Arghya Mukherjee, PhD

Senior Data Scientist | AI Engineer | Founder, ChainDiligence

Target roles: Principal Data Scientist | Forward Deployed Engineer (FDE) | Security Research

Tulsa, Oklahoma | 918.330.1803 | arghyamukherjee1991@gmail.com | linkedin.com/in/arghyamukherjee | github.com/amukherjee1991 | chaindiligence.com

PROFESSIONAL SUMMARY

Data scientist and AI engineer with 10+ years across government, energy, industry, and research. Builds machine learning and AI systems end to end, including vision-language models (VLMs), generative AI and agentic AI, retrieval-augmented generation (RAG), Model Context Protocol (MCP) servers, ontologies, and computer vision. Deep expertise in fraud detection, anti-abuse, and anti-money laundering (AML) from a PhD on cryptocurrency crime, peer-reviewed publications with 236+ citations, and founding ChainDiligence, a blockchain intelligence company.

Core competencies: Generative & agentic AI, Vision-language models, MCP servers, Ontology development, Computer vision, Fraud detection & AML, Blockchain intelligence, Digital twins, MLOps & data pipelines, Cybersecurity research.

PROFESSIONAL EXPERIENCE

Founder, ChainDiligence May 2026 – Present

Tulsa, OK

- Founded a research-driven blockchain intelligence company that helps compliance, fraud, security, and investigation teams act on blockchain activity.
- Leading a suite of eight connected products in active development, covering wallet screening, transaction monitoring, fraud prevention, and crypto investigations.
- Turning peer-reviewed research on pump-and-dump schemes, exchange failure, and cryptocurrency scams into explainable risk signals.
- Designing risk scoring that shows its evidence: labels, fund-flow paths, confidence, and sources behind every decision.

Key projects: Address screening; Transaction monitoring; Investigation workspace; Market-integrity alerts.

Computer Vision Engineer, Chevron Oct 2025 – Apr 2026

USA

- Built a computer vision inspection system that detects damage to energy infrastructure from imagery.
- Replaced slow manual review with automated detection, so inspectors can focus on the assets that need attention.
- Took models from data collection and labeling through training, evaluation, and deployment into inspection workflows.

Data Scientist III, Office of Management and Enterprise Services, State of Oklahoma Jan 2025 – Present

Oklahoma, USA

- Design, build, and deploy machine learning and AI solutions for state agencies, from scoping with stakeholders to production and monitoring.
- Build vision-language models that interpret infrastructure imagery and answer questions about it in plain language.
- Led two generations of roadway asset detection, moving from a GIS-based first version to a computer vision system that cuts manual inspection time and cost.
- Built a custom image viewer so staff can review imagery and model detections side by side.
- Shape AI projects in digital twins and noise mitigation, from early design through delivery.

Key projects: Vision-language models; Digital twin design; Noise mitigation AI; Roadway assets v1; Roadway assets v2; Custom image viewer.

Data Scientist, ML/AI Engineer, Workonward Inc Apr 2024 – Dec 2024

Remote

- Designed voice search and an AI chatbot that let customers reach key information hands-free, increasing user interaction by 200% within three months of launch.
- Designed and implemented the backend for ML features that score résumés against job descriptions.
- Built a custom recruiter and job-seeker chatbot on an LLM using retrieval-augmented generation.
- Worked with IT and DevOps to ensure quality assurance and data integrity.

AI Software Engineer, Anti-Phishing Working Group Apr 2024 – Present

Remote (part-time)

- Develop ontologies that give APWG's eCrime data a shared structure, linking scams, wallets, domains, and campaigns.
- Build MCP servers that let AI assistants query and reason over APWG data through controlled, auditable tools.

- Apply language models to classify, enrich, and connect incoming threat reports for fraud, anti-abuse, and AML use cases.

Key projects: eCrime ontology; MCP servers.

Data Scientist, ML/AI Engineer, Robbie AI May 2023 – Apr 2024

Remote

- Developed computer vision models to identify body structures, and neural networks that detect pain level from FACS and PSPI scores.
- Raised accuracy of existing production models by at least 23%.
- Built a real-time video pipeline with Azure Blob Storage for live feeds and Apache Kafka streaming data to deep learning models.

Software Engineer (DevOps), Anti-Phishing Working Group Oct 2020 – Mar 2024

Remote (part-time)

- Built and ran data services for APWG's Cryptocurrency Working Group, used by members in academia, industry, and law enforcement-facing research.
- Maintained a curated database of roughly 1.2 million Bitcoin wallet addresses linked to scams and ransomware, used by members for fraud prevention, anti-abuse, and AML screening.
- Kept client-facing systems at 99.999% uptime through testing, monitoring, and CI/CD with Git, Jenkins, and Apache Airflow.

Key projects: Cryptocurrency scam address feed; Member data APIs; eCrime data lake; Workflow automation.

Research Assistant & Data Scientist, The University of Tulsa Jan 2018 – Dec 2024

Tulsa, OK

- Built fraud detection research end to end: pump-and-dump schemes, scams, market manipulation, and the failure of coins and exchanges.
- Measured anti-abuse and anti-money-laundering signals in blockchain data, tracing how illicit funds move through exchanges and services.
- Developed ETL pipelines for Bitcoin, Ethereum, and Solana data with Python, Snowflake, and SQL, plus web crawlers and API clients for market and scam data.
- Combined econometric models with neural networks and NLP to detect manipulation and classify large volumes of text.
- Published peer-reviewed work at ACM CCS workshops, APWG eCrime, HICSS, and Information Processing and Management.

