Table Representation Learning Workshop @NeurIPS 2024

Expertise-Centric Prompting Framework for Financial Tabular Data Generation using Pre-trained Large Language Models

Subin Kim[1], Jungmin Son*[1], Minyoung Jung*[2], Youngjun Kwak[1]*

[1] KakaoBank
[2] Korea Electronics Technology Institute

Subin Kim,

kakaobank

1. Motivation
2. Proposed methods

- Expertise−Centric Prompting Framework
- Evaluation Metrics

3. Experiment Results
4. Application : Today’s mini diary
5. Future work

1. Motivation
2. Proposed methods
- Expertise-Centric Prompting Framework
- Evaluation Metrics
3. Experiment Results
4. Application : Today's mini diary
5. Future work

1. Motivation

- Shortage of authentic and accessible financial tabular datasets
  - Privacy and security concerns limit access to real-world data
  - Anonymized public datasets lose realism and fail to reflect real-world scenarios
  - Impedes AI development using financial tabular data

- Challenge of generating pseudo-financial datasets
  - Complex attribute relationships and diverse data ranges
  - Requires expert knowledge for accurate data validation

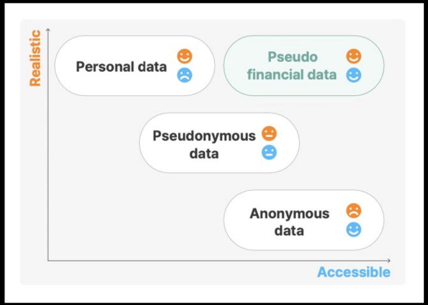


Figure 1. (a)
Comparison of financial data

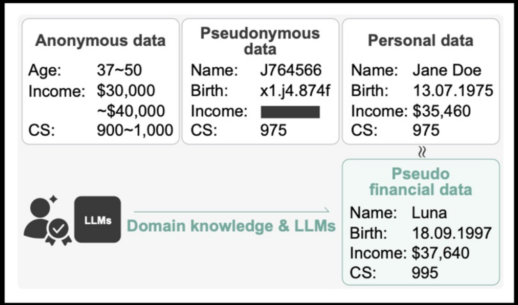


Figure 1. (b)
Example of pseudo-financial data

1. Motivation

- Goals → *How?*
  a. Generate realistic and accessible pseudo-financial datasets
     → *Collaboration between financial experts and LLM*

  b. Minimize experts' efforts in generating and validating financial data
     → *Prompt-based dataset generation framework with quantitative evaluation metrics*

