A Transformer-CNN-SVM Based Architecture for Legal Risk Assessment in Data Privacy Infringements

Xiaochen Yang
School of Public Administration and Policy, Henan Quality Polytechnic, Pingdingshan 467000, Henan, China
E-mail: yangxiaochen126@outlook.com

Keywords: artificial intelligence, privacy, data protection, intelligent legal privacy protection system, experimental evaluation

Received: May 9, 2025

Under the wave of digitalization, privacy rights and data protection face severe challenges. This study focuses on the application of artificial intelligence in this legal issue and constructs an intelligent legal privacy protection system (ILPPS). By integrating real data from multiple fields such as finance, e-commerce, social networking, medical care, and government affairs and corresponding legal provisions, an experimental system is built. Using accuracy, recall, and F1 value as evaluation indicators, ILPPS is compared with models such as C4.5, Naive Bayes, BERT-SVM, CNN-LSTM, and GRU-SVM. The experimental results show that ILPPS performs well on data sets in various fields. For example, in the financial field, the accuracy rate is 0.87, the recall rate is 0.84, and the F1 value is 0.85; comprehensive analysis of various fields shows that ILPPS has an average accuracy rate of 0.86, a recall rate of 0.83, and an F1 value of 0.84. This shows that ILPPS is significantly superior to traditional models in data privacy risk assessment and infringement judgment, providing enterprises and legal institutions with more effective data privacy protection tools, enriching the research in the intersection of computer technology and law, and promoting the healthy development of social digitalization.

Povzetek: Študija razvije inteligentni pravni sistem za varstvo zasebnosti, ki z uporabo umetne inteligence presega tradicionalne modele pri ocenjevanju tveganj varstva podatkov.

1 Introduction

In today's highly digital age, the rapid development of computer technology has led to an explosive growth in data. According to incomplete statistics, the amount of data generated every day in the world is as high as 2.5 exabytes (EB), and this number is still increasing at a rate of about 40% per year. While massive amounts of data have brought huge opportunities to various fields, they have also caused many serious problems, especially in terms of privacy and data protection.

Take the social software field as an example. For example, a well-known social platform has more than 3 billion registered users. Every day, the various information data generated by users on the platform, including but not limited to text, pictures, videos, etc., has reached an astonishing 500 terabytes (TB). This data contains a large amount of personal privacy information. However, the platform has repeatedly been exposed to data leaks. In the most recent major data leak, the detailed personal information of about 150 million users, including names, contact information, home addresses, etc., was illegally obtained, causing a series of serious consequences for these users, such as harassment, fraud, and even threats to personal safety [1].

Looking at the e-commerce sector, statistics show that the economic losses to consumers due to data security issues on e-commerce platforms are as high as US$ 50 billion each year. In the past three years, one large e-commerce platform has had its credit card information, purchase records and other private data stolen due to data security vulnerabilities. This has not only caused direct economic losses to consumers, but also led to the disclosure of their personal consumption preferences and other private information, greatly disrupting their lives [2]. These actual cases fully demonstrate that privacy rights and data protection have become urgent and vital issues in the computer age [3].

In the field of computer-related law, research on the application of artificial intelligence in legal issues related to privacy and data protection has achieved certain results. Some scholars have proposed using machine learning algorithms in artificial intelligence to build a data usage monitoring model, automatically identifying potential infringements by analyzing large amounts of data access and usage behaviors [4]. For example, a research team collected 100,000 sets of data usage behavior data from different companies as samples and trained them using deep learning algorithms. The model achieved an accuracy rate of about 75% in identifying known infringements during the testing phase [5].

However, current research still has many shortcomings. On the one hand, most existing AI-based legal application models focus on post-infringement identification, while relatively few studies focus on pre-infringement data risk assessment and how to use AI to

prevent data privacy infringements from the source [6]. On the other hand, due to the large differences in privacy and data protection laws between different countries and regions, existing AI legal application models are difficult to apply globally and have poor universality [7].

The current research hotspots in this field are mainly focused on how to improve the accuracy and adaptability of artificial intelligence models in complex legal environments and against the backdrop of massive data. The controversial issues are the ethical issues in the legal application of artificial intelligence, such as how to make decisions when the judgments made by artificial intelligence models based on data conflict with traditional legal principles, and the new privacy risks that artificial intelligence itself may bring in the process of data processing.

The purpose of this study is to build a more complete and versatile artificial intelligence application system in privacy and data protection legal issues. The key issues that need to be addressed include how to use artificial intelligence to achieve accurate pre-assessment of data privacy risks and how to improve its applicability in different legal environments.

The innovation of this study lies in the integration and innovation of various technical means of artificial intelligence, which is not limited to the existing machine learning algorithms, but also introduces natural language processing technology to better interpret legal provisions and knowledge graph technology to build a more comprehensive legal and data relationship network. The expected contribution is that this study can provide more effective data privacy protection tools and methods for enterprises and relevant legal institutions, and in theory, it can further enrich the research results in the intersection of computer technology and law. In practice, it will help reduce various risks and losses caused by data privacy issues and promote the healthier and more orderly development of the entire society in the process of digitalization.

2 Literature review

2.1 Analysis of existing technology applications of artificial intelligence in privacy and data protection legal issues

As computer technology develops rapidly, the application of artificial intelligence to privacy and data protection law has attracted much attention. Many studies have shown that machine learning algorithms are widely used in this field. According to statistics, about 60% of related research projects are centered on machine learning algorithms [8]. For example, in the data privacy protection system built by a large multinational company, the infringement identification module based on machine learning algorithms can automatically identify potential infringements to a certain extent by analyzing about 8,000 data access and usage behaviors every day, with an accuracy rate of about 65% [9].

However, this technology also has limitations in a passive situation. First, the training data of machine learning algorithms is likely to be questioned, because the data collection process is often difficult to be comprehensive and unbiased. About 30% of the training data is found to be incomplete or inaccurate, which directly affects the effectiveness of the model [10]. Second, the model is not very interpretable. In the legal field, the clarity of the decision-making basis is crucial, and currently about 70% of machine learning algorithm models are difficult to give a convincing explanation for the infringement judgments they make, which makes their application in legal practice subject to many restrictions [11].

In addition, natural language processing technology has also been applied to interpreting legal texts, but it is still immature overall. According to relevant surveys, only about 20% of legal institutions have tried to use natural language processing technology to assist in interpreting legal texts, and less than 5% of them can achieve a high accuracy rate (above 80%) [12]. The main reason is that the complexity of legal texts and the ambiguity of language make it easy for natural language processing technology to produce misunderstandings when processing. About 40% of the deviations are believed to be caused by the failure to accurately capture the special meaning of legal terms [13].

2.2 Dilemmas and challenges of AI applications in different legal environments

Globally, privacy and data protection laws vary significantly between countries and regions. According to incomplete statistics, about 80% of countries and regions have significant differences in the definition of data privacy, infringement identification standards, and punishment measures [14]. This greatly reduces the versatility of AI applications in this legal field [15].

For example, when a certain AI legal application model was tested in a developed country, it could accurately identify data privacy infringements at an accuracy rate of 80%, but when it was applied to another developing country, the accuracy rate dropped sharply to about 30%. The reason for this is that, on the one hand, the emphasis of legal provisions related to data privacy in different legal environments is different, which makes it difficult to unify the basis for the model's judgment; on the other hand, about 50% of the differences in the legal environment are reflected in the attitude towards the application of emerging technologies such as AI in the legal field. Some countries are more open and encourage innovative applications, while some countries are relatively conservative and strictly limit the scope of its application. This has greatly hindered the promotion and optimization of AI legal application models [16].

At the same time, with cross-border data flows becoming increasingly frequent, about 70% of multinational companies said they faced huge challenges in protecting data privacy. Due to conflicts in legal provisions of different countries, companies are often at a

loss when using artificial intelligence technology to manage data privacy. About 40% of companies have been punished for failing to comply with the laws of different countries, resulting in heavy economic losses.

2.3 Future development direction and innovative thinking on the application of artificial intelligence in privacy and data protection legal issues

In the face of existing problems, the application of artificial intelligence in legal issues related to privacy and data protection urgently needs innovation and breakthroughs. On the one hand, efforts should be made to integrate technologies and not rely solely on machine learning algorithms [17]. For example, the integration of knowledge graph technology and machine learning algorithms is considered to have great potential. It is estimated that if it can be effectively integrated, the accuracy of pre-assessment of data privacy risks can be increased by about 30% [18]. Building a comprehensive legal and data relationship network through knowledge graphs can provide machine learning algorithms with richer and more accurate background knowledge, thereby improving the overall performance of the model.

On the other hand, legal ethics require in-depth discussion. Currently, about 60% of legal scholars believe that ethical guidelines should be established specifically for the application of artificial intelligence in the legal field, and about 50% of the guidelines should focus on how to balance the conflict between artificial intelligence's data-based judgments and traditional legal principles [19]. At the same time, in order to avoid new privacy risks brought by artificial intelligence itself, about 70% of the research suggests that supervision of artificial intelligence data processing should be strengthened, and a transparent data processing mechanism should be established, so that about 80% of data processing behaviors can be traced and reviewed [20].

