

# Fine-tuning Arabic-English NMT Engines for Investment Laws

Najat Al-Huthali  
University of Surrey, UK  
[n.alhuthali@surrey.ac.uk](mailto:n.alhuthali@surrey.ac.uk)

Yuan Zou  
University of Surrey, UK  
[y.zou@surrey.ac.uk](mailto:y.zou@surrey.ac.uk)

Constantin Orasan  
University of Surrey, UK  
[c.orasan@surrey.ac.uk](mailto:c.orasan@surrey.ac.uk)



## The Problem

- Generic MT struggles with Legal Arabic
- Scarcity of Saudi Legal Corpora
- Growing Need for AI Legal Translation

## Our Contributions

- Built one of the largest Arabic-English parallel corpora of Saudi Investment laws and regulations
- Proposed a practical pipeline for extraction, cleaning, and AI-assisted alignment of Arabic legal texts
- Provided initial empirical evidence that domain-specific fine-tuning improves Arabic-English legal MT quality

## Building the Saudi Laws Corpus

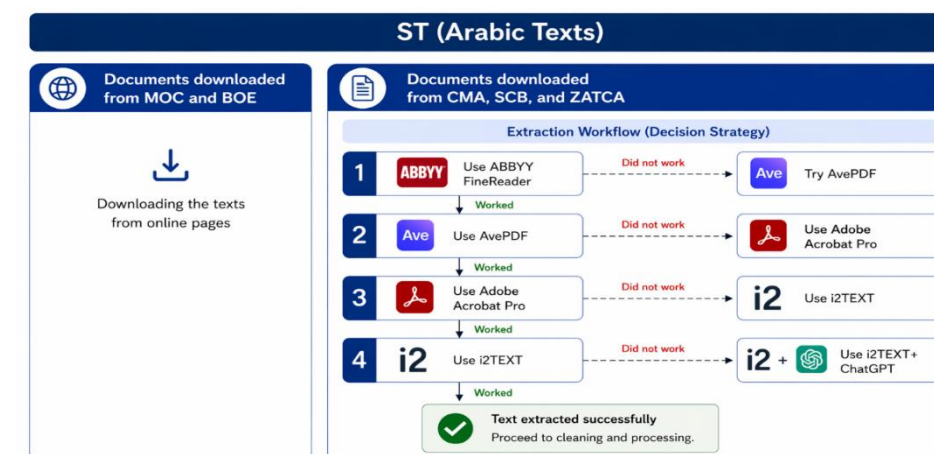
### A: Collecting Texts from Governmental Websites

- Ministry of Commerce
- Capital Market Authority
- Bureau of Experts
- Saudi Central Bank
- Zakat, Tax, and Customs Authority

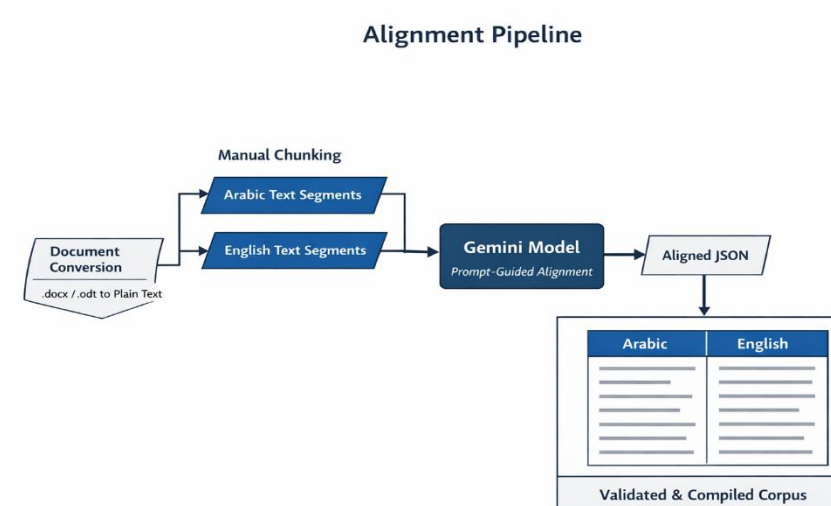
Docs	193
ض	763,765
A	958,581

### B: Extracting Texts Using OCR Programs and PDF Converters

Adobe Acrobat Pro was used to extract text from all English documents. However, Arabic documents were more challenging. The following figure explains the workflow adopted to extract texts from Arabic documents



