TMID: A Comprehensive Real-world Dataset for Trademark Infringement Detection in E-Commerce

Tongxin Hu[1,a], **Zhuang Li**[2,a],
Xin Jin[1], **Lizhen Qu**[2,b], **Xin Zhang**[1]
Ant Group[1], Monash University[2]
[1]{tongxin.htx, king.jx, evan.zx}@antgroup.com
[2]{zhuang.li1, lizhen.qu}@monash.edu

Abstract

Annually, e-commerce platforms incur substantial financial losses due to trademark infringements, making it crucial to identify and mitigate potential legal risks tied to merchant information registered to the platforms. However, the absence of high-quality datasets hampers research in this area. To address this gap, our study introduces TMID, a novel dataset to detect trademark infringement in merchant registrations. This is a real-world dataset sourced directly from Alipay, one of the world's largest e-commerce and digital payment platforms. As infringement detection is a legal reasoning task requiring an understanding of the contexts and legal rules, we offer a thorough collection of legal rules and merchant and trademark-related contextual information with annotations from legal experts. We ensure the data quality by performing an extensive statistical analysis. Furthermore, we conduct an empirical study on this dataset to highlight its value and the key challenges. Through this study, we aim to contribute valuable resources to advance research into legal compliance related to trademark infringement within the e-commerce sphere.

1 Introduction

E-commerce companies are required to register in an online platform before conducting any business activities in that platform. However, the registration information may breach trademark laws if e.g. their registration names are similar to protected trademarks. However, it is expensive and time-consuming to check registration information manually when the number of daily registrations is large. To avoid trademark infringements and reduce manual costs, it is desirable for those online platforms to build tools to check legal compliance of the registration information *automatically*. However, there is no dataset to evaluate such tools rigorously.

[a]The authors contribute equally to this work.
[b]Corresponding author.

A trademark is an easily recognizable combination of signs, designs, letters, words and sounds that differentiates products or services of a company from those of others in a marketplace (Act, 2000). Detecting trademark infringement in a registration requires understanding the trademark laws in the corresponding country, identifying relevant issues and legal rules based on the understanding of the registration and relevant merchant information, and perform reasoning to draw a conclusion if there is an infringement or not. However, prior studies on trademark infringement simplify it either as a task of recognizing similar logos (Trappey et al., 2020) *without considering any contexts and laws* or focus on constructing trademark ontologies from precedents (Trappey et al., 2021b).

The recently released large language models (LLMs) demonstrate strong abilities in reasoning and document understanding (Huang and Chang, 2022). Hence, they are applied to tackling a variety of legal tasks (Katz et al., 2023). However, researchers find out that LLMs often yield different or wrong intermediate reasoning steps than humans despite the outcomes being the same (Tang et al., 2023; Paul et al., 2023). Although it is crucial for the users of infringement detection systems to understand how and why models draw particular conclusions, it lacks studies to understand the alignments between LLMs and human experts w.r.t. legal reasoning for trademark infringement.

To promote the research in the areas of trademark protection and legal reasoning, we build the *first* dataset on trademark infringement detection in registrations, coined TMID. The dataset consists of 17,365 pairs of merchant registration and trademark data, collected from Alipay, an e-commerce and online payment platform that primarily operates in China. Additionally, it includes a comprehensive database of auxiliary information relevant to merchant registrations, a collection of relevant trademarks and their auxiliary information, and a com-

Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: Industry Track, pages 176–184
December 6-10, 2023

pilation of statutes from Chinese trademark law. Each registration and trademark pair is annotated with a binary judgment indicating whether there is an infringement or not. To understand the alignments between LLMs and legal experts in terms of reasoning traces, a group of legal experts has manually codified the complete reasoning traces for 192 randomly selected registrations. In addition, we conduct the empirical study with BERT (Devlin et al., 2019), ChatGLM (Zeng et al., 2022), and GPT3.5[1] with varying settings for infringement detection on TMID, as well as compare the usefulness of reasoning traces created by humans and GPT3.5, and obtain the following novel findings:

- Our dataset is valuable for boosting the performance of LLMs. Both BERT and ChatGLM fine-tuned on our dataset outperform GPT3.5 and a rule-based baseline by more than 30% in terms of F1 scores.

