

Eryclis Silva

☎ 217-418-2042 ✉ ersilva2@illinois.edu [in linkedin.com/in/eryclissilva](https://www.linkedin.com/in/eryclissilva) github.com/Eryclis eryclis.com

Education

- University of Illinois at Urbana-Champaign** Expected May 2027
PhD in Information Science (Data Science and Engineering concentration) Champaign, IL
- **Relevant Coursework:** trustworthy machine learning, advanced natural language processing, statistical learning, network analysis, software engineering.
- Federal University of Pernambuco** Aug 2013 - May 2018
LL.B. Bachelor of Law Recife, Brazil

Experience

- Princeton University - Research Computing** May 2024 – Aug 2024
Software Engineering Intern Princeton, NJ
- Developed Tiger Data Metrics, a comprehensive web application using Flask and Dash for analyzing storage metrics across a 100+ Petabyte research database.
 - Implemented efficient queries using Asset Query Language, automating data collection and manipulation.
 - Optimized performance by creating a SQLite database for efficient storage and retrieval of processed metrics.
 - Applied SOLID principles and design patterns such as MVC, implementing unit tests with pytest to ensure code maintainability and scalability.
- Governance Lab on Sociotechnical Systems** Aug 2023 – Present
Research Assistant Champaign, IL
- Developed advanced NLP pipelines using Python and state-of-the-art models like BERT, RoBERTa, and T5 to analyze over 100 international AI policy documents, enhancing semantic understanding and feature extraction.
 - Engineered a custom topic modeling solution combining BERTopic and LDA, and implemented a Retrieval-Augmented Generation (RAG) system using Langchain and Hugging Face Transformers, significantly improving accuracy and context-awareness in AI policy analysis.
 - Identified regional trends and risk considerations in AI regulation by quantifying qualitative patterns with attention mechanisms and contextual embeddings, and created an interactive dashboard using Streamlit for stakeholders to explore regulatory landscapes.

Projects

- GraphExtract** | *Langchain, Ollama, NetworkX, PyVis*
- Developed the GraphExtract project to extract relevant information from unstructured texts and visually present them as interactive graphs, highlighting entities, relations and contextual proximity.
 - Applied Langchain to collect text chunks and large language models (Gemini, Mistral) for entity and relationship extraction, using Ollama for executing LLMs locally and Networkx for graph construction. Implemented interactive visualizations with Pyvis to facilitate the understanding of results.
 - The adopted approach allows for a comprehensive analysis of the data, identifying patterns and trends clearly and concisely.
- Fund Allocation for Countries in Need (Clustering Analysis)** | *Python, Scikit-learn, Pandas, Matplotlib, Seaborn*
- Developed fund allocation model for needy countries using unsupervised learning techniques (clustering) to identify groups of countries with similar needs.
 - Applied clustering algorithms, including K-means, DBSCAN, and hierarchical clustering, to group countries based on socioeconomic indicators. Conducted data pre-processing, including normalization and dimensionality reduction using PCA.
 - Successfully clustered countries using K-means, obtaining a Silhouette Score of 0.332, Davies-Bouldin Score of 1.133, and Calinski-Harabasz Score of 85.015.

Technical Skills

Languages: Python, R, SQL, HTML, CSS
Technologies: Scikit-learn, NumPy, Matplotlib, Pandas, Seaborn, NLTK, SciPy, Langchain
Frameworks: TensorFlow, PyTorch, Django, Flask, Dash, Ollama, NetworkX, Pyvis