Jie Gao

44(0) 741 27 82628
jiegaouk@gmail.com
http://jerrygaolondon.github.io

General

- I'm specializing in a wide range of computational techniques for the detection of disinformation and misinformation. With over a decade of hands-on experience in large-scale distributed systems and several years in mobile app development, I possess a solid technical foundation.
- In my role as a **Research Scientist**, I bring more than ten years of experience in academic research in **Natural Language Processing** (NLP) and **Social Media Analytics**. I have authored over 10 publications in prestigious conferences and journals (e.g., ICLR, KDD, ASONAM, LREC, IEEE/ACM, etc) with **h-index and i10-index 9**. My expertise encompasses **information system architecture**, **AI algorithm optimization**, **NLP**, **machine learning**, **knowledge engineering** and **knowledge graph** technology with good hands-on experience and knowledge of the following languages and formats: *SPARQL 1.1*, *RDF / RDFS*, *OWL*, *SKOS*. Ref.: ORCID: 0000-0002-3610-8748, RID (Web of Science ResearcherID): AAB-4270-2020; Google ScholarID: 1zoqXaMAAAAJ
- I am proficient in both traditional and cutting-edge NLP and machine learning techniques, utilizing libraries such as *Pytorch*, *Keras*, *Huggingface*, *LangChain*, *spaCy*, *NLTK*, and more. Furthermore, I am an adept software developer with a background in developing multi-tier web applications in *JEE*, *Node JS*, and *Python*.
- My experience extends to cloud technologies like *AWS* and *GCP*, *ETL* processes, Big Data technologies such as *MapReduce* and *Spark*, and graph technology. I have also made contributions to open-source tools and frameworks.
- Earlier in my career, I was involved in the development of enterprise applications and web systems, catering to clients from Fortune 500 companies, as well as banks, insurance companies, and government entities. I gained rich experience in fundamental web technologies like *HTML*, *JSP*, *CSS*, *XML*, *JSON*, *JavaScript*, and *Ajax*, applying various design patterns and frameworks.
- In the academic realm, I have actively contributed to the field by serving as a reviewer for esteemed conferences and journals, including ACL Rolling Review (ARR), EMNLP, Plos One, Journal of NLP, and Journal of Intelligent Systems, showcasing my dedication to advancing our knowledge in this domain.

RELEVANT EXPERIENCE

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

- With extensive expertise and involvement in misinformation AI, I lead a collaborative AI team dedicated to fact-checking, evidence retrieval techniques, rumor resolution on social media, and information credibility assessment. My work encompasses computational methods for analyzing disinformation, detecting coordinated inauthentic behavior on social media, and addressing online threats and influences.
- I am also skilled and actively leads in various projects and people management, strategic planning/direction, human-and-model-in-the-loop strategy, machine learning engineering standards, test-driven ML framework, data quality methodologies, and **cross-disciplinary** & **cross-functional collaboration** (with sales, OSINT experts, fact-checkers, product).
- work closely with teammates and help to upskill junior colleagues with various aspects including SoTA methodology, rigorous evaluation, code quality, code of conduct to maintain transparency and accountability with professional work standardizations for whole ML development lifecycle covering dataset development workflow, model development practice, code quality, iterative iteration, service deployment, version control of data and model, knowledge transferring skills and high quality documentations;
- My role also involves conducting research and coming up with novel algorithms and improvements, authoring papers, company white papers, research proposals, and presenting at academic conferences, etc.
- This technical lead role focusses more on **human(client/user)-centered AI architecture,** pipeline planning and ambitious research agenda for areas of threat intelligence such as election integrity, national security and public health. This entails a thorough understanding of end-to-end ML and deep learning workload workflows, with expertise in the components and architectural trade-offs inherent in data pipelines, data management (*e.g., with DVC, GCP Bucket*), governance, model building, training and evaluation (*e.g., Mlflow, model registry*) for both supervised and unsupervised techniques, and the deployment and production (CI/CD) workflows of AI.
- Extensive hands-on work involved in current deep learning framework and contribute to data quality and code hygiene with comprehensive documentations to ensure the project is easy to maintain and reusable for real-world impact.
- **Key skills**: Applied Machine Learning; Research Skills; Distributed System; Large Language Models; AI system architecture; Natural Language Processing; Leadership; Semantic Technologies; Graph processing and Modelling; Social Media Analysis; Multimodal analysis and modelling; Information/Knowledge Retrieval (both sparse and dense methods), bi-encoder/cross-encoder models/transformer based re-rankers/LLMs for IR, Knowledge Graph & Engineering, Disinformation and TTP framework; Misinformation and automated fact

checking; Agile Methodology and Test Driven-Development; Cross-disciplinary and Cross-functional team communication and collaboration; Project Management;

