John Chester

New York, NY | John.Anthony.Chester@Gmail.com
Website: johnchester.dev | linkedin.com/in/johnchester212 | GitHub: @JohnAChester

SUMMARY STATEMENT

Data scientist with 7+ years in healthcare analytics at Cigna and Aetna, guiding negotiation strategy for a $2.3B provider portfolio and building Python dashboards adopted nationwide. Pairs payer-side expertise in reimbursement and value-based contracting with graduate training in deep learning, Bayesian modeling, and NLP. Recently shipped a production Bayesian public-health platform that anchored a $1M HHS grant bid, first-authored a GameSec 2026 paper on pay-for-performance design, and is now building citation-grounded enterprise RAG and RLHF LLM alignment systems.

EDUCATION

FORDHAM UNIVERSITY **New York, NY**
Master of Science in Data Science ***2025 – December 2026***
Selected Coursework: Deep Learning, NLP, Machine Learning, Artificial Intelligence, ML in Finance **GPA 4.0**

UNIVERSITY OF MARYLAND, COLLEGE PARK **College Park, MD**
Bachelor of Science in Economics ***2017***
Specialization in Econometrics, Statistics, and Behavioral Economics

PUBLICATIONS & PRESENTATIONS

Who Pays to Secure Shared Components? A Pay-for-Performance Model for Cyber Insurance
Chester, J. A., & Hastings, A. K. Published in the proceedings of *GameSec 2026: 17th Conference on Game Theory and AI for Security*, Ann Arbor, MI, Oct. 2026. Presenting author.

PROFESSIONAL EXPERIENCE

TICK MITT **New York, NY**
Data Science & AI Engineering Intern ***May 2026 – October 2026***

- **TiCK Heatmap (HHS grant finalist):** As sole engineer, designed, built, and shipped a live, community-powered platform tracking tick encounters and Lyme disease prevalence, with a serverless AWS backend (CDK, Lambda, Aurora Serverless v2, DynamoDB) and a React/Vite frontend; became the company's central pitch for a $1M HHS TOPxHHS Tech Sprint grant that reached final-stage review
- **Bayesian Modeling:** Built an empirical-Bayes hierarchical engine with CDC & NOAA priors (Poisson-Gamma sighting density, Binomial-Beta disease prevalence), geographic pooling, design-effect weighting for within-submission correlation, and tick-selection bias correction; privacy-by-design via separated PII stores and zip-only geocoding
- **Social Media Agent:** Deployed a low-code comment-management agent (n8n, Google Sheets, Claude API) that analyzes and drafts replies to 1K+ monthly comments across YouTube, Facebook, and Instagram, saving 15 hrs/month at under $10/month in inference cost
- **Automated Sales Figures Reporting:** Built a weekly Python/GitHub Actions pipeline with failsafes that pulls Shopify and Amazon API data and posts Plotly trend summaries to Slack for team standups
- **Financial Dashboard:** Built Claude Skills with data dictionaries and modular ETL scripts so the CEO could run ad-hoc analytics conversationally and get interactive dashboards, saving 1–2 hrs/week of manual data collection
- **Customer Insights:** Connected OuterSignal customer-intelligence and Zigpoll survey data via MCP and surfaced the top ten post-purchase sentiments weekly, revealing product skepticism that informed a marketing education push

FORDHAM UNIVERSITY **New York, NY**
Graduate Research Assistant, Economics of Cybersecurity ***2025 – Present***

- Identified an unexplored parallel between health insurers' pay-for-performance provider contracts and the problem of incentivizing security in shared software components, drawing on 7 years of payer-side contracting experience
- Led the mechanism design of a pay-for-performance incentive schema for cyber insurance and contributed to its mathematical derivation, numerically demonstrating viability within plausible parameter zones
- First-authored the resulting paper, accepted to GameSec 2026; presented October 2026

CIGNA **New York, NY**
Business Analyst – Hospital System Negotiations ***2021 – 2025***

- Guided data-driven negotiation strategies for a $2.3B healthcare portfolio across key healthcare systems, physician groups, and ancillary facilities throughout New England, boosting Cigna's competitive market position by 1.7%

