Tab-MIA: A Benchmark Dataset for MIAs on Tabular Data in LLMs

Eyal German, Sagiv Antebi, Daniel Samira, Asaf Shabtai and Yuval Elovici

The Fourteenth International Conference on Learning Representations

CBG Cyber@ Ben-Gurion University of the Negev

Motivation

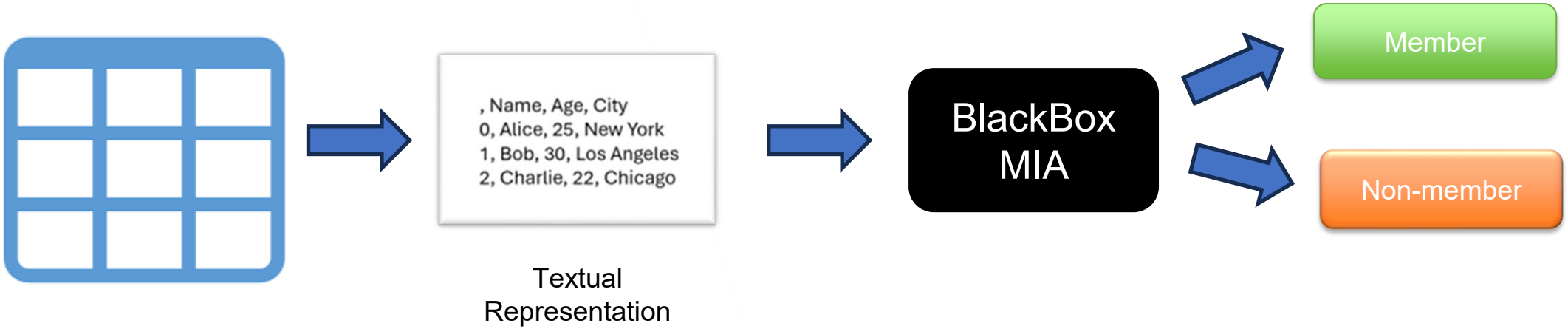
Large language models are now more frequently trained using tabular data from fields like finance, healthcare, and census information.

In contrast to text, tables hold explicit personally identifiable information and exhibit structured patterns.

Current benchmarks include BookMIA, WikiMIA, and MIMIR, which are focused on text, while MIDST targets diffusion models for generating synthetic tables.

Tab-MIA: A Benchmark Dataset for MIAs on Tabular Data in LLMs

- **Key Question**
- 👉 How vulnerable are LLMs to MIAs on tabular data?
- Evaluation of existing MIAs on tabular datasets.
- Investigate different Membership Inference Attacks (MIAs) in a black-box setting.



Tab-MIA Datasets

- Includes tables with both short and long contexts.
- Short-context datasets: WTQ, WikiSQL, TabFact.
- Long-context (CSV) datasets: Adult (Census), California Housing.
- All datasets have been processed to develop the Tab-MIA benchmark.

Table 1: Summary of datasets used in Tab-MIA.

Name	Short/Long	# Records	# After Filter	# Features	Based On
WTQ	Short	2,108	1,290	≥ 5	Wikipedia
WikiSQL	Short	24,241	17,900	≥ 5	Wikipedia
TabFact	Short	16,573	13,100	≥ 5	Wikipedia
Adult (Census Income)	Long	48,842	2,440	15	US Census
California Housing	Long	20,640	1,030	10	US Housing Survey

The Different Encodings

(a) JSON

```
[
  {"Name": "Alice", "Age": 30},
  {"Name": "Bob",   "Age": 25},
  {"Name": "Carol", "Age": 28}
]
```

(b) HTML

```
<table>
  <tr><th>Name</th><th>Age</th></tr>
  <tr><td>Alice</td><td>30</td></tr>
  <tr><td>Bob</td><td>25</td></tr>
  <tr><td>Carol</td><td>28</td></tr>
</table>
```

(c) Markdown

```
| Name  | Age |
|-------|-----|
| Alice | 30  |
| Bob   | 25  |
| Carol | 28  |
```

(d) Key-Value Pair

```
Name: Alice | Age: 30
Name: Bob   | Age: 25
Name: Carol | Age: 28
```

(e) Key-is-Value

```
Name is Alice. Age is 30.
Name is Bob.   Age is 25.
Name is Carol. Age is 28.
```

