# CHARLES YEE, Ph.D.

DIRECTOR OF DATA SCIENCE | PRODUCT OWNER | MACHINE LEARNING & AI

## PROFESSIONAL SUMMARY

Internationally recognized and award-winning leader with 10+ years of progressive experience translating scientific rationale into commercial products across NLP & Healthcare domains. Strong command of data science with significant exposure to backend software development, statistical modeling, artificial intelligence, machine learning, product ownership, project management, end-to-end UAT and OAT,

Successfully leading multi-million dollar projects throughout Fortune 500's including United Health, Philips, AstraZeneca. and co-founder of multiple Series-A and Series-B startup companies. Strategically focused with proven aptitude in leading high-impact, high-performing teams in delivering end-to-end projects while managing quality deliverables.

## AREAS OF EXPERTISE

- Enabling strategy execution, road map creation and alignment to help deliver short-term performance improvement and long-term enterprise value.
- Owning decisions for project prioritization, workflows, tooling, integrations, modeling, and algorithms.
- Developing, scaling, and maturing Operations, Technology, Analytics, Engineering, and Data Science teams in order to deploy ML models into production.
- Directing reviews of tools and technologies to ensure technology currency, improving performance, reducing costs and time to market.
- Smoothly adapting to both technical and development constraints, feedback, and test results while also proactively readjusting the direction needed.
- Fostering trusted relations and high levels of engagement among internal teams, stakeholders, board of directors, and regulatory bodies.

## EDUCATION

**Ph.D., Computational Linguistics, NLP**
University of Stuttgart, Germany

**M.Sc, Computational Linguistics, NLP**
King's College London, UK

**B.Sc, Computer Science**
Ohio University

## CONTACT ME

Phone: (720) 771-1828
Email: charles.wj.yee@gmail.com
Website: http://www.charlesyee.com/

## CAREER EXPERIENCE

### AI, Machine Learning Director

**Inovalon | 2023 - Present**

*Provide strategic leadership for the company's 80+ product portfolio. Responsibilities include prototyping and integrating machine learning solutions into various business use cases: Coordinating across (hospital) Provider, (insurance) Payer, and (Speciality) Pharmacy business units and engineering teams. Regularly communicate with executives to identify market opportunities and ensure alignment of product vision*
*Attend industry conferences. Represent the company as an event/panel speaker*

- Implement domain-specific Q&A chatbot to assist customers and internal staff in installing, deploying, and triaging Inovalon's flagship pharmacy SaaS product (ScriptMed Cloud). This utilizes extensive Retrieval Augmented Generation (RAG), prompt engineering, AWS Bedrock (LLM), and Kendra.
- Build XGBoost models to predict patients' medication adherence to flag those not taking their medications (on time) for intervention plan, with 95% accuracy covering hundreds of thousands of patients.
- Build optimization software to reduce the number of False Positives from AWS Comprehend Medical output. This saves the medical coding team significant time and improves the accuracy of insurance Risk Adjustment.
- Trained gradient boosting models to minimize the number of "Member not found" errors during Insurance Eligibility Verification. This saves hospital customers money as Payers financially penalize large volumes of non-existing member queries.

### Principal Scientist

**AstraZeneca | 2021 - 2023**

*Spearheading R&D efforts on COVID-19 vaccine adverse event detection and prevention. Responsible for AZD1222's pharmacovigilance reporting, particularly those related to myocarditis, coagulopathy, and thrombocytopenia. Results reach directly corporate executives, CMO, and regulatory agencies (FDA, ECDC).*

- Implement microservices to discover and detect adverse event signals from vaccine manufacturing lots, including lots that are misspelled.
- Develop automation tools for relevant entity detection of drug-induced liver injury used by IQ DILI Consortium.

### Director of Data Science

**iQuartic (Acquired, 12M USD) | 2019 - 2020**

*Responsible for building the company's front and backend microservice architecture. Business focuses include: Health insurance risk adjustment, AI-assisted ICD10 coding. Streamline the tagging of >10,000 pages of Electronic Medical Records (EMR) daily with encounter and ICD tags using Named Entity Recognition.*

- Managed 7 direct reports while being the product owner for all NLP, OCR, and Machine Learning-related company offerings.
- Led code reviews, supervised evaluation and hyperparameter tuning of deep learning models, orchestrated Continuous Integration and Deployment (CI/CD), end-to-end User Acceptance (UAT), and Operational Acceptance Testing (OAT).

### Director of Data Science

**Mixfit Inc. Boston (Series A, 8M USD) | 2018 - 2019**

*Authoring invention disclosures which lead to two IP's for the key value propositions of the company. Managing 1 vendor company and 6 developers in the team.*

- Guided by the company's scientific board of nutritionists and physicians, implement microservice to calculate user nutritional intake recipe based on API feeds from wearable activity trackers, photo food journals, vital sign readings, gene biomarkers, and historical data, as well as NIH/NHANES metrics fit for the user's demographic cohort.
- Implemented a microservice that translates the nutrition recipe to device dispensation mechanics. Utilizing dynamic programming, the algorithm maximizes dispensation efficiency and minimizes shipping costs for the company.

