A neurosymbolic approach to Fraud Detection on financial data in the public administration.

Michele Vitale[1]

[1]Department of Matematics and Computer Science, DEMACS, University of Calabria Rende, Italy

Abstract

This doctoral research, in collaboration with Invitalia, the Italian National Development Agency, focuses on using a neurosymbolic approach to detect fraud in public administration. The main goal is to create a decision support system that can both identify past fraudulent activities and proactively flag suspicious requests for human review. The project addresses the challenges of limited labeled data and the need for interpretability in AI models, which is crucial in the public investment field. The initial phase involved creating a data lake from public funding documents using Large Language Models (LLMs). This data will be used to construct a knowledge graph to identify fraudulent patterns. The research proposes using neurosymbolic approaches to overcome existing challenges. Symbolic AI methods will be used for transparent dataset labeling by incorporating expert domain knowledge and legal regulations. Additionally, a neural ensemble approach is proposed where individual classifiers for specific fraud indicators are built, and their outputs are combined using a symbolic program. The author is seeking guidance on integrating these neurosymbolic paradigms into their research.

Keywords

Neurosymbolic AI, Fraud Detection, Knowledge Graph, Public Administration

1. Introduction

My research work focuses mainly on Fraud Detection. The project is carried out in cooperation with Invitalia, the Italian National Development Agency, which is responsible for authorizing, distributing, and managing public incentives on behalf of various key ministries of the Italian Republic. The primary goal is to engineer a robust decision support system capable of serving a dual purpose: first, to accurately pinpoint fraudulent activities that have already occurred, and second, to proactively identify and flag suspicious requests in real-world operational environments, thereby enabling human operators in the background to conduct further, in-depth reviews and intervene before potential fraud materializes.
Fraud, in its essence, is intentional deception or misrepresentation for personal gain or to cause loss to another party. It is a deliberate act designed to mislead or trick someone into giving up something of value, whether it is money, property, information, or even a service. In the context of public funds, fraud is the intentional misuse, misappropriation, or deception involving governmental financial resources for unauthorized personal gain, causing a loss to the national community.
This short presentation outlines my proposed doctoral research on fraud detection, mainly using neural AI technologies. My aspiration is to take advantage the power of neural networks to identify complex patterns indicative of fraudulent activities on financial and public individual data. While the field has seen significant advancements, I am particularly interested in exploring the potential of neurosymbolic AI to further enhance the capabilities and interpretability of these systems, which is a crucial requirement in the public investment field.
My motivation for applying to this doctoral consortium is to gain valuable insights and guidance on how to effectively bridge the gap between deep learning and symbolic approaches in my research, especially since my project is in its very initial stages - at writing time, it has been starded three months ago. I believe that the interdisciplinary discussions and expert feedback offered by the Consortium will be crucial in shaping the foundational steps in said research direction, providing me with the necessary perspectives to integrate neurosymbolic paradigms into my upcoming work on *Fraud Detection*.

ICLP DC 2025: 21st Doctoral Consortium on Logic Programming, September 2025, Rende, Italy.
michele.vitale@unical.it (M. Vitale)

2. Project Overview

The project goal is to build a complete decision support system that can help domain experts, mainly with economic and legal backgrounds, to analyze and identify frauds and to report potential fraudulent intents in new fund requests. A simple schema can be seen in Figure 1.
To achieve this, the initial step is to build a data collection from all instances of funding requests and revocations directed to Invitalia, the Italian National Development Agency. A revocation occurs when, following thorough checks, Invitalia identifies irregularities in an entity's acquisition of funds, after that entity had previously received financing. Revocations can stem from numerous reasons, including but not limited to: non-compliance with conditions, failure to submit crucial documentation, deviation from the spending plan, non-payment of installments for loans, or ongoing criminal proceedings. Thus, it is important to note that a revocation do not necessarily imply that the beneficiary has fraudulent purposes: a variety of different causes might lead the agency to retire the funds, such as unmet deadlines or loss of eligibility for funding. Furthermore, not every fraud has been identified because the manual verification process is long and complex. Therefore, a fraud might not have an associated revocation document, but only an approval and access to funds. It is clear that the main step is the identification of a fraud, based on the data that have been collected in the internal protocol and, thus, in the data lake previously built.
The next step involves creating a knowledge graph from the extracted data, integrating it with financial data collections from paid databases. Based on this knowledge graph, a deep learning model needs to be built to identify recurring patterns in organizations prone to fraud. Indeed, these organizations are often identifiable by various components, such as a network of interconnected people or entities, or a recurring pattern in transaction amounts. Another strategy could involve calculating similarity metrics with previously recognized frauds. This module will be the primary focus of my contribution and, consequently, the core of my future research.