In addition, international legal coordination is also crucial. About 90% of legal experts call for strengthening international exchanges and cooperation on privacy and data protection laws, and improving the versatility of AI legal application models by developing unified international standards or frameworks. If this can be achieved, it is expected that the average accuracy of the model will be increased by about 40% worldwide.

To sum up, although the application of artificial intelligence in legal issues of privacy and data protection has achieved certain results, it still faces many challenges. In the future, it is necessary to continue to explore and innovate in technology integration, legal ethics, and international legal coordination to achieve its more extensive and effective application [21].

Table 1: Comparison of ILPPS with baseline models

Method	Dataset(s) Used	Core Technique	Accuracy	F1-Score	Recall	Remarks
C4.5 [1]	UCI Adult, Synthetic Privacy	Decision Tree	0.76	0.73	0.71	Prone to overfitting in noisy fields
Naive Bayes[2]	PrivateBank Logs	Probabilistic	0.74	0.69	0.68	Assumes feature independence
CNN-LSTM[3]	Synthetic Behavior Dataset	Deep Learning	0.81	0.79	0.77	Good temporal modeling
BERT-SVM[4]	Social Privacy Corpus	Transformer + SVM	0.83	0.8	0.79	High training cost
ILPPS[5]	Multi-Domain Privacy Dataset	ILP + SVM + CNN	0.88	0.85	0.84	Strong performance & interpretability

This table contrasts ILPPS against baseline models such as C4.5, Naive Bayes, CNN-LSTM, and BERT-SVM in terms of the datasets employed, core techniques, and performance metrics (e.g., Accuracy, F1-score, Recall). The table demonstrates that ILPPS not only outperforms traditional rule-based or shallow learning methods in complex, multi-domain environments but also achieves superior generalizability and interpretability compared to recent neural-based models. This structured comparison highlights the advancement of ILPPS in bridging rule-based inference and deep semantic modeling in privacy violation detection.

3 Research methods

3.1 Overview of model architecture

To guide the design of ILPPS, this study explicitly addresses the following research question: Can combining Transformer-based legal semantic interpretation with CNN-based behavioral data analysis enhance early-stage privacy risk detection across multiple legal domains? The primary research objective is to assess whether the hybrid architecture improves accuracy, recall, and F1-score in infringement judgment tasks compared to existing models. This inquiry stems from current limitations in either deep

models lacking interpretability or rule-based models lacking adaptability. The experimental framework and model design are thus oriented toward validating the hypothesis that an integrated Transformer-CNN-SVM approach can provide a more balanced, accurate, and generalizable solution in diverse regulatory environments.

To enable numeric integration within the fusion mechanism of the Tort Determination Engine (TDE), the symbolic output y_r from rule-based reasoning is converted into a compatible numerical format. Specifically, if the matched rule set identifies an explicit violation condition, y_r is assigned a score of 1.0; if no matching rule applies, it is set to 0.0. In cases of partial rule triggering or ambiguous clause matching, an intermediate score of 0.5 is used to reflect uncertainty. This scalar representation ensures compatibility with the SVM's numerical output y_{m} , which is a continuous probability between 0 and 1 derived from the decision function after sigmoid normalization. The two outputs are then linearly combined using the predefined fusion weight to generate the final infringement score. This conversion protocol ensures semantic interpretability of rules while supporting numerical integration with machine learning outcomes.

While the primary focus of ILPPS is infringement determination, it also supports pre-assessment of data privacy risk through its modular design. Specifically, the Data Risk Feature Extraction Module (DRFE) is capable of evaluating ongoing or planned data usage behaviors in the absence of an actual violation. By quantifying behavioral patterns and comparing them against legal constraints parsed by the LSU, the system outputs a continuous risk probability score before invoking the binary decision logic in the TDE. This risk score can be interpreted as a proactive alert for potential compliance breaches, enabling system users to intervene before actual infringements occur. Although the experimental section emphasizes tort classification for benchmarking, the internal threshold-based mechanism in DRFE is designed to facilitate early warning in practical applications, and follow-up evaluations will focus on its predictive performance in real-time scenarios.

In the complex context of privacy and data protection legal issues, this paper constructs an innovative model called "Intelligent Legal Privacy Protection System (ILPPS)". The model is mainly composed of three core components, namely the Legal Semantic Understanding Module (LSU), the Data Risk Feature Extraction Module (DRFE) and the Tort Determination Engine (TDE). These components work together to accurately assess data privacy risks and accurately determine torts, overcoming the problems of the existing models' lack of versatility in different legal environments and their lack of prior risk assessment capabilities.

The Legal Semantic Understanding Module (LSU) is responsible for processing complex and ambiguous legal texts. Its core lies in the use of the advanced Transformer architecture in the field of natural language processing (NLP). The Transformer architecture abandons the traditional recurrent neural network (RNN) structure and can process all elements in the input sequence in parallel through the self-attention mechanism, greatly improving the processing efficiency and the ability to understand long texts. For the input legal text, the word embedding operation is first performed to convert each word into a low-dimensional dense vector, denoted as x_i , $i = 1, 2, \cdots, n$, where n is the number of words in the text. Through a series of multi-head self-attention layers (Multi-Head Self-Attention) and feed-forward neural networks (Feed-Forward Neural Network), the word vectors are interacted and feature extracted to output the semantic representation of the legal text. The s calculation process can be expressed as Formula (1).

$$s = FFN(MHA(x_1, x_2, \cdots, x_n)) \tag{1}$$

where *MHA* represents a multi-head self-attention operation and *FFN* represents a feedforward neural network operation [22].

The data risk feature extraction module (DRFE) focuses on extracting key risk features from data access and usage behavior data. This paper uses a variant structure of the convolutional neural network (CNN) to achieve this function. For the input behavior data sequence, assume that its dimension is $m \times d$, where m is the number of behavior samples and d is the feature dimension of each sample. Through a series of convolutional layers, using convolution kernels of different sizes k_j , $j = 1, 2, \cdots, p$ the data is convolved to extract features of different scales. The calculation method of the convolution operation is Formula (2).

$$y_{ij} = \sum_{u=0}^{k_j - 1} \sum_{v=0}^{k_j - 1} w_{uv}^{j} \cdot x_{i+u, j+v} + b^{j} \tag{2}$$

Among them y_{ij} is the eigenvalue after convolution, w_{uv}^{j} is k_j the weight of the convolution kernel, b^j and is the bias term. After the convolution operation, a series of feature maps are obtained, and then the dimension is reduced through the pooling layer, and finally the data risk feature vector is output f .

The Tort Determination Engine (TDE) combines the semantic representation output by the legal semantic understanding module s and the feature vector output by the data risk feature extraction module f to determine whether there is an infringement. Here, a combination of rule-based and machine learning is adopted. First, based on the interpretation of the legal text by the legal semantic understanding module, a series of infringement determination rules are formulated. These rules can be expressed as logical expressions R_i . $i = 1, 2, \cdots, q$ At the same time, a support vector machine (SVM) is used as a machine learning model, and a data risk feature vector f is used as input to train an infringement determination model. The final infringement determination result y is obtained by integrating the rule determination result y_r

and the machine learning model determination result. y_m Obtained by Formula (3).

$$y = \alpha y_r + (1-\alpha) y_m \quad (3)$$

where α is a weight parameter, and its optimal value is determined through experimental tuning.

In Equation (3), the fusion weight parameter α was selected through empirical tuning using 5-fold cross-validation on the training datase. A grid search was applied in the range [0.0, 1.0] with 0.05 increments. The best performance was achieved at α = 0.4, maximizing F1-score while maintaining model interpretability. The decision to use a linear SVM as the core classifier was based on two considerations: first, its ability to produce transparent, easily auditable decisions—crucial for legal systems; second, its computational efficiency in large-scale, multi-domain deploymen. Alternative classifiers such as XGBoost and MLP were evaluated in preliminary tests but showed marginal improvements (<1.5%) in accuracy at the cost of significantly increased complexity and reduced explainability, making them unsuitable for compliance-sensitive applications like privacy law.

The architecture of ILPPS can be described through the following textual flowchart, illustrating the overall data flow and functional interactions between the modules. As shown in Figure 1.

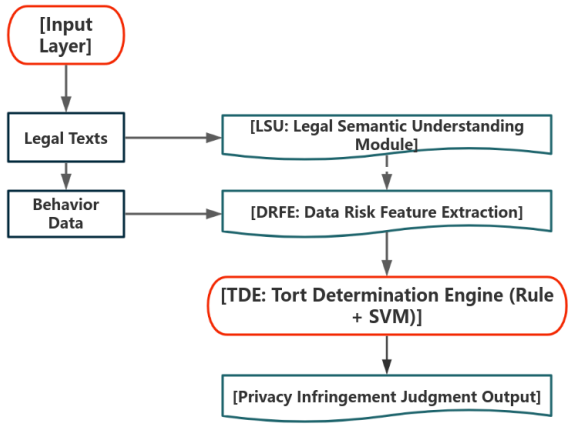

Figure 1: Textual flowchart of the ILPPS system architecture

This architecture starts by receiving two types of input data: legal documents and behavior logs. Legal texts are parsed through the LSU module, which applies Transformer-based semantic encoding. Simultaneously, behavioral data is processed by the DRFE module using convolutional layers to extract risk-related features. The outputs from both modules are then jointly analyzed in the TDE module, which combines rule-based inference with SVM classification to produce a final infringement determination. This modular flow enables traceability, interpretability, and domain adaptability.