- Both statutes and the auxiliary information relevant to the merchants in registrations provide particularly useful contextual information for LLMs. As a result, they improve the performance of ChatGLM by more than 10% in terms of F1 scores.

- The reasoning traces curated by legal experts provide highly valuable information for LLMs. By providing the first 33% of each human-crafted reasoning trace as inputs, the F1 score of GPT3.5 is improved by 18% and reaches 95.68%.

- In contrast, the reasoning traces generated by GPT3.5 degrade its performance by approximately 8%. A manual inspection by legal experts finds that only 25% of them are complete, and the reasoning steps in 42.5% of them are correct.

2 Background and Problem Definition

A registration in a Chinese online e-commerce platform breaches the Chinese trademark law (outlined in Appendix A.1), if it violates the corresponding statutes to protect the IP rights of the existing trademark owners, who can be individuals, businesses, or other legal entities. Statutes are the legal rules codified in legislation. Because China employs a civil law system, legal decisions are made mainly based on legal rules.

[1]https://chat.openai.com/

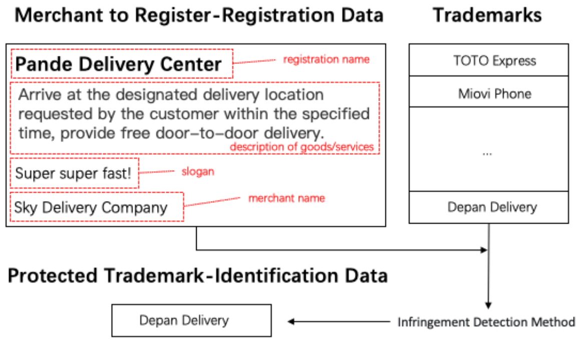

Figure 1: A registration example translated into English.

An example registration is depicted in Fig. 1. It consists of a registration name "Pande delivery center", a description of provided goods or services, the registered merchant name, and optionally also a slogan. It is an infringement because i) the registration name is similar to a protected trademark "Depan delivery" by only swapping two Chinese characters *de* and *pan*, ii) the service provided by the merchant is almost the same as the one provided by the company of the protected trademark, and iii) the company "sky delivery company" is not affiliated with or has a business relation with the company "Depan delivery".

Our legal experts usually consider four factors to determine if there is an infringement or not.

- **Similarity to a Trademark:** whether a registration has a name or any words and phrases in its description that bears a resemblance to other registered trademarks in terms of pronunciation, meaning, or the overall combination of elements.

- **Similarity between Goods:** whether the merchant to register sells goods that serve the same purpose, have the same use, target similar consumers, are produced by the same manufacturers, and are sold via the same distribution channels.

- **Similarity between Services:** whether the merchants to register provide services with the same objective, contents, methods, or targeting the same consumers.

- **Business Relations:** whether the merchant to register has an existing business relation with or is affiliated with the trademark owners.

when a registration name is similar to a protected trademark, the similarity between goods and services is also important, because it may not be an infringement if the provided services and goods are substantially different from the trademark owners. Moreover, if the merchant is a subsidiary of the trademark owner, it is legal to use a registration name similar to the trademark. Therefore, trademark infringement detection is beyond only measuring similarity between logos, as done in prior works (Trappey et al., 2020).

We manually analyze 192 registrations annotated with reasoning traces, which do not involve business relations with any trademark owners. Among them, 139 registrations involve infringements because those registrations are filtered by a recall-oriented deployed ensemble method detailed in Section 3.3. All of the infringed registrations use registration names or words in descriptions similar to protected trademarks. As the majority are service providers, 88.49% of infringements provide similar services, while only 10.07% of them provide similar goods.

Formally, the target task is to predict if a registration x_s^r infringes the trademark law or not.

$$\pi_\theta : \mathcal{X}s \times \mathcal{T}p \times \mathcal{R} \to \{0, 1\} \tag{1}$$

where $x_s \in \mathcal{X}s$ represents both the registration data x_s^r provided by a merchant who registers on the platform and x_s^p, which denotes the auxiliary information of this merchant collected by us. Herein, $t_p \in \mathcal{T}p$ corresponds to the information of a protected trademark, and R denotes the relevant statutes from Chinese Trademark Law. More detailed explanations regarding each type of information will be provided in Section 3.

