SHAWN ALBERT – DIRECTOR OF ARTIFICIAL INTELLIGENCE

[GitHub](#) • [LinkedIn](#)

SUMMARY

Director of Artificial Intelligence with an established record of driving transformative initiatives in the healthcare sector. A strategic thinker and innovative problem-solver, skilled in leveraging cutting-edge artificial intelligence (AI) and machine learning technologies to deliver end-to-end solutions. Known for leading interdisciplinary teams to develop pioneering products, reinventing legacy systems, and achieving millions in annual savings and revenue. An adept communicator and technical mentor, excelling in translating complex data into actionable insights, bridging the gap between data science innovation and production deployment, and cultivating a culture of continuous learning and growth. Equipped with a unique blend of technical expertise, strategic leadership, and entrepreneurial acumen, uniquely positioned to drive innovation and revolutionize healthcare through data-driven solutions and transformative approaches in AI and machine learning.

EDUCATION

The University of Chicago Booth School of Business
Master of Business Administration, MBA
Concentrations in Entrepreneurship and Strategic Management

- **Strategy Coursework:** New Venture Strategy, Strategy and Structure: Markets and Organizations
- **Entrepreneurship Coursework:** Building the New Venture, Entrepreneurship in Technology
- **Finance Coursework:** Financial Accounting, Entrepreneurial Finance and Private Equity
- **Analytics Coursework:** Managerial Decision Modeling, Data-Driven Marketing, Game Theory

The University of Chicago
Master of Science, MS
Computer Science

- **Mathematics Coursework:** Discrete Mathematics, Algorithms
- **Programming Coursework:** Python Programming, Software Engineering
- **Systems Coursework:** Computer Systems, Databases, Big Data Application Architecture

Vanderbilt University
Bachelor of Arts, BA
Double Major - Medicine, Health and Society | Molecular and Cellular Biology

EXPERIENCE

Belle   **2023 - 2024**
Director of Artificial Intelligence | Data & Innovation

- Engineered agentic workflows with Amazon Bedrock and Step Functions to automate call transcription and clinical note generation, reducing administrative overhead while enhancing AI/ML model features
- Architected and built Belle's MLOps foundation and modern data platform, implementing a data lakehouse using dbt-Athena-SparkSQL integrated with AWS Glue catalog and LakeFormation, orchestrated through Dagster
- Established comprehensive CI/CD framework for infrastructure development and deployment using GitHub Actions, Terraform and the AWS CDK, delivering end-to-end automation for the data platform
- Designed and implemented extensible SageMaker Pipeline framework for causal inference model deployment using econml package with BYOC approach, incorporating EMR serverless for data preprocessing and centralizing model management through SageMaker Model Registry

Major Projects

Agentic AI Workflows with Amazon Bedrock
- Developed serverless AI/ML pipeline leveraging AWS services (S3, Transcribe, Bedrock, Comprehend) to transform patient call recordings into structured clinical documentation, significantly reducing manual documentation overhead
- Engineered HIPAA-compliant workflows using AWS Step Functions to orchestrate end-to-end processing of health records, automating extraction of ICD codes, prescription data, and social determinants of health
- Deployed secure integration between Amazon Bedrock and Comprehend to generate comprehensive SOAP notes and extract SNOMED clinical terms, streamlining clinical documentation while maintaining accuracy
- Designed fault-tolerant audio processing system that automatically transcribes care manager calls and generates clinical documentation, enabling providers to maximize direct patient care time

Modern Data Platform & Cloud Infrastructure
- Architected and implemented Belle's data lakehouse using AWS Glue for metadata cataloging and LakeFormation for data governance, orchestrated through Step Functions and Dagster
- Engineered robust data processing pipelines using dbt-Athena-SparkSQL, enabling efficient data ingestion, transformation, and analytics workflows
- Established comprehensive CI/CD framework with GitHub Actions, Terraform, and AWS CDK, standardizing infrastructure development workflow and streamlining team collaboration
- Implemented end-to-end data quality framework combining Great Expectations and dbt tests, ensuring data reliability through automated validation checks and schema enforcement

Causal Inference Model Deployment
- Implemented comprehensive causal inference modeling framework using SageMaker and econml for patient outcome prediction, incorporating EMR Serverless for distributed data preprocessing
- Orchestrated end-to-end ML workflows through SageMaker Pipelines and Model Registry, streamlining model versioning and deployment processes
- Configured and deployed SageMaker Studio environment for team-wide ML development, enabling collaborative model experimentation and deployment
- Designed modular training architecture using multi-layer Docker containers and Ray clusters on SageMaker, enabling data scientists to rapidly adapt the pipeline for various causal analyses including difference-in-differences studies

---

Healthfirst   **2022 - 2023**
Lead Data Scientist & Senior Manager | Enterprise Analytics

- Managed a team of 3 machine learning engineers, bridging the gap between data science innovation and deploying and maintaining models in production
- Spearheaded the MLOps initiative at Healthfirst, reducing time-to-value of going from idea to production-ready model by leveraging AWS technologies (Glue, SageMaker Feature Store, SageMaker Pipelines, SageMaker Clarify, and SageMaker Model Monitor) implemented in GitHub Actions with a Continuous Integration, Continuous Deployment, and Continuous Training framework