Research Associate, OAK Group, University of Sheffield — Mar 2014 – June 2020

- Involved in various research project, including European Foundation funded Europeana DSI-3, two European FP7 projects (WeSenseIt and SETA), two Innovative UK (prev. TSB) funded projects (SPEEAK-PC 101947 with TATA Steel UK and KTP project 007326 for ActiveStandards, the market-leading SaaS platform), three industry/government/academic collaboration projects (with K-Now, JustGiving, MoveMore and Active10)
- Doctoral Candidate in NLP and computational terminology
- 2 years as Teaching Assistant for "COM3504 The Intelligent Web" module (>70 students)
- Invited speaker/presentation (4 times) in international conference / workshop.
- Leading the development and am the manager of the scalable, computational and economically efficient infrastructure that collects the data in large-scale from the mobile phones in EU project SETA and Active10 projects (over 129K monthly active users and over 600,000 downloads over past 10 months) on behalf of Public Health England (as featured in BBC).
- As a founding researcher/inventor, played a key role in developing licensed university-owned intellectual property (IP) that formed the core technology for Aeqora Ltd, a university spin-off providing mobility service.
- co-PI of Europeana Digital Service Infrastructure (DSI) DSI-3 (addendum project) for Metis and Solr services
- leading & coordinating the research for work package (WP) 3 ("Data Representation, Mining & Analysis") and responsible for the delivery for two work packages (WP3 and WP4 – terminology driven text mining and knowledge discovery application) in SPEEAK-PC collaboration project with TATA Steel funded by Innovative UK
- peer reviewer for academic journals / conferences / workshops, such as ACL ARR, Journal of Natural Language Engineering, ESWC, USES, PLOS ONE
- providing cloud computing environment administration and maintenance for EU/National funded research projects and our group VMs for PhD research in the lab
- technical support for PhD students/colleagues (e.g., data collection via low-cost sensors, database systems, software engineering, web servers and service hosting)
- **Key Skills**: Research Skills; Team Leadership; Scientific paper and research grants writing, Project Management; Social Network Analysis & Modelling; Data Quality for Machine Learning; Natural Language Processing; Mobile Applications; Time Series Data Modelling & Analysis; Knowledge Management and Engineering; Amazon Web Service (AWS); Cloud Security; Distributed System; Information System Architecture; 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 — Oct 2011 – Oct 2013

To undertake the Knowledge Transfer Partnership (KTP) objectives as defined in the KTP proposal covering the development of new semantic systems and services on ActiveStandards platform, in order to assist businesses and organizations to discover, understand and act on the actionable knowledge within their information repositories.

The R&D project (KTP project 007326) is funded by the UK Technology Strategy Board. I was the KTP associate and the lead R&D engineer in the project. The main technologies in this project covers the area of :

- **Knowledge extraction** (more specifically cross-domain NER and relation extraction with GATE),
- **Knowledge management and retrieval** (RDF, RDFS, OWL) with OWLIM-SE (now rebranded as GraphDB)
- **Entity linking and knowledge base enrichment** (via linked open data) and Enterprise Information Integration (ontology normalisation)
- **Web compliance** (SEO and fact checking)

I became Professional membership (990394505) of The British Computer Society (**MBCS**) since 2011.

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 case study published in Uni Westminster.