1. Motivation
2. Proposed methods

- Expertise-Centric Prompting Framework
- Evaluation Metrics

3. Experiment Results
4. Application : Today's mini diary
5. Future work

2. Proposed methods

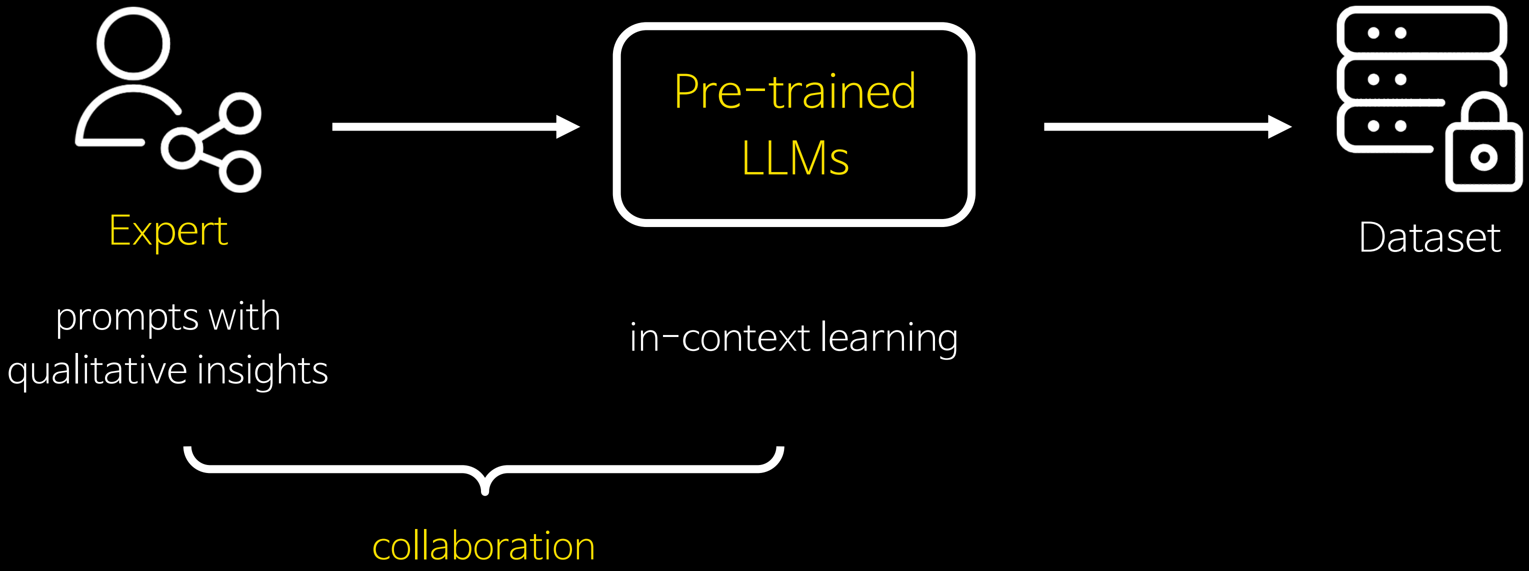
- Expertise-centric prompting (ECP) framework
  - Addressing limited accessibility of realistic financial data

- Evaluation metrics
  - Validating the quality of generated financial tabular datasets in terms of:
    - Dataset diversity
    - Constraint satisfaction

1. Motivation
2. Proposed methods
- Expertise-Centric Prompting Framework
- Evaluation Metrics
3. Experiment Results
4. Application : Today's mini diary
5. Future work

2. Proposed methods

- Expertise-Centric Prompting (ECP) Framework
  - collaboration between financial experts and LLM
    - i) schema calibration
    - ii) attribute constraints



2. Proposed methods

- Expertise-Centric Prompting (ECP) Framework
  - collaboration between financial experts and LLM
    - i) schema calibration
    - ii) attribute constraints