- Engineered Python dashboards with data storytelling and customizable visualizations for high-stakes contract negotiations, driving savings, improving price transparency, and enhancing cost prediction; adopted nationwide
- Developed and modeled over 35 deals with hospital systems delivering $21M in savings above market expectations
- Leveraged patient and sales data using SQL to create dynamic models which increased planning efficiency by 20%
- Achieved more favorable contract terms in 65% of negotiations completed, for a projected savings of $6.4M
- Evaluated 8 hospital systems for fee-for-value programs, shifting 14% of dollars to a risk-sharing arrangement

CIGNA **New York, NY**
Contract Negotiator – Hospital Systems ***2019 – 2021***

- Negotiated high-value healthcare contracts with leading hospital institutions including Mount Sinai Health System and Hospital for Special Surgery, both strengthening Cigna's network and restructuring existing agreements
- Spearheaded surgery center renegotiation efforts in the NYC region, achieving $1.8M savings over one year
- Constructed and executed competitive analysis strategies for hospitals, securing favorable terms in 80+% of deals

AETNA **New York, NY**
Project Manager ***2018 – 2019***

- Managed over 10 projects with a team of 15, ensuring timely and efficient completion of each project, while leveraging advanced analytics to optimize processes and increase overall project savings by 15%
- Led the termination process, coordinating across Aetna to ensure regulatory-compliant communications to 350K+ patients

PROJECTS

Lumbr: Fine-Tuned vs. Hybrid Retrieval for an Enterprise RAG Assistant | *CISC 6000, Deep Learning; with P. Angeles & J. Jeong (in progress, expected Dec 2026)*

- Leading a controlled comparison of six retrieval setups (BM25, small and ~35× larger embedders, hybrid fusion, cross-encoder reranking, and a contrastively fine-tuned embedder trained on label-free synthetic questions) over EnterpriseRAG-Bench's 512K-document corpus, holding the generator fixed
- Building the end-to-end RAG pipeline, from structure-aware ingestion of nine enterprise sources to citation-grounded generation and evaluation (Recall@10, nDCG, LLM-judge faithfulness), plus a live onboarding-simulation demo

Context-Sensitive RLHF for Contextual Moral Value Alignment | *CISC 6210, NLP (in progress, expected Dec 2026)*

- Extending a published moral value alignment system by fine-tuning its LLM aggregator with RL (GRPO), using a counterfactual-profile reward that penalizes responses whose moral-value scores ignore user-profile changes
- Using the Moral Integrity Corpus: LoRA SFT baseline replication, KL-regularized RL against five moral-value reward classifiers, and evaluation by ROUGE, profile-weighted alignment, and a profile-swap test scored by a held-out classifier

Comparative Analysis of ML Classifiers for Type 2 Diabetes Prediction | *CISC 5800, Machine Learning*

- Built a reproducible preprocessing pipeline over the CDC BRFSS survey (441K records, 330 columns), narrowing to 21 literature-grounded features and yielding 253K clean records for modeling
- Trained and compared four model families (logistic regression, linear/kernel SVM, PyTorch MLP), tuning the precision/recall trade-off via pos_weight and F2-driven early stopping to raise minority-class recall from 76% to 86% at a screening-appropriate operating point, with all four converging at ROC-AUC ≈ 0.82

FinBERT Sentiment vs. Price-Based ML for Post-Earnings Prediction | *CISC 5352, ML in Finance; with P. Angeles*

- Built an NLP pipeline that classified ~27,000 text chunks from 480 SEC 10-K/10-Q filings (24 S&P 500 companies, 2021 to 2025) with zero-shot FinBERT and confidence-weighted majority voting to produce 2,400 filing-level sentiment labels
- Engineered 16 price-based features and trained four classifiers under walk-forward validation to predict 5-day post-earnings abnormal returns; found only 44.9% sentiment-price agreement (below the 50% baseline) and near-zero correlation (r = -0.043), consistent with the semi-strong Efficient Market Hypothesis

TECHNICAL SKILLS

- Programming Languages: Python, SQL
- ML & Statistics: PyTorch, scikit-learn, Hugging Face, NLP, RL (GRPO), LoRA, Bayesian modeling, classification
- Databases: PostgreSQL (Aurora Serverless v2, PostGIS), DynamoDB
- Data & Visualization: Pandas, NumPy, SciPy, Plotly, Matplotlib, Seaborn, Excel
- Cloud & Engineering: AWS (Lambda, CDK, serverless architecture), GitHub Actions (cron scheduling), n8n
- LLM Engineering & Evals: RAG, retrieval evaluation (Recall@k, nDCG), LLM-as-judge, Claude API, MCP, agentic workflows, prompt engineering