LONDON, UNITED KINGDOM • JIEGAOUK@GMAIL.COM

JIE GAO

PROFESSIONAL SUMMARY

A highly experienced and accomplished AI Scientist and Engineer with over 18 years in information technology spanning both industry and academia. Specializing in knowledge-intensive systems, Generative AI, Natural Language Processing (NLP), Machine Learning (ML), and large-scale information processing. Expertise in developing and deploying AI-driven solutions for misinformation and disinformation detection, fact-checking, and social network analysis. A proven track record of leading cross-functional teams, driving innovation, and delivering impactful results in high-stakes domains

RELEVANT EXPERIENCE

Staff NLP Engineer, Logically AI, London	May 2022 –
Senior Research Engineer, Logically AI, London	July 2020 – May 2022

- Led an AI team specializing in multimodal fact-checking, evidence retrieval, rumor detection, and information credibility assessment on social media.
- Developed computational models (as **IC role**) to analyze disinformation, detect <u>coordinated inauthentic</u> <u>behavior</u> (CIB) and Foreign Information Manipulation & Interference (FIMI), and counter online threats.
- Played a key role in strategic planning, project management, and AI model lifecycle governance, ensuring high standards in machine learning engineering, data quality, and <u>human-in-the-loop AI</u>.
- Mentored junior colleagues, improving expertise in SoTA NLP methodologies, rigorous evaluation, and best coding practices, while maintaining transparent and accountable AI development workflows.
- Designed and optimized AI architectures focused on threat intelligence, including election integrity, national security, and public health applications. Managed ML pipeline design, data governance, and deployment strategies for both generative, supervised and unsupervised learning models.
- Published research papers and authored research/grant proposals (e.g., MoD, DSIT), presenting novel AI solutions at academic conferences.
- Contributed to deep learning frameworks, data quality improvements, and documentation standards, ensuring scalability, maintainability, and real-world impact.

Key Skills: Applied Machine Learning | NLP | AI System Architecture | Large Language Models | Fine-tuning | Post-Training | Model Alignment | Agentic AI | RAG | Social Media Analysis | Graph Processing & Knowledge Engineering | Disinformation & Misinformation AI | Information Retrieval | Distributed Systems | Agile & Test-Driven Development | Cross-Functional Collaboration | Project Management

Research Associate, OAK Group, University of Sheffield	Mar 2014 – June 2020
rebearen ribboeratet orbiet, on biennera	inter 2011 builde 2020

- Contributed to multiple EU and UK-funded research projects, including <u>Europeana DSI-3</u>, <u>WeSenseIt</u>, <u>SETA</u>, and SPEEAK-PC (TATA Steel UK), collaborating across industry, government, and academia.
- Led and managed large-scale data infrastructure for SETA and Active10 projects (serving 129K+ monthly users, 600K+ downloads) for Public Health England, featured by <u>BBC</u>.
- Co-Principal Investigator (Co-PI) for the <u>Europeana DSI-3</u> (Metis & Solr services) and led research on terminology-driven text mining and knowledge discovery (SPEEAK-PC, TATA Steel UK).
- Played a key role in developing university-licensed intellectual property (IP), forming the core technology of Aeqora Ltd, a mobility service spin-off.
- Invited speaker (4+ times) at international conferences and workshops; peer-reviewed for ACL ARR, Journal of NLP, ESWC, PLOS ONE, and more.
- Taught NLP and Intelligent Web technologies (<u>COM3504</u>) as a Teaching Assistant for two years to 70+ students.
- Provided cloud infrastructure administration and technical support for EU/National research projects, assisting PhD students and colleagues with data collection, database systems, and software engineering.

Key Skills: Research & Grant Writing | Project Management | NLP | Machine Learning Data Quality | Social Network Analysis | Mobile Applications | Knowledge Engineering | AWS & Cloud Security | Distributed Systems

| Agile Methodology

Semantic R&D developer, Crownpeak (prev. ActiveStandards), London Oct 2013 – Jan 2014

Follow-up research and development of KTP project to build innovative semantic products on ActiveStandards platform, covering:

- Semantic Annotation & Indexing (GATE platform, Sesame/OWLIM 5.x (now GraphDB), Ontotext KIM platform)
- Semantic technologies for Enterprise Information/content integration (Named Entity Recognition, entity linking, RDF, Ontology)
- Semantic SEO and Web Compliance (corpus pre-processing & indexing, keywords extraction, RDFa, Web Accessibility)

KTP Associate, University of Westminster, London

- Undertake the Knowledge Transfer Partnership (KTP) objectives as defined in the KTP proposal.
- Development of new semantic systems and services on ActiveStandards platform to assist businesses and organizations in discovering, understanding, and acting on actionable knowledge within their information repositories.
- The partnership project has been graded as second higher level which is shortlisted as one of KTP case-studies among 250 national funded projects. See <u>case study</u> published in Uni Westminster.

Senior Software Engineer/Java developer	Aug 2006 – Jun 2010
EPAM System (prev. Dextrys), Suzhou, China	

- Engage in large scale enterprise application development with complicated business requirements.
- Involved in projects such as Chartis Brazil Insurance System, Trinity Project for Tripwire, Hanover Insurance system for Hanover, EA Back Office System for Electric Arts Inc., Mayban Fortis Insurance System for eBaoTech.com, Non-Domestic management System for Hong Kong Housing Authority, and Greece Reporting SMS System for Proctor & Gamble.