prompts with
qualitative insights

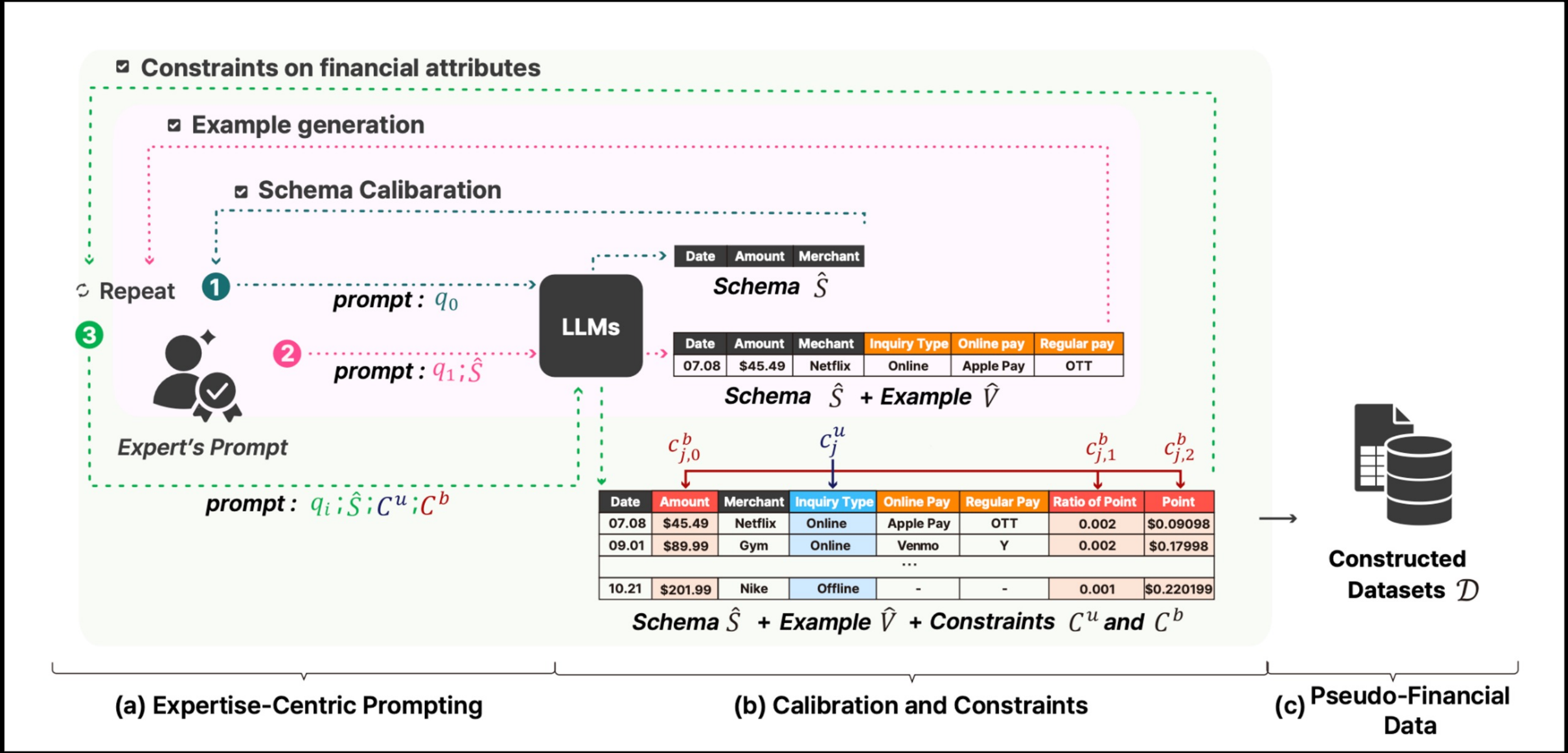
in-context learning

✓ schema calibration

✓ attribute constraints

2. Proposed methods

- Expertise-Centric Prompting (ECP) Framework



2. Proposed methods

- Expertise-Centric Prompting (ECP) Framework
  - prompt template

Expert

Prompting

```
## prompt template
```

$$t(q_i) = \begin{cases} q_i, & if\ i = 0 \\ q_i; C, & otherwise \end{cases}$$

Pre-trained LLMs

$, where$

$t(\bullet)$: *prompt template (LLM input)*

q_i : i^{th} *prompt*

C: *a set of K conditions* $\{c_{j=0,\ldots,K-1}\}$

2. Proposed methods

- Expertise-Centric Prompting (ECP) Framework
  - schema estimation

Expert

Prompting

```
## prompt template
You are a table data generator. Please create
a table schema for card transactions. Make
sure to include the Customer ID column.
```

1 $t(q_0) = q_0$

Pre-trained LLMs

Estimated attributes $\hat{S}$

```
## output
| Date       | Amount   | Merchant      |
|------------|----------|---------------|
| YYYY-MM-DD | Currency | Text          |
```

$$\hat{S} = LLM\big(t(q_0)\big)$$

$$, where\ \hat{S} = \{\hat{a}_{i=0,\ldots,|\hat{S}|-1}\}$$

2. Proposed methods

- Expertise-Centric Prompting (ECP) Framework
  - schema calibration

Expert

Prompting

$S = \{a_{i=0,\dots,|S|-1}\}$

```
## prompt template
```

① $t(q_2) = q_2; h(\hat{S})$

```
You are a table data generator. Please create a table schema
for card transactions. Make sure to include the Customer ID
column.

| Customer ID | Date       | Amount   | Merchant      |
|-------------|------------|----------|---------------|
| ID NO.      | YYYY-MM-DD | Currency | Text          |

Please add the columns Inquiry Type, Online pay, and Regular
pay to the table above.
```