Major Projects

Machine Learning Operations (MLOps)
- Spearheaded MLOps initiative of Enterprise Analytics Data Science Team (20 FTEs) over the course of 9 months, focusing on overcoming manual workflows, mitigating bottlenecks in building, training, and deploying ML models, and reducing the number of disparate deployment patterns for models in production
- Planned and implemented an MLOps roadmap segmented into 4 key phases--Initial, Repeatable, Reliable, and Scalable--using Amazon SageMaker's purpose-built tools and integrations to streamline the ML lifecycle
- Designed the Initial phase to create an experimentation environment for initial learnings and introduction to the AWS tech stack, enabling data scientists to experiment with SageMaker notebooks, and integrated services like Amazon Glue and Athena to leverage serverless technologies for data processing and analysis
- Developed the Repeatable phase to productionize ML solutions, transitioning from notebooks to workflows developed and orchestrated with SageMaker Pipelines
- Expanded the MLOps foundation to the Reliable and Scalable phases, introducing automatic testing, quality gates for manual evaluation at key milestones, templatization of solutions, and automation of best practices with GitHub Actions, achieving a reduction in development time for new production solutions from weeks to days

Large Language Model (LLM) Database Query Assistant

- Led a team of 2 FTEs and 2 undergraduate interns to develop a first-of-its-kind capability at Healthfirst, successfully developing a database querying chatbot for both relational and graph databases over the course of a 10-week summer internship
- Enabled the company to achieve business objectives more efficiently through generative AI, abstracting the complexities of writing SQL and Cypher queries
- Architected a comprehensive solution using cutting-edge technologies
  - Model Development: Utilized Llama2 open-source LLM and QLoRA for efficient fine-tuning, enabling innovative chatbot capabilities
  - Data Integration and Querying: Leveraged AWS Glue Catalog, Amazon Athena, Neo4J, and Cypher to enable natural language querying of both relational and graph databases
  - User Interface and Visualization: Employed Streamlit to develop an intuitive user interface, enhancing accessibility for non-technical users
  - Deployment and Scalability: Hosted the solution on an internal SageMaker endpoint and used ECS for frontend deployment, ensuring robustness and scalability

---

Aetna, a CVS Health Company   **2020 - 2022**
Senior Data Scientist | Clinical Product Analytics

- Led strategic analytics projects, including business problem solving, guidance on methodologies, innovation, and developing strategies throughout projects
- Developed campaigns promoting healthy behavior change through targeted interventions that leverage unsupervised machine learning for customer segmentation and predictive modeling for early outreach
- Implemented experimental design principles to design pilot studies and conduct hypothesis testing to evaluate internal campaigns and external product partnerships
- Spearheaded product development with demonstrated end-to-end project ownership, interfacing with clinicians and business stakeholders to size value and influence product strategy by communicating data-derived insights

Major Projects
Enhanced Maternity Program
- Led a team of 3 FTEs in day-to-day project management and serve as the primary data science liaison for business stakeholders
- Launched 6 data science maternity campaigns with projected savings of $7.1 million annually. Campaigns implemented in randomized controlled trial (RCT) and A/B frameworks to evaluate clinical impact and outreach channel efficacy in nudging members toward healthier behaviors
- Developed the Enhanced Maternity Program, sold to 9.7 million annual lives in Aetna Commercial book of business with $37 million in annual revenue and a projected $4.7 million NPV over 5 years

---

Blue Cross Blue Shield of IL, MT, NM, OK & TX   **2018 - 2020**
Data Scientist | Provider & Network Data Science
- Developed and maintained a provider cost efficiency metric derived by comparing actual to predicted medical episode costs, incorporating 500+ variables for risk adjustment that demonstrated over 30% improvement in cost estimation capabilities
- Created visualization dashboards in Tableau for communicating provider clinical quality and cost efficiency metrics and translating metrics into business solutions for development of high-performance provider networks

Major Projects
Health Data Exchange
- Led a team of 4 FTEs to develop a PySpark data pipeline that transformed and consumed 15.2+ million Athenahealth electronic health record (EHR) files to enable enterprise-wide reporting, analytics, and machine learning use cases
- Improved performance of XGBoost medical episode cost regression model through integration of clinical data (e.g., unstructured provider notes and semi-structured family history data) through more accurate risk score assignment at the member-episode level

CORE COMPETENCIES

- **Amazon Web Services (AWS):** S3, EC2, Glue, EMR, ECS, SageMaker, Redshift, RDS, Athena, Lambda, SNS, SQS, Cognito, Bedrock
- **Programming**: Python, SQL, Colima, Docker, Bash, Unix, PySpark, SparkSQL, Git, Terraform, Typescript
- **Software**: GitHub, GitHub Actions, Docker Desktop, Apache Airflow, Dagster, dbt, Tableau, Microsoft Office
- **Business:** Valuation, Financial Modeling, Financial Analysis, Project Planning, Strategic Planning
- **Technical Proficiency:** Machine Learning, Artificial Intelligence, Machine Learning Operations (MLOps), Data Engineering