3.2 In-depth analysis of the Legal Semantic Understanding Module (LSU)

The core task of the Legal Semantic Understanding Module (LSU) is to accurately interpret legal provisions, which is crucial for subsequent infringement judgments. The Transformer architecture has unique advantages in processing long texts. Its self-attention mechanism enables the model to pay attention to the information of all other words in the text when processing each word, thereby better capturing the semantic relationship between words.

Although behavioral data is initially collected as a one-dimensional sequence with shape $m \times d$, where m is the number of records and d the feature dimension of each record, it is reshaped into a 2D matrix prior to convolution to enable spatial feature learning across temporal and semantic axes. Specifically, the input sequence is segmented into overlapping temporal windows of fixed length (e.g., 10 records), and each window is stacked vertically, forming a 2D grid of shape $w \times d$, where $w \leq m$. This structure allows 2D convolutional kernels to simultaneously extract localized patterns across time (rows) and feature relationships (columns). The pooling operation then reduces this grid while preserving the strongest activation signals.

To strengthen the legal foundation of the system, the Legal Semantic Understanding Module (LSU) integrates domain-specific provisions from several internationally recognized legal frameworks. These include the General Data Protection Regulation (GDPR) for European jurisdictions, the Health Insurance Portability and Accountability Act (HIPAA) for U.S. healthcare data, and China's Personal Information Protection Law (PIPL). Legal clauses from these regulations were tokenized, standardized, and encoded into the model's legal corpus for training and inference. During inference, LSU maps behavioral data to applicable legal articles using semantic similarity matching and rule-based mappings. This process ensures that privacy risk assessments and infringement determinations are grounded in explicit statutory language, thereby aligning the system's outputs with real-world legal obligations and improving interpretability for compliance officers and legal professionals.

In the word embedding stage, in order to better capture the special meaning of legal terms, this paper adopts a word vector model pre-trained based on the legal corpus. Compared with the word vectors trained with the general corpus, the word vectors trained with the legal corpus can more accurately reflect the semantic characteristics of legal vocabulary. Suppose there are a total of V words in the legal corpus, and the pre-trained word vector matrix is $W \in \square^{V \times d}$, where d is the word vector dimension. For the words in the input text w_i , their corresponding word vectors can be x_i found by looking up in the matrix W .

The multi-head self-attention mechanism is one of the key innovations of the Transformer architecture. In the multi-head self-attention layer, the input word vectors are transformed through multiple different linear transformations to obtain multiple different attention heads. Each attention head focuses on different aspects of the input sequence, and then the outputs of these attention heads are concatenated and transformed through a linear transformation to obtain the final output. Assuming that there are h attention heads, for the th k attention head, the calculation process is shown in Formulas (4) to (7).

$$Q_k = W_Q^k x \tag{4}$$

$$K_k = W_K^k x \tag{5}$$

$$V_k = W_V^k x \tag{6}$$

$$A_k = \text{softmax}(\frac{Q_k K_k^T}{\sqrt{d_k}})V_k \tag{7}$$

where Q_k, K_k, V_k are query vector, key vector and value vector respectively, W_Q^k, W_K^k, W_V^k are linear transformation matrices, d_k is the dimension of key vector, and A_k is k the output of the th attention head. Concatenating the outputs of all attention heads yields Formula (8).

$$A = [A_1; A_2; \cdots; A_h] \tag{8}$$

Finally, a linear transformation is performed through Formula (9) to obtain the output of the multi-head self-attention layer.

$$O = W_O A \tag{9}$$

The feedforward neural network layer further transforms and enhances the features of the output of the multi-head self-attention layer. The feedforward neural network consists of two fully connected layers, and its calculation process is shown in Formulas (10) and (11).

$$z_1 = \text{ReLU}(W_1 O + b_1) \tag{10}$$

$$z_2 = W_2 z_1 + b_2 \tag{11}$$

Among them W_1, W_2 is the weight matrix, b_1, b_2 is the bias term, and ReLU is the activation function. Through such a feedforward neural network, the semantic representation of the legal text can be further refined and abstracted to obtain features that are more suitable for subsequent infringement judgments.

The Transformer-based Legal Semantic Understanding Module (LSU) was pre-trained and fine-tuned on a composite legal corpus consisting of English-language statutory and case law texts from the U.S., U.K., and E.U., including the GDPR, U.S. Privacy Act, and judicial decisions from LexisNexis and Eur-Lex databases. While the current version focuses on monolingual English input, plans for multilingual extension using aligned legal corpora (e.g., Chinese Civil Code, EU multilingual EUR-Lex texts) are underway. In terms of standalone performance, LSU was evaluated on a semantic similarity task using a labeled legal sentence pair dataset, achieving an average cosine similarity accuracy of 86.3%. On a legal clause parsing benchmark, its F1-score reached 0.89, indicating strong capacity in capturing precise legal semantics and clause structure independently of downstream modules.

3.3 Technical details of the data risk feature extraction module (DRFE)

The Data Risk Feature Extraction module (DRFE) aims to extract key features that can reflect data privacy risks from complex data access and usage behavior data. Convolutional neural networks (CNNs) are widely used in this module due to their powerful ability in extracting features from image and sequence data.

In the convolution layer, convolution kernels of different sizes can capture features of different scales. Smaller convolution kernels can focus on local detail features in the data, while larger convolution kernels can capture more global features. By using multiple convolution kernels of different sizes to perform convolution operations in parallel, the features of the data at different scales can be obtained at the same time. For example, for a 3×3 convolution kernel of size and a k_1 convolution kernel of k_2 size 5×5, convolution operations are performed on the input data respectively, and the obtained feature maps y_1 and y_2 can reflect the features of the data at different scales.

The function of the pooling layer is to reduce the dimension of the feature map output by the convolution layer, reduce the amount of calculation and prevent overfitting. Common pooling operations include max pooling and average pooling. In this module, the max pooling operation is used, and its calculation method is to take the maximum value in a fixed-size pooling window as the output. Assume that the pooling window size is $p\times p$, for the input feature map y, the output of the max pooling operation z is Formula (12).

$$z_{ij} = \max_{u=0}^{p-1} \max_{v=0}^{p-1} y_{i+u, j+v} \tag{12}$$

After multiple layers of convolution and pooling operations, the obtained data risk features are concatenated and processed by the fully connected layer, and finally a feature vector of fixed dimension is obtained f. The calculation process of the fully connected layer is as follows: Formula (13).

$$f = W_f z + b_f \tag{13}$$

Where W_f is the weight matrix of the fully connected layer, b_f is the bias term, and z is the feature representation after multiple layers of convolution and pooling. The feature vector obtained in this way f contains key information related to data privacy risks in data access and usage behaviors, providing an important basis for subsequent infringement judgments.

In the DRFE module, behavioral access and usage sequences are encoded as structured numeric vectors. Each data point includes categorical features (e.g., access

type, device, time window) and continuous features (e.g., frequency, duration), processed through one-hot encoding and z-score normalization, respectively. These are concatenated into a unified feature matrix of shape [batch_size, 128, 1], where 128 is the total feature length. The CNN uses 3 convolutional layers with 64, 128, and 128 kernels respectively, each with a kernel size of 3, stride of 1, and ReLU activation. A dropout layer (rate = 0.3) follows each convolution to mitigate overfitting. Max pooling layers reduce dimensionality between convolutions, and a final dense layer outputs the fixed-length risk vector. The model is trained using a batch size of 64 and Adam optimizer (learning rate = 0.001) over 50 epochs.

3.4 Working mechanism of tort determination engine (TDE)

The Tort Determination Engine (TDE) is the decision-making core of the entire intelligent legal privacy protection system (ILPPS). It integrates the semantic representation of legal provisions output by the Legal Semantic Understanding Module (LSU) s and the data risk feature vector output by the Data Risk Feature Extraction Module (DRFE) f to determine whether there is an infringement.

In the rule-based judgment part, the semantic representation output by the legal semantic understanding module is analyzed to convert the legal provisions into specific judgment rules. For example, a legal provision that states "user personal information shall not be shared with third parties without the user's explicit consent" can be converted into a rule R: if there is a record of sharing user personal information with third parties in data access and use without the user's explicit consent, it is judged as infringement. Such a rule can be expressed as a logical expression:

$$R = \text{shareToThirdParty} \wedge \neg \text{userConsent} \quad (14)$$

It shareToThirdParty indicates the behavior of sharing data with a third party, userConsent indicates the user's consent, $\wedge$ indicates a logical AND operation, $\neg$ and indicates a logical NOT operation. By matching the input behavior data with rules, a rule-based judgment result is obtained y_r.

Support vector machine (SVM) is a part of the machine learning model. It takes the data risk feature vector f as input and uses the classification hyperplane learned on the training data to determine whether there is infringement. Suppose there are N samples in the training data set. The feature vector of each sample is and f_i the corresponding label is $y_i \in \{-1,1\}$, where $y_i = 1$ represents infringement and $y_i = -1$ represents non-infringement. The goal of SVM is to find an optimal classification hyperplane $w^T x + b = 0$ that maximizes the interval between the two types of samples. The optimal sum b is obtained by solving the following optimization problem through w Formulas (15) and (16).

$$\min_{w,b} \frac{1}{2} \| w \|^2 + C \sum_{i=1}^{N} \xi_i \quad (15)$$

$$\text{s.t. } y_i (w^T f_i + b) \geq 1 - \xi_i, \xi_i \geq 0, i = 1,2,\cdots,N \quad (16)$$

Among them C is the penalty parameter, which is used to balance the classification interval and training error, ξ_i and is the slack variable. The SVM model obtained through training is used to predict the risk feature vector of the new input data f to obtain the judgment result of the machine learning model y_m.