Pre-trained LLMs

Calibrated attributes $h(\hat{S})$

```
## output

| Customer ID | Date       | Amount   | Merchant      | Inquiry Type | Online pay | Regular pay |
|-------------|------------|----------|---------------|--------------|------------|-------------|
| ID NO.      | YYYY-MM-DD | Currency | Text          | Text         | Yes/No     | Yes/No      |
```

$$h(\hat{S}) = \begin{cases} \hat{S}, & if\ S - \hat{S}, = \emptyset \\ S, & otherwise \end{cases}$$

2. Proposed methods

- Expertise-Centric Prompting (ECP) Framework
  - example generation

Expert

Prompting

```
## prompt template
```

2 $t(q_0) = q_0; \hat{S}$

```
Please create a table with 3 rows where every column has
diverse values. Please start customerID from A001.

The table schema is as follows:

| Customer ID | Date       | Amount   | Merchant      |
|-------------|------------|----------|---------------|
| ID NO.      | YYYY-MM-DD | Currency | Text          |

| Inquiry Type | Online Pay | Regular Pay |
|--------------|------------|-------------|
| Text         | Yes/No     | Yes/No      |
```

Pre-trained LLMs

```
## output
```

Customer ID	Date	Amount	Merchant	Inquiry Type	Online Pay	Regular Pay
A001	2024-12-01	150.00	CoffeeShop	Payment Issue	Yes	No
A002	2024-11-28	300.50	SuperMart	Refund	No	Yes
A003	2024-11-30	79.99	BookStore	Inquiry	Yes	Yes

$$\hat{V} = LLM(t(q_1) = q_1; \hat{S})$$

2. Proposed methods

- Expertise-Centric Prompting (ECP) Framework
  - attribute constraints

Expert

Prompting

- *unary constraints* $C^u = \{c^u_{j=0,\dots,K^u-1}\}$
- *binary constraints* $C^b = \{c^b_{j=0,\dots,K^b-1}\}$

```
## prompt template
```

③ $t(q_i) = q_i; C = q_i; \hat{S}; C^u; C^b$

```
Please create a table with 3 rows where every column has
diverse values. Please start customerID from A001.

The table schema is as follows:
| Customer ID | Date       | Amount   | Merchant       |
|-------------|------------|----------|----------------|
| ID NO.      | YYYY-MM-DD | Currency | Text           |

| Inquiry Type | Online Pay | Regular Pay |
|--------------|------------|-------------|
| Text         | Yes/No     | Yes/No      |

Constraints for each column:

Inquiry Type: Indicates whether the payment is online or
offline.

Regular Payment Status: Whether the payment was a regular
payment or not (Y/N). When the regular payment status is 'Y',
the payment is made on the same date every month. For example,
insurance fees, OTT services like Netflix, or apartment
maintenance fees are regular payments.

Ratio Point: The payment accrual rate for the payment amount.
The rate is a 3-digit decimal between 0 and 1, and the accrual
rate will be less than 5%.

Point: The points accrued from the payment. This is calculated
by multiplying the values in the Amount and Ratio Point
columns, with the unit in dollars.
```

Pre-trained LLMs

```
## output

| Customer ID | Date       | Amount | Merchant       |
|-------------|------------|--------|----------------|
| A001        | 2024-12-01 | 150.00 | CoffeeShop     |
| A002        | 2024-11-28 | 300.50 | SuperMart      |
| A003        | 2024-11-30 | 79.99  | BookStore      |

| Inquiry Type | Online Pay | Regular Pay |
|--------------|------------|-------------|
| Online       | Yes        | No          |
| Offline      | No         | Yes         |
| Online       | Yes        | No          |

| Ratio Point | Point |
|-------------|-------|
| 0.035       | 5.25  |
| 0.045       | 13.52 |
| 0.02        | 1.60  |
```

c^u_j $c^b_{j,1}$ $c^b_{j,2}$

1. Motivation
2. Proposed methods
   - Expertise-Centric Prompting Framework
   - Evaluation Metrics
3. Experiment Results
4. Application : Today's mini diary
5. Future work