### C-Aligning Texts



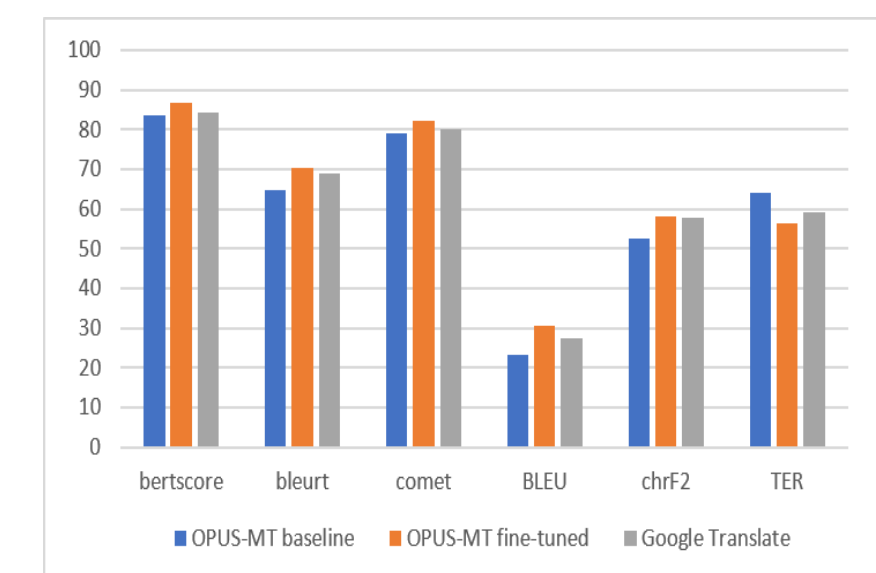
### Fine-tuning OPUS-CAT Tool

OPUS-CAT tool was installed on the computer. The Arabic-English model was downloaded and fine-tuned using 90% of the created corpus.

## Results

Translation quality was evaluated using both automatic metrics and qualitative analysis.

### Automatic Evaluation



The fine-tuned OPUS-MT consistently outperforms the baseline OPUS-MT and Google Translate across all automatic metrics, demonstrating that domain adaptation with the created corpus improves translation quality.

### Qualitative Analysis

The fine-tuned model showed improved consistency in translating legal terms and legal modality, and better correspondence of article numbering.

ST	TT	OPUS	Fine-tuned OPUS
نظام ضريبيه الدخل	Income Tax Law	Income tax system	Income Tax Law
شركة الأموال : الشركة المساهمة ، أو الشركة ذات المسئولية المحدودة ، أو شركة التوصية بالأسهم	Capital company, A joint stock company, limited liability company, or a company limited by shares.	Money Company: Contributing Company, Limited Liability Company, or Equity Recommendation Company	Money Company: A joint-stock company, limited liability company, or a joint-stock company
تنظم اللوائح التنفيذية العلاقة بين	The implementing Regulations shall regulate the relationship between	The implementing regulations regulate the relationship between	The Implementing Regulations shall regulate the relationship between
تخضع لأحكام النظام: المنشآت المالية، والشركات القابضة	This Law shall apply to financial institutions, holding companies,	They are subject to the provisions of the regulations: financial establishments, holding companies,	The following shall be subject to the provisions of this Law: financial institutions, holding companies,
المادة الأولى	Article 1	Article I	Article 1
المادة الثانية	Article 2	Article II	Article 2
المادة الثالثة	Article 3	Article III	Article 3

## Acknowledgements

This study is funded by the Ministry of Education in Saudi Arabia and the Saudi Cultural Bureau.

This study is a pilot investigation. Future work will explore other NMT models and conduct human evaluation with professional legal translators

Better Data. Better Models. Better Legal Translation