Senior Software Engineer/Java developer, Dextrys, Suzhou, China Aug 2006 – Jun 2010

- Engage in **large scale enterprise application development** with complicated business requirements
- Involved projects include Chartis Brazil Insurance System, Trinity Project for Tripwire, Hanover Insurance system for Hanover, EA Back Office System for Electric Arts Inc., Mayban Fortis InsuranceSystem for eBaoTech.com, Non-Domestic management System for Hong Kong Housing Authority and Greece Reporting SMS System (GRSMS) for Proctor & Gamble
- Main duties
  - Design, enhance application solution and work closely with business analysis team.
  - Transform clients' idea and business concepts into carefully designed and well-authorised software engineering solutions.
  - Work closely with UI design team to ensure the usability of the software.
  - Coordinate with QA Test Engineers on Module business testing, System Integration Test, Regression Test, and so forth during the whole software development lifecycle.

EDUCATION & RELEVANT COURSEWORK

PhD candidate in NLP, University of Sheffield, Sheffield, England Sept 2014 – Sept 2019

- Completed coursework and research in Computational Terminology.
- Published 4 related papers in peer-reviewed journals or conferences.
- Presented at 2 international conferences related to this PhD study.

M.S. in Web Technology, University of Southampton, Southampton, England Sept 2011

- Thesis: "Linked Data based enterprise content management"
- **Graduated with Merit**
- **Subject studies**: Enterprise Web Development, Advance Topics on Web Services, Hypertext and Web technologies, Assistive Technologies and Universal Design, Intelligent Agent, Semantic Web Technologies, Rich Internet Applications.
- **Skills developed**: Enhanced the knowledge and understanding of important technologies and issues in designing and building enterprise web systems; decent English written and oral communication; research, interpretation and presentation; ability to analysis and evaluate issues regarding universal design and assistive technologies; Enhanced hands-on experience with building web services in Java; ability to design an agent-based system and evaluate the suitability of an agent-based approach to a given problem; Enhanced the ability to evaluate difference RIA technologies;

B.S. in Computer Science and Technology, Yancheng Institute of Technology, Yancheng, China Sept 2006

- Dissertation: "Struts based online shopping system"
- Grade: 2:1

Certificates and Training

- Generative AI with Large Language Models (License RSKBJ5B8W4SM), Coursera(deeplearning.ai), Aug. 2023
- Convolutional Neural Networks (License DWLCCDHKCE2N), Coursera (deeplearning.ai), April 2018
- Structuring Machine Learning Projects (License 8JUA2EYXDHAN), 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 BXJWXH56N6TZ), Coursera (deeplearning.ai), October 2017
- Machine Learning (SoA 973756~189167), Stanford University through Coursera Inc., May 2015
- Data Mining with Weka, University of Waikato, August 2014
- Associate Certificate (License KTP007326-A02), UK Technology Strategy Board, Oct 2013
- M101J: MongoDB for Java Developers, "10gen, The MongoDB Company", July 2013
- Certified ScrumMasters (License 000285406), GoodAgile, Oct 2013
- 5th GATE Training Course Track 3 (Advanced GATE), University of Sheffield – 2012
- CMI Level 5 in Management, Ashorne Hill Management College, 2011 - 2012

PUBLICATIONS

- 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 https://ceur-ws.org/Vol-3555/paper15.pdf (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 arXiv:2112.09253. [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., Ireson, N., Copeland, R., Walsh, J., & Lanfranchi, V. (2019, May). "Active 10: Brisk Walking to Support Regular Physical Activity." *In Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare* (pp. 11-20). ACM. Access via https://dl.acm.org/doi/abs/10.1145/3329189.3329208

- 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 WRRO

- Zhang, Z., Gao, J., & Ciravegna, F. (2018). SemRe-Rank: Incorporating Semantic Relatedness to Improve Automatic Term Extraction Using Personalized PageRank. *ACM Transactions on Knowledge Discovery from Data (TKDD).* [Code]

- Yuan, Y., Gao, J., & Zhang, Y. (2017, October). Supervised Learning for Robust Term Extraction. In *The proceedings of 2017 International Conference on Asian Language Processing (IALP)*. IEEE. Access via WRRO

- Gao, Jie, and Suvodeep Mazumdar. "Exploiting linked open data to uncover entity types." In *Semantic Web Evaluation Challenge*, pp. 51-62. Springer International Publishing, 2015. [code][Slides]

- Zhang, Z., Gao, J. and Ciravegna, F. (2016) *JATE2.0: Java Automatic Term Extraction with Apache Solr.* 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]