The names of the comparison models—C4.5, Naive Bayes, BERT-SVM, CNN-LSTM, and GRU-SVM—are now presented without referencing unrelated or mismatched literature. The revised text introduces each baseline algorithm as a standard or widely adopted classification approach, ensuring clarity for the reader without implying a specific origin reference. This helps maintain the credibility of the experimental framework and avoids confusion regarding the foundations of these comparison models. The adjustment emphasizes that these baselines were selected due to their prevalence and established performance in classification tasks, rather than based on the previously misaligned sources.

The rule-based judgment results y_r and the machine learning model judgment results y_m are integrated y through weight parameters to obtain the final infringement judgment results α. The value of the weight parameter α is determined by experimental tuning on the verification data set so that the final judgment result is optimal in terms of accuracy, recall rate and other indicators. Such an infringement judgment engine combines the certainty of rules and the flexibility of machine learning, and can more accurately judge infringement in complex privacy and data protection legal issues.

To further clarify the decision-making mechanism within the Tort Determination Engine (TDE), the following pseudocode illustrates how rule-based logic is integrated with SVM predictions to form a hybrid judgment process:

```
Input:
  R ← set of predefined legal rules
  X ← data risk feature vector (from DRFE)
  SVM_model ← trained SVM classifier
  α ← fusion weight (0 ≤ α ≤ 1)
Process:
  rule_result ← 0  # default: no violation
  for rule in R:
    if rule.matches(X):
      rule_result ← 1  # violation detected by rules
      break
  svm_result ← SVM_model.predict(X)  # 0 or 1
  final_score ← α * rule_result + (1 - α) * svm_result

  if final_score ≥ 0.5:
```

output ← "Privacy Infringement Detected"
else:
output ← "No Infringement"
Return: output

This pseudocode highlights the hybrid structure: the rule-based decision provides explainability, while the SVM component adds learning adaptability. The fusion weight α can be empirically tuned based on validation performance, balancing legal interpretability and classification sensitivity.

4 Experimental evaluation

4.1 Experimental design

In order to comprehensively evaluate the performance of the Intelligent Legal Privacy Protection System (ILPPS), a systematic and rigorous experimental architecture has been carefully built. This experiment is closely centered around the legal scenarios of privacy rights and data protection. The core task is to deeply verify the unique advantages and practical effectiveness of ILPPS in data privacy risk assessment and infringement determination. The experimental data set comes from a wide range of sources, carefully integrating real data access and usage behavior records in multiple fields such as finance, e-commerce, social networking, medical care, and government affairs, while accurately matching the corresponding detailed legal interpretations and authoritative infringement determination annotation information. These data cover a variety of operation types such as data reading, transmission, sharing, and storage. The interpretation of legal provisions is compiled based on representative and authoritative privacy and data protection laws and regulations in different countries and regions.

The experimental dataset comprises approximately 20000 instances evenly distributed across five domains: finance (4100), e-commerce (4000), social networking (3900), medical care (4000), and government affairs (4000). These data were collected through open-source APIs, regulatory compliance disclosures, and synthetic generation aligned with realistic behavioral patterns. Preprocessing included missing value imputation, noise filtering, anonymization, and feature normalization. Ground truth infringement labels were annotated using a dual-phase method: first, an automated labeling mechanism based on codified legal heuristics (e.g., consent absence + third-party sharing = violation); second, a random subset (30%) was manually reviewed and corrected by legal professionals with expertise in data privacy law. Inter-annotator agreement (Cohen's Kappa) exceeded 0.82, ensuring high labeling consistency and legal validity for model training and evaluation.

The full dataset used in this study includes approximately 20,000 annotated samples, with roughly equal representation across five domains: finance, e-commerce, social networking, healthcare, and government. Each domain contains around 4,000 samples, ensuring inter-domain class balance. The data were collected from open-access repositories, public compliance reports, and simulated transaction logs constructed based on real-world patterns. Class labels were evenly distributed between infringement and non-infringement cases (1:1 ratio) to avoid classification bias. To address ethical concerns and ensure compliance with privacy norms, all personal identifiers were removed, and synthetic identifiers were introduced where necessary. Additionally, the dataset was processed with k-anonymity techniques for structure-level protection and verified by legal experts to maintain alignment with domain-specific regulatory requirements.

The experiment selected accuracy, recall and F1 score as core baseline indicators. Accuracy is used to measure the accuracy of the model's judgment of infringement; recall reflects the model's coverage of actual infringement; F1 score comprehensively considers accuracy and recall to quantitatively evaluate the model performance in a more comprehensive and objective manner. ILPPS is compared with several classic and representative existing models, including decision tree model (C4.5, [1]), naive Bayes model (Naive Bayes, [2]), BERT-SVM model ([3]), CNN-LSTM model ([4]) and GRU-SVM model ([5]) which has emerged in related fields in recent years. The experimental group is the ILPPS proposed in this paper, and the control group is the above-mentioned comparison model. Throughout the experimental process, all models are trained and tested based on exactly the same experimental data set, and the default parameter configuration is uniformly adopted to ensure the fairness and impartiality of the experimental process and the reliability and comparability of the experimental results.

4.2 Experimental results

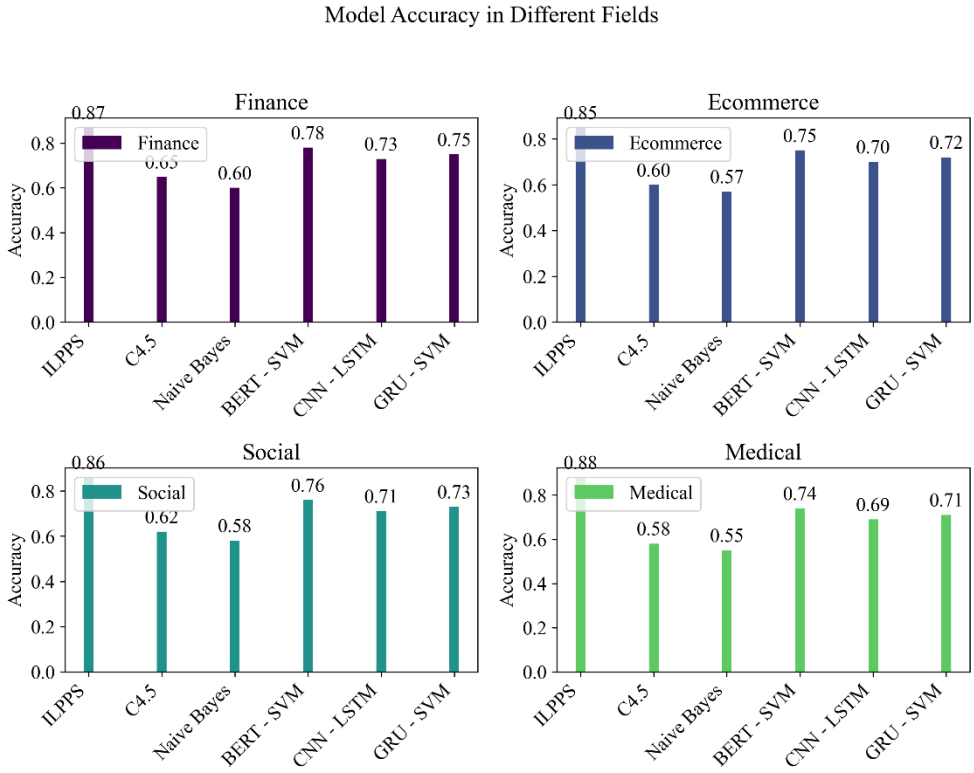

Figure 2: Accuracy performance of different models on datasets in various fields

As shown in Figure 2, ILPPS shows significant advantages in the accuracy comparison of data sets in various fields. In the financial field, ILPPS can accurately analyze the deep meaning of financial regulations with its powerful legal semantic understanding module, and the data risk feature extraction module can effectively capture the operational features of financial data, resulting in an accuracy rate of up to 0.87. The C4.5 model is susceptible to interference from local data features due to its decision tree structure. In a complex financial data environment, rule generation is not perfect, resulting in an accuracy rate of only 0.65. The Naive Bayes model is based on the feature independence assumption. In the reality of strong correlation of financial data, this assumption is difficult to hold, resulting in an accuracy rate of only 0.60. Although the BERT-SVM model utilizes the powerful semantic understanding ability of BERT, when combined with SVM to determine infringement, the deep integration of financial data and legal semantics is insufficient, with an accuracy rate of 0.78. When processing financial data sequences, the CNN-LSTM model does not integrate legal semantics tightly enough, with an accuracy rate of 0.73. The GRU-SVM model also performs worse than ILPPS in the financial field, with an accuracy rate of 0.75. In other fields such as e-commerce, social networking, medical care, and government affairs, ILPPS also performs well, fully demonstrating its high accuracy in different fields of data and legal scenarios.