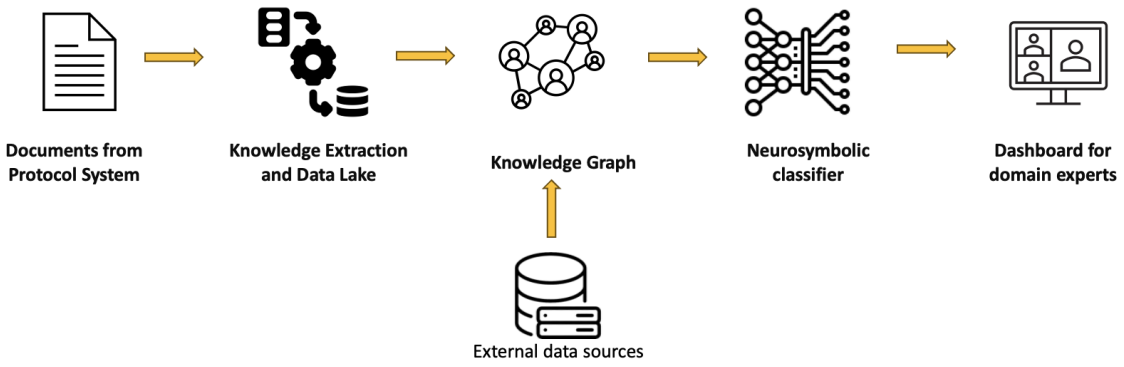

Figure 1: A high-level representation of the final project idea.

3. Data and Knowledge Extraction

The first part of the project, which is already complete, involved developing a knowledge extraction pipeline. Its goal was to create a data lake containing all relevant information from public funding approval and revocation documents granted to third parties. We achieved this by processing a large set of PDF documents resulting from the period 2018 - 2025 protocol registrations, and using LLMs (Large Language Models) for querying the files content.
A general overview of the architecture can be seen in Figure 2. The module implements a pipeline flow, in which both old and new documents are processed in the I/O Documents Handler object, that via a

set of API calls to the Data Query Server can retrieve the predefined set of information that are relevant to the construction of the Data Lake. The Document Loader takes advantage of the caching mechanism built in the project to reduce the number of tokens used for each processed document, thus to reduce the cost of the infrastructure. After many different strategies that have been tested, I ended up choosing the one that has a main instruction prompt that defines the overall context, the background, and the general instructions that the agent must keep clear and one single and straightforward prompt for each of the fields that we want to populate in the final tabular structure in which the files are stored. This drastically reduced the LLM misinterpretation of data and requests, achieving an high level of data quality metrics, verified manually on a consistent set of about 100 documents. The prompt engineering also includes zero-shot, few-shot techniques, with refinements for each single piece of data that must be extracted[1].
The data lake will be expanded with external data from paid databases as sources and represented on a knowledge graph, built tracking all the relations among the documents, the involved entities (people, business), and the presented projects.

