Ahmed Boutar

+1(984) 244-8924 • ahmed.boutar@duke.edu • https://ahmedboutar.com/ • https://github.com/ahmed-boutar

Full-stack and AI engineer with 1+ years of professional experience building scalable software and applied ML solutions across industry, academia, and freelance projects. Proven impact in deploying end-to-end applications and research-driven AI systems. Skilled in Java, Python, React, FastAPI, AWS, and PyTorch, with a track record of reducing operational costs, automating workflows, and delivering user-focused products

EDUCATION

Duke University   Durham, NC
Master of Engineering, Artificial Intelligence, Cumulative GPA 4.0/4.0   *Aug 2024 – Dec 2025*
Coursework: Modeling Process & Algorithms, Explainable AI, Deep Learning, Sourcing Data for Analytics, Large Language Models
Teaching Assistant: Explainable AI (Spring 2025, Fall 2025)

University of Rochester   Rochester, NY
Bachelor of Science, Computer Science, Cumulative GPA 3.62/4.0   *Aug 2018 – May 2022*
Minors: Mathematics, Audio Music Engineering; Certifications: AWS Cloud Practitioner
Coursework: Artificial Intelligence, Design & Analysis of Efficient Algorithms, Data Mining, Database Systems

WORK EXPERIENCE

Software Development Engineer Intern | Amazon   New York, NY
Technologies Used: React, TypeScript, Java, S3, Lambda, CloudFormation, DynamoDB   *May 2025 – Aug 2025*
- Built a custom UI and backend workflow for the Billing Escalation team to auto-generate tickets and reject invalid requests, reducing manual validation effort by 7% and eliminating 153 false escalations/year, saving over 300 hours annually
- Spearheaded an end-to-end automation feature for generating Requests for Payment, cutting processing time by 87.5% and reducing manual effort by 93%, while eliminating 100% of human errors
- Collaborated with 3+ TPMs and cross-functional teams across disconnected workflows to gather requirements, align feedback, and drive adoption across monetization and billing operations, resulting in teams' alignment on the project goals

Applied ML Researcher | Duke CREATE Lab   Durham, NC
Technologies Used: React, Azure, FastAPI, Git, OpenAI APIs, MongoDB, TypeScript   *Nov 2024 – Present*
- Led the end-to-end development of a personalized learning platform that used LLMs to dynamically generate quizzes and provide AI-powered feedback, improving student engagement and content retention
- Built full-stack architecture to support real-time AI quiz generation from uploaded lecture material, including an automated grading pipeline using an LLM, reducing instructor's quiz prep time
- Spearheaded research into LLM evaluation strategies to ensure feedback quality and instructional value for academic use
- Designed and deployed RESTful APIs for reliable and secure content storage, integrating LLM inference and scoring logic directly into backend services, ensuring scalability and a seamless user experience

Lead Developer | Freelance   Tunis, Tunisia
Technologies Used: React Native, Figma, TestFlight, Expo, Git   *Oct 2023 – Aug 2024*
- Led a team of 4 to develop a mobile app for Tunisia's leading radio company, achieving 100,000+ downloads and enhancing user engagement through features like live radio streaming and account management
- Directed the design phase on Figma, managed client communications and iterative feedback, and ensured successful app deployment on TestFlight, driving continuous improvements in UI/UX and functionality
- Spearheaded API integrations for online radio and live video streaming, coordinating with the client to modify and create API endpoints, resulting in enhanced functionality and ensuring scalability and adaptability to evolving client requirements

SELECTED PROJECTS

- **Agentic Chatbot (Duke):** Developed a LangChain-based multi-agent chatbot to answer curriculum, events, and campus queries using Duke APIs and Pinecone RAG. Architected a router + specialist agents design for accurate, low-latency responses, hosted on Flask (EC2) with Lambda request filtering. Built an evaluation harness (BLEU, ROUGE, BERTScore, user ratings) to benchmark performance, achieving >4.1 average user score while minimizing inference costs
- **Legal-BERT:** Collected and labeled congressional bill metadata (93rd–118th congress) and fine-tuned DistilBERT to classify bills into subject categories using only titles and factual legal knowledge. Built preprocessing and training pipeline in Colab with W&B monitoring and evaluated on an independent 119th congress test set. Achieved strong generalization across 31 policy areas

SKILLS

- **Languages:** Python, Java, TypeScript, JavaScript, SQL; **Frameworks & Tools:** React, React Native, Node.js, Next.js, FastAPI, Flask, LangChain; **AI/ML:** PyTorch, TensorFlow, Hugging Face, Scikit-learn W&B; **Cloud & DevOps:** AWS (S3, Lambda, CloudFormation, CloudWatch, StepFunctions, SQS), Azure, Docker, Git; **Databases:** MongoDB, PostgreSQL, DynamoDB