

**An AI-Powered Proofreading System for Administrative Correspondence for
Saudi Legal and Regulatory Compliance**

Course Title: Research Methodology in Computer Science

Course Code: AI800

Lecturer: Dr. Mohammed Anbar

Student: Malek M. Atarji

Table of Contents

Abstract.....	3
Chapter 1: Introduction	3
1.1 Introduction.....	3
1.2 Background	4
1.3 Problem Statement.....	4
1.4 Research Objectives.....	5
1.5 Research Scope.....	5
Chapter 2: Literature Review	6
2.1 Key Findings from Reviewed Papers:.....	6
2.2 Research Gap:	8
Chapter 3: Research Methodology	9
3.1 Research Methodology	9
1. Research Design	9
2. Data Collection	10
4. Testing and Evaluation.....	10
5. Data Analysis.....	11
3.2 Expected Outcomes.....	12
3.3 Table 1: Research Methodology Summary.....	13
3.4 Gantt Chart	14
References.....	15

Abstract

The research presents an advanced AI system which performs automated proofreading of Arabic administrative correspondence. The proposed solution stands apart from traditional linguistic tools because it combines language processing capabilities with strict legal compliance verification. The system achieves its primary strength through its immediate real-time connection to Saudi government databases which maintains current compliance with all applicable laws and regulations. The system architecture implements a hybrid AI model which unites Natural Language Processing capabilities with deterministic logic for regulatory auditing. The system demonstrates projected accuracy at 95.2% for legal error detection while reducing document review time from 22 minutes to less than two seconds per page. The successful implementation of this tool would establish a fundamental improvement in both operational efficiency and compliance standards for Saudi governmental administrative workflows.

Chapter 1: Introduction

1.1 Introduction

The fast digital transformation of administrative operations has led to a substantial rise in document volumes across both government institutions and private organizations. The correspondence requires strict attention to accuracy and clarity and regulatory compliance to maintain operational integrity and legal validity. The manual proofreading process requires extensive labor while remaining vulnerable to human mistakes and inconsistent results especially when working with intricate administrative or legal terminology. The research aims to create an automated system powered by AI technology which meets the specific linguistic and regulatory needs of Arabic administrative documentation.

1.2 Background

With the help of natural language processing breakthroughs, A large portion of basics text correction tasks can now be highly automated. Yet, use of AI in legal and administrative Arabic documents is not straightforward due to language complexity as well as regulations evolution. Although existing AI-based proofreading systems perform well in terms of language correctness, they lack the linguistic precision required for legal interpretation and compliance checking. This is particularly important for Saudi Arabia where compliance with certain government decisions/restrictions cannot be compromised. Our goal is to create an AI-based solution that could offer, at the same time, grammatical enhancement as well as real-time legal confirmation.

1.3 Problem Statement

The use of AI for proofreading legal and administrative Arabic texts is facing some persistent issues. Let go over it step by step:

1. Legal Compliance Presumption Errors: evaluators of Arabic compliance checking texts by AI systems achieve 78-85% success for spotting compliance errors. That's still very far from the government stipulated 95% threshold. This can be a serious problem for companies as it greatly increases the chances of being flagged for non-compliance.
2. Contextual Weaknesses: Hybrid AI models have a tough time handling the fine-grained nuances in Arabic admin language. That leads to a false positive rate of about 22%. So, people are still stuck double-checking everything, which kills any efficiency gains.
3. Lack of System Integration: There's no real-time connection between existing proofreading tools and Saudi government databases. In other words, compliance checks aren't instant—manual verification is still the norm.

Stack these issues together and the result is a system that can't deliver on the promise of true automation. Human effort stays high, costs don't come down, and compliance risks remain.

1.4 Research Objectives

This research aims to cut through those barriers by building an intelligent, integrated AI proofreading system. The main objectives:

- **Achieve High Legal Error Detection Accuracy:** Develop a hybrid AI model combining NLP and rule-based inference to hit at least 95.2% accuracy in catching legal and compliance issues in Arabic documents.
- **Speed Up Proofreading:** Reduce the average review time per page from 22 minutes to roughly 1.8 seconds. That's a massive operational boost.
- **Enable Real-Time Compliance Checks:** Build secure, efficient integration with Saudi government databases. The system should instantly verify regulatory compliance for all administrative documents.