EDUCATION

PhD candidate in NLP University of Sheffield

M.S. in Web Technology University of Southampton

B.S. in Computer Science and Technology Yancheng Institute of Technology Sept 2014 – Sept 2019 Sheffield, England

Oct 2011 – Oct 2013

Sept 2010 - Sept 2011 Southampton, England

Sept 2002 - Sept 2006 Yancheng, China

Certificates and Training (Selected)

- Generative AI with Large Language Models (License <u>RSKBJ5B8W4SM</u>), Coursera(deeplearning.ai), Aug. 2023
- Convolutional Neural Networks (License <u>DWLCCDHKCE2N</u>), Coursera (deeplearning.ai), April 2018
- Structuring Machine Learning Projects (License <u>8JUA2EYXDHAN</u>), Coursera (deeplearning.ai), November 2017
- Improving Deep Neural Networks: Hyperparameter tuning, Regularization and Optimization (License K39M5D7B9Q6Z), Coursera (deeplearning.ai), November 2017
- Neural Networks and Deep Learning (License <u>BXJWXH56N6TZ</u>), Coursera (deeplearning.ai), October 2017
- Machine Learning (SoA 973756~189167), Stanford University through Coursera Inc., May 2015
- Certified ScrumMasters (License <u>000285406</u>), GoodAgile, Oct 2013
- CMI Level 5 in Management, Ashorne Hill Management College, 2011 2012

PUBLICATIONS (Selected)

• Verschuuren, P. J., Gao, J., Van Eeden, A., Oikonomou, S., & Bandhakavi, A. (2023). Logically at Factify 2: A

multi-modal fact checking system based on evidence retrieval techniques and transformer encoder architecture. In *AAAI'23*: *De-Factify 2: 2nd Workshop on Multimodal Fact Checking and Hate Speech Detection*, Washington, DC, USA. Access via arXiv: 2301.03127 [Slides]

- Gao. J., Hoffmann, H. F., Oikonomou, S., Kiskovski, D., & Bandhakavi, A. (2022). "Logically at Factify 2022: Multimodal Fact Verification.", In: *AAAI'22: First Workshop on Multimodal Fact-Checking and Hate Speech Detection*, Vancouver, BC, Canada, access via arXiv preprint <u>arXiv:2112.09253</u>. [Slides]
- Gao. J., Han S., Song X., Ciravegna, F. (2020). "RP-DNN: A Tweet level propagation context based deep neural networks for early rumor detection in Social Media", In: *The LREC 2020 Proceedings. The International Conference on Language Resources and Evaluation*, 11-16 May 2020, Marseille. *LREC 2020* (accepted for an Oral presentation). [Code/Data]
- Han S., Gao, J., Ciravegna, F. (2019). "Neural Language Model Based Training Data Augmentation for Weakly Supervised Early Rumor Detection", *The 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2019)*, Vancouver, Canada, 27-30 August, 2019 [Acceptance Rate: 14%] [Dataset]
- Han S., Gao, J., Ciravegna, F. (2019). "Data Augmentation for Rumor Detection Using Context-Sensitive Neural Language Model With Large-Scale Credibility Corpus", *Seventh International Conference on Learning Representations (ICLR) LLD*, New Orleans, Louisiana, US [Code] [Dataset]
- Ciravegna, F. Gao, J., Ingram, C., Ireson, N., Lanfranchi, V., & Humasak S. (2018). "Mapping Mobility to Support Crisis Management." 15th Proceeding of International Conference on Information Systems for Crisis Response and Management (ISCRAM). ISCRAM Association. 2018. Access via <u>WRRO</u>
- Zhang, Z., Gao, J., & Ciravegna, F. (2018). <u>SemRe-Rank: Incorporating Semantic Relatedness to Improve</u> <u>Automatic Term Extraction Using Personalized PageRank</u>. *ACM Transactions on Knowledge Discovery from Data (TKDD)*. [Code]
- Zhang, Z., Gao, J. and Ciravegna, F. (2016) <u>JATE2.0: Java Automatic Term Extraction with Apache Solr.</u> In: *The LREC 2016 Proceedings. The International Conference on Language Resources and Evaluation, 23-28 May* 2016, Slovenia. *LREC 2016*. [Code] [poster presentation][Jate Solr Plugin Demo] [Jate4WeakSupervision Demo code] [Industrial use case demo]
- Gao, J., and Suvodeep Mazumdar. "Exploiting linked open data to uncover entity types." In Semantic Web Evaluation Challenge at ESWC 2015, pp. 51-62. Springer International Publishing, 2015. [code][Slides]

SKILLS & ATTRIBUTES

- AI & Machine Learning (pre-training, post-training, LLM, Multimodal AI, Agentic AI, Transformer Networks, few shot learning, contrastive learning)
- Natural Language Processing, Misinformation/Disinformation Detection
- Knowledge Engineering (knowledge representation, knowledge graph, Semantic Web)
- Data Science & Engineering
- Cloud Computing (AWS, GCP), Software engineering (Python, Java, SQL)
- Libraries & Frameworks (Pytorch, Huggingface, LlamaIndex, LangChain, Sklearn, Mlflow, DVC, etc.)
- Graph Technology (graph processing & modeling, Social Network Analysis)
- Database (Vector DB, Solr/Elasticsearch, SQLite, MySQL, GraphDB, MongoDB, etc)
- Agile Methodologies (Scrum, JIRA, Confluence, Test-Driven Development, CI/CD, docker, Jenkins, etc)

LANGUAGES

Chinese (Native), English (Highly proficient)

LINKS

Personal Website: jerrygaolondon.github.io, Google Scholar: scholar.google.com, Github: github.com

LinkedIn: linkedIn.com, ORCID: 0000-0002-3610-8748, Semantic Scholar: author/Jie-Gao

WORK ELIGIBILITY

Settlement -- INDEFINITE LEAVE TO REMAIN