COLING 2025 • Abu Dhabi
The 31st International Conference on Computational Linguistics

CHIFRAUD: A Long-term Web Text Dataset for Chinese Fraud Detection

Min Tang, Lixin Zou*, Shujie Cui, Shiuan-Ni Liang, Zhe Jin, Weiqing Wang

Contact: min.tang@monash.edu

1. Motivation

Fraudulent Text

- Concise texts deliberately crafted to **disseminate deceptive or illegal trading**, thereby contravening Chinese laws and regulations.

(a) A demo fraud-text in June 2022.

(b) The fraud-text in June 2023.

Orange: homograph variants, e.g., O to 0.
Red: homophone variants, e.g., 芯 to 信.
Teal: confounding characters, e.g., ☆ ※ 叶 stt.
Gray: the normal text.

Fraud-text Detection

- Determine whether a given text is normal content or fraudulent and assigns it to the appropriate fraud category.

☐ **Camouflage techniques**

- homophone variants
- confounding characters
- interspersing deceptive content with normal text

- Perpetually shifting

2. CHIFRAUD Dataset

- First anonymous public Chinese fraud-text detection dataset.
  - extensive expert annotations (**59,106** fraudulent texts)
  - Long-term data collection duration
- Dataset Comparison

Dataset	Source	Language	Availability	Ethic	Duration	# Total	# Annotated	# Target	# Category	Pub-Year
SpamAssassin	Email	Multilingual	Public	Yes	4 years	6,047	6,047	1,874	5	2002
Enron Email	Email	Multilingual	Public	/	/	33,716	33,716	17,171	2	2002
SpamHunter	SMS	Multilingual	Private	Yes	4 years	21,918	947	/	8	2022
Spam SMS	SMS	Chinese	Public	/	/	11,358	11,358	11,358	1	2022
360 Spearphishing	SMS	Chinese	Private	Yes	3 months	31,956,437	10,399	90,801	10	2021
CHIFRAUD	Web	Chinese	Public	Yes	1 year	411,934	411,934	59,106	11	2024

- **CHIFRAUD Construction**

- **CHIFRAUD Composition**

Subset	Total	Normal	Gambling	Whoring	Credentials	Bank	Drugs	Cash-out	Certification	SIM	Loan	New
CHIFRAUD$_{train}$	193,567	167,914	3,629	11,637	542	951	1,616	1,499	4432	486	861	/
CHIFRAUD$_{t2022}$	96,766	83,951	1,732	6,003	303	485	748	746	2,139	221	438	/
CHIFRAUD$_{t2023}$	121,101	100,463	5,332	8,547	536	401	2,764	572	502	698	223	1,063
Total	411,434	352,328	10,674	26,187	1,381	1,837	5,128	2,817	7,073	1,405	1,522	1,063

