

# Pedro Henrique Rocha Moy

**AI, Machine Learning, Data Science & Engineering** Miami, FL

## Education

### **Executive MBA in Business Administration**

University of Miami - Miami, 2021 - 2022

### **Master's Degree in Computer Science (Machine Learning)**

Georgia Institute of Technology - Atlanta, GA, 2015 - 2017

### **Master's Degree in Earth Science and Engineering (Geophysics)**

King Abdullah University of Science and Technology - Saudi Arabia, 2010 - 2012

### **Bachelor's Degree in Mechanical Engineering (Math Minor)**

University of Massachusetts Lowell - Lowell, MA, 2008 - 2010

## Experience

### **Chief Architect | Rocha Moy Trading & Capital**

#### **2017 - Present**

- Developed the API for probabilistic and algorithmic options trading with Interactive Brokers and TD Ameritrade
- Specialties include data integration, task automation, portfolio simulations, risk mitigation, and strategy validation
- Integrated many different data sources from APIs to web scraping
- Automated trade execution, trade scheduling, and fund releases for trading
- Technologies: Python, AWS, REST APIs, Web Scraping, Machine Learning, Probability Theory, Statistical Modeling, Time Series Analysis, Quantitative Finance, Options Trading, Portfolio Optimization, Risk Management, Backtesting, Simulations, Data Integration, Data Pipelines, Task Automation

### **AI Consultant | Toptal Client**

#### **2025**

- Developed a prototype platform that automated LLC formation workflows, including a detailed regulatory study for Florida, reducing manual research time for new founders by well over 50%

- Built a back-end LLM system that validated user inputs and flagged compliance gaps during the LLC creation process, ensuring accuracy of filings
- Designed a domain-specific knowledge base and implemented an RAG pipeline to deliver precise, context-aware guidance to users
- Technologies: Python, LLMs, RAG, NLP, Prompt Engineering, Vector Databases, Knowledge Base Design, Document Processing, Regulatory Compliance, Prototype Development

### **AI Software Developer | Toptal Client**

**2023 - 2024**

- Built a full-stack Django and React application from scratch that parsed contract requirement PDFs into structured suggestions and clarification questions, enabling teams to accelerate project scoping significantly
- Engineered an autoscaling cloud infrastructure using ECS and IaC best practices to support stable, high-throughput processing of large procurement documents, ensuring near-zero downtime
- Designed the end-to-end LLM architecture with OpenAI models and served as the primary expert for all LLM-related issues, improving extraction accuracy and capabilities, shaping model-driven workflows across both the Python and JavaScript stacks
- Technologies: Python, JavaScript, Django, React, OpenAI API, LLMs, NLP, Prompt Engineering, AWS ECS, Infrastructure as Code, Docker, PDF Processing, Autoscaling, Full-stack Development

### **Lead Data Scientist | Deloitte Contractor**

**2021 - 2022**

- Designed, implemented, and deployed different natural language processing models
- Worked with stakeholders to understand use cases, the pathway to product development, and implementation using deployed models
- Mentored and supported junior data scientists on the team
- Technologies: Python, NLP, Machine Learning, Model Deployment, Stakeholder Management, Data Analysis, Mentoring

### **Enterprise Lead Data Architect | Toptal Client**

**2020 - 2022**

- Handled the architecture, development, and automation of distributed computing pipelines and data storage in the cloud for the enterprise
- Automated scalable infrastructure in the cloud to respond to development and consumer demand

- Co-managed and supervised a team of engineers from designing and delegating tasks, mentoring, and overseeing work
- Technologies: Python, AWS, Amazon EMR, Spark, Distributed Computing, Data Lake Architecture, Cloud Infrastructure, Data Pipelines, Data Storage, Team Leadership

### **Enterprise Senior ETL and Data Engineer | Toptal Client**

**2019 - 2020**

- Designed, implemented, and deployed to production fully-fledged distributed ETL jobs in Spark/Scala API
- Worked with various sources and sinks of data including disparate files, Hive tables, Mongo collections, and Kafka brokers
- Served as the senior engineer and tech lead of the team strengthening engineering and development processes, improving software quality control, and helping design stories for sprints
- Technologies: Scala, Python, Spark, MongoDB, Hive, Kafka, Distributed Computing, ETL, Data Pipelines, Agile/Scrum, Code Review, Tech Leadership