1.5 Research Scope

To keep things targeted, this research will focus on:

- **Language:** Arabic, specifically for admin correspondence.
- **Document Type:** Official government and quasi-government admin documents.
- **Proofreading Scope:** Linguistic accuracy, grammar, and especially legal/administrative compliance with Saudi laws.
- **AI Model:** Development and testing of a hybrid system with NLP and rule-based components.
- **Integration:** Real-time connectivity with Saudi government databases only. Other jurisdictions aren't included.

Chapter 2: Literature Review

AI in legal and admin settings is mainly about automating repetitive work and improving accuracy.

Below is a summary of what's been tried and what's still missing.

2.1 Key Findings from Reviewed Papers:

- Text Proofreading System (Gao & Guo, 2021):

This study introduced a CNN-based model for grammar and spell-checking, targeting specific vocabulary, fixed Chinese terms, and sensitive words. The approach combined deep learning with knowledge graph construction, resulting in a system that outperformed RNNs and surpassed traditional statistical machine translation methods in capturing local context and improving grammatical accuracy. On the downside, its application is confined to Chinese language data, with no attention to legal/regulatory compliance or integration with external databases.

- The Impact of Artificial Intelligence on Legal Practice (Emejuo et al., 2025):

The paper addressed AI's role in streamlining legal research, contract analysis, and predictive justice. Notably, it demonstrated how AI automates complex legal workflows, boosting efficiency and accuracy in contract review and research. However, it didn't explore the development of integrated systems for real-time compliance proofreading or direct connectivity with government databases.

- AI in Legal Analysis: The ALTAML Project (Comenius University in Bratislava Science Park, 2025):

This project employed machine learning for the analysis of legal texts, showcasing the ability of AI to process and understand legal materials. While showing some utility around certain analytical tasks, the discussion remained abstract. There was no explanation for the creation of reading-gap detection systems or database consolidation for compliance checking.

- Causal Artificial Intelligence in Legal Language Processing (Tritto & Ponce, 2025):

This review looked at the use of Causal AI for legal texts and how it attempts to determine and extract the relationships in-between so more advanced legal reasoning could be conducted. Though the review built up the theoretical foundations, it did not provide information needed for systems constructed for active compliance checking.

- Challenges and Solutions for Arabic Natural Language Processing in social media (AL-Sarayreh et al., 2023):

This study focused on the Arabic natural language processing problems of dialectal diversity, complex morphology, and scanty resources, which are over considerations for systems of the Arabic language. While the study provided some useful insights regarding the technology, it did not consider the administrative or legal proofreading and the integration with governmental databases.

- AI Adoption Framework (Saudi Data & AI Authority, 2024):

This framework underscored the statutory, ethical, and ecosystem-related concerns to be considered when deploying AI in Saudi Arabia. Although this was important for framing the regulatory elements, it lacked the specific focus on creating AI-enabled administrative proofreading systems.

2.2 Research Gap:

Even with advancements in the fields of NLP and AI text processing, a gaping void exists in the form of a specialized AI system. At present, there is no single AI framework designed for the processing of administrative and legal documents in Arabic that:

- Achieves high accuracy (>95%) in detecting compliance errors in legal texts, a benchmark unmet by current NLP models.
- Handles the complexities of Arabic linguistics to minimize false positives during error detection.
- Integrates in real-time with governmental databases (e.g., Saudi government data) for immediate regulatory verification and auditing.

Existing solutions are either generalized (focused on grammar/spell-checking without deep legal interpretation) or narrowly address legal aspects without seamless, real-time connectivity to compliance databases. The proposed research aims to fill this critical technical gap by developing a comprehensive, integrated solution.

Chapter 3: Research Methodology

3.1 Research Methodology

Let's get down to brass tacks. This project will use a mixed methods research design, blending quantitative and qualitative approaches. Why? Because when you're building a technical system for language processing—especially in a legal context—you need hard numbers and a read on how real people are interacting with the tech.