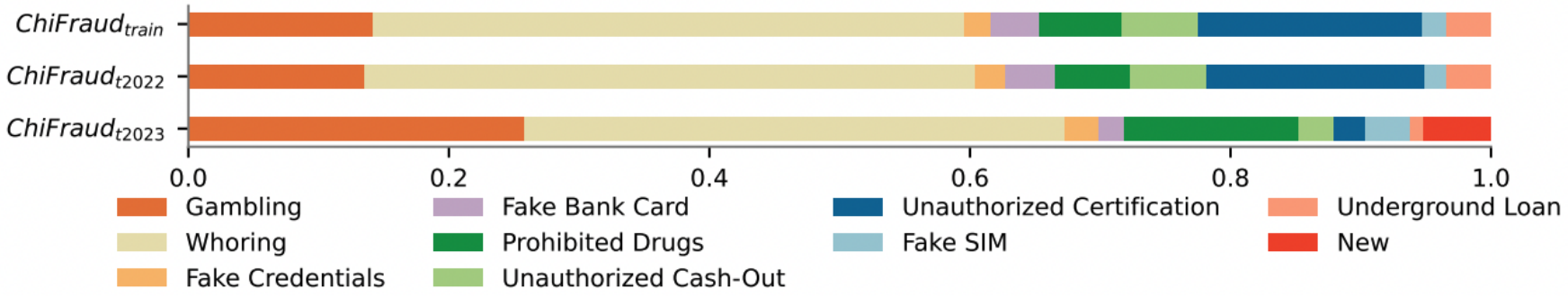
2. CHIFRAUD Dataset

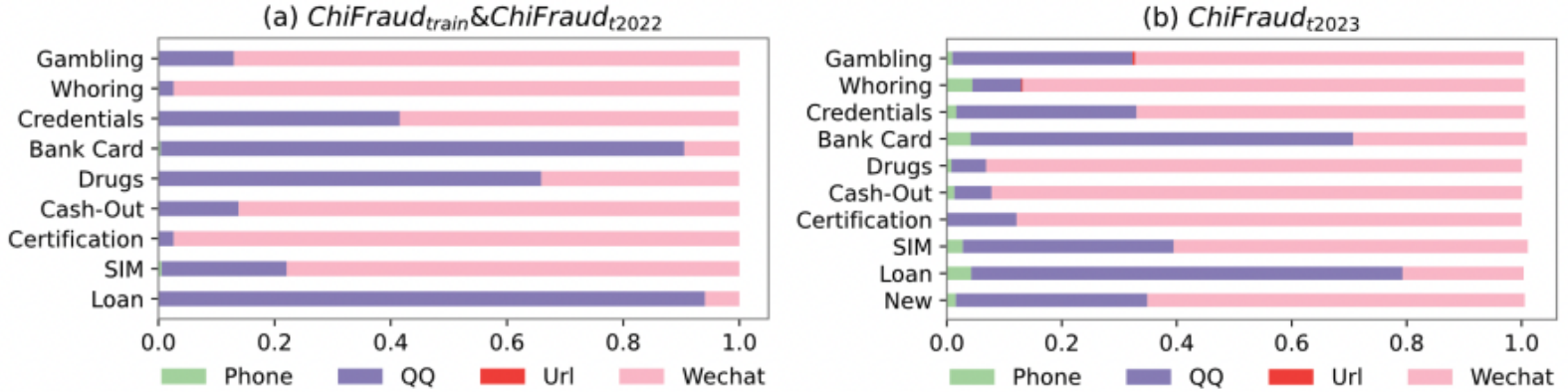
Unbalanced Fraud

- Fraudulent texts constitute a small portion compared to normal texts.
- Significantly skewed distribution across different fraud categories.

Distribution Shifts

- Distribution of fraud fluctuates significantly over time.
- Distribution of contacts in fraudulent texts varies over time.
- Newly emergent fraudulent texts.





3. Bechmark Detectors

- **Large language model-based detectors**
  - Llama2-D
  - Qwen-D
  - ChatGPT-D
- **Pre-trained language model-based detectors**
  - Bert-D
  - ChineseBert-D
- **Deep learning-based detectors**
  - Transformer-D

Noticeable performance degradation over time!