(f) Line-Separated

```
Name,Age
Alice,30
Bob,25
Carol,28
```

Evaluation and Results

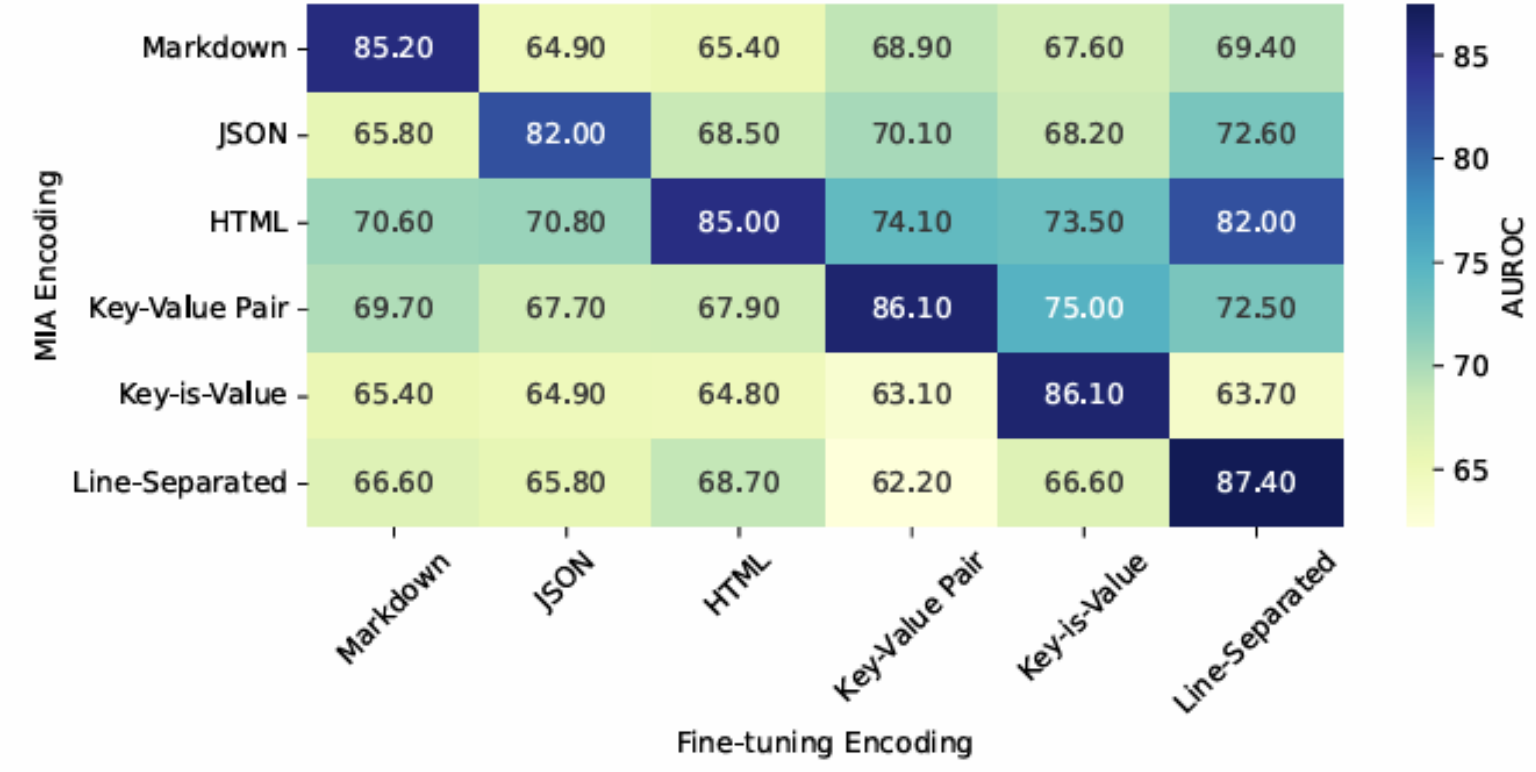
Results: Effect of Different Encoding

- AUROC results for multiple LLMs evaluated with various MIAs on the WikiSQL dataset, comparing the different encoding formats.

	Llama-3.2 3B			Mistral 7B			Gemma-3 4B		
Encoding Method	PPL	Min-K 20.0%	Min-K++ 20.0%	PPL	Min-K 20.0%	Min-K++ 20.0%	PPL	Min-K 20.0%	Min-K++ 20.0%
Markdown	60.60	60.90	72.00	65.60	73.10	**80.00**	59.10	64.10	67.80
JSON	59.60	59.60	53.00	**61.40**	**61.40**	54.50	58.40	58.40	55.00
HTML	59.70	59.70	55.80	**61.70**	**61.70**	50.60	59.10	61.20	55.40
Key-Value Pair	62.80	62.80	78.70	72.40	74.70	**92.60**	59.30	60.80	67.00
Key-is-Value	60.20	60.20	55.10	63.70	65.00	**74.90**	59.20	60.60	66.70
Line-Separated	61.60	64.90	77.20	69.70	84.90	**86.80**	62.30	72.10	73.80

Results: Cross-Format Generalization

- AUROC results showing how attacks transfer across different formats.
- Gemma-3 4B evaluated on WTQ using the Min-K++ 20% MIA.



Key Findings

LLMS **MEMORIZE TABULAR DATA**

MIAS ACHIEVE **HIGH ACCURACY**

ENCODING STRONGLY AFFECTS LEAKAGE

NEED **PRIVACY-AWARE TRAINING METHODS**

Thank You!

Eyal German

The Fourteenth International Conference on Learning Representations