3 Dataset Construction

3.1 Data Description

Assessing the risk of trademark IP infringement during a merchant's registration on an e-commerce platform is a complex reasoning process. It depends not only on the registration details from merchants and legal rules but also leverages supplementary data from both merchants and protected trademarks. We gather abundant information from both parties to facilitate accurate judgment of infringement risk, providing a rich context for both human annotators and automated machine learning models for their decision-making.

Merchants to Register. For the merchants to register on the platform, there are two types of data from different sources.

- *Registration Data* includes the registration name used by the merchants to register, a description of their goods/services, the names of the merchants, as well as the slogan and enterprise credit code.

- *Auxiliary Data* includes publicly available information that cannot be directly acquired from the e-commerce platform. This includes names of legal shareholders - either individuals or companies - linked to the merchants, which reveals ownership structures and details about the merchants' subsidiaries, as well as the industry category of the merchant. Such data aids in identifying whether the merchant's company is a subsidiary of a protected trademark owner or not.

Protected Trademarks. Data regarding the protected trademarks includes two main components.

- *Identification Data* include the distinctive names of a trademark registered with the China National Intellectual Property Administration[2]. This includes both their Chinese and English names, which serve as identifiers to differentiate the trademark from other ones.

- *Auxiliary Data* covers supplementary details, such as the type of the trademark defined by the platform and its registered industry category.

Trademark Law. The dataset includes statutes in the Chinese Trademark Act and the regulations related to trademark protection in China[3]. Detailed legal rules are listed in Appendix A.1.

3.2 Annotation

Judgements. A registration is aligned with protected trademarks and is annotated with a binary label, indicating whether the registration poses a risk of violating legal rules (labelled as 1) or not (labelled as 0).

Reasoning Traces. A reasoning trace is a sequence of reasoning steps leading up to the final judgement. Each intermediate step is a statement in natural language articulating the conclusion drawn upon previous steps and the relevant legal rules. As interpretability of judgements is crucial for legal

[2]https://english.cnipa.gov.cn/

[3]http://ip.people.com.cn/n1/2019/1106/c179663-31440313.html

The text [Pande Delivery Center] uses expressions that are similar or easily confused with the trademark [Depan Delivery], or uses expressions that are similar or related to the trademark [Depan Delivery].
Considering the registration name [Pande Delivery Center] and the description [Arrive at the designated delivery location requested by the customer within the specified time, provide free door-to-door delivery.], it could be determined that the registration information is in the same industry category as the services or products provided by the trademark [Depan Delivery].
The merchant name [Sky Delivery Company] has no information association with the brand [Depan Delivery].

Figure 2: An example of 3-step reasoning traces. The Chinese reasoning steps have been translated into English for a better understanding.

applications, reasoning traces provide insights into why and how a judgement is made. In light of this, our legal experts manually annotate their reasoning traces on a random sample of high-risk registrations. Herein, each reasoning step reflects their interpretation of the law and application of the relevant legal rules. As a result, we obtain 192 reasoning traces with 2.49 reasoning steps on average. Although this explicit annotation of reasoning traces is resource-intensive and time-consuming, it provides invaluable insights into the decision-making process of legal experts and assists in studying the alignments between human and machine reasoning. The three reasoning steps in the reasoning trace of the infringing case in Figure 1 are displayed in Figure 2.

3.3 Data Collection Process

We offer comprehensive details on the process of collecting the merchant and trademark information, along with instructions for annotating these pairs.

Merchant Data. We extract the registration information of all merchants from the database of the Chinese e-commerce platform.

The auxiliary data of the merchants is acquired via two steps: i) using web crawlers to extract information from enterprise websites to obtain shareholder and legal representatives information for the merchants registered on the enterprise websites, ii) linking the enterprise credit code of the merchant on the e-commerce platform with the data obtained from the enterprise data websites, and obtaining their legal representatives and shareholders.