2. Proposed methods

- Evaluation Metrics

Diversity

- Variety of generated datasets
  - avoid redundancy in data caused by repetition
  - ensure a variety of customer behaviors reflecting real-world scenarios

Constraint Satisfaction

- Alignment between generated datasets and constraints
  - ensure that data properties of specific attributes are met (e.g., value ranges, categories, etc.)
  - ensure logical and numerical consistency among the attributes

2. Proposed methods

- Evaluation Metrics

Diversity

- Variety of generated datasets
  - Inter-Instance Diversity
    : diversity among instances
    - uniformity metric
    - principal component (PC) analysis
  - Intra-Instance Diversity
    : diversity within individual attributes
    - entropy

Constraint Satisfaction

- Alignment between generated datasets and constraints
  - Unary constraint satisfaction
  - Binary constraint satisfaction

2. Proposed methods

- Evaluation Metrics

Diversity

- Variety of generated datasets
  - Inter-Instance Diversity
    : diversity among instances
    - uniformity metric
    - principal component (PC) analysis
  - Intra-Instance Diversity
    : diversity within individual attributes
    - entropy

$$Unif \simeq f_{unif}\left(\underset{h}{\text{concat}}[e_h]\right)$$

e_h : $embedding\ of\ h^{th}\ row$

```
def f_unif(e, t=2):
  pdst = torch.pdist(e, p=2).pow(2)
  return pdst.mul(-t).exp().mean().log()
```

Figure 3: Implementation of f_{unif} using PyTorch.

Constraint Satisfaction

- Alignment between generated datasets and constraints
  - Unary constraint satisfaction

2. Proposed methods

- Evaluation Metrics

Diversity

- Variety of generated datasets
  - Inter-Instance Diversity
    : diversity among instances
    - uniformity metric
    - principal component (PC) analysis
  - Intra-Instance Diversity
    : diversity within individual attributes
    - entropy

$$\mathcal{H} = -\sum_{i \in I} p_i * \log p_i$$

I : *a set of unique examples on a certain attribute*

p_i : *probability of each distinct value i*

Constraint Satisfaction

- Alignment between generated datasets and constraints
  - Unary constraint satisfaction
  - Binary constraint satisfaction

2. Proposed methods

- Evaluation Metrics

Diversity

- Variety of generated datasets
  - Inter-Instance Diversity
    : diversity among instances
    - uniformity metric
    - principal component (PC) analysis
  - Intra-Instance Diversity
    : diversity within individual attributes
    - entropy

Constraint Satisfaction

- Alignment between generated datasets and constraints
  - Unary constraint satisfaction