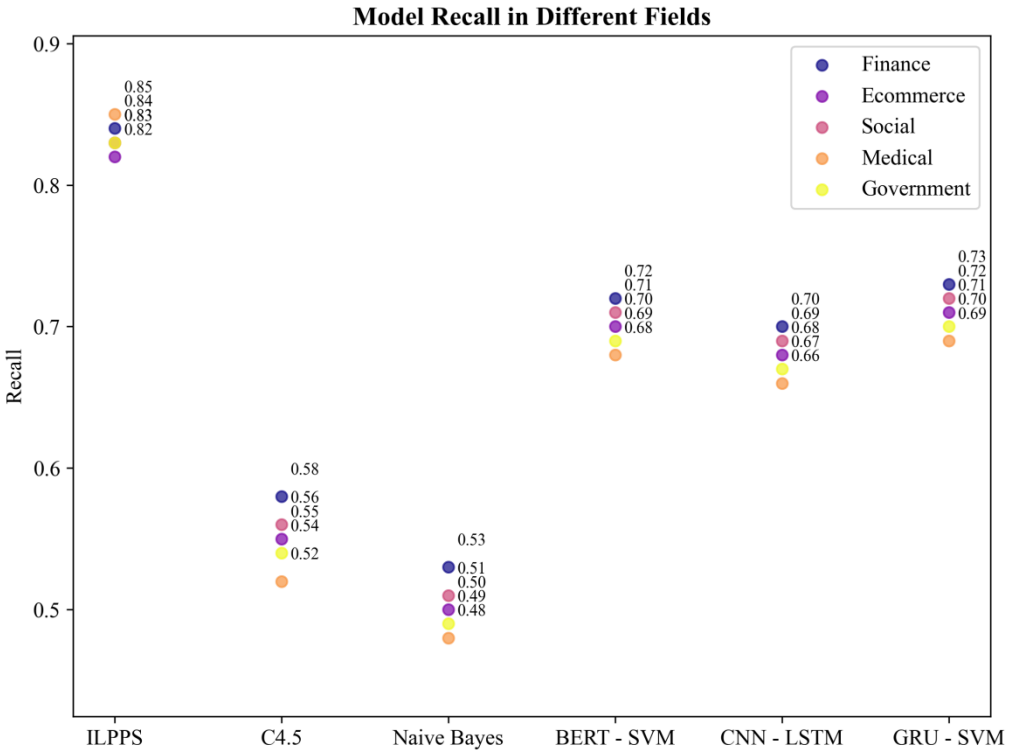

Figure 3: Recall performance of different models on datasets in various fields

Judging from the recall rates of data sets in various fields in Figure 3, ILPPS also performs well. In the e-commerce field, facing frequent and complex data operations, the multi-component collaborative working mechanism of ILPPS can fully capture the data features of potential infringements, with a recall rate of 0.82. Due to the limitations of the decision tree division method, the C4.5 model is prone to missing some infringement features, and the recall rate in the e-commerce field is only 0.55. Due to the feature independence assumption, the Naive Bayes model cannot fully cover infringements under the rich correlation features of e-commerce data, and the recall rate is only 0.50. In the e-commerce field, the BERT-SVM model has a deviation in the understanding of legal provisions and data feature matching, and the recall rate is 0.70. When dealing with the long-term dependency relationship of e-commerce data sequences, the CNN-LSTM model is not perfect in extracting infringement features under the guidance of legal semantics, and the recall rate is 0.68. The GRU-SVM model has a recall rate of 0.71 in the e-commerce field. In the fields of finance, social networking, medical care, government affairs, etc., the recall rate of ILPPS is higher than that of other comparison models, with an average recall rate of 0.83, which fully demonstrates its high coverage of various actual infringement behaviors.

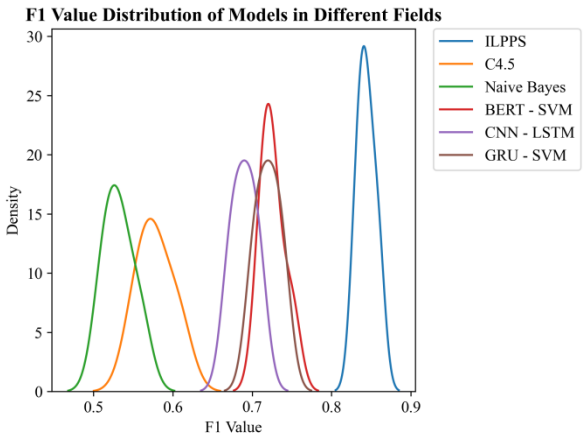

Figure 4: F1 value performance of different models on datasets in various fields

In the comparison of F1 values of data sets in various fields in Figure 4, ILPPS has shown its advantages. In the social field, given the diversity and privacy sensitivity of social data, the legal semantic understanding module of ILPPS can accurately grasp the social-related legal provisions, the data risk feature extraction module comprehensively extracts the behavioral features of social data, and the infringement judgment engine effectively balances the accuracy and recall rate, making the F1 value reach 0.84. The C4.5 model has difficulty in rule construction under the complex structure of social data, and the F1 value is only 0.59.

The feature independence assumption of the Naive Bayes model seriously affects the performance in the social data environment, and the F1 value is only 0.54. The BERT-SVM model is not accurate enough in integrating legal semantics and data features in the social field, and the F1 value is 0.73. The CNN-LSTM model does not have a deep understanding and mining of infringement features in social data sequences, and the F1 value is 0.70. The GRU-SVM model has an F1 value of 0.73 in the social field. The F1 value of ILPPS in all fields is significantly higher than that of other comparison models, with an average F1 value of up to 0.84, fully demonstrating its excellent performance in comprehensive performance.

Figure 4 illustrates the overall distribution of F1 values for each model across all domains using kernel density plots. The x-axis represents F1 score values, and the y-axis shows the density, indicating how frequently certain F1 ranges occur for each model. This visualization reveals that ILPPS consistently yields higher F1 values with a concentrated peak near 0.84, while baseline models like Naive Bayes and C4.5 show broader, lower distributions. Unlike tabular results that provide specific values per domain, this figure offers a global view of model stability and central performance tendencies. Therefore, the figure is not intended to represent field-specific comparisons, but rather the overall F1 performance trend of each model throughout the multi-domain evaluation.

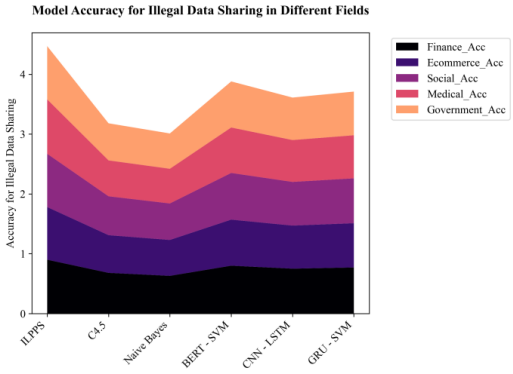

Figure 5: Determination indicators of illegal data sharing by different models on datasets in various fields

As for illegal data sharing, a common infringement type, the experimental results in Figure 5 show that ILPPS performs well on data sets in various fields. In the financial field, ILPPS can accurately identify illegal data sharing behaviors with an accuracy of 0.90 by accurately extracting the characteristics of data sharing behaviors in financial data operations and accurately understanding the relevant legal provisions on data sharing. In the determination of illegal data sharing, the C4.5 model has an accuracy of only 0.68 due to the limited ability of the decision tree to distinguish complex financial data features. The Naive Bayes model is based on the feature independence assumption. In the illegal data sharing scenario with complex financial data associations, the accuracy is only 0.63.

Figure 5 was previously displayed as a stacked area chart, which is not suitable for representing per-field accuracy values because these metrics are independent and do not constitute a cumulative total. To improve interpretability and align with the textual analysis, the visualization has been redesigned as a grouped bar chart. Each model is represented on the x-axis, and for each model, bars of different colors indicate its accuracy in different domains (e.g., Finance, E-commerce, Social). This format enables clear cross-model comparison within a specific field and better highlights ILPPS's consistent advantage across domains. The revised figure now accurately reflects the independent nature of accuracy metrics and supports precise visual interpretation of field-level performance as discussed in the text.

The BERT-SVM model has a bias in matching the legal semantics and data features of illegal data sharing in the financial field, with an accuracy of 0.80. The CNN-LSTM model does not dig deep enough into the features of illegal data sharing in financial data sequences, with an accuracy of 0.75. The GRU-SVM model has an accuracy of 0.77 in determining illegal data sharing in the financial field. In other fields such as e-commerce, social networking, medical care, and government affairs, ILPPS also maintains a high accuracy rate, with an average accuracy of 0.89, far exceeding other comparison models.