Trademark Data. We curate a list of trademarks sourced from a variety of backgrounds. This includes trademarks of prestigious luxury brands, those owned by business entities that proactively seek protection of their intellectual properties from the platform, and trademarks representing proprietary brands owned by the e-commerce platform. The data of protected trademarks is obtained by crawling the official websites of the trademarks.

Annotation. Considering the large number of merchants on the e-commerce platform and the fact that the majority of them do not violate trademark statute laws, annotating all possible pairs of merchants and trademarks would incur significant costs. To address this, we have adopted a *recall-oriented* infringement detection ensemble method to identify pairs of potentially infringed registrations and the related trademarks. The approach is an ensemble algorithm encompassing various techniques like text classification, entity linking, edit distance, and keyword extraction, designed to reflect the logical rules inherent in trademark laws. This method is currently deployed in the e-commerce platform, processing over two million registration document-trademark pairs. Our human annotators then carefully label 17,365 high-risk cases identified by that method.

To ensure the dataset's quality, each pair selected for annotation undergoes a voting process involving two trained annotators but with only amateur-level legal backgrounds. If both annotators agree on the annotation, it is retained. However, if there is a discrepancy between the annotations, a legal expert with rich legal knowledge performs quality control and makes the final decision by selecting one of the annotators' results as the final annotation. The inter-annotator agreement rate between the two regular annotators is recorded as 89.6%.

Regarding the reasoning traces, due to the costly nature of annotation, we employ random sampling to select 192 pairs from the entire dataset. Legal experts are then asked to annotate these selected pairs using the rules applied during their reasoning trails. All annotation fees are paid in the monthly salary of the annotators and legal experts working for the e-commerce platform.

3.4 Data Statistics

Out of two million evaluated pairs, 17,365 were selected, with 2,836 labelled as infringing cases, 14,694 as non-infringing cases, and 192 annotated with reasoning traces. Each data instance has nine fields: six from merchants and three from trademarks. For each pair, there is a minimum of one field available from both merchant registration and

Merchant to Register						Protected Trademark		
Registration Data				Auxiliary Data		Identification Data	Auxiliary Data	
Registration Name	Service Description	Slogan	Merchant Name	Shareholder Name	Industry Category	Trademark Name	Industry Category	E-commerce Type
98.48%	95.58%	92.26%	99.01%	33.76%	41.48%	100%	52.38%	21.40%

Table 1: Data statistics of different data fields.

trademark identification data for infringement detection. However, not all fields are completely filled. Table 1 demonstrates that the data obtained from the e-commerce platform is more comprehensive, with over 90% coverage compared to the less-complete data scraped from public sources.

4 Experiments

4.1 Experimental Setup

Baselines. Four baselines are considered.

- **Deployed Ensemble Method**: As mentioned in Section 3.3, this approach, which is currently embedded in the online system of the e-commerce platform, is used to identify potential law violators during data annotation. The algorithm is based on heuristic rules, devised in accordance with legal statute laws.
- **BERT** (Devlin et al., 2019): This masked pretrained language model is primarily used for language understanding tasks. Here, we use its Chinese version[4]. We fine-tune BERT on TMID to enhance its performance in our specific task.
- **ChatGLM** (Zeng et al., 2022): With 6 billion parameters, ChatGLM is a large pretrained language model specifically designed for Chinese natural language generation tasks. To utilize this model, we transform classification tasks into generation tasks by instructing ChatGLM to generate the word “Infringe/Non-Infringe”. We further fine-tune ChatGLM with LORA technique (Hu et al., 2021) on TMID.
- **GPT3.5**[5]: GPT3.5 is an immensely powerful large language model with 175 billion parameters. However, direct access to its parameters is restricted, allowing us to only utilize zero/one-shot learning for our tasks.

Evaluation Metrics. In evaluating the performance of each baseline in the infringement detection task, we employ precision, recall and F1-measure, on the test sets.

