Charles Yee

(720) 771-1828 . charles.wj.yee@gmail.com . www.charlesyee.com

Summary

NLP, data science product owner and technical manager. Computational linguistics expert in semantic representations, ontology, machine learning and deep learning approaches for information classification and extraction. Over a decade specialisation in the healthcare domain. Provide leadership, orchestration and implementation to materialise domain experts' scientific rationale into commercial products. Uniquely versed in clinical trial protocol, oncology, EMR/EHR, NLP-Assisted ICD coding. Headed various companies' R&D and infrastructure teams: Code review, model diagnosis, product and project management, end-to-end UAT and OAT, DevOps, recruitment-onboarding, aspirational speeches, and business acumen

Well acclimated to large corporate culture. Nevertheless, full track record of entrepreneurship initiatives.

Work Experiences

1/2021 – Present, **Principal Scientist**, AstraZeneca. Remote

- Spearhead 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 to corporate executives, CMO, and regulatory agencies (FDA, ECDC)
- Implement microservices to discover and detect misspelt vaccine manufacturing lots related to adverse events
- Develop automation tools for relevant entity detection of drug-induced liver injury used by IQ DILI Consortium

10/2019 – 11/2020, **Director of Data Science**, iQuartic. Boston

- Responsible for building and maintaining 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**
- As hiring manager: Recruit, tech-screen, and onboard roles ranging from full-stack devs, NLP engineers, and DevOps
- Project manage 7 direct reports. Product owner for all NLP, OCR, and Machine Learning related company offerings
- Lead code reviews. Supervise evaluation and hyperparameter tuning of deep learning models. Orchestrate Continuous Integration and Deployment (CI/CD), end-to-end User Acceptance (UAT) and Operational Acceptance Testing (OAT)

08/2018 – 10/2019, **Director of Data Science**, Mixfit Inc. Boston (Series-A, 8 million USD in investment)

- Manage software developer direct reports and vendors. Set technical standards and best practises across the company
- Author invention disclosures which lead to two IP's for the key value propositions of the company
- Guided by company scientific board of nutritionists and physicians, implement microservice to calculate user nutritional intake recipe based on API feeds from wearable activity tracker, photo food journals, vital sign readings, genomic and historical data, as well as NIH metrics fit for the user's demographic cohort
- Implement microservice that takes the nutrition recipe and translates it to device dispensation mechanics. Utilize dynamic programming, the algorithm maximises dispensation efficiency and minimises shipping cost for the company

03/2016 – 08/2018, **Senior Biomedical Informatics Scientist, Project Leader**, Philips North America. Boston

- Research team leader at Philips Healthcare: Team delivers all of Philips' oncology informatics and NLP solutions. Responsible for IntelliSpace Clinical Trial Matching; configuration, release management to research staff of 8
- Close collaboration with software architect to devise trial match methodology, product logic, and workflow
- Coordinate with the development team, internal business venture, and senior executives on new product feature offerings
- 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 Centre); 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
- Provide business development insights leveraging technological know-how to business ventures and hospital customers

07/2016 – 08/2018, **Co-founder, CTO,** and **Advisor,** Twyla. Berlin, Germany (Series-A, 4 million EUR in investment)

- Chatbot AI architecture design: Finite-state automaton with hybrid transition model- Through both rule-based pattern recognition, and machine learning that approximates semantic similarity with historical chat logs

- Hands-on implementation of linguistic pre-processing, textual feature selection and extraction. Utilising Scikit-learn, build **regression, gradient boosting, random forest** modules for customer chat intention detection
- Provide big data analysis (NGram, tf-idf, cosine, Word2Vec) to enterprise clients including T-Mobile, HTC, Heineken, Cebu Pacific, that yields insights on their customer behaviour, product issues and marketability

03/2014 – 03/2016, **NLP Engineer**, Optum Analytics, UnitedHealth Group. Boston

- Design 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
- Specialised in extracting concepts ranging from genetic mutation, chromosomal structural rearrangement, cancer staging, tumour sizes. Other topics include neurostimulator, various pain scores, tumour and pain locations
- Preprocess, train, and conduct 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.

08/2011 – 12/2013, **Research Scientist, Computational Linguist**, ACT. Denver

- Design and implementation of Automated Scoring engine for essays in K12 level standardised testing, with accuracy equivalent to human scorers. Approaches include: Random forest, gradient boosting, **Latent Dirichlet Allocation (LDA)**
- Participated and won top place among all vendors on behalf of the company in the **Kaggle** Essay Scoring competition

Notable Qualifications

- Player-coach manager. Excellent communicator and eloquent public speaker. Understand how to sell ideas to executives, prioritise customers' needs, articulate needs to the team, then rally talents to come together and deliver
- Daily interaction with Jira, Confluence, Git/ Bitbucket, and Agile development process. **SAFe Certified**
- Fluency in **Java** and **Python**. Familiarity with popular libraries (Pandas, Numpy, Matplotlib), design patterns and data structures. Years of developing unit and integration tests and test automations (JUnit, Mockito, unittest/PyUnit, Pytest)
- Professional competency in **Statistical NLP, Machine Learning and Deep Learning**, particularly Supervised Learning-Document classification, information extraction, and Named Entity Recognition using Recurrent Neural Networks (**RNN**)
- Working experience with probabilistic, NLP open source, including WordNet, Elasticsearch, Stanford CoreNLP, NLTK, SpaCy, OpenNLP, Spark NLP, **Scikit-learn**, **TensorFlow Keras, and Flair (PyTorch, BERT, ELMo Embeddings)**
- Solid understanding of common IaaS and PaaS such as **AWS**, Azure, Heroku. Tech stack including Kubernetes, Terraform, Docker, Jenkins, Ruby on Rails, Flask, PostgreSQL, MongoDB, GraphQL, Spring Boot. Experiences in both **creating and using REST API endpoints** (Swagger and Postman)

Education

05/2003 - 01/2011, **Ph.D.**, Computational Linguistics, NLP, University of Stuttgart, Germany. Graduated with Distinction
10/2001 - 01/2003, **M.Sc.**, Computational Linguistics, NLP, King's College London, UK. Graduated with Distinction

Selected Publications, Talks, and 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
Further patents, patents pending, and publication references are available upon request

Scholarship and Awards

Graduiertenkolleg "Sprachliche Repräsentationen und ihre Interpretation", German Research Foundation, 2005 –
2009 King's College London Taiwan Scholarship, from King's College London, 2003 – 2006

Additional Information

Native bilingual in English and Mandarin. Proficient German speaker
Seasoned motorcyclist, violinist, Tanguero, Iaido-ka, and user of the Oxford comma