# CHARLES YEE, Ph.D.

DIRECTOR OF DATA SCIENCE | PRODUCT OWNER, MACHINE LEARNING & AI

## LEADERSHIP CAPACITY

**Strategic Leader** - Proven ability to make memorable and polished presentations for executive audiences and boards with an uncanny ability to engage and sell groups of people on unconventional ideas.

**Performance Driven** - Targeting problematic cost areas, assessing root causes, and redesigning operating models to help drive bottom-line growth.

**Growth** - Spearheading execution blueprint through market segmentation, engagement strategies, and positioning for digital brand equity across wide spectrum of industries.

**Innovation** - Managing strategic corporate process planning such as standardization, streamlining, and automation to enhance efficiency and brand effectiveness.

## PERSONALITY ASSESSMENT

**Social Intelligence:** Showcasing accurate self-assessment and assessment of others, framing connectedness in teams. Intuition, and connective bridge between divergent groups.

**Love:** Valuing solid connections, caring about team and the company. and looking to find common ground.

**Prudence:** Providing order and precision. creating structure. and gaining all the facts before moving forward.

**Self Regulation:** Balancing needs with expectations. staying productive with multiple demands, and not compromising long-term success for short-term gain.

## CORE SKILLS

Data Science, Natural Language Processing (NLP), ETL, Deep Learning, Transformers. Large Language Models (LLM). Retrieval Augmented Generation (RAG). AWS SageMaker and BedRock. Foundation Models, Engineering team management, Product ownership, Cross Functional Management, Go-To-Market Strategy, Pitching and briefing, Data Management, Programming, Artificial Intelligence, Backend Development, Sprint Planning, and SAFe/ Agile

## TECHNICAL SKILLS

Java, Python, Pandas, Numpy, SciPy, Matplotlib, Tableau. Scikit-learn, XGBoost, TensorFlow Keras, PyTorch, RNN, LSTM, BERT, ELMo, Word Embeddings, LLM, RAG. LangChain. Elasticsearch, spaCy, Stanford CoreNLP, NLTK, OpenNLP, Spark NLP, JUnit, Mockito, unittest/PyUnit, Pytest, AWS, Azure, Kubernetes, Terraform, Docker, Jenkins, Flask, PostgreSQL, MongoDB, GraphQL, Spring Boot, Swagger, Postman APIs

## Senior Biomedical Informatics Scientist

### Philips North America | 2016 - 2018

*Research team leader at Philips Healthcare: The team delivers all of Philips' oncology informatics and NLP solutions. Responsible for IntelliSpace Clinical Trial Matching; configuration, release management to research staff of 8.*

- Liaised with the development team, internal business venture, and senior executives on new product feature offerings.
- Managed validation and incremental precision-recall improvements of the final solution with pathologists and oncologists.
- Set up research exhibits at partner hospitals (MD Anderson, Dana-Farber, Westchester Medical Center); benchmark product performances and usability with clinicians.
- Technical lead in NLP algorithm implementation for prototype clinical trial matching tool. Provide direction and hands-on expertise to key features such as clinical phenotype Named Entity Recognition (NER) using Long Short-Term Memory Networks (LSTM) and Conditional Random Field (CRF); Bayesian genome spelling correction and acronym canonicalization; contextual interpretation of negations using NegEx; concept extraction for patient profiling using cTakes.

## NLP Engineer

### Optum Analytics, UnitedHealth Group | 2014 - 2016

*Designing ontology-consistent feature structures and syntax-semantic interface to capture and harvest new concepts from unstructured data ranging from physician notes, claims, EMR, EHR- spanning over 130 million American patients.*

- Specialized in extracting concepts ranging from genetic mutation, chromosomal structural rearrangement, cancer staging, and tumor sizes. Other topics include neurostimulators, pain scores, tumors, and pain locations.
- Preprocessed, trained, and conducted diagnostics on Support Vector Machine (SVM) classification solutions for linguistic issues related to EMR and physician notes/prescriptions. Topics include: Drug change action rationale (such as cost, side effects, and efficacy), sentence boundary vs. abbreviation recognition, generalities vs. patient-centric data, etc.

## Research Scientist, Computational Linguist

### ACT | 2011 - 2013

*Designed and implemented an Automated Scoring engine for essays in K12 level standardized testing, with accuracy equivalent to human scorers. Approaches include: Random forest, gradient boosting, and Latent Dirichlet Allocation (LDA).*

- Participated and won the top place among all vendors on behalf of the company in the Kaggle Essay Scoring competition.

## SELECT PUBLICATIONS, TALKS & PATENTS

- "A Domain Knowledge-Enhanced LSTM-CRF Model for Disease Named Entity Recognition", Journal of the American Medical Informatics Association (JAMIA), 2019:761–770. US PCT/EP2019/071747, Nomenclature Informed Gene Name Recognition, Aug 13, 2019
- US PCT/EP2018/075263, Natural Language Processing using Ontology Mapping, Sept 19, 2018
- US PCT/EP2019/066322, A Method for Genome Spelling Correction and Acronym Standardisation, May 21, 2018
- US WO/2018/060838, A Method and System for Matching Subjects to Clinical Trials, May 4, 2018
- "Human-Algorithm Interaction to Define Variables from Free-Text Notes in Electronic Health Records", presented at the 31st International Conference on Pharmacoepidemiology and Therapeutic Risk Management (ICPE), Aug 2015

## LANGUAGES

Native bilingual in English and Mandarin. Proficient German speaker.