$$\rho = \frac{1}{K^u} \sum_{c_j^u \in C^u} 1\left(\Psi(c_j^u)\right)$$

  - Binary constraint satisfaction

$$\tau = \frac{1}{K^b} \sum_{c_j^b \in C^b} 1\left(\Phi(c_j^b)\right)$$

3. Experiment Results

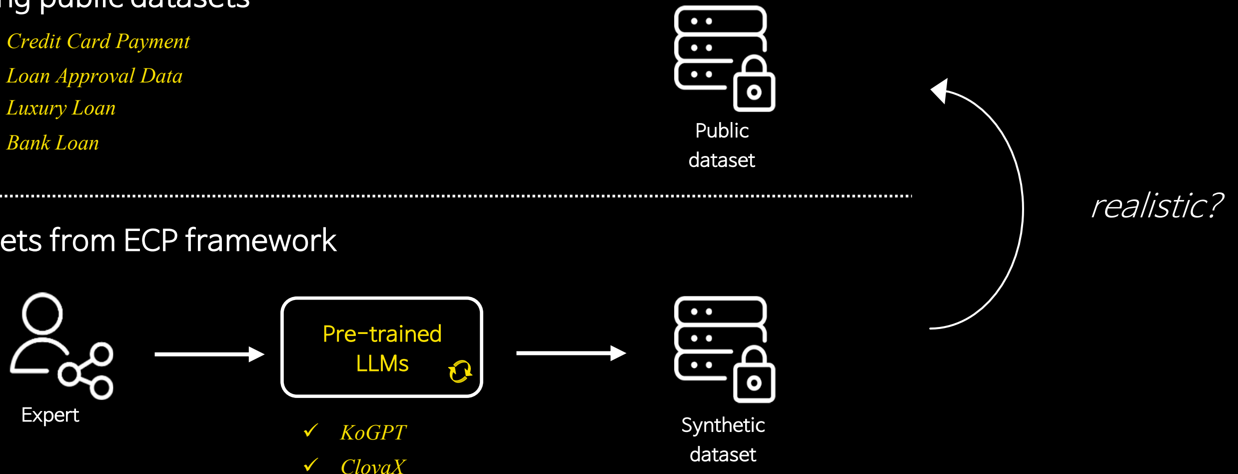
● Experimental Setting

- We evaluated the generated datasets using various *backbone LLMs and existing public datasets* while keeping both the framework and the evaluation metrics fixed

Existing public datasets

- ✓ *Credit Card Payment*
- ✓ *Loan Approval Data*
- ✓ *Luxury Loan*
- ✓ *Bank Loan*

Datasets from ECP framework



- ✓ *KoGPT*
- ✓ *ClovaX*
- ✓ *LLAMA 2-Chat*
- ✓ *GPT-3.0*
- ✓ *ChatGPT-3.5/4.0*

3. Experiment Results

- Evaluation results - diversity

[Table 2]

Language	Card Transactions Dataset	ECP	Diversity of Instance		Diversity of Attribute↑					
			Embedding-Level		Categorical-Level		Binary-Level		Numerical-Level	
			Uniformity↓	PC↑	Avg($\mathcal{H}$)	Max($\mathcal{H}$)	Avg($\mathcal{H}$)	Max($\mathcal{H}$)	Avg($\mathcal{H}$)	Max($\mathcal{H}$)
English	*Credit Card Payment* [12]		0.351	0.056	3.714	3.714	N/A	N/A	8.287	8.287
	KoGPT [52]	✔	0.365	0.040	2.652	5.174	1.072	1.284	4.823	4.823
	ClovaX [44]	✔	0.369	0.003	1.953	2.574	1.093	1.119	2.574	2.574
	LLAMA 2-Chat [45]	✔	0.339	0.036	3.737	4.934	0.971	0.971	5.057	5.057
	GPT-3.0 [19]	✔	0.340	0.060	3.190	5.461	0.963	1.000	3.874	3.874
	ChatGPT-3.5 [20]	✔	0.320	0.080	4.627	6.115	0.998	1.000	4.689	4.689
	ChatGPT-4.0 [53]	✔	0.327	0.081	3.987	6.483	0.954	0.990	5.751	5.751
Korean	KoGPT [52]	✔	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	ClovaX [44]	✔	0.396	0.006	1.807	2.722	0.961	1.000	3.322	3.322
	LLAMA 2-Chat [45]	✔	0.379	0.028	3.718	4.663	1.526	1.640	3.715	3.715
	GPT-3.0 [19]	✔	0.396	0.053	2.743	4.941	0.993	1.133	4.060	4.060
	ChatGPT-3.5 [20]	✔	0.388	0.107	3.412	6.723	1.024	1.101	6.096	6.096
	ChatGPT-4.0 [53]	✔	0.387	0.082	3.879	7.085	0.321	0.584	6.176	6.176

  - Dataset generated through our approach is comparable in diversity to existing datasets
  - ChatGPT-based series outperformed other LLMs
  - Robustness of our framework in generating multilingual datasets

3. Experiment Results

- Evaluation results – <u>constraints satisfaction</u>

[Table 5]

