan Jiacheng...	-1,500.00	82,589.71	Transfer Payment	Expense	Mobile Banking	Packaging m...
Nov.10	Ele.me Network...	6,700.00	89,289.71	Agency Payment	Income	E-commerce	Platform sett...
Nov.12	Changsha Wate...	-450	88,839.71	Alipay Payment	Expense	E-commerce	Water bill - S...
Nov.12	State Grid Huna...	-530	88,309.71	Alipay Payment	Expense	E-commerce	Electricity - S...
Nov.15	Hunan Tianyua...	-2,300.00	86,009.71	Transfer Payment	Expense	Mobile Banking	Frozen ingre...
Nov.18	UnionPay Merc...	3,800.00	89,809.71	UnionPay Credit	Income	UnionPay	POS receipt -...
Nov.22	Yiyang Dongshe...	3,200.00	93,009.71	Interbank Trans...	Income	Online Banking	Wholesale c...
Nov.25	Changsha Huito...	-1,800.00	91,209.71	Transfer Payment	Expense	Online Banking	Employee all...
Nov.30	ICBC Co., Ltd.	3.7	91,213.41	Interest Credit	Income	Bank System	Demand dep...
Dec.01	Changsha Yufen...	-5,600.00	85,613.41	Transfer Payment	Expense	Mobile Banking	Raw material...
Dec.03	Meituan Dianpi...	8,000.00	93,613.41	Agency Payment	Income	E-commerce	Platform sett...
Dec.05	Xingsha Wangd...	-2,000.00	91,613.41	Transfer Payment	Expense	Online Banking	Store rent -...
Dec.07	Hunan Jiacheng...	-1,600.00	90,013.41	Transfer Payment	Expense	Mobile Banking	Packaging m...
Dec.10	Ele.me Network...	6,900.00	96,913.41	Agency Payment	Income	E-commerce	Platform sett...
Dec.12	Changsha Wate...	-480	96,433.41	Alipay Payment	Expense	E-commerce	Water bill - S...
Dec.12	State Grid Huna...	-540	95,893.41	Alipay Payment	Expense	E-commerce	Electricity - S...
Dec.15	Hunan Xingtu M...	-2,600.00	93,293.41	Transfer Payment	Expense	Online Banking	Year-end adv...
Dec.18	UnionPay Merc...	4,000.00	97,293.41	UnionPay Credit	Income	UnionPay	POS receipt -...
Dec.22	Huaihua Hengxi...	3,400.00	100,693.41	Interbank Trans...	Income	Online Banking	Wholesale c...
Dec.25	Hunan Tianyua...	-2,200.00	98,493.41	Transfer Payment	Expense	Mobile Banking	Year-end inv...
Dec.28	Changsha Huito...	-2,000.00	96,493.41	Transfer Payment	Expense	Online Banking	Employee be...
Dec.31	ICBC Co., Ltd.	4.1	96,497.51	Interest Credit	Income	Bank System	Demand dep...

Question

Compare operating expenses (ingredients, packaging, rent, utilities) with operating income in the same period. Which description cannot reflects the expense rationality?

a. Generally matches monthly income, no obvious imbalance

b. Significantly higher than monthly income, facing substantial cash pressure

c. Significantly lower than monthly income, may affect supply stability

d. Cannot be determined

Answer

Correct Answer

B, C, D

Question Type: Multiple Choice
Category: Income & Expense Stability / Expense Rationality
Skills: Transaction categorization, Financial ratio analysis, Pattern recognition
Cognitive Level: Memory & Recognition (Low)
Tests: Ability to identify expense categories and assess overall financial

Figure 12: Sample Multiple Choice Question about Memory & Recognition of Micro-enterprise Clients Log Information

M Usage of Large Language Models

The authors used large language models only for polishing prose of text where the complete draft was fully written by the authors initially and polished later with the help of LLM-based assistants including ChatGPT, Gemini, and Perplexity. The authors' used code assistants including Cursor and Copilot to implement the authors' original design and ideas. The scientific contributions, technical methods, ideas and core results are entirely the original work of the authors.