

Oro: A Programming Language Designed for Document Compliance

Nolan McKenna '25

Prof. Johnson

Department of Computer Science, Trinity College



Introduction

Oro is a programming language built from the ground up for simplifying document compliance across industries.

With a wide range of native functions designed to redact sensitive information, classify document risk based on detected personal information, and much more.

Development

Oro is based off Robert Nystrom's "jlox" interpreter from his book, *Crafting Interpreters*. Additional features were developed in addition to Nystrom's implementation, including but not limited to enhanced error reporting, file handling, f-strings, and data structures.

Example

FILL OUT BELOW TO DISCLOSE/OBTAIN				
I authorize	Nolan McKenna	to disclose /obtain health information to:	Kathleen McKenna	
Address	300 Summit Street	Hartford	CT	06106
Tele#:	123-456-7890	Fax#:		

redactPDF.oro

```
def pdf = PDFDocument("../scripts/medical-records-release-form-test.pdf");  
redactHIPAA(pdf, "../scripts/medical-records-release-form-test_HIPAA.pdf");
```

FILL OUT BELOW TO DISCLOSE/OBTAIN				
I authorize		to disclose /obtain health information to:		
Address				
Tele#:		Fax#:		

This specific function, redactHIPAA, uses Stanford CoreNLP and Tesseract OCR to accurately detect and redact all necessary Personally Identifiable Information (PII) from a given healthcare document according to HIPAA privacy rules.

Applications

Many existing compliance solutions from Microsoft, Relativity, and DocuSign are expensive, complex, or only specialize in one type of compliance. Oro provides a simple, lightweight, and scriptable language that empowers companies across legal, healthcare, and financial industries to fine-tune their document compliance pipelines.

Future Work

- Develop functions for compliance validation and document watermarking
- Expand redaction capabilities
- Fine-tune NLP for improved redaction

Acknowledgements

Prof. Johnson, Prof. Kousen, Travelers