Ben Bubnick

AI DATA CURATION ARCHITECT · ANALYTICS TRAINING SPECIALIST

3419 Superior Park Drive, Cleveland Heights, Ohio, 44118, U.S.

☐ (+1) 216-556-1612 | ✉ ben.bubnick@gmail.com | 🔗 benbubnick

"With 13+ years experience in data analysis and curation, I have leadership aspirations because I enjoy helping others grow and succeed. I find problem solving and working with other business units engaging. I have several years of experience doing this, and if the company sees value in me, I would be able to work in a leadership role and hit the ground running."

Summary

Professional Summary	Data Curation, Data Science, Analytics Training, Machine Learning, Scientific Analysis
Core Competencies	Cognitive Practitioner, Metrics Improvement, Agile Development, Training Development
Industries	Health Care, Financial & Insurance, Chemicals Manufacturing, R&D/Scientific Research
Demonstrated Past Success	Subject Matter Expert, Lab Services Award Winner, Eminence and Excellence Award Winner

I manage client resources and drive technical solutions to timely completion. I identify, analyze, and extract key data from heterogeneous systems; and lead technical requirements gathering, data discovery, documentation, and implementation of assigned client integrations. I also lead the technical delivery of assigned client integrations which has included the ordering, installation, and configuration of required hardware, including VPN devices to support client integrations. Configuration and development of software necessary to support client integrations, including ETL and Data Validation.

Experience

AI Data Curation Architect
IBM

WATSON HEALTH
Jan. 2018 - PRESENT

Lead member of Data Onboarding & Governance, member of the Watson Foundation for Health Solutions Architecture team. Remediation of data issues during UAT. Builds the data onboarding platform and associated infrastructure, and develop improvements to reducing customer implementation cycle time and data triage activities, and report on the "health of the data" during data onboarding. Tune machine learning classificiation algorithms for patient matching, and the internal review of similar work by my peers.

- Reduced mapping rework by 45% with through new knowledge-driven initiatives
- Developed new data validation and review process that reduced total rework by 25%
- Managed and developed ambitious multi-org enterprise with over 70 interconnected, project-sized components
- Earned AI certification to peer review tuning algorithm and reviewed over 50% of projects in the last 2 quarters
- Earned 7 badges during IBM Re-skilling Initiative and logged over 200 hours of training in 2018
- Curated 25 pages of AI workflow documentation and mentored new hires in AI tuning process
- Awarded Lab Services Award for coordinating internal project teams to deliver client needs on time
- Awarded GDP bonus for improving revenue per SaaS client by at least 1.34%

Data Scientist
IBM, Explorys

WATSON HEALTH, EXPLORYS AN IBM COMPANY
Jun. 2015 - May. 2018

Data Ingestions, pre-analytic enrichments, analytic augmentation, and data governance. Identified and extracted key data from multiple heterogeneous systems; and led technical requirements gathering, data discovery, documentation, and implementation of assigned client integrations. Ordered and oversaw installation of & configuration of required hardware, including VPN devices to support client integrations. Configuration and development of software necessary to support client integrations.

- Eliminated need for a $30,000 annual contract by developing data curation tool
- Principal integration specialist on over 60 projects for 20 different clients
- Chosen as team lead for ≈12 remote data science contractors for 36 integration projects throughout 2015 - 2017
- Went above my duties to close project documentation gaps, organizing more than 100 past projects
- Created over 125 new knowledge base articles, making up 30% of the team's online content
- Reduced time spent on knowledge transfer by ≈80% by reorganizing project documentation
- Awarded Eminence and Excellence Award in 2017 for analytics project
- Awarded GDP bonus for improving revenue per SaaS client by at least 1.17%

Data Integration Training Specialist
IBM

WATSON HEATLH
Sep. 2016 - PRESENT

Project Management for technical training, providing both basic and advanced software training, along with consulting through cross-team and other training/presentation settings. Designed and managed the internal training, an effective onboarding program, and cross-team development training. Focus on HDFS technologies: Pig, Hive & Impala, Ruby & Java.