- Zhang, Z., Gao, J., and Gentile, A.L. *The LODIE team (University of Sheffield) Participation at the TAC2015 Entity Discovery Task of the Cold Start KBP Track.* In: Proceedings of the 2015 Text Analysis Conference. TAC Knowledge Base Population (KBP) 2015, 16-17 Nov 2015, Gaithersburg, Maryland USA. [Slides][Code]

SKILLS & ATTRIBUTES

- **Strong written & oral communication, presentation and demonstrating skills** with confident, articulate and professional speaking ability.
    - Two "**Best Paper**" awards and invited talks in three workshops co-located with top AI conferences including ICLR 2019, AAAI 2022 & 2023
    - Deliver regular presentations and demonstrations in project meetings and international conferences with **six invited presentations** in the past few years.
    - **Work as PI** in Europeana DSI-3 addendum project
    - Work as one of **core developers** (both mobile and backend server) and coordinator in PHE active10 project
    - Work as the sole associate & lead researcher in two Innovate UK (prev. TSB) funded projects for four years, both of which are rated as second higher score
- Strong **collaborative mindset**: able to communicate effectively with the diverse range of people and **strong teamwork ability in multicultural and cross-disciplinary work environment.**

- Many years' experiences of actively and effectively working with data scientists, data engineers, data analysis, QA, MLops, other architects, business unit leaders, OSINT team, fact checking team, sales and product to help the organization adopt an AI-driven solution and ensure AI platforms deliver both business and technical requirements and harmonize the relationships among all the stakeholders @Logically.
- At Logically, I built effective data science teams through a well-established and streamlined squad-level (4 teammates) workflows.
  - helped team to deliver outcomes in a more transparent way e.g., by introducing and demonstrating the way of using confluence, JIRA, bi-weekly sprint review/retro/planning meeting, and peer-review process from data engineering, model/experiments and code quality.
  - lead the team to identify the most important deliverables for misinfo team per sprint by combining prod-led and research-led priority.
  - Helped team to decide on a clear evaluation metric up front for individual but interdependent projects.
  - foster a research focus culture in misinfo team to ensure **quality rather than speed** per project deliverables.
  - raise standards by helping the team to establish code standards, new work culture and set a higher standard for
    - **completeness** (e.g., via automated test, code testability, code level documentation, confluence page documentation
    - **transparency** (trackability with archived work and jira tickets, improved documentation of key projects, meeting notes, regular mid-term sprint check-in meeting)
  - make sure misinfo team to regularly resolve technical debt and log everything, and retro/ review periodically.
  - Worked side-by-side with project owners, using our rock-solid data science process to double-check results and think through the "what ifs".
- Good team player with >10 years' experience in multicultural and cross-disciplinary working environment for both commercial and academic research projects funded by EU (FP7) and UK gov (e.g., PHE, MoD, DIST) @Sheffield.
- Work as an embedded member of ActiveStandards team and as a KTP associate in Semantic Computing lab of University of Westminster to complete **two-years KTP project** with a series of challenging project milestones (see case study of this project).
- Work in large **EU projects** (WeSenseIt and SETA) involving partners from more than 5 countries
- Complete **CMI Level 5 training** in Management in Ashorne Hill Management College
- Trained as a **Certified Scrum Master** (CSM)

