
COSE474-2024F: Final Project Proposal

“AI Lease Fraud Defense System”

Donghyun Ahn, Department of Data Science, Korea University

1. Introduction

As a student majoring in Law and Data Science, I am dedicated to making a positive impact on society. With the prevalence of lease fraud on the rise, it is crucial to empower tenants with the knowledge and tools to protect themselves. This AI-Based program aims to provide a comprehensive and user-friendly resource for individuals seeking to navigate the rental process safely. By raising awareness and equipping potential renters with essential information, we can effectively reduce the risk of fraud and create a more secure rental environment for everyone. .

2. Problem definition & challenges

Problem Definition:

In recent years, the incidence of lease fraud has been increasing, causing significant financial losses and distress for many individuals seeking rental housing. According to recent data, over 16,000 people nationwide have suffered approximately 2.5 trillion won in damages due to lease fraud over the past two years. This crime burdens individuals with debts amounting to hundreds of millions of won. Lease fraud in Korea manifests in a variety of highly specialized forms. In particular, 35.3% of lease fraud cases involve the abuse of the financial institution's lease loan system, where fraudsters illegally obtain lease deposits through false guarantees or insurance. Additionally, various forms exist, such as zero-capital gap investments (24.0%), which systematically collect deposits and referral fees, and illegal brokerage and appraisal practices (18.9%). There are also frequent instances of lease fraud where individuals are unable to recover their deposits due to encumbrances, tax delinquencies, or excessively high lease prices compared to purchase prices.

These issues heighten the risk of tenants falling victim to fraud due to a lack of access to information and awareness. My goal is to develop a comprehensive program that provides citizens, who are unfamiliar with lease contracts, with the knowledge and tools needed to systematically identify and prevent lease fraud, thereby creating a safer rental environment. Additionally, I aim to establish a public benefit

system that simplifies complex lease agreements and offers free consultations to individuals affected by this issue.

Challenges:

- Dataset Construction:** Legal matters require specialized AI capabilities. My ultimate goal is to create a model that allows individuals who lack knowledge of law and contracts to articulate their situations and obtain the necessary information, just like consulting a real expert. This necessitates the construction of a diverse and legally specific dataset.
- Complexity of Lease Fraud:** The tactics employed by fraudsters continuously evolve, making it challenging to keep the checklist up to date. The dynamic nature of lease fraud requires ongoing adaptation to emerging trends and continuous updates.
- Diverse Rental Markets:** Rental practices and regulations vary significantly across regions, making it difficult to create a universal checklist. I plan to develop a prototype primarily focused on Seoul or specific areas within Seoul.
- User Convenience:** My ideal outcome is to implement a model that guides users step-by-step in a manner similar to a real expert. Therefore, if feasible, I aim to develop an intuitive and user-friendly interface for the program.

3. Related Works

Yufeng Kou, Chang-Tien Lu, S. Sirwongwattana and Yo-Ping Huang, "Survey of fraud detection techniques," IEEE International Conference on Networking, Sensing and Control, 2004, Taipei

Hwang, Se Eun, Jang, H, S (2023). Analysis and Solution of the lease of a house on a deposit basis fraud by type. RESIDENTIAL ENVIRONMENT : JOURNAL OF THE RESIDENTIAL ENVIRONMENT INSTITUTE OF KOREA, 21(1), 21-36.

4. Datasets

Given the limitations of existing datasets, I plan to create a custom dataset tailored to the unique characteristics of lease fraud in Korea. To achieve this, I will utilize a Retrieval-Augmented Generation (RAG) system, which will help aggregate and retrieve relevant information effectively.

1. **Diverse Data Sources:** The dataset will be constructed from various sources, including:
 - Types of lease fraud and documented case studies.
 - Lists of essential checks and precautions for preventing lease fraud.
 - Expert opinions and insights on identifying fraudulent activities.
 - Other Pictures and documents about lease fraud.
2. **RAG System Implementation:** The RAG system will serve as a core component of the dataset-building process.
 - Aggregation of data from diverse sources to create a comprehensive knowledge base on lease fraud types and prevention strategies.
 - Real-time retrieval of relevant information, ensuring that the system remains updated with the latest insights and expert opinions.
 - Contextual generation of recommendations and guidelines for users based on retrieved data.

5. State-of-the-art methods and baselines

Starting this year, the government has employed artificial intelligence (AI) to investigate suspected lease fraud transactions, resulting in a 2.3-fold increase in notifications to law enforcement agencies. However, there remains a significant gap in implementing fundamental preventive measures that utilize AI to provide tenants with convenient and useful information. To address this, I plan to develop a Retrieval-Augmented Generation (RAG) system based on foundational generative models such as GPT or Llama, enhanced by BERT and various information processing techniques. This model will be trained to create a user-friendly platform for tenants to easily access relevant information. Additionally, having recently taken a course on time series analysis, I see an opportunity to explore the use of LSTM for analyzing trends in lease price changes.

6. Schedule

1st Week of November: Meeting with professors regarding Deep learning, transfer learning, criminal law, and criminal procedure law, and extracting necessary information.

2nd Week of November: Basic system design and construction.

3rd Week of November: Building and integrating the RAG system, and conducting final analysis.