- Eliminated average of 3 weeks from employee onboarding process
- Reduced cost-per-hire by $5,210 by re-developing education program
- Spearheaded new-hire mentoring program that led to ≈4,500 increase in total billable hours
- Trained 60% of current data curation teams, including 2 managers

Software Engineer
AmTrust Financial Services

COMMERCIAL PACKAGE POLICIES
Mar. 2014 - Jun. 2015

Developed web forms for insurance industry using ISO standards and created multi-tier MVC components, Web APIs, and UIs using VB, ASP, Javascript, & HTML5

- Development on two software architecture projects in 2015 and principal developer on offering product in 2014
- Developed over 65 ISO rating based web forms for insurance industry
- Collaborated with developers and business analysts to prioritize needs
- Steered development group into a unit testing project on multi-tier MVC components in 2014

Color Scientist
PPG Industries

AUTOMOTIVE REFINISH
Apr. 2012 - Jun. 2013

Developed software and radiative transfer models on complex paint mixtures to determine formulation methods and for spectroscopic profile analysis on color match samples.

- Improved color matching classification algorithm performance by 40% through new numerical methods
- Wrote and filed application for patent on the application of the Radiative Transfer Equation (Chandrasekhar) to automotive coatings classification models
- Developed cross-instrument correlation software for risk analysis project
- Awarded Color Lab Automation Award in 2012 for color matching laboratory opening

Adjunct Professor
various colleges

LORAIN COUNTY COMMUNITY COLLEGE, MIAMI UNIVERSITY, CINCINNATI STATE TECHNICAL AND COMMUNITY COLLEGE
Aug. 2010 - May. 2019

Developed curricula in introductory physics & astronomy courses, and technical presentations of complex physical concepts for major, non-major, and remedial students. Wrote concise laboratory instruction manuals, trained in precise measurement techniques & data analysis using DataStudio, and coordinated analysis documentation.

- Exceeded feedback evaluation goals by an average of 16% in 2017 and by 17% in 2018
- Tracked a 28% reduction in student confusion by reorganizing online education materials
- Uploaded over 700 hours of lecture material for convenient consumption

Scientific Research
University of Cincinati

DEPARTMENT OF PHYSICS
Sep. 2006 - Sep. 2010

Gathered data using SpeX, a 0.8–5.4 μm spectrograph, at the NASA Infrared Telescope Facility. Observations in the short-wavelength, crossdispersed mode SXD. Multiple analysis projects comparing 2MASS with measured data, statistical classification without independent photometry for cluster properties, and photometric analysis bootstrapped to independent measurements.

- Volunteered 120 hours of Quality Control for the Cambridge Astronomical Survey Unit Variable Stars database
- Published 3 scholarly articles on near-infrared astrophysics over a 24-month period
- Awarded Hanna Research Fellowship in 2010

Talks

IBM Watson Care Insights Reach Initiative
Watson Health

TRAINER
Apr. - Sep. 2019

- Watson Care Insights aims to analyze longitudinal patient records and real-time data points through HL7 feeds to provide practitioners insights into the patient's data. Watson Care Insights seeks to overcome physician-reported challenges incorporating shared decision making technology into their direct patient interactions, and address these challenges by integrating within the physician's workflow in the electronic health record system.

Content & Data Analytics Learning Exchange
Watson Health

PEER PRESENTER
Apr. & May 2018

- Bayesian Statistics using the Monty Hall problem example
- Naive Bayesian classification for patient matching
- Splines regression for population vitals ditributions

2018 IRI Fall Networks Conference
Innovation Research Interchange

PRESENTER FOR THE INFORMATION SERVICES/INFORMATION TECHNOLOGY NETWORK TALKS
Sep. 2018

- Cognitive computing can use NLP to extract insights from the unstructured data that holds most of the relevant information, but also it can also perform the necessary transformation accurately at the earliest stages of the data integration pipeline. Combined with the scalability and speed of the data lake architecture, this and other cognitive tools can effectively close the gaps on some of the most formidable flat file curation challenges.

