

CHINMAY ARORA

DATA SCIENCE | ANALYTICS | APPLIED AI | RESPONSIBLE AI

Minneapolis, MN | 737-340-7827 | chinmayarora2001@gmail.com | [LinkedIn](#) | [GitHub](#) | [Portfolio](#)

MS Data Science graduate with a computer science foundation and applied experience across public-sector analytics, experimentation, geospatial health research, ML evaluation, retrieval systems, operations intelligence, and responsible AI. Strongest fit for early-career Data Scientist, Data Analyst, Analytics Engineer, Applied AI/ML Engineer, AI Governance Analyst, and technical AI strategy roles. Long-term orientation toward technical product building and venture exploration.

CORE CAPABILITIES

Data Science & Statistics: Python, SQL, R, Pandas, NumPy, scikit-learn, XGBoost, regression, classification, clustering, experimental design, A/B testing, cohort analysis, calibration, feature engineering, SHAP/LIME, subgroup analysis, fairness and drift evaluation

Analytics & Data Systems: PostgreSQL, DuckDB, Azure SQL, MongoDB, ETL/ELT, dimensional modeling, KPI design, Power BI, Tableau, Looker Studio, Qlik Sense, Excel/VBA, data-quality testing

Applied AI & Retrieval: RAG, embeddings, FAISS, ChromaDB, Pinecone, LangChain, OCR ingestion, Text-to-SQL, retrieval/generation evaluation, FastAPI, Flask, Docker, MLflow, Evidently, Redis, GCP Cloud Run

Responsible AI & Product Work: model cards, human review, audit trails, risk registers, deployment gates, NIST AI RMF concepts, privacy/fairness analysis, stakeholder research, product requirements, market mapping, workflow design, technical communication

PROFESSIONAL EXPERIENCE

Graduate Research Assistant - CURA / Hennepin County Digital Experience May 2025 - Jan 2026 | Minneapolis, MN

- Built resident segmentation analyses using census, accessibility, service, survey, and behavioral indicators to support more inclusive digital-service planning.
- Created segment profiles, issue heatmaps, mini journey maps, use-case inventories, and implementation measures for leadership, research, design, and technical stakeholders.
- Evaluated public-sector AI opportunities using evidence on resident needs, digital-literacy barriers, privacy concerns, fairness, human review, and operational feasibility.
- Designed internal AI-readiness and public AI-knowledge surveys and translated findings into implementation recommendations and presentation materials.

Founder - Aegis AI 2025 - Present | Remote, USA

- Prototyped AI-readiness scoring, risk registers, evaluation gates, benchmark templates, audit-trail artifacts, and decision-trace workflows for compliance-sensitive AI use cases.
- Researched retrieval reliability, trust provenance, model evaluation, agent identity, and responsible deployment; separated implemented prototypes from concept-stage architectures.
- Explored managed AI-transformation services spanning diagnostic audits, pilot orchestration, governance artifacts, and operational reporting for public-sector, healthcare, education, and back-office workflows.

Data Science Intern - MarketMakerCRE Jul 2024 - Sep 2024 | Remote, USA

- Built asynchronous Python scraping and ETL pipelines using BeautifulSoup, Selenium, aiohttp, MongoDB, and Azure SQL, reducing a full extraction cycle from roughly eight hours to two hours.
- Developed property valuation and segmentation analyses using XGBoost, Random Forest, KMeans, PCA, and SHAP/LIME; documented model behavior and limitations for investor-facing analysis.
- Worked with real-estate and market data pipelines, database integration, feature engineering, and model interpretation in a fast-moving startup setting.

Data Analyst Intern - Confetti AI May 2024 - Jul 2024 | Remote, USA

- Analyzed campaign and product engagement, ran A/B tests, and built KPI reporting used to refine audience targeting and personalization decisions.
- Partnered with product and engineering teams to investigate feature performance, define success measures, and communicate findings through dashboards and concise recommendations.

Data & AI Solutions Intern - Luthra Corporate Advisors Dec 2021 - Aug 2023 | New Delhi, India

- Prototyped an internal knowledge assistant using embeddings and Pinecone semantic retrieval; returned source context and confidence cues to improve answer traceability.
- Supported reporting, finance, accounting, and workflow-automation projects using Python, SQL, dashboarding, and operational data from logistics and professional-services workflows.
- Built domain familiarity with CargoWise-supported freight operations and later earned CargoWise professional certification.

Business Analytics Intern - Daikin India Jun 2022 - Jul 2022 | Gurugram, India

- Built Qlik Sense sales dashboards and automated recurring reporting to improve visibility into regional sales performance and operational trends.
- Prototyped Python/Flask NLP and workflow tools to improve internal information access and reduce repetitive manual work.

