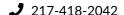
Eryclis Silva



J 217-418-2042 ■ ersilva2@illinois.edu in linkedin.com/in/eryclissilva

github.com/Eryclis

eryclis.com

Education

University of Illinois at Urbana-Champaign

Expected May 2028

PhD in Information Science (with concentration in Data Science and Engineering)

Champaign, IL

• Relevant Coursework: Trustworthy Machine Learning, Data Structures and Algorithms, Statistical Learning, Network Analysis, Linear Algebra

Federal University of Pernambuco

Aug 2013 - May 2018

LL.B. Bachelor of Law

Recife, Brazil

Experience

Governance Lab on Sociotechnical Systems

Aug 2023 - Present

Research Assistant

Champaign, IL

- Utilize BERTopic and LDA techniques in Python to cluster over 100 international documents, including Al National Strategies and Bills, for analysis in global AI regulation.
- BERT-based models for document representations, pre-processing data for clustering, and extracting contextualized embeddings. Leverage NLP techniques such as tokenization, word embeddings, and attention mechanisms to enhance semantic understanding and feature extraction.
- Identify qualitative and quantitative patterns, such as priorities and risk considerations, to understand regional trends in AI regulation.

Attorney General's Office

Jan 2020 - Aug 2023

Legal Clerk

Umuarama, Brazil

- Supervised and actively participated in combating tax and financial crimes, including monitoring police investigations. Led over 70 agreements of non-prosecution, resulting in an average of \$5 million recovered for public coffers, and initiated over 20 criminal prosecutions.
- Collaborated with interdisciplinary teams in forensic data analysis, resulting in a significant reduction of 60% in the number of police inquiries and investigative procedures concluded.

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 (Zephyr and Mistral) for entity and relationship extraction, using Ollama for executing language models 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, TensorFlow, PyTorch, Langchain

Frameworks: Django, Ollama, NetworkX, Pyvis