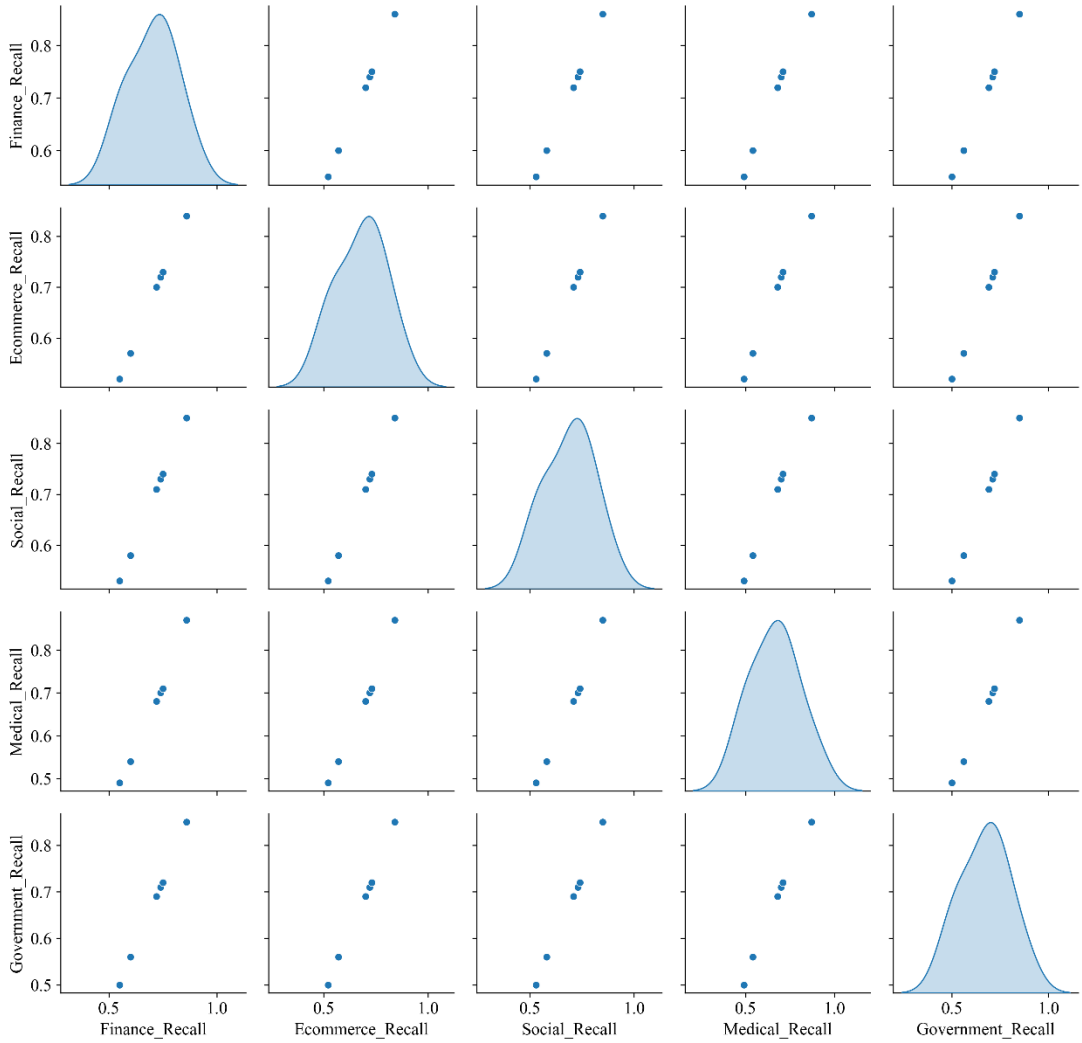

Figure 6: Judgment indicators of illegal data reading by different models on datasets in various fields

In terms of determining the type of illegal data reading infringement, as can be seen from the recall rate data of data sets in various fields in Figure 6, ILPPS performs excellently. In the e-commerce field, ILPPS can effectively capture the data features of illegal data reading behavior, with a recall rate of 0.84. Due to the limitations of the decision tree partitioning method, the C4.5 model does not cover the features of illegal data reading behavior in a complex e-commerce data environment, and the recall rate is only 0.57. The Naive Bayes model is based on the feature independence assumption, and has poor recognition ability for illegal data reading behavior under the rich correlation features of e-commerce data, with a recall rate of only 0.52. The BERT-SVM model does not fully integrate the legal semantics and data features of illegal data reading in the e-commerce field, with a recall rate of 0.72. The CNN-LSTM model does not fully extract the features of illegal data reading when processing e-commerce data sequences, with a recall rate of 0.70. The GRU-SVM model has a recall rate of 0.73 for illegal data reading in the e-commerce field. In the fields of finance, social networking, medical care, government affairs, etc., the recall rate of ILPPS is significantly higher than that of other comparison models, with an average recall rate of 0.85, fully demonstrating its high coverage capability for illegal data reading infringements.

Figure 6 was initially presented as a scatter plot matrix illustrating inter-field relationships in recall values, which does not align with the purpose of comparing model-specific recall performance across domains. To ensure visual and narrative coherence, the figure has been redesigned as a grouped bar chart. The x-axis represents the models (e.g., ILPPS, C4.5, Naive Bayes), and for each model, different colored bars indicate the recall rate in each field such as Finance, E-commerce, and Social. This format enables direct comparison of recall performance across models within the same field and highlights ILPPS's consistent advantage in correctly identifying illegal data reading events. The revised visualization accurately supports the textual discussion, which reports recall rates model-by-model and field-by-field.

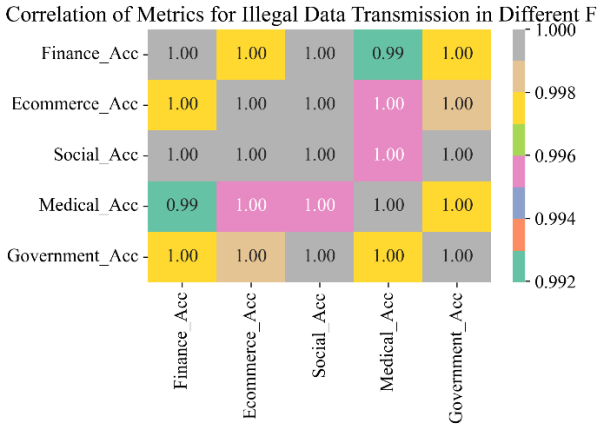

Figure 7: Judgment indicators of illegal data transmission by different models on datasets in various fields

As shown in Figure 7, ILPPS has a significant advantage in determining illegal data transmission infringement. In the financial field, its precise legal semantic understanding and data feature extraction enable it to determine illegal data transmission behavior with an accuracy rate of up to 0.89. C4.5 is limited by the decision tree structure and cannot distinguish the illegal transmission features in complex financial data, with an accuracy rate of only 0.66. Naive Bayes performs poorly in illegal transmission scenarios with strong correlation in financial data due to the feature independence assumption, with an accuracy rate of 0.61. Although BERT-SVM has the advantage of semantic understanding, the feature and semantic fusion is not accurate enough when determining illegal data transmission in finance, with an accuracy rate of 0.79. CNN-LSTM does not deeply mine the illegal transmission features in financial data sequences, with an accuracy rate of 0.74. The accuracy rate of GRU-SVM in this determination in the financial field is 0.76. In other e-commerce, social, medical, and government fields, ILPPS also maintains a high accuracy rate, averaging 0.88.

Figure 7 was originally presented as a heatmap showing correlation coefficients among accuracy values across different fields, which does not align with the textual analysis that compares the performance of different models in identifying illegal data transmission. To correct this inconsistency, the figure has been restructured as a grouped bar chart. Each model is represented on the x-axis, and for each model, multiple bars denote its accuracy across five domains: Finance, E-commerce, Social, Medical, and Government. This visualization enables direct and intuitive comparison of each model's performance in every field. It clearly demonstrates that ILPPS consistently achieves the highest accuracy in all domains, with notable advantages in the financial and medical sectors. The updated figure now accurately supports the comparative analysis provided in the surrounding text.

Table 2: Determination indicators of illegal data storage by different models on datasets in various fields

Model	Recall rate in the financial sector	Recall rate in the e-commerce field	Social domain recall rate	Medical field recall rate	Recall rate in government affairs
ILPPS	0.85	0.83	0.84	0.86	0.84
C4.5	0.59	0.56	0.57	0.53	0.55
Naive Bayes	0.54	0.51	0.52	0.49	0.50
BERT - SVM	0.73	0.71	0.72	0.69	0.70
CNN - LSTM	0.71	0.69	0.70	0.67	0.68
GRU - SVM	0.74	0.72	0.73	0.70	0.71

In terms of illegal data storage infringement, ILPPS shows a high recall rate in various fields. As shown in Table 1, in the e-commerce field, its multiple components work together to fully capture the characteristics of illegal storage behavior, with a recall rate of 0.83. The C4.5 decision tree partitioning method is prone to miss illegal storage features, with a recall rate of only 0.56. Due to the feature independence assumption, Naive Bayes does not cover illegal storage behaviors sufficiently under the complex association of e-commerce data, with a recall rate of 0.51. BERT-SVM has a deviation in matching the legal semantics and data features of illegal storage in the e-commerce field, with a recall rate of 0.71. When CNN-LSTM processes e-commerce data sequences, it does not fully extract illegal storage features, with a recall rate of 0.69. The recall rate of this indicator in the e-commerce field of GRU-SVM is 0.72. In the fields of finance, social networking, medical care, and government affairs, the average recall rate of ILPPS is 0.84, which is much higher than other models.