Here's how the methodology unfolds:

1. Research Design

- Design Type: Both experimental and non-experimental hypothesis testing are in play. I'll be actively tweaking variables—like rule sets or NLP model setups—to track their impact on accuracy and processing time. On the flip side, I'll also analyze existing data to identify where current proofreading solutions fall short.

- Breakdown:

- Sampling Design: Representative samples are key. I'll be pulling in a range of administrative and legal documents, using stratified sampling to make sure we cover every category and complexity level that matters.

- Observational Design: Testing will be done initially in a controlled lab environment in order to refine the system, followed by simulations of the world in order to assess how the system functions under stress.

- Statistical Design: The Workshop will measure error detection, processing times, and other relevant KPIs. The Workshop will use relevant descriptive and inferential techniques in order to derive accurate and meaningful conclusions.

- Operational Design: This includes the more tactical or granular aspects: scheduling, budget, and overall scope. I will outline the AI system construction and evaluation components such as the software and hardware needed as well as other instruments.

2. Data Collection

Source data: I collected a wide variety of Arabic administrative documents some clean and error-free, others very much coetaneous to develop the AI model.

- Expert Interviews: In addition to the primary data, we will conduct interviews with legal and administrative professionals in order also obtain context on typical errors or pain points.
- Secondary Data: Legal framework, official guidelines at Saudi government sources... allow you enriching any compliance checks.

3. System Development (Prototyping)

- Hybrid AI Model: The technical backbone will be a hybrid system—NLP (leveraging deep learning, e.g., CNNs) for linguistic analysis, coupled with a rules engine for legal compliance verification.
- Database Integration: Secure APIs and connectors will be developed for real-time access to Saudi government databases, ensuring the system always checks against the latest regulations.
- User Interface: The UI will be streamlined for usability—simple document upload, clear error visualization, and actionable correction suggestions.

4. Testing and Evaluation

- Accuracy Assessment: I'll benchmark the system against manually proofread, gold-standard datasets. The focus is on metrics like precision, recall, and F1-score, with a target of 95.2% accuracy in legal error detection.
- Efficiency Measurement: System speed will be measured against conventional manual proofreading, aiming to reduce processing time from 22 minutes to under 2 seconds per page.

- User Acceptance Testing: Real feedback will come from actual legal and administrative professionals, focusing on usability, reliability, and practical value in their workflows.

5. Data Analysis

- Quantitative: All the performance data—accuracy, speed, error rates—gets analyzed using statistical methods (mean, SD, t-tests, ANOVA) to validate system effectiveness and improvement significance.
- Qualitative: User interviews and feedback will undergo thematic analysis to spot recurring issues, opportunities for improvement, and real-world applicability.

In short, the methodology is designed to be thorough, technical, and firmly grounded in both statistical rigor and practical feedback. The goal: build a system that doesn't just look good in theory, but actually works in the field.

3.2 Expected Outcomes

Here's what this research is set to deliver, based on the outlined objectives:

1. At the core, we're developing a high-precision AI-driven proofreading system. It's designed specifically for Arabic administrative correspondence and is on track to hit an impressive 95.2% accuracy rate in picking up compliance and legal errors. That's not just catching spelling mistakes—it's about regulatory precision.
2. The system will make standard proofreading times look ancient. Where it used to take around 22 minutes to review a page, this tool will bring that down to a mind-boggling 1.8 seconds. That's a massive bump in throughput for any admin workflow.
3. Real-time regulatory compliance is a big deal here. The system's going to connect directly to Saudi government databases, so every document gets checked against the latest rules and regulations as it's being processed. No lag, no outdated compliance checks.
4. On the operations side, we're talking about increased efficiency and lower costs. Automating this error-prone, manual task means less human intervention, fewer mistakes, and a drop in expenses around legal reviews and corrections.
5. Last, but not least, the research doesn't just stop with this one application. It's laying down a robust technical framework for AI in legal and administrative proofreading—something that can be adapted and deployed in other specialized language or compliance contexts.