- **Demonstrated thought leadership through self-direction, dynamism and adaptability**:
  - Proven record of quickly adapting to new projects or tasks
  - Leading the ideation and write-ups (end to end) of scientific papers company white papers and research proposals (e.g., MoD DASA innovation funding) at Logically AI Innovation Solution team and one of core researchers in a number of EU and UK **multi-million projects proposal** writing (e.g., SETA);
  - Lead the conceptualization, development, and deployment of LLM projects tailored for misinformation & disinformation intelligence (e.g., CIB system, *RAG* system for automated fact checking, LLM *prompt-based data augmentation*) @Logically since early 2023;
  - Build a complex data driven CIB system as sole contributor with comprehensive conceptual and technique documentations (on company confluence) and presentations for company-wide after three years in when the concept of CIB was initially introduced by Meta and not widely known in this field
    - The initial iteration addressed the fundamental issue of platform and content independence and behaviour centred analysis through a modular and extensible architecture. This architecture incorporates an industry standard compliant framework of efficient and reusable techniques that enable the seamless implementation and integration of novel activity types and analytical capabilities.
    - The system is built from scratch, and I have demonstrably led the system's iterative development and ongoing optimization, employing state-of-the-art technologies and industry-recognized best practices (*OSINT/TTP/DISARM*) to guarantee its sustained efficacy and relevance.
    - Core technique covers large scale co-occurrence modelling and fast computation, web scale record linkage algorithms, language & modality independent design, vector-based content indexing & clustering with vector store, *quantization* and *HNSW32* technique, integration of multiple database systems (*mongodb + Elasticsearch + sqlite + vector DB*), SoTA unsupervised community detection for large social network graph, large scale graph processing, computation efficient image processing and embedding for both GPU and CPU env, memory efficient algorithm and data structure (e.g., *Trie*), LLM with advanced prompt engineering techniques for domain specific (disinfo) analysis (e.g., narrative detection),

summarisation (e.g., narrative, inauthentic network) and custom (automated) adversarial threats reporting;

- I champion an **engaged leadership** approach, rolling up my sleeves and working alongside my team members.
  - For instance, in the 1st month joining misinfo team in Logically, I helped team to re-design and refactor existing misinfo pipeline, key algorithms end to end.
  - By leading through active demonstration of best practices, I contributed to a measurable improvement in both data quality, code quality and system performance.
  - with Faiss engine upgraded Fact Matching (FM) system and OOP refactored code base, the FM system is 9x faster than previous version in PROD.
    - I developed an easy to use and extensible Faiss based vector store and ANN based retrieval + clustering engine as standalone module adaptable for multi-projects and across multimodality applications in 2021 when autofaiss and most recent vector store techniques (e.g., Weaviate, milvus, pinecone) have not been not available
  - I actively mentored and upskilled junior colleagues, guiding them on SoTA methodologies, rigorous evaluation practices, and maintaining high-quality code of conduct throughout whole ML development lifecycle.
    - Through a comprehensive codebase rewrite guided by my commitment to cutting-edge ML practices (e.g., *DVC*, *MLflow*, *PyTorch Lightning*, *few-shot learning*, *contrastive learning*) during first few months leading misinfo AI team, I instilled demonstrably optimal standards of code hygiene and a strong foundation for the team's continued success.
    - To ensure successful adoption and standardization of best practices, I developed comprehensive technical documentation, incorporated informative inline comments, and created detailed step-by-step demo documents and presentation slides. These resources facilitated peer collaboration and knowledge transfer, leading to the formalization of the desired best practices.
  - Actively review and identify key improvements per sprint review with concrete solutions and performance weakness as follow-up for iterative improvement via team effort.
  - Always willing to invest time in coaching, hands-on-approach, pair-programming, constructive suggestions amongst the team in both engineering and workflow strategies.
  - I drove the development of claim detection dataset development (v2.x) by collaborating with teammate and product teams to identify, prioritise and secure new data source in this field.
    - The resulting dataset is >34x time larger than v1.x which covers multi-domains, multi-topics, and multi-categories, and eventually improves model robustness and efficiency in real-world applications, w.r.t. adversarial examples, deceptive language, and unsupported claims (e.g., general world knowledge)
    - This project has also paved the way for future approaches to claim detection data collection by establishing a well-designed data strategy and framework.
    - This work demonstrated my rich experience in dataset ablations and other sorts of data quality efforts to ensure data diversity, coverage, and consistency.
  - I took over the responsibility and ownership of the evidence retrieval capability when previous owner was left without meeting deliverable goals.
    - With limited time, I delivered an end-to-end standardised solution for automated fact-checking from scratch aligning with industry and academic standards.
    - The delivered solution showcased strong SoTA baseline performance (a hybrid architecture combining sparse and dense retrieval techniques) while provided with clear use case.
    - A large collection of datasets has been compiled with reusable and extensible data pipeline in one unified format that is easily accessible to test our evidence retrieval model and support end-to-end fact checking pipeline performance, which is highly adaptable & extensible to new dataset with minimum effort.
    - This work demonstrated my rich experience since Sheffield Uni. in developing and optimize data pipelines that efficiently handle the ingestion, annotation, and integration of large datasets for model training and evaluation
  - In the first few months, when leading misinfo AI team in Logically, I step in with *hands-on management* to document our existing work with different levels of details with well defined project templates serving for various different purposes including *cross-functional* and *team-level collaborations*, team level decisions & rationale, *code of conduct*, *best practices* and demos are regularly maintained and published in our team confluence space openly accessible, encourage teammates always to involve whole team with relevant expertise in code review via pull request and have regular team meeting to encourage engagement and involve everyone into decision making and planning process;
  - In another case, one teammate was working on a ML task with dataset trained by another colleague. After going through the model training, evaluation and system test, I found that dataset building method was not justifiable and dataset also has critical quality issues. Much