[4]https://huggingface.co/bert-base-chinese
[5]https://chat.openai.com/

Baselines	F1 score	Precision	Recall
Deployed Ensemble	13.19	7.1	**92.8**
BERT	63.18	68.56	58.58
ChatGLM	**63.58**	**74.14**	55.66
GPT3.5			
zero-shot	35.46	22.76	80.20
one-shot	35.06	24.86	59.42

Table 2: Main results of different baselines in trademark IP infringement detection using full data information.

Implementation Details. In Table 2, BERT is configured with a batch size of 32 and a learning rate of 5e-5, and it is trained for 20 epochs on a V100-32G. The best-performing model, selected based on the validation set, is evaluated directly on the test set. ChatGLM is trained for 3 epochs on an A100 using LoRA fine-tuning, with a batch size of 2 and an input length of 2048. This model is tested on the test set using the trained weights.

The ChatGPT zero-shot experiments directly utilize the OpenAI gpt-3.5-turbo interface. To standardize the output, we append an additional prompt instructing the model to provide a response as either ‘yes’ or ‘no’. The GPT3.5 one-shot experiments add a single data point from the training set to the input prompt.

For Table 3, we employ the same ChatGLM LoRA fine-tuning configuration as in Table 2. For Table 4, we utilize the OpenAI gpt-3.5-turbo interface for zero-shot inference, yielding binary classification results. Table 5 uses the OpenAI gpt-3.5-turbo interface as well, prompting it to generate reasoning traces.

4.2 Main Results and Analysis

Settings. We partition the data into training, validation, and test sets, with 13,864, 1,488, and 2,013 instances. The input for the baseline models is formed by filling the texts of data fields and laws into the slots of a text template.

Analysis. Table 2 demonstrates that while the online deployed ensemble method achieves the highest recall, indicating its great ability to capture most infringement cases, its precision is low, dropping below 10. This discrepancy causes substantial human intervention to filter out false positives in practice. Furthermore, the deployed system falls

Merchant Auxiliary	Trademark Auxiliary	Law	F1 score	Precision	Recall
×	×	×	50.20	**87.70**	34.62
×	×	✓	62.72	80.30	51.46
×	✓	×	51.43	80.14	37.86
✓	×	×	63.63	81.73	52.10
✓	✓	×	59.71	81.11	47.25
×	✓	✓	58.87	78.07	47.25
✓	×	✓	**67.65**	78.88	**59.22**
✓	✓	✓	63.58	74.14	55.66

Table 3: Ablation study results on the impacts of auxiliary information on ChatGLM performance in trademark IP infringement detection. Checkmarks indicate the corresponding auxiliary information is included in the model input during training and inference.

significantly behind all baseline models that utilize TMID, showing a substantial gap of at least 15 points in terms of both F1 scores and precision. This underlines an immense opportunity for improving the current online system. By leveraging TMID, we can potentially enhance the effectiveness of the system and simultaneously reduce human effort.

Among the fine-tuned models, ChatGLM achieves the highest F1 score. We speculate that, although both models are pre-trained on billions of Chinese corpus, ChatGLM's larger model size grants it a more comprehensive understanding of Chinese legal knowledge compared to BERT. Interestingly, GPT3.5, which employs zero-shot or in-context learning, performs even worse than BERT, the fine-tuned model with significantly fewer parameters (only 1/500 of GPT3.5's size). This suggests that zero-shot and one-shot learning methods are inadequate for GPT3.5 to leverage knowledge from TMID effectively. However, despite this limitation, incorporating legal knowledge during pre-training still ensures that GPT3.5 outperforms the deployed ensemble method regarding F1.

4.3 Influence of Auxiliary Data

Settings. For infringement detection, our system always treats the pairing of merchant registration data and trademark identifiers as primary inputs, while auxiliary data and related statute laws serve as additional inputs for the model. Therefore, this experiment investigates how various auxiliary information influences ChatGLM performance.