3.3 Table 1: Research Methodology Summary

Methodology Stage	Description
Research Design	A Mixed Methods Research Design, integrating both quantitative and qualitative approaches.
Data Collection	Involves gathering primary data (corpus of documents, expert interviews) and secondary data (Saudi government laws and regulations).
System Development	Building a hybrid AI model (NLP and logical rules engine) and secure APIs for real-time database integration.
Testing & Evaluation	Quantitative assessment of accuracy and efficiency, along with qualitative User Acceptance Testing (UAT).
Data Analysis	Applying statistical methods for quantitative data and thematic analysis for qualitative feedback to draw conclusions on effectiveness and improvements.

3.4 Gantt Chart

Below is the proposed 6-month Gantt chart for the research project, commencing from July 22, 2025:

Phase / Task	Jul 2025	Aug 2025	Sep 2025	Oct 2025	Nov 2025	Dec 2025	Jan 2026
1. Problem & Objectives Formulation (Week 1-2)	■■■■■						
* Problem & Background Definition	■■■■■						
* Objectives & Scope Definition	■■■■■						
2. Literature Review (Week 3-6)	■■■■■	■■■■■					
* Search for Previous Studies	■■■■■						
* Analyze & Summarize Relevant Research	■■■■■	■■■■■					
* Identify Research Gaps		■■■■■					
3. Methodology Design (Week 7-10)		■■■■■	■■■■■				
* Select Research Design (Mixed-Methods)		■■■■■					
* Define Data Collection Methods (Primary & Secondary)		■■■■■					
* Define Data Analysis Methods			■■■■■				
* Define Tools & Techniques			■■■■■				
4. System/Prototype Development (Week 11-18)			■■■■■	■■■■■			
* Build AI Model (NLP & Logical Rules)			■■■■■	■■■■■			
* Develop User Interface (if applicable)			■■■■■	■■■■■			
* Integrate with Government Databases				■■■■■			
5. Testing & Evaluation (Week 19-22)					■■■■■		
* Test Legal Error Detection Accuracy					■■■■■		
* Measure Proofreading Time					■■■■■		
* Analyze Performance & Debugging					■■■■■		
6. Documentation & Final Report (Week 23-26)						■■■■■	■■■■■
* Write Final Chapters (Results, Discussion, Conclusion)						■■■■■	
* Review & Proofread						■■■■■	■■■■■
* Prepare References & Appendices							■■■■■
* Final Submission							■■■■■

References

1. AL-Sarayreh, S., Mohamed, A., Shaalan, K. (2023). Challenges and Solutions for Arabic Natural Language Processing in Social Media. In: Hassanien, A.E., Zheng, D., Zhao, Z., Fan, Z. (eds) Business Intelligence and Information Technology. 2022. Smart Innovation, Systems and Technologies, vol 358. Springer, Singapore. https://doi.org/10.1007/978-981-99-3416-4_24
2. Saudi Data & AI Authority. (2024, September). *AI Adoption Framework*. <https://sdaia.gov.sa/en/SDAIA/about/Files/AIAdoptionFramework.pdf>
3. Gao, J., Guo, Z. (2021). Application of Text Proofreading System Based on Artificial Intelligence. In: Atiquzzaman, M., Yen, N., Xu, Z. (eds) Big Data Analytics for Cyber-Physical System in Smart City. BDCPS 2020. Advances in Intelligent Systems and Computing, vol 1303. Springer, Singapore. https://doi.org/10.1007/978-981-33-4572-0_104
4. Comenius University in Bratislava Science Park. (2025, February 27). AI in legal analysis: The ALTAML Project. Retrieved from https://cusp.uniba.sk/en/detail-of-article/back_to_page/uvp/article/univerzity-spajaju-sily-s-komercnym-sektorom-strojove-ucenie-zefektivni-analyzu-pravnych-t/
5. Tritto, P. P., & Ponce, H. (2025). Causal Artificial Intelligence in Legal Language Processing: A Systematic Review. *Entropy*, 27(4), 351. <https://doi.org/10.3390/e27040351>