Cross Team Training Series
Explorys

LEAD PRESENTER FOR THE METL GROUP
Sep. 2016

- Seven part series in data ingestion approach using the Hadoop software framework. Data can be aggregated much more quickly and cost-effectively using a data lake framework than in traditional data warehouses, because of the speed and low-cost of massively parallel computing. All data is stored in the data lake in its native format until it is needed, and each data element has a metadata tag for easy retrieval. As a result, responses are ad-hoc rather than predetermined, and reports based on new requirements can be delivered in days or weeks rather than months or years.

Writing

Process and Requirements for CCD Data Integration
IBM

Lead Integration Developer
Oct. 2017

- Technical paper on processes and requirements, overcoming the many difficulties with flat file data curation of CCDs.

Characterization of Surface Coatings using Radiative Transfer Models for Color Matching
PPG

Radiative Transfer Physics
Dec. 2010

- Approximation of Chandrasekhar radiative transfer model for atmospheric scattering is applied to color matching. (U.S Patent Application)

Massive Stellar Clusters in the Disk of the Milky Way Galaxy
OhioLINK ETD

Graduate Researcher
Dec. 2010

- Using spectroscopic and photometric analysis of two open clusters to determine the structure of the plane of the Milky Way Galaxy.

Near-Infrared Spectroscopy of Candidates Members of the Galactic Cluster [BDS2003] 107
PASP

Undergraduate Research Assistant
Feb. 2008

- New near-infrared classification spectra for nine candidate members and comparative 2MASS photometry for several cluster members.

Volunteer

2018/19	**Instructor**, Computer Science supplemental high school course hosted by BioEnterprise	*CSforCLE/CMSD*
2018/19	**Volunteer Coach**, Data Science High School Competition hosted by Case Western Reserve University	*HIT in the CLE*
2015	**ETL Volunteer**, RTAHeatMap project for Open Cleveland, a Code for America Brigade	*Open Cleveland*
2008/09	**Tutor**, On-site after school tutoring for T.C.P. World Academy	*Community School*

Education

Master of Science in Physics
University of Cincinnati

McMicken College of Arts and Sciences
2008 - 2010

- Thesis - Massive Stellar Clusters in the Disk of the Milky Way Galaxy

Bachelor of Fine Arts in Printmaking
Ohio University

College of Fine Arts
2000 - 2004

- Thesis - A Novel Method for Low-Toxic Photo-Etch Printing

Certificate of Specialization in Data Science (MOOC)
Johns Hopkins University

Bloomberg School of Public Health
2015 - 2016

- Capstone - A Machine Learning Algorithm for Paint Formula Prediction from Spectroscopic Measurements

Certifications in Data Science
IBM

Cognitive Class
2019

- Machine Learning with Python - Machine Learning, Regression, Classification, Unsupervised Learning, and Recommender Systems

Badges in Data Science
IBM

Cognitive Class
2018

- Cognitive Practitioner - Foundational understanding of IBM Cognitive solutions and offerings
- Hadoop Data Access - Hive and Hive QL data access on Hadoop

Badges in Software Architecture
IBM

Infrastucture & Architecture
2018

- IBM Cloud Private - Architecture and infrastructure requirements for deploying microservices-based cloud native applications
- Blockchain Essentials - Blockchain principles & practices, Distributed ledger systems, and asset transfer

Badges in Business
IBM

Offering Management
2018

- Offering Manager: Scaling the Business - Scale a business through Digital Channels, Expansive Ecosystems, and Value-based Service Delivery
- Offering Manager: Manage the Business - Holistic view of Health of Business and Customer Sentiment
- Offering Manager: Define, Build & Deliver - Offering Launch/Delivery (OD) discipline for delivering an offering to market

Badge in Agile Development
IBM

Agile Development
2017

- Agile Explorer - Foundational understanding of Agile values, principles, and practices