Analysis. Table 3 reveals that integrating auxiliary information and legal rules generally improves the model's performance, as measured by the F1 score, with varying degrees of efficacy based on

	F1 score	Precision	Recall
w/o RT	77.30	75.69	78.99
w. 33% RT	88.08	81.60	**95.68**
w. 67% RT	**88.66**	**83.77**	94.16
w. 100% RT	87.37	83.12	92.09
w. 100% GPT3.5 RT	69.38	70.68	68.12

Table 4: Zero-shot performance of GPT3.5 under different configurations, utilizing either human or GPT3.5-generated reasoning traces.

the data blend. Auxiliary information about the merchant yields the most substantial enhancement, boosting the F1 score by approximately 13 points over the model that includes no auxiliary data or legal rules. The trademark auxiliary information, however, only modestly improves performance by about a point. Notably, when trademark auxiliary information is combined with other data types, ChatGLM's performance decreases compared to when using any of them individually or the other two without trademark auxiliary, showing a gap of at least 4 points.

Interestingly, ChatGLM can achieve the highest F1 by leveraging only the name phrases of trademark identifiers without using any trademark auxiliary data. We speculate that ChatGLM has already assimilated comprehensive background knowledge related to the corresponding trademarks, which might explain why it obtains limited benefits from the trademark auxiliary data. In contrast, other models lacking such inherent capabilities can still benefit from including trademark auxiliary data. In the case of BERT and zero-sot GPT3.5, F1 drops by 5 and 8 points, respectively, when no trademark auxiliary data is used, compared to using the full data setting. Please see Appendix A.2 for details.

4.4 Influence of Reasoning Traces

Settings. To explore the potential benefits of reasoning traces in infringement detection, we incorporate different proportions (33%, 67%, and 100%) of *each text* from 192 reasoning traces (RTs) into the input. We aim to assess whether this inclusion could enhance the zero-shot performance of GPT3.5. Furthermore, we apply a chain-of-thought approach (Wei et al., 2022) to GPT3.5 to evaluate whether GPT3.5 could be improved by using reasoning traces generated by GPT3.5 in a zero-shot manner, in contrast to those generated by humans.

Analysis. Table 4 reveals the substantial impact of incorporating RTs into the input of GPT3.5. In-

	Correctness	Completeness
GPT3.5 R.T.	42.5%	25%

Table 5: Human evaluation results on 20 GPT3.5-generated reasoning traces.

cluding merely 33% of text in RTs has significantly boosted the F1 score, precision, and recall for GPT3.5 in a zero-shot setting (+10.78%, +5.91%, and +16.69%, respectively). However, including 33% of the RT text can enable GPT3.5 to perform comparably to those fed with 67% and 100% of the text. We observe that the initial stages of reasoning often convey the most crucial information for detecting infringements.

We further evaluate whether the GPT3.5-generated RTs can help GPT3.5 in a chain-of-thought manner. However, integrating the GPT3.5-generated RTs led to a significant performance decline of 8 points. Subsequently, we examine the alignment of 20 selected GPT3.5 RTs with human RTs by two legal experts. In Table 5, our findings reveal that, on average, experts reach a consensus that only a mere 42.5% of GPT3.5 RTs can result in the final correct judgments as determined by human judgment and only 25% of GPT3.5 RTs show the complete set of reasoning steps observed in human-written RTs. The poor RT quality could degrade the overall performance of GPT3.5.

5 Related Work

Existing research in automatic trademark IP infringement detection typically simplify the problem definition to logo image similarity (Peng and Chen, 1997; Alshowaish et al., 2022; Trappey et al., 2020; Li et al., 2023; Mao et al., 2023; Trappey et al., 2021a; Tursun et al., 2019) or textual similarity detection (Trappey et al., 2020), subsequently proposing methods using diverse machine learning models like convolutional neural networks (Gu et al., 2018) or recurrent neural networks (Hochreiter and Schmidhuber, 1997). Other studies delved into constructing trademark ontologies (Trappey et al., 2021b) or developing logo similarity detection datasets (Hou et al., 2021; Wang et al., 2022). Distinct from these studies, our work directly addresses a real-world issue of trademark IP infringement detection on the e-commerce platform, providing comprehensive textual data with legal annotations based on statute laws.