Metric	Method	Normal	Gambling	Whoring	Credentials	Bank	Drugs	Cash-out	Certification	SIM	Loan	All
						ChiFraud$_{t2022}$						
Recall	Transformer	0.9969	0.6859	0.9179	0.6436	0.6165	0.8396	0.6863	0.9528	0.7285	0.7854	0.7853
	Bert	0.9961	0.9527	0.9678	**0.8482**	**0.9196**	**0.9746**	0.9142	0.9598	0.7149	**0.9498**	0.9198
	ChineseBert	**0.9977**	**0.9590**	**0.9825**	0.8053	0.8887	0.9505	0.9491	**0.9874**	0.7421	0.9452	0.9208
	Qwen0.5B-D	0.9967	0.9365	0.9775	**0.8482**	0.9134	0.9263	**0.9780**	0.9780	**0.7873**	0.9452	**0.9235**
	Llama2-D	0.9970	0.8482	0.9202	0.8119	0.7361	0.9144	0.8646	0.9579	0.5792	0.8425	0.8472
Precision	Transformer	0.9836	**0.9827**	0.9467	0.9374	**0.9374**	0.8556	**0.9626**	0.9115	0.4850	**0.9972**	0.9000
	Bert	0.9969	0.9531	**0.9764**	0.8371	0.8544	0.8555	**0.9459**	**0.9884**	0.7784	0.7955	0.8982
	ChineseBert	**0.9979**	0.9684	0.9759	**0.9839**	0.8979	**0.9455**	0.9427	0.9561	0.8962	0.9221	**0.9487**
	Qwen0.5B-D	0.9971	0.9387	0.9679	0.9346	0.8754	0.9207	0.9050	0.9753	0.8366	0.9538	0.9363
	Llama2-D	0.9904	0.8426	0.9691	0.8978	0.8381	0.8735	0.9035	0.9753	**0.9013**	0.9867	0.9178
F1-score	Transformer	0.9902	0.8079	0.9321	0.7632	0.7438	0.8475	0.8013	0.9317	0.5823	0.8787	0.8279
	Bert	0.9965	0.9529	0.9721	0.8426	0.8858	0.9112	0.9298	0.9739	0.7453	0.8658	0.9076
	ChineseBert	**0.9978**	**0.9637**	**0.9792**	0.8857	0.8933	**0.9480**	**0.9459**	0.9715	**0.8119**	0.9335	**0.9331**
	Qwen0.5B-D	0.9969	0.9376	0.9727	**0.8893**	**0.8940**	0.9235	0.9401	**0.9767**	0.8112	**0.9495**	0.9292
	Llama2-D	0.9937	0.8454	0.9440	0.8527	0.7838	0.8935	0.8836	0.9665	0.7052	0.9089	0.8777
						ChiFraud$_{t2023}$						
Recall	Transformer	0.9969	0.0294	0.3305	0.0896	0.2145	0.2627	0.4003	0.6733	0.5072	0.3946	0.3899
	Bert	0.9963	0.0940	0.5350	**0.3190**	0.5362	0.3122	0.7692	0.7311	0.5158	**0.7489**	**0.5558**
	ChineseBert	**0.9977**	0.1080	0.5250	0.1884	0.4564	0.2688	**0.8759**	**0.8685**	0.5072	0.6726	0.5469
	Qwen0.5B-D	0.9969	**0.1262**	0.4810	0.2351	**0.5461**	0.3412	0.7133	0.8008	**0.6132**	0.6143	0.5468
	Llama2-D	0.9973	0.0986	**0.5495**	0.2836	0.4564	0.3788	0.6836	0.8167	0.4642	0.5381	0.5267
Precision	Transformer	0.8694	0.8805	0.8434	0.8661	0.7048	0.9080	**0.8982**	0.5577	0.7865	**0.9779**	**0.8293**
	Bert	0.8915	0.8664	0.9550	0.7037	0.5457	0.8874	0.7074	**0.8716**	0.8631	0.4134	0.7705
	ChineseBert	0.8945	**0.9063**	0.9360	**0.9366**	0.6080	0.9445	0.5382	0.4861	**0.9670**	0.6819	0.7899
	Qwen0.5B-D	0.8927	0.8449	0.8700	0.7073	0.6366	**0.9473**	0.6667	**0.6722**	0.9030	0.7327	0.7873
	Llama2-D	**0.8954**	0.8016	**0.9554**	0.8084	0.5429	0.9075	0.4962	0.6084	0.9100	0.8957	0.7821
F1-score	Transformer	0.9288	0.0569	0.4749	0.1624	0.3289	0.4075	0.5538	0.6101	0.6167	0.5623	0.4702
	Bert	0.9410	0.1696	0.6858	**0.4390**	0.5409	0.4619	**0.7370**	**0.7952**	0.6457	0.5327	0.5949
	ChineseBert	0.9433	0.1930	0.6727	0.3137	0.5214	0.4185	0.6667	0.6233	0.6654	**0.6772**	0.5695
	Qwen0.5B-D	0.9419	**0.2196**	0.6195	0.3529	**0.5879**	0.5017	0.6892	0.7309	**0.7304**	0.6683	**0.6042**
	Llama2-D	**0.9436**	0.1756	**0.6977**	0.4199	0.4959	**0.5345**	0.5750	0.6973	0.6148	0.6723	0.5827

4. Research Directions

❑ Robust Detection

- Detectors need further improvement to consistently achieve high recall and precision rates.

	Method	Recall	Method	Recall
Tuning	Transformer	0.0555	Bert	0.0630
	ChineseBert	0.1340	Qwen0.5B-D	<u>0.1467</u>
	Llama2-D	**0.2144**		
ICL	Qwen0.5B-D	0.2895	Qwen1.8B-D	0.3490
	Qwen7B-D	<u>0.7845</u>	Qwen14B-D	**0.9586**

❑ Attacks on Detectors

- LLM-based detectors demonstrate limitations in countering carefully designed deceptions.

Type	ASR(%)	Type	ASR(%)
Gambling	<u>70.68</u>	Whoring	23.79
Fake Credentials	32.90	Fake Bank Card	40.46
Prohibited Drugs	39.60	Unauthorized Cash-Out	63.01
Unauthorized Certification	14.76	Underground Loan	69.48
Fake SIM	**72.98**	Overall	35.37

❑ Efficient Detection

- Billion-parameter detectors pose challenges for practical implementation in industry applications.

Detector	Transformer	Bert	ChineseBert	Qwen0.5B-D	Qwen1.8B-D	Qwen7B-D	Qwen14B-D	Llama2-D
Seconds	0.0009	0.0110	0.0066	0.4737	3.7501	12.2031	15.9008	7.4200

Thank you

Q & A