of my team leading work was spent on examining data samples and profiling quality of dataset, reviewing code, architecture design and evaluation metrics, do regular check-ins and proactively provide hands-on help for teammates who may stuck in a task or spend more time than expected.

- always stay on top of latest technology trends and leading misinfo AI team working on building innovative solutions with LLM
  - 1st working prototype with ChatGPT powered evidence-based fact checking system based on *in-context learning paradigm* has been made available for evaluation and feedback within 1 months in Jan 2023 when ChatGPT was released to public.
  - The system is built based on solid theoretic principle and clear understanding of its hype, limitation and advantage for business exploitation, and presentation and evaluation report has been made available for informed decision making.
- Leading Logically's participation and paper write-ups in Multimodal Fact Verification (Factify) challenge at AAAI 2022 and AAAI 2023 (twice) and ranked first in leaderboard in 1ST workshop (over 14 citations), ranked 3rd in 2nd workshop and **"Best Paper" award.**
- Lead and manage rumor detection project in OAK group (Sheffield) and successfully delivered three scientific papers and opensource dataset published in top tier AI conferences (**ICLR 2019, ASONAM 2019 and LREC 2020**) within 1 year without previous experience in rumor detection field.
- Developed two open source libraries of terminology recognition technique (JATE2 and jgTextrank) with research publications and comprehensive technical documentation along with many use case demo apps.
- Researched and developed a machine learning based entity typing algorithm in python and got **2nd place** in **OKE 2015 challenge** in less than 2 months, despite no previous python experience.
- Researched and developed an end-to-end distant supervision learning based sentiment classifier for Italian in one month and further enhance the prototype as a commercial product for K-Now (a spin-out company from OAK Group) in 2015, despite no previous sentiment analysis experience.

- **Experience** in Pytorch, Keras, LangChain, LlamaIndex, Numpy, scikit-learn, Python 3.x, Mlflow, Data Version Control (DVC), LLMs, Database (Faiss, Solr 5.x/6.x/7.x, Elasticsearch, MySQL 5.x, OWLIM/GraphDB, Sqlite), Web Service (Flask/fast.ai/unicorn/PM2, SOAP, WSDL, RESTful), NLTK, pytorch-geometric, Graph technique (igraph, networkx), Information Retrieval (Elasticsearch, Lucene, Apache Solr 5.x), ORM (Hibernate 3.x, SQLAlchemy), Java >6.x, Java EE, Spring 2.x/3.x (DI/IoC), Matlab, Javascript, NodeJS, HTTP Server(Flask, Tomcat 7.x), RDF framework (Sesame or Jena), Cloud computing (GCP, AWS, SolrCloud, Zookeeper, security),
- Good **skills in agile methodologies**: TDD (Junit, mockito), CI/CD (Bamboo, Jenkins)
  - More than 6 years commercial experience in an AGILE Scrum based environment (KTP project with ActiveStandards development team, PHE active10 @Sheffield, DS team@Logically AI)
- **Software/OS proficiency**: IntelliJ / PyCharm, Eclipse, Git, Docker, SVN, MATLAB, WEKA, GATE Platform, protégé, Linux/Ubuntu, AWS EC2, GCP, LaTex/Overleaf, Gephi
- **Languages**: Chinese (Native), English (working proficiency)
- **Work eligibility**: Settlement – INDEFINITE LEAVE TO REMAIN