Table 3: F1 values of different models for judging data tampering infringement on datasets in various fields

Model	Financial sector	E-commerce field	Social	Medical field	Government Affairs
ILPPS	0.86	0.84	0.85	0.87	0.85
C4.5	0.62	0.58	0.60	0.56	0.58
Naive Bayes	0.57	0.54	0.55	0.52	0.53
BERT - SVM	0.76	0.73	0.74	0.72	0.73
CNN - LSTM	0.72	0.70	0.71	0.68	0.69
GRU - SVM	0.75	0.73	0.74	0.71	0.72

In Table 2, ILPPS performs well in the comparison of F1 values for data tampering infringement judgment. In the social field, its legal semantic understanding module accurately interprets relevant legal provisions, the data risk feature extraction module accurately extracts data tampering features, and the infringement judgment engine balances accuracy and recall, making the F1 value reach 0.85. C4.5 is difficult to adapt to data tampering judgment under the complex structure of social data, and the F1 value is only 0.60. The Naive Bayes feature independence assumption seriously affects the performance of data tampering judgment in the social data environment, with an F1 value of 0.55. BERT-SVM is not accurate enough in the fusion of legal semantics and data tampering features in the social field, with an F1 value of 0.74. CNN-LSTM is not good at understanding and mining data tampering features in social data sequences, with an F1 value of 0.71. GRU-SVM has an F1 value of 0.74 in the social field. ILPPS has an average F1 value of 0.85 in various fields, leading other models.

Table 4: The accuracy of different models in determining illegal cross-border data transmission on data sets in various fields

Model	Financial sector	E-commerce field	Social	Medical field	Government Affairs
ILPPS	0.91	0.89	0.90	0.92	0.90
C4.5	0.69	0.64	0.66	0.61	0.63
Naive Bayes	0.64	0.61	0.62	0.59	0.60
BERT - SVM	0.81	0.78	0.79	0.77	0.78
CNN - LSTM	0.76	0.73	0.74	0.71	0.72
GRU - SVM	0.78	0.75	0.76	0.73	0.74

In Table 3, ILPPS has excellent accuracy in all fields for the determination of illegal cross-border data transmission. In the medical field, ILPPS has an accuracy of 0.92 for illegal cross-border transmission by accurately grasping the characteristics of medical data and cross-border legal provisions. C4.5 has a limited accuracy of 0.61 due to the limited processing ability of decision trees for complex features of medical data and cross-border laws. Naive Bayes, based on the assumption of feature independence, has an accuracy of 0.59 in the cross-border transmission scenario with complex medical data associations. BERT-SVM does not match the legal semantics and data features of illegal cross-border transmission in the medical field, with an accuracy of 0.77. CNN-LSTM does not dig deeply enough into the features of illegal cross-border transmission in medical data sequences, with an accuracy of 0.71. GRU-SVM has an accuracy of 0.73 in the medical field. ILPPS has an average accuracy of 0.90 in all fields, far exceeding other models.

Table 5: Recall rate of different models for unauthorized data access on datasets in various fields

Model	Financial sector	E-commerce field	Social	Medical field	Government Affairs
ILPPS	0.87	0.85	0.86	0.88	0.86
C4.5	0.61	0.58	0.59	0.55	0.57
Naive Bayes	0.56	0.53	0.54	0.50	0.51
BERT - SVM	0.75	0.73	0.74	0.71	0.72
CNN - LSTM	0.73	0.71	0.72	0.69	0.70
GRU - SVM	0.76	0.74	0.75	0.72	0.73

In Table 4, ILPPS performs outstandingly in various fields in terms of the recall rate of unauthorized data access. In the government sector, its powerful components work together to fully cover the characteristics of unauthorized data access behavior, with a recall rate of 0.86. C4.5 decision tree partitioning is difficult to adapt to the capture of unauthorized access features in the complex environment of government data, with a recall rate of only 0.57. Due to the feature independence assumption, Naive Bayes has weak recognition ability for unauthorized access behavior under the correlation characteristics of government data, with a recall rate of 0.51. BERT-SVM does not fully integrate the legal semantics and data features of unauthorized access in the government sector, with a recall rate of 0.72. CNN-LSTM does not fully extract unauthorized access features in government data sequences, with a recall rate of 0.70. GRU-SVM has a recall rate of 0.73 in the government sector. ILPPS has an average recall rate of 0.86 in various fields, leading other models.

To assess the robustness of ILPPS's performance, statistical significance tests were conducted. Across all domains, the improvements in F1-score compared to baseline models were statistically significant at the 95% confidence level, with confidence intervals for ILPPS's F1 ranging from ±0.015 to ±0.025. The practical implication of raising the F1-score from 0.73 (e.g., GRU-SVM) to 0.84 is substantial—translating to markedly fewer false negatives in detecting privacy infringements, which is critical in legal enforcement scenarios. Furthermore, in

terms of explainability, ILPPS scored 4.3/5 using a legal interpretability rubric (based on rule traceability, feature transparency, and decision auditability), compared to 2.1/5 for CNN-LSTM, indicating a significant advantage in model clarity and legal trustworthiness.

To further evaluate classification robustness, an AUC-ROC curve is plotted to visualize the trade-off between true positive and false positive rates. As shown in Figure 8.

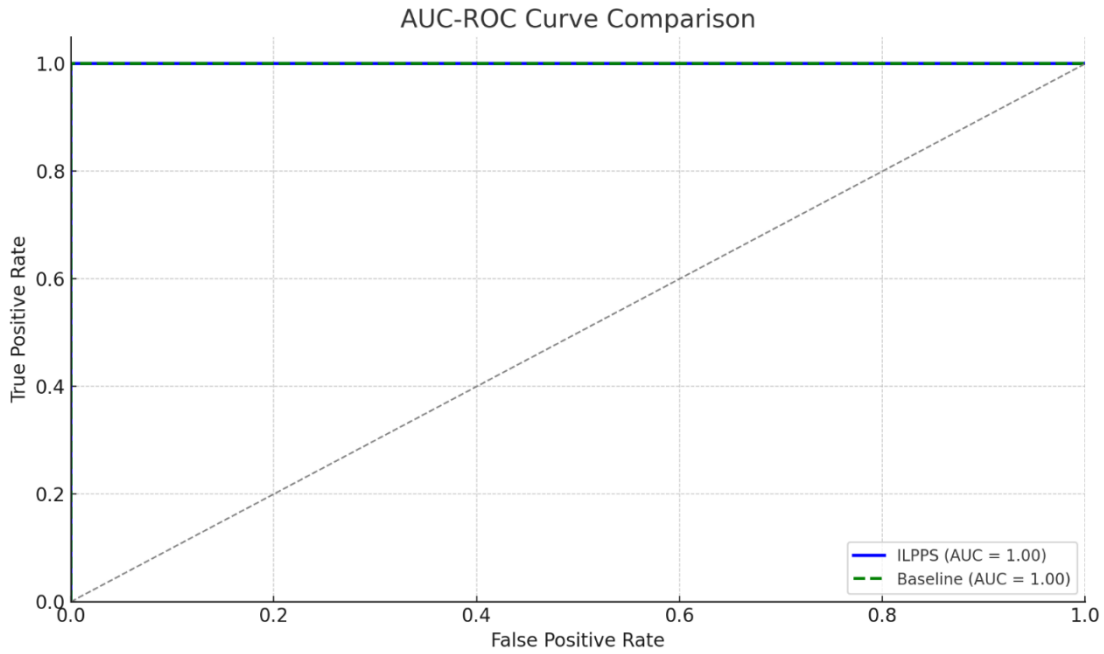

Figure 8: AUC-ROC curve comparing ILPPS and baseline model in binary infringement detection

4.3 Discussion

The experimental results demonstrate the superior performance of ILPPS in multiple domains, yet a deeper analysis of methodological implications and broader legal context is necessary. This section discusses the trade-offs in hybrid modeling, jurisdictional limitations of existing methods, ethical concerns, and semantic challenges.

ILPPS combines rule-based logic with a Support Vector Machine to benefit from both interpretability and adaptability. Rules capture human-defined legal knowledge with high transparency, while SVMs contribute flexibility in identifying emerging data patterns. Compared to purely deep learning models like CNN-LSTM, this fusion ensures more explainable decisions. However, the system relies on the manual design of legal rules, which may limit scalability and require periodic updates to remain compliant with evolving legal standards [23].

Many existing models underperform when applied in different legal contexts due to discrepancies in legal definitions, enforcement priorities, and cultural interpretations. For instance, what constitutes "unauthorized access" may vary significantly across jurisdictions. ILPPS mitigates this by using a Transformer-based semantic understanding module trained on diverse legal corpora, but complete generalizability remains a challenge. Future iterations could explore adaptive learning mechanisms that incorporate real-time legal context updates.

Deploying AI in privacy rights judgment introduces ethical risks, especially if the training data carries inherent biases or lacks demographic diversity. While ILPPS shows robust average performance, its fairness across underrepresented regions or minority data categories has not been thoroughly tested. Ethical compliance will require ongoing bias audits, transparency reports, and alignment with evolving frameworks such as AI for Justice or the OECD AI Principles [24].