RESEARCH EXPERIENCE

Graduate Research Volunteer - Data Analysis & Management Group, University of Minnesota | Spring 2025 | Minneapolis, MN

- Benchmarked direct retrieval, Tree-of-Thought, and Graph-of-Thought context strategies for small-language-model and RAG workflows.

- Worked with ChromaDB, FAISS, Pinecone, LangChain, embeddings, and similarity search to compare context construction and retrieval behavior.

Graduate Capstone Research - University of Minnesota | 2025 - 2026 | Minneapolis, MN

- Studied neighborhood deprivation and diabetes outcomes using EMR-linked HbA1c data, Area Deprivation Index, geospatial controls, and temporal prediction at six-, twelve-, and twenty-four-month horizons.
- Compared logistic regression, Random Forest, and XGBoost models; evaluated calibration, subgroup performance, temporal validation, and the incremental value of geographic variables.
- Documented weak incremental predictive value from neighborhood deprivation in the available data and emphasized limitations, overfitting risk, and responsible interpretation.

SELECTED TECHNICAL PROJECTS

- **Healthcare Risk, Fairness & Geography** [Completed analysis] - Built leakage-controlled readmission models on approximately 101K encounters with calibration, subgroup false-negative analysis, SHAP, geospatial context, and a model card; surfaced a 0.44 age-group FNR gap and documented why the model should not be deployed without remediation.
- **Tabular ML Reliability Benchmark** [Completed benchmark] - Compared five pipelines across three real domains using random and shifted splits, AUC, calibration, runtime, MLflow tracking, and evidently drift reports; found preprocessing produced larger gains than tuning or AutoML in the tested settings.
- **Operations Control Tower** [Completed analytics project] - Modeled 98,666 orders in a DuckDB star schema, defined delivery and supplier KPIs, added data-quality checks, and built a five-page Looker Studio dashboard that surfaced a 6.5x higher bad-review rate for late orders.
- **PolyRAG** [Deployed demo] - Built a multi-provider RAG service over AI-governance documents with shared FAISS retrieval, side-by-side citation-grounded answers, OCR/web ingestion, FastAPI, Docker, tests, and Cloud Run deployment.
- **Geospatial SDOH & Diabetes Outcomes** [Graduate capstone] - Analyzed ADI, SVI, and EMR-linked HbA1c outcomes using Python, R, GeoPandas, PySAL, and ArcGIS Pro; used temporal evaluation and subgroup analysis to test whether geography improved prediction.
- **SKM Football Analytics** [Research project] - Developed an event-value framework using StatsBomb data and VAEP probability changes with difficulty, context, and role adjustments; packaged reproducible analysis, sensitivity considerations, and reporting artifacts.
- **Logistics Finance BI Dashboard** [Applied analytics project] - Built SQL, Power BI, and Excel/VBA reporting for freight-forwarding finance and operations, improving visibility into job profitability, receivables, and operational trends.
- **CAI Architect** [Interactive research prototype] - Explored structured AI-assisted opportunity and positioning analysis using user context, company information, market pressures, and human-reviewed recommendations.
- **Karya Survey AI** [Product exploration] - Oversaw product planning for a multilingual conversational survey tool using chat, voice, and video interfaces, with Hindi ASR/TTS support, consent, quality controls, and interviewer-style follow-up logic.
- **Agent Identity / Passport Infrastructure** [Venture thesis and prototype exploration] - Researched agent authentication, addressability, auditability, certification, and inter-agent trust for enterprise AI workflows, including potential collaboration with auditing or infrastructure providers.

LEADERSHIP, PRODUCT & VENTURE EXPERIENCE

- Project Lead, Enactus SRM - led ReStub, a social-enterprise initiative exploring conversion of cigarette waste into construction materials while involving local brickmakers and low-cost housing use cases.
- Developed and documented venture explorations across logistics intelligence, responsible AI infrastructure, survey automation, visual guidance, sports analytics, food systems, and applied enterprise AI.
- Experienced translating technical ideas into product requirements, client-facing explanations, market maps, outreach strategies, and pilot concepts across public-sector, logistics, finance, hospitality, and startup contexts.
- Languages: English, Hindi, French.

EDUCATION

University of Minnesota–Twin Cities | M.S. Data Science | Sep 2023 – May 2026 | GPA 3.7/4.0

Selected coursework: Machine Learning, Data Mining, Advanced Database Systems, Applied Regression and Multivariate Analysis, Categorical Data Analysis, Spatial Data Science, Responsible AI, Visualization with AI, AI for Sequential Decision Making.

SRM Institute of Science and Technology | B.Tech. Computer Science & Engineering | 2019 – 2023 | GPA 8.98/10

Additional: CargoWise professional certification; project leadership through Enactus SRM and technical extracurricular work.