### **Hadoop Proof of Concept for Atmospheric Sciences Project | Toptal Client**

**2019 - 2020**

- Built a cluster from scratch, adhering to the client's needs to work with the home cluster
- Designed and implemented generic and specific data architectures meeting the client's query complexity and performance needs
- Built PySpark and Python software layers of abstraction to allow the client to build on top of the current infrastructure
- Technologies: PySpark, Python, Hadoop, HDFS, Big Data, Data Architecture, Cluster Management, Atmospheric Sciences

### **Research Data Engineer | Nicklaus Children's Hospital**

**2018 - 2019**

- Developed existing analytical and data workflows for users of R, Python, and Impala establishing best engineering practices
- Provided ad hoc and systematically developed ETL and big data pipelines, validation, and integration of varying data sources
- Liaised for the research department to IT and BI departments providing guidance and expertise on analytical and data needs
- Technologies: Spark, Scala, Python, R, Impala, ETL, Big Data, Healthcare Data, Data Validation, Business Intelligence

## **Technical Advisor | Insight Data Science**

**2018**

- Worked with fellows and their data engineering projects on problem definition, systems architecture, and execution
- Advised on technologies such as Spark, Kafka, Redis, HBase, Cassandra, and PostgreSQL
- Conducted mock interviews with fellows on scalability concepts, algorithms, and CS fundamentals
- Technologies: Spark, Kafka, Redis, HBase, Cassandra, PostgreSQL, Distributed Systems, Systems Architecture, Mentoring, Algorithms

## **Senior Software Engineer | NexHealth**

**2016 - 2017**

- Developed and deployed software to the client's site to perform data collection and server sync
- Performed both database and web-based data integrations of electronic medical records back to NexHealth servers
- Developed a smart SMS response system allowing the user to interact with NexHealth products via SMS
- Technologies: Python, Scala, JavaScript, Redis, PostgreSQL, Apache Spark, REST APIs, Data Integration, SMS/Messaging, Electronic Health Records, Web Development

## **Data Scientist | QuaEra Insights**

**2016**

- Served as the lead data scientist in a consulting project overseeing data management and modeling strategy
- Used natural language processing to transform unstructured data into features and extract business intelligence
- Built a recommendation engine as business rules potentially yielding savings on up to 50% of the business
- Technologies: Python, NLP, Machine Learning, Recommendation Systems, Feature Engineering, Business Intelligence, Data Management

## **Data Engineering Fellow | Insight Data Science**

**2015**

- Built themidgame-tube, a platform designed to discover YouTube influencers on brand names worldwide
- Deployed Amazon's EMR Spark with HBase processing and ingesting billions of data tuples
- Attained linear scalability performance tested with up to 20 nodes

- Technologies: Python, AWS, Apache Spark, EMR, HBase, Big Data, Distributed Computing, Data Ingestion, Scalability Testing

## **Data Analyst | Cartesian**

**2015**

- Aided managed analytics efforts promoting best practices within batch workflows and data management
- Conducted independent research into big data workflows considering data mining and BI integration
- Built short data pipelines consuming APIs, transforming, loading, and exposing data connections to BI tools
- Technologies: Python, PostgreSQL, REST APIs, ETL, Data Pipelines, Business Intelligence, Data Mining, Data Analytics

## **Data Analytics Engineer | Daktari Diagnostics**

**2013 - 2015**

- Worked as the lead developer of mainstream data processing and data analysis applications in Python for Windows/Mac
- Developed a calibration model for the Daktari CD4 testing device improving the system's accuracy by 20-30%
- Deployed machine learning models embedded in standalone applications to end users for data classification
- Technologies: Python, Machine Learning, Statistical Modeling, Calibration, Data Classification, Desktop Applications, Medical Devices, Data Processing

## **Portfolio**

### **Pastoral Conscience AI**

[https://rocha-moy-engineering-technology.github.io/pastoral\\_conscience\\_site/](https://rocha-moy-engineering-technology.github.io/pastoral_conscience_site/)

Built an “Artificial Conscience” AI system that generates scripture-grounded spiritual reflections using RAG retrieval from Psalms and DSPy-based conscience checkers. Designed and implemented a governed reasoning pipeline with three verification layers: Helpfulness, Psalm Grounding, and Citation Integrity checks. The system enforces strict alignment of beliefs; every response is scored, verified, and corrected before delivery. Architecture: Go back end with hexagonal architecture, Gemini File Search API for RAG retrieval, Python and DSPy services for conscience checkers with model fallback, Next.js front end with real-time SSE progress updates, and PostgreSQL for data persistence.