Datasets	Lang.	Pre-trained LLMs	ρ	τ
Card Transactions	EN	KoGPT [52]	0.92	0.86
		ClovaX [44]	0.97	0.97
		LLAMA 2-Chat [45]	1.00	1.00
		GPT-3.0 [19]	1.00	1.00
		ChatGPT-3.5 [20]	1.00	0.99
		ChatGPT-4.0 [53]	1.00	1.00
	KR	KoGPT [52]	N/A	N/A
		ClovaX [44]	0.95	0.95
		LLAMA 2-Chat [45]	0.82	0.67
		GPT-3.0 [44]	0.97	0.96
		ChatGPT-3.5 [20]	0.97	0.94
		ChatGPT-4.0 [53]	1.00	1.00

  - Close to 1.0 indicates strong adherence to these constraints
  - Datasets generated by GPT series demonstrate strong alignment with unary constraints
  - Binary constraints lower than unary constraints
  - A need for further development for understanding the complex operations and calculations of the current LLMs

3. Experiment Results

- Ablation study

[Table 6]

		Diversity of instance		Diversity of attribute			Constraint Satisfaction	
		Embedding-Level		Categorical-Level	Binary-Level	Numerical-Level		
Dataset	ECP	Uniformity↓	PC↑	Avg($\mathcal{H}$)	Avg($\mathcal{H}$)	Avg($\mathcal{H}$)	ρ	τ
Loan Statements Dataset		0.42	0.06	5.33	0.90	3.82	0.6	0.5
	✔	0.43	0.06	6.33	0.93	3.39	**1.0**	**0.7**
Deposits and Savings Dataset		0.39	0.06	2.99	1.48	4.48	0.1	0.4
	✔	0.41	0.08	3.47	0.95	3.46	**1.0**	**0.9**

  - Our approach increased instance and attribute diversities but slightly decreased binary and numerical diversities
  - These adjustments were necessary to improve constraint satisfaction through ECP
  - This demonstrates the effectiveness of our ECP framework in balancing diversity with realistic dataset adjustments

3. Experiment Results

- Ablation study

[Table 7]

Method	$h(\hat{S})$	C^u	C^b	D_{ls} ρ	D_{ls} τ	D_{ds} ρ	D_{ds} τ	Avg(Std)
ChatGPT-3.5 [20]	✗	✔	✔	1.00	0.73	0.93	0.92	0.90‡($\pm$0.12)
	✔	✗	✔	1.00	0.76	1.00	0.88	0.91($\pm$0.15)
	✔	✔	✗	1.00	0.74	1.00	0.73	0.87($\pm$0.11)
	✔	✗	✗	1.00	0.75	0.66	0.39	0.70†($\pm$0.25)
	✔	✔	✔	1.00	0.77	1.00	0.90	**0.92**($\pm$0.11)

- Schema calibration $h(\hat{S})$, unary constraints C^u, binary constraints C^b
- The highest average values of ρ and τ are achieved when all components are utilized across all combinations
- The average constraint satisfaction decreased by 23.91% (from 0.92 to $0.70^{\dagger}$) when both constraints were removed, whereas the average decreased by only 2.17% (from 0.92 to $0.90^{\ddagger}$) when schema calibration was removed

1. Motivation
2. Proposed methods
- Expertise-Centric Prompting Framework
- Evaluation Metrics
3. Experiment Results
4. Application : Today's mini diary
5. Future work

4. Application: *Today's mini diary*

- Develop an LLM-powered financial service

Expertise-Centric Prompting Framework

Synthetic Dataset Design

- Diverse financial scenarios
- Exclusion from privacy regulations
- Realistic attributes for customer behavior modeling