Key projects: Pump-and-dump detection; Exchange risk; Scam & fraud measurement; AML & anti-abuse.

IT Risk Assurance Associate, PricewaterhouseCoopers Jan 2015 – Dec 2016

Birmingham, UK

- Delivered IT risk advisory work: risk assessments, security reviews, compliance assessments, and IT governance assessments.
- Evaluated client IT environments for risks and control gaps, and recommended remediation.

CYBERSECURITY EXPERTISE

Security experience spans a full career, from IT risk advisory at PwC to eCrime data at the Anti-Phishing Working Group, a PhD measuring cryptocurrency crime, and now building blockchain intelligence at ChainDiligence.

- **Cryptocurrency crime measurement:** Empirical studies of pump-and-dump schemes, scams, exchange closures, and the lifespan of coins, published at venues including ACM CCS workshops, APWG eCrime, and HICSS.
- **Blockchain intelligence & investigations:** Wallet screening, entity attribution, fund-flow tracing across chains, and explainable risk scoring for compliance and investigation teams.
- **Fraud, anti-abuse & AML:** Detection of pump-and-dump schemes, scams, and illicit fund flows across University of Tulsa research and APWG work, built on labeled data from research and industry feeds.
- **Phishing & eCrime data:** Collecting, enriching, and serving threat data on scam and ransomware wallets for APWG's global membership.
- **Security economics:** Measuring the impact of breaches and security shocks on markets, and modeling how organizations should invest in security.
- **IT risk & governance:** Risk assessments, security reviews, compliance testing, and control-gap analysis for enterprise clients at PwC.

PUBLICATIONS

236+ citations on Google Scholar (h-index 4). Research covered by Bloomberg and Schneier on Security.

1. **Arghya Mukherjee**, Tyler Moore. "Unicorns in the Wild West: Empirical Analysis of Cybercrime Facilitated by Cryptocurrencies." *APWG Symposium on Electronic Crime Research (eCrime)*, 2025.
2. **Arghya Mukherjee**, Tyler Moore. "Beyond the Hype: Empirical Evaluation of Cryptocurrency Unicorn Success." *Hawaii International Conference on System Sciences (HICSS)*, 2025.
3. **Arghya Mukherjee**, Tyler Moore. "Cryptocurrency Exchange Closure Revisited (Again)." *APWG Symposium on Electronic Crime Research (eCrime)*, 2022.
4. JT Hamrick, Farhang Rouhi, **Arghya Mukherjee**, Amir Feder, Neil Gandal, Tyler Moore, Marie Vasek. "An examination of the cryptocurrency pump-and-dump ecosystem." *Information Processing and Management*, 2021.
5. JT Hamrick, Farhang Rouhi, **Arghya Mukherjee**, Amir Feder, Neil Gandal, Tyler Moore, Marie Vasek. "Analyzing Target-Based Cryptocurrency Pump and Dump Schemes." *ACM CCS Workshop on Decentralized Finance and Security*, 2021.
6. JT Hamrick, Farhang Rouhi, **Arghya Mukherjee**, Amir Feder, Neil Gandal, Tyler Moore, Marie Vasek. "The Economics of Cryptocurrency Pump and Dump Schemes." *Workshop on the Economics of Information Security (WEIS)*, 2019.

PRESENTATIONS

- "Unicorns in the Wild West: Empirical Analysis of Cybercrime Facilitated by Cryptocurrencies." *APWG Symposium on Electronic Crime Research (eCrime 2025)*, Nov 2025.
- "Beyond the Hype: Empirical Evaluation of Cryptocurrency Unicorn Success." *Hawaii International Conference on System Sciences (HICSS-58)*, Jan 2025.
- "Cryptocurrency Exchange Closure Revisited (Again)." *APWG Symposium on Electronic Crime Research (eCrime 2022)*, Nov 2022.
- "The Economics of Cryptocurrency Pump and Dump Schemes." *Workshop on the Economics of Information Security (WEIS 2019)*, Jun 2019.

EDUCATION

PhD, Computer Science, The University of Tulsa 2018 – Dec 2025
Dissertation: Control + Alt + Deceive: Empirical analysis of cryptocurrency ecosystem and cyber crime. Advisor: Dr. Tyler Moore

MSc, University of St Andrews 2014
Thesis: Measuring enterprise IEEE 802.11 (Wi-Fi) standards. Advisor: Dr. Colin Allison

Bachelor of Computer Applications, West Bengal University of Technology 2012
Kolkata, India

TECHNICAL SKILLS

AI & machine learning: PyTorch, TensorFlow, scikit-learn, Hugging Face Transformers, Vision-language models, OpenCV
Generative & agentic AI: LLMs, RAG, AI agents, MCP servers, LangChain, Amazon Bedrock, Fine-tuning
Knowledge & data modeling: Ontology development, Knowledge graphs, Data modeling, SQL, GraphQL
Data & analytics: Python, R, Pandas, PySpark, Stata, Power BI, Tableau, ESRI
Cloud & MLOps: AWS, Azure, Google Cloud, Docker, Kubernetes, Apache Airflow, Snowflake, CI/CD
Security & blockchain: Blockchain analysis, Fraud detection, AML, Anti-abuse, Threat intelligence, SIEM, Elasticsearch, IT risk assessment