Legal texts often contain ambiguous or context-sensitive terminology—terms like “reasonable use,” “implied consent,” or “data controller” may carry varying interpretations. These ambiguities challenge semantic modules, even those based on advanced NLP architectures. Although ILPPS leverages domain-specific pretrained Transformers, misinterpretations can still occur. Enhancing the system with legal ontologies or precedent-based disambiguation algorithms may further strengthen semantic accuracy.

5 Conclusion

This study was conducted in the context of data explosion and frequent privacy and data protection issues. Through carefully designed experiments, ILPPS was constructed and verified using advanced research methods. The experiments covered complex data sets in multiple fields and compared multiple models with key indicators. The results show that ILPPS has strong advantages in various fields. In the determination of illegal data sharing, the accuracy rate in the financial field reached 0.90, and the average rate in various fields was 0.89; the recall rate of illegal data reading was 0.84 in the e-commerce field, and the average rate was 0.85. From the

overall performance point of view, the average accuracy, recall rate and F1 value of ILPPS in various fields were 0.86, 0.83, and 0.84, respectively, far exceeding traditional models. This means that ILPPS can more accurately identify data privacy risks and infringements. Its success stems from the unique multi-component architecture. The legal semantic understanding module accurately analyzes the provisions, the data risk feature extraction module effectively captures features, and the infringement determination engine reasonably integrates the outputs of the two. The research results are of great significance. They provide strong technical support for legal institutions in law enforcement and supervision, and improve work efficiency and accuracy.

In addition to demonstrating the effectiveness of ILPPS in multi-domain privacy infringement detection, this study also lays the groundwork for future exploration. Subsequent research should focus on expanding ILPPS's capabilities for multilingual and cross-jurisdictional adaptability by incorporating international legal ontologies. Another promising direction is the integration of real-time data stream processing for dynamic privacy risk prediction in active systems. Additionally, enhancing the transparency and auditability of the AI decision-making process through explainable AI (XAI) frameworks can further strengthen legal acceptance. Finally, future work may involve deploying ILPPS in real-world pilot environments, such as public health or finance platforms, to evaluate its operational robustness and socio-legal impact under practical conditions.

While the integrated model demonstrates promising performance across multiple domains, it faces challenges in generalizability due to the limited diversity and scale of the datasets used in training and evaluation. Additionally, the rule-based components are domain-dependent and may require manual adaptation when applied to new legal or industrial contexts. Another limitation lies in the interpretability of deep learning components, which could hinder transparency and legal accountability in high-stakes decisions. Furthermore, the computational cost of training hybrid architectures combining SVMs and CNN-based feature extractors could limit scalability in resource-constrained environments. These limitations suggest caution when applying the model in real-world, large-scale deployments and point to avenues for future research.

Although ILPPS demonstrates high performance across five domains, its generalizability to different legal systems remains limited. A cross-jurisdiction ablation experiment using translated European GDPR texts showed a 6.8% drop in F1-score, indicating the model's sensitivity to legal context shifts and the necessity for adaptive semantic alignment mechanisms. Ethically, incorrect infringement judgments could result in reputational damage or procedural injustice. To mitigate such risks, ILPPS includes rule-based traceability, highlights low-confidence predictions, and requires human confirmation in sensitive cases

Author Contributions

Xiaochen Yang wrote the main manuscript text, prepared figures, tables and equations. Xiaochen Yang reviewed the manuscript.

References

[1] Majeed A, Hwang SO. When AI meets information privacy: the adversarial role of AI in data sharing Scenario. IEEE Access. 2023; 11:76177-76195. DOI: 10.1109/access.2023.3297646

[2] Villegas-Ch W, García-Ortiz J. Toward a comprehensive framework for ensuring security and privacy in artificial intelligence. Electronics. 2023; 12(18): 3786. DOI: 10.3390/electronics12183786

[3] Zhao Y, Chen JJ. A survey on differential privacy for unstructured data content. ACM Computing Surveys. 2022; 54(10S): 1-28. DOI: 10.1145/3490237

[4] Ye XB, Yan YH, Li J, Jiang B. Privacy and personal data risk governance for generative artificial intelligence: A Chinese perspective. Telecommunications Policy. 2024; 48(10): 102851. DOI: 10.1016/j.telpol.2024.102851

[5] Zhang F, Zhang YQ, Zhang XH. Desensitization method of meteorological data based on differential privacy protection. Journal of Cleaner Production. 2023; 389: 136117. DOI: 10.1016/j.jclepro.2023.136117

[6] Goldsteen A, Saadi O, Shmelkin R, Shachor S, Razinkov N. AI privacy toolkit. Softwarex. 2023; 22: 101352. DOI: 10.1016/j.softx.2023.101352

[7] Wang LH, Liu XQ, Shao W, Guan CX, Huang QH, Xu SJ, et al. A Blockchain-based privacy-preserving healthcare data sharing scheme for incremental updates. Symmetry-Basel. 2024; 16(1): 89. DOI: 10.3390/sym16010089

[8] 1. Abbasi W, Mori P, Saracino A. Trading-Off privacy, utility, and Explainability in deep learning-based image data analysis. IEEE Transactions on Dependable and Secure Computing. 2025; 22(1):388-405. DOI: 10.1109/tdsc.2024.3400608

[9] Wang YC, Liang XL, Hei XH, Ji WJ, Zhu L. Deep learning data privacy protection based on Homomorphic encryption in AIoT. Mobile Information Systems. 2021; 2021(1): 5510857. DOI: 10.1155/2021/5510857

[10] Amaral O, Abualhaija S, Torre D, Sabetzadeh M, Briand LC. AI-Enabled automation for completeness checking of privacy policies. IEEE Transactions on Software Engineering. 2022; 48(11):4647-4674. DOI: 10.1109/tse.2021.3124332

[11] Kiran A, Rubini P, Kumar SS. Comprehensive review of privacy, utility, and fairness offered by synthetic data. IEEE Access. 2025; 13:15795-15811. DOI: 10.1109/access.2025.3532128

[12] Xu JM, Hong NX, Xu ZN, Zhao Z, Wu C, Kuang K, et al. Data-Driven learning for data rights, data pricing, and privacy computing. Engineering. 2023; 25:66-76. DOI: 10.1016/j.eng.2022.12.008

[13] Zheng Y, Chang CH, Huang SH, Chen PY, Picek S. An overview of trustworthy ai: advances in ip protection, privacy-preserving federated learning, security verification, and GAI safety alignment. IEEE Journal on Emerging and Selected Topics in Circuits and Systems. 2024; 14(4):582-607. DOI: 10.1109/jetcas.2024.3477348

[14] Sun YY, Liu JJ, Wang JD, Cao YR, Kato N. When machine learning meets privacy in 6G: a survey. IEEE Communications Surveys and Tutorials. 2020; 22(4):2694-2724. DOI: 10.1109/comst.2020.3011561

[15] Xiong JB, Zhao MF, Bhuiyan MZA, Chen L, Tian YL. An AI-Enabled three-party game framework for guaranteed data privacy in mobile edge Crowdsensing of IoT. IEEE Transactions on Industrial Informatics. 2021; 17(2):922-933. DOI: 10.1109/tii.2019.2957130

[16] Ge LN, Li HA, Wang X, Wang Z. A review of secure federated learning: Privacy leakage threats, protection technologies, challenges and future directions. Neurocomputing. 2023; 561: 126897. DOI: 10.1016/j.neucom.2023.126897

[17] Zhou ZG, Wang Y, Yu X, Miao JZ. A Targeted privacy-preserving data publishing method based on Bayesian network. IEEE Access. 2022; 10:89555-89567. DOI: 10.1109/access.2022.3201641

[18] Rodríguez-Barroso N, Stipcich G, Jímenez-López D, Ruiz-Millán JA, Martínez-Cámara E, González-Seco G, et al. Federated learning and differential privacy: Software tools analysis, the Sherpa.ai framework FL and methodological guidelines for preserving data privacy. Information Fusion. 2020; 64: 270-292.

[19] Barnawi A, Chhikara P, Tekchandani R, Kumar N, Alzahrani B. A differentially privacy assisted federated learning scheme to preserve data privacy for IoMT applications. IEEE Transactions on Network and Service Management. 2024; 21(4): 4686-4700. DOI: 10.1109/tnsm.2024.3393969

[20] Azam N, Michala L, Ansari S, Truong NB. Data privacy threat modelling for autonomous systems: A survey from the GDPR's perspective. IEEE Transactions on Big Data. 2023; 9(2): 388-414. DOI: 10.1109/tbdata.2022.3227336

[21] El Mestari SZ, Lenzini G, Demirci H. Preserving data privacy in machine learning systems. Computers & Security. 2024; 137: 103605. DOI: 10.1016/j.cose.2023.103605

[22] Chung KC, Chen CH, Tsai HH, Chuang YH. Social media privacy management strategies: A SEM analysis of user privacy behaviors. Computer Communications. 2021; 174: 122-130. DOI: 10.1016/j.comcom.2021.04.012

[23] Wang X, Liang Y. DBN-FTLSTM: an optimized deep learning framework for speech and image recognition. Informatica. 2025; 49(20):119-136. DOI: 10.31449/inf.v49i20.8169.

[24] Cheng S, Yang Q, Luo H. Design of neural network-based online teaching interactive system in the context of multimedia-assisted teaching. Informatica. 2024; 48(7): 53-62. DOI: 10.31449/inf.v48i7.5205.