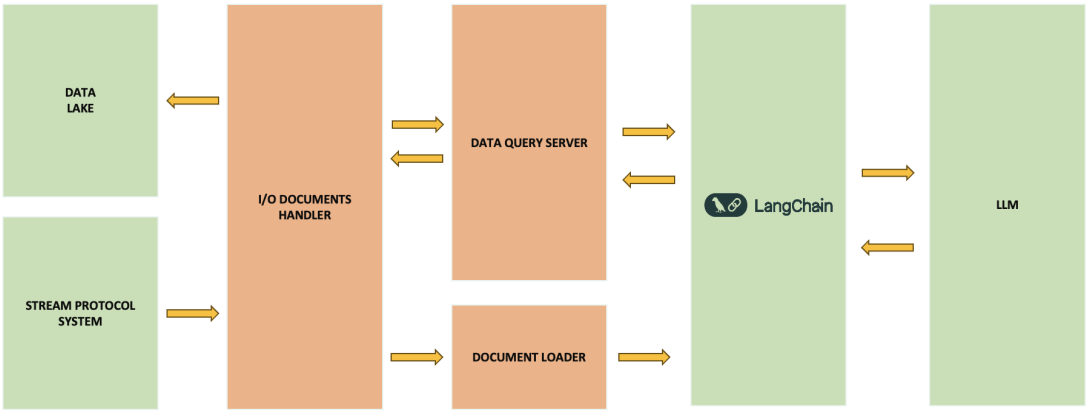

Figure 2: The Knowledge Extraction module. In green, the components that are in the overall project architecture, and in orange the components that have been developed specifically for the task.

4. Anomaly and Fraud Detection

Anomaly detection and fraud detection are closely related fields within data science and machine learning, both aiming to identify unusual patterns or behaviors that deviate significantly from what is considered ”normal”. Although anomaly detection is a broader concept, fraud detection is a specific application of anomaly detection in the context of criminal deception.
Anomaly detection, also known as outlier detection, is the process of identifying data points, events, or observations that do not conform to an expected pattern or other items in a dataset. These deviations are often indicative of some underlying problem or rare event that requires further investigation.
Fraud detection is a specialized application of anomaly detection aimed at identifying and preventing deceptive activities undertaken to gain illegal financial or other benefits. It focuses on recognizing patterns, anomalies, or suspicious activities in transactional or behavioral data that indicate fraudulent intent [2].

5. Neurosymbolic Approaches

In the neural approach to fraud detection, a significant challenge is the lack of an empirical, readily interpretable method to identify fraud. Given the inherent complexity of the domain, building a neural

classifier from scratch can be prohibitively expensive and ultimately ineffective. This is because neural networks often act as ”black boxes,” making it difficult to understand why a particular decision was made, which is crucial in regulated fields like financial fraud detection. Furthermore, they typically require vast amounts of labeled data, that not only are not present in the domain, but also are often scarce and imbalanced for rare events like fraud.
Two Neurosymbolic approaches are presented in the following sections, with the goal of my application for the Doctoral Consortium being to have the important chance to explore other possible Neurosymbolic approaches to fulfill the task and also find some new research streams to expand in the field.

5.1. Dataset Construction and Labeling

In this context, symbolic approaches can contribute significantly to solving the task of labeling the data already collected in the knowledge graph to build a labeled dataset. Unlike neural networks, symbolic AI methods offer transparency and explainability. They allow us to explicitly define and reason about the patterns and relationships indicative of fraudulent activity. This means that we can encode domain expertise, legal regulations, and known fraud indicators directly into the system. Furthermore, it can give an explanation on the model's behaviour, starting from the domain knowledge given by experts. For instance, an example of domain-expert description might be:

> If the financing originates from more than two funds and the aggregate amount is greater than €20,000, then it is a potential fraud.

Starting from the description, some simple rules with the Answer Set Programming formalism [3] might be inferred, as follows.

```
at_least_3_funds :- #count{A : fund(A,I)} > 2.

total_over_20k :- #sum{I,A : fund(A,I)} > 19999.

potential_fraud :- at_least_3_funds, total_over_20k.
```

The predicate potential_fraud can also be expressed in a probabilistic manner, with weighted atoms based on the seriousness of its semantic meaning in the domain.

5.2. Neural Ensemble

The task of fraud detection is large and complex, including several subtasks that naturally emerge from decomposing the initial domain. A fraud is a dynamic entity, potentially composed of diverse elements such as collusion among a group of individuals, specific patterns within transaction amounts, or a project proposal that includes certain scopes or elements. Each one of these subtask has an impact on the final classification, asserting whether a fund application is a fraud or not. A neural model might not be able to leverage the importance of each of the sub-components that, put together, represent a fraud. An interesting approach to this problem could involve identifying each potential fraud indicator, with the help of domain experts. Instead of building a single and monolithic model that tries to ingest every information in the dataset, we could construct a series of individual classifiers for each specific subtask. The results could then be combined at the end using a symbolic program that considers the output of each model, in an approach very similar to semantic loss [4]. Such an ensemble might be more capable to adapt to the patterns of the data, with each task having a precise weight defined from the practical knowledge of the domain given by experts. A simple schema of this idea can be seen in Figure 3.