6 Conclusion

In this work, we present the *first* dataset, coined TMID, on trademark infringement detection in the merchant information registered to online e-commerce platforms. The target task requires legal reasoning over registrations, information about the merchant to register, statutes in Trademark laws, protected trademarks and auxiliary information about trademark owners. Our empirical study shows that i) LLMs greatly benefit from the training data and the contextual information from our dataset; ii) powerful GPT3.5 still fails to generate reasoning traces aligning with those from legal experts but are able to reach an F1 over 95% if the reasoning traces are correct. This work does not only provides a useful resource but also sheds light on the limitations of LLMs on complex reasoning.

Limitations

The primary limitation of this study arises from incomplete data. Some data fields, notably those related to trademarks, are incomplete in a subset of the instances, which may undermine the value of our data for training language models towards infringement detection. Moreover, the collection of reasoning traces is a labour-intensive process, resulting in a relatively small dataset. This scarcity may impede further studies into infringement detection using reasoning traces.

Ethics Statement

We ensure all relevant studies are carefully reviewed and approved by an internal ethics board, focusing on privacy and legal considerations.

Privacy of Personal Information. To improve privacy standards and mitigate the risk of personal identification disclosure, we've implemented anonymization measures on our dataset. The characters of the personal names, including those of individual shareholders, have been scrambled based on a predefined vocabulary mapping to ensure anonymity.

Misuse of Data. It is important to note that this dataset is strictly reserved for academic research. Its deployment in real-world business environments or for commercial pursuits is expressly forbidden.

References

Trade Marks Act. 2000. What is a trade mark?

Hayfa Alshowaish, Yousef Al-Ohali, and Abeer Al-Nafjan. 2022. Trademark image similarity detection using convolutional neural network. *Applied Sciences*, 12(3):1752.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Jiuxiang Gu, Zhenhua Wang, Jason Kuen, Lianyang Ma, Amir Shahroudy, Bing Shuai, Ting Liu, Xingxing Wang, Gang Wang, Jianfei Cai, et al. 2018. Recent advances in convolutional neural networks. *Pattern recognition*, 77:354–377.

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation*, 9(8):1735–1780.

Qiang Hou, Weiqing Min, Jing Wang, Sujuan Hou, Yuanjie Zheng, and Shuqiang Jiang. 2021. Foodlogodet-1500: A dataset for large-scale food logo detection via multi-scale feature decoupling network. In *Proceedings of the 29th ACM International Conference on Multimedia*, pages 4670–4679.

Edward J Hu, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, Weizhu Chen, et al. 2021. Lora: Low-rank adaptation of large language models. In *International Conference on Learning Representations*.

Jie Huang and Kevin Chen-Chuan Chang. 2022. Towards reasoning in large language models: A survey. *arXiv preprint arXiv:2212.10403*.

Daniel Martin Katz, Dirk Hartung, Lauritz Gerlach, Abhik Jana, and Michael J Bommarito II. 2023. Natural language processing in the legal domain. *arXiv preprint arXiv:2302.12039*.

Xingzhuo Li, Sujuan Hou, Baisong Zhang, Jing Wang, Weikuan Jia, and Yuanjie Zheng. 2023. Long-range dependence involutional network for logo detection. *Entropy*, 25(1):174.

KeJi Mao, RunHui Jin, KaiYan Chen, JiaFa Mao, and GuangLin Dai. 2023. Trinity-yolo: High-precision logo detection in the real world. *IET Image Processing*.

Debjit Paul, Mete Ismayilzada, Maxime Peyrard, Beatriz Borges, Antoine Bosselut, Robert West, and Boi Faltings. 2023. Refiner: Reasoning feedback on intermediate representations. *arXiv preprint arXiv:2304.01904*.

Hsiao-Lin Peng and Shu-Yuan Chen. 1997. Trademark shape recognition using closed contours. *Pattern Recognition Letters*, 18(8):791–803.

Ruixiang Tang, Dehan Kong, Longtao Huang, and Hui Xue. 2023. Large language models can be lazy learners: Analyze shortcuts in in-context learning. *arXiv preprint arXiv:2305.17256*.

Amy JC Trappey, Charles V Trappey, and Samuel Shih. 2021a. An intelligent content-based image retrieval methodology using transfer learning for digital ip protection. *Advanced Engineering Informatics*, 48:101291.