Train, fine-tune with synthetic tabular datasets

- Ensuring data privacy and compliance with synthetic data

Performance Evaluation

- Latency optimization
- Quality assurance test

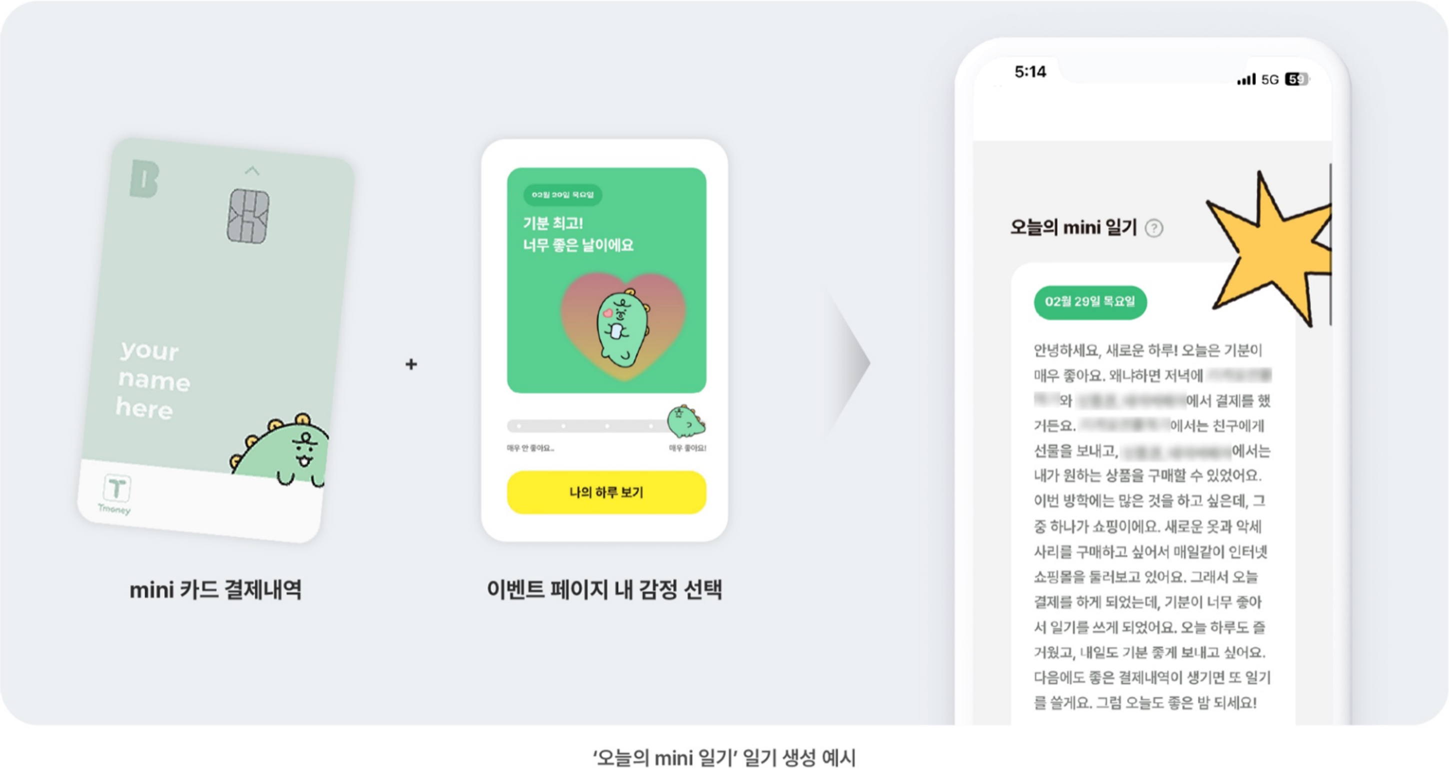


Figure 6: Example of the *Today's Mini Diary* service in action.

1. Motivation
2. Proposed methods
- Expertise-Centric Prompting Framework
- Evaluation Metrics
3. Experiment Results
4. Application : Today’s mini diary
5. Future work

5. Future work

LLM Safety and Privacy Issue

- Integrating guardrail technologies to prevent sensitive data output, enhancing robustness and security
- Prevent risks of harmful content or training data leakage in LLMs

Customize evaluation metrics in framework

- Tailored aspects for target dataset
- Apply customized formulas for constraint satisfaction validation
  - Constraint satisfaction evaluation metrics potentially cover more real-world arithmetic formulas, conceptual relationships, and specific statistical distributions

EOD

Paper

kakaobank