CS269 Final Project Proposal: Building a Framework for Explainability and Accountability for Albania's AI Minister, "Diella"

Chenyang Zhao
UCLA
zhaochenyang20@gmail.com

Genglin Liu
UCLA
genglinliu@gmail.com

Mingqi Zhao
UCLA
mizhao34@g.ucla.edu

Jaelyn Fan
UCLA
jfan981@ucla.edu

Abstract

Using Albania's AI Minister, "Diella," as a case study, this project addresses the risks of opaque "black box" models in high-stakes public governance. We argue that such systems must be inherently interpretable, not merely post-hoc explainable. We propose a framework mandating both technical interpretability and democratic governance principles. To demonstrate feasibility, we will build a prototype interpretable model for public procurement that provides transparent, human-readable justifications for its decisions, enabling true accountability.

1 Motivation

Albania's AI Minister, "Diella," who oversees public procurement, exemplifies the risks of using "black box" models for high-stakes governance. Such opaque systems lack accountability and can perpetuate hidden biases. Inspired by Cynthia Rudin's work (Rudin, 2019), we argue that post-hoc explanations are insufficient for decisions with major political and economic consequences. This project's goal is to design a framework mandating the use of **inherently interpretable models**, ensuring that AI in public service is transparent, fair, and subject to democratic scrutiny.

2 Data

Our prototype will use public government procurement data, such as from the **World Bank's Open Data platform**, or a **simulated dataset** if real-world data is inaccessible. The dataset will include key features like bid price, company qualifications, financial history, and past performance records.

3 Methodology

Our research combines policy analysis with technical prototype development.

1. **Policy Framework Analysis:** Based on Rudin's theory, we will propose a governance framework with two core principles:
   - **Technical Principle:** Mandating the use of inherently interpretable models (e.g., decision trees, EBMs).
   - **Governance Principle:** Ensuring democratic oversight through public audits, participation, and disclosure of decision logic.
2. **Technical Prototype Development:** We will build an interpretable model to classify procurement bids. To demonstrate our framework's feasibility, the model's output will not be just a prediction, but a **clear, human-readable reason** for its decision (e.g., "Reject: bid price is too low [weight: -25] and company history is too short [weight: -15]").

4 Expected Timeline

Weeks	Task
Week 1-2	Literature review
	Dataset identification and acquisition
Week 3-4	Draft initial policy framework
	Data preprocessing and analysis
Week 5-6	Build interpretable model prototype;
Week 7-8	Evaluate model performance
Week 9-10	Integrate and revise full paper
	Final review, formatting, and submission
Week 11	Finalize & Submit Report

Table 1: Project Timeline

References

Cynthia Rudin. 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Preprint*, arXiv:1811.10154.