

# Denis Leang

📞 646-749-5489    ✉ dsl2179@columbia.edu    🔗 linkedin.com/in/denis-leang    🐙 github.com/YeriAddict    🌐 yeriaddict.github.io

## Education

### Columbia University

Sep 2023 - Dec 2024

M.S. in Data Science

New York, NY

- **Relevant Coursework:** Machine Learning, Artificial Intelligence, Computer Vision, Statistical Inference, Exploratory Data Analysis & Visualization, Computer Systems

### Telecom Saint-Etienne

Sep 2020 - Aug 2023

B.S./M.S. in Computer Science & Mathematics (Diplome d'ingenieurs)

Saint-Etienne, FRA

- Prepared for the French national entrance examinations to Grandes Ecoles at Lycee du Parc (2017-2020 – Lyon, FRA)
- **Relevant Coursework:** Data Structures & Algorithms, Database and Information Systems, Operating Systems, Distributed Computing, Cloud Computing, Big Data, Computer Graphics, Probability & Statistics, Optimization & Estimation methods

## Experience

### Columbia University Mailman School of Public Health

Mar 2025 - Now

Graduate Research Assistant

New York, NY

- Designed and implemented a RAG AI agent in **Python** that retrieves and synthesizes public health data from research databases, improving information retrieval efficiency and reducing search time

### Analysis Group

Jun 2024 - Aug 2024

Data Scientist Intern

Boston, MA

- Revamped a CRUD patient survey management app built with **C# (.NET)** and **React**, enabling more efficient distribution of clinical drug testing surveys by honing workflows and strengthening vendor engagement
- Reduced user navigation time by 50% by redesigning the application's structure using scalable **React** components and optimizing API calls, improving overall user efficiency
- Executed **Python** scripts leveraging the Etherscan API to analyze cryptocurrency transactions across multiple clusters, tracking billions of dollars on Binance and facilitating investigations into suspicious wallet activity

### Bionomous SA

Feb 2023 - Aug 2023

Full-Stack Software Engineer Intern

Villaz Saint-Pierre, CHE

- Developed a CRUD client update and support portal written in **Java Spring Boot** and **Vue.js** to automate software updates for in-house products, eliminating manual engineer involvement and improving customer experience and release management
- Implemented a secure authentication system with **Microsoft Azure** Active Directory and JWT tokens, boosting login efficiency and strengthening security for users and administrators
- Upgraded company products (**Java** for **Android**, **Flutter/Dart** for **Android/iOS**) by incorporating client-requested functionalities, increasing usability and customer satisfaction

## Projects

### Spotify Dashboard

- Built a Spotify listening habits dashboard in **React** with real-time, dynamically fetched data from the Spotify and Last.fm APIs, visualizing trends through interactive charts
- Refined data fetching for real-time music analytics by executing an efficient caching strategy and a CI/CD pipeline with GitHub Actions, reducing memory usage and query time by 4x

### End-to-end Fake News Classifier Pipeline

- Trained a Random Forest model using **scikit-learn** to classify news articles, reaching 70% test accuracy
- Deployed the model locally by building a REST API with **FastAPI** in **Python** and containerizing it with **Docker**, making it scalable and production-ready for cloud services

### Iris Recognition Model

- Engineered an iris recognition package with **OpenCV** and **Python**, incorporating preprocessing, feature extraction, and LDA-based matching to support one-to-one and one-to-many biometric recognition modes
- Achieved a consistent 95% Correct Recognition Rate (CRR) by implementing a Gabor-based multichannel filtering technique, enhancing feature extraction across L1, L2, and Cosine similarity metrics

### Electronic Health Record (EHR) Study on Interstitial Lung Disease with Columbia University Irving Medical Center

- Conducted data wrangling in **Python** to harmonize large-scale EHR data across multiple health systems
- Fine-tuned Named Entity Recognition (NER) and BERT-based models on medical text to analyze chest CT reports, achieving a 30% improvement in accuracy for identifying drug-disease associations, contributing to enhanced clinical treatment guidelines

## Skills

**Coding:** Proficient in Python, SQL, TypeScript. Experience in Java, C#, Dart, Flutter.

**Technologies:** Proficient in Git, Docker, pandas, scikit-learn, Tensorflow, Spring Boot, React. Experience in Linux, Shell Scripting, Azure, AWS, Streamlit, OpenCV, PyTorch, Spark, Vue.js. Used GCP, Kubernetes, Kafka, MongoDB, OpenGL, Leaflet, FastAPI.

**Spoken Languages:** French, Khmer, German