6. Conclusion

This paper presented my doctoral research on fraud detection within public funds, specifically focusing on the decision support system being developed for Invitalia, the Italian National Development Agency.

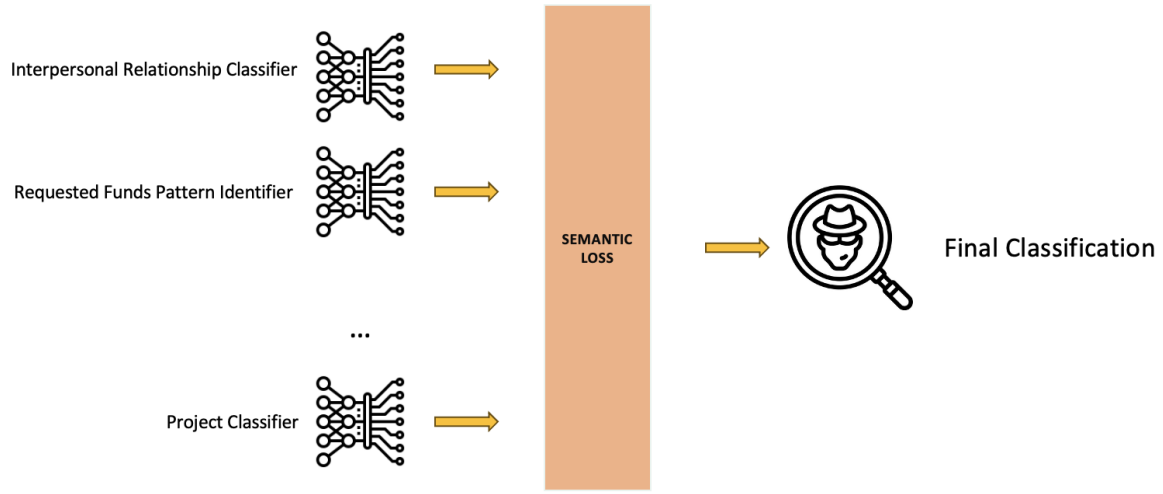

Figure 3: A simple schema on the idea underlying the neurosymbolic ensemble.

We highlighted the project's dual aim: to both identify past fraudulent activities and proactively flag suspicious requests for human review, recognizing the complex, multi-faceted nature of fraud.

My research begins with building a rich knowledge graph from Invitalia's funding data, augmented by external financial databases. This will constitute the dataset for a deep learning model designed to pinpoint recurring fraudulent patterns, a module central to my future research contribution.

A key challenge lies in the interpretability of AI models and the scarcity of labeled data in this sensitive domain. I am particularly enthusiastic about exploring neurosymbolic AI approaches to address these issues. Symbolic methods can provide transparent means for dataset labeling by incorporating expert domain knowledge, while a neural ensemble approach could offer greater adaptability and explainability by combining task-specific classifiers.

My motivation for attending this Doctoral Consortium is to gain crucial insights and guidance on these neurosymbolic avenues. As my project is still in its early stages, the consortium's interdisciplinary discussions and expert feedback will be invaluable in shaping my research direction and effectively integrating neurosymbolic paradigms into my work on Fraud Detection.

Declaration on Generative AI

The author has not employed any Generative AI tools.

References

[1] L. Boonstra, Prompt engineering, 2024.

[2] D. Carneiro, P. Veloso, A. Ventura, G. Palumbo, J. Costa, Network Analysis for Fraud Detection in Portuguese Public Procurement, 2020, pp. 390–401. doi:10.1007/978-3-030-62365-4_37.

[3] G. Brewka, T. Eiter, M. Truszczyński, Answer set programming at a glance, Commun. ACM 54 (2011) 92–103. URL: https://doi.org/10.1145/2043174.2043195. doi:10.1145/2043174.2043195.

[4] J. Xu, Z. Zhang, T. Friedman, Y. Liang, G. V. den Broeck, A semantic loss function for deep learning with symbolic knowledge, 2018. URL: https://arxiv.org/abs/1711.11157. arXiv:1711.11157.