Charles V Trappey, Ai-Che Chang, and Amy JC Trappey. 2021b. Building an internet-based knowledge ontology for trademark protection. *Journal of Global Information Management (JGIM)*, 29(1):123–144.

Charles V Trappey, Amy JC Trappey, and Sam C-C Lin. 2020. Intelligent trademark similarity analysis of image, spelling, and phonetic features using machine learning methodologies. *Advanced Engineering Informatics*, 45:101120.

Osman Tursun, Simon Denman, Sabesan Sivapalan, Sridha Sridharan, Clinton Fookes, and Sandra Mau. 2019. Component-based attention for large-scale trademark retrieval. *IEEE Transactions on Information Forensics and Security*, 17:2350–2363.

Jing Wang, Weiqing Min, Sujuan Hou, Shengnan Ma, Yuanjie Zheng, and Shuqiang Jiang. 2022. Logodet-3k: A large-scale image dataset for logo detection. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 18(1):1–19.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems*, 35:24824–24837.

Aohan Zeng, Xiao Liu, Zhengxiao Du, Zihan Wang, Hanyu Lai, Ming Ding, Zhuoyi Yang, Yifan Xu, Wendi Zheng, Xiao Xia, et al. 2022. Glm-130b: An open bilingual pre-trained model. *arXiv preprint arXiv:2210.02414*.

A Appendix

A.1 Chinese Trademark Protection Laws

The Intellectual Property Protection Law, encompassing patent, trademark, copyright, and network security laws, outlines the framework for intellectual property rights protection. Trademark infringement protection is primarily based on the Trademark and Anti-Unfair Competition Laws of the People's Republic of China, which define the scope of protection, rights of holders, determination of infringement, and legal consequences. Figure 3 shows the detailed statute law rules in both Chinese and English.

Chinese	English
未经商标注册人的许可，在同一种商品上使用与其注册商标相同的商标的	using a trademark identical to a registered trademark on the same goods without the permission of the trademark registrant
未经商标注册人的许可，在同一种商品上使用与其注册商标近似的商标，或者在类似商品上使用与其注册商标相同或者近似的商标，容易导致混淆的	using a trademark similar to a registered trademark on the same or similar goods without the permission of the trademark registrant, which is likely to cause confusion
销售侵犯注册商标专用权的商品的	selling goods that infringe the exclusive right to use a registered trademark
伪造、擅自制造他人注册商标标识或者销售伪造、擅自制造的注册商标标识的	counterfeiting or manufacturing the registered trademark of others without authorization or selling counterfeit or unauthorized registered trademark
未经商标注册人同意，更换其注册商标并将该更换商标的商品又投入市场的	replacing the registered trademark of others without their consent and putting the goods with the replacement trademark back on the market
故意为侵犯他人商标专用权行为提供便利条件，帮助他人实施侵犯商标专用权行为的	intentionally providing convenience for others' infringement of trademark exclusive rights and helping others to commit infringement of trademark exclusive rights
给他人的注册商标专用权造成其他损害的，中华人民共和国商标法第五十八条规定，将他人注册商标、 未注册的驰名商标作为企业名称中的字号使用， 误导公众，构成不正当竞争行为的，依照《中华人民共和国反不正当竞争法》处理。	causing other damages to the exclusive right to use a registered trademark of others. Article 58 of the Trademark Law of the People's Republic of China stipulates that using others' registered trademarks or unregistered well-known trademarks as the name of an enterprise to mislead the public and constitute unfair competition shall be handled in accordance with the "Anti-Unfair Competition Law of the People's Republic of China."

Figure 3: Trademark IP Laws of the People's Republic of China.

A.2 Influence of Trademark Auxiliary Data

	F1 score	Precision	Recall
BERT	58.09	67.23	51.13
ChatGLM	67.65	78.88	59.22
ChatGPT			
zero-shot	27.34	21.22	38.44
one-shot	36.77	28.08	53.25

Table 6: Performance comparison of BERT, ChatGLM, and GPT3.5 zero/one-shot models on full data types, excluding the trademark auxiliary data.

Table 6 illustrates the performance of various models, evaluated with all data types as input but excluding the trademark auxiliary data.