Georgios Spithourakis, PhD

Email: geospith@gmail.com Homepage: http://geospith.github.io

SUMMARY

Machine Learning Scientist with hands-on experience in big tech, academia, and startups. Expertise in neural networks, deep learning, and conversational AI, with a track record of deploying state-of-the-art spoken dialogue systems. Passionate about integrating NLP, ML, and software engineering to create innovative AI solutions. An active researcher with multiple publications, patents, and contributions to top-tier conferences. Recognized for persistence and creativity in tackling high-impact problems.

WORK EXPERIENCE

Semblar AI, London, UK
Founder & Tech Lead, Sep 2023 – Dec 2024
Built MVPs for AI-driven, voice-based, intelligent assistants, securing early LOIs from 3 local councils and 2 healthcare clinics ⬩ Led technical development from prototyping to deployment

Entrepreneur First, London, UK
Founder-in-residence, Feb 2023 – Jun 2023
Selected as one of 64 participants from 1000+ applicants for the LD20 cohort ⬩ Iterated over early-stage startup ideas and conducted market research with 100+ customer interviews to validate business opportunities

PolyAI, London, UK
Senior Machine Learning Scientist, Dec 2021 – Jan 2023
Designed and deployed novel ML models to infer user metadata (satisfaction levels, demographics), enabling the company's first large-scale user segmentation reporting and unlocking new client insights ⬩ Published 2 papers and 1 patent
Machine Learning Scientist, Dec 2018 – Dec 2021
Built and deployed a state-of-the-art knowledge-based identification and verification system for spoken dialogue users, expanding the company's product offerings and contributing to securing 5 new business clients within a year ⬩ Trained, evaluated and deployed NLU and ASR ML models for dialogue systems ⬩Published 3 papers and 1 patent
Machine Learning Engineer, Aug 2018 – Dec 2018
Developed, trained, and deployed NLU models for dialogue systems ⬩ Built core ML infrastructure (data collection, text processing, intent classification) in a fast-paced, early-stage startup (team of <10)

Amazon, Cambridge, UK
Research Scientist Intern, Jul 2016 – Feb 2017
Developed neural network models for question-answering in dialogue systems, contributing to the Alexa product

Microsoft Research, Redmond, Washington, US
Research Intern, Jun 2015 – Aug 2015
Developed neural network models for conversational AI, contributing to the Cortana product ⬩ Published 2 papers

Department of Computer Science, University College London, UK
Postgraduate Teaching Assistant, *Oct 2013 – Dec 2018*
Led practical computer lab sessions for BSc/MSc software engineering courses, mentored student projects, and assessed coursework ⬩ Promoted to senior TA in 2015 (top 10% of assistants) and selected to lead a student exchange program with a Japanese university in 2016

Forecasting and Strategy Unit, National Technical University of Athens, Greece
Research Associate, Aug 2011 – Aug 2012
Developed and deployed a forecasting simulation platform, adopted by the university for MSc-level teaching ⬩ Conducted and published research on supply chain forecasting and assisted in research proposal development
Research Software Engineer, Sep 2009 – Aug 2011
Developed web-based tools for time-series forecasting

EDUCATION

University College London, UK
PhD in Natural Language Processing, 2018
Pioneered the concept of "numeracy" (understanding of numbers) for "language models" and catalysed an increased interest in the topic (after my original publication, research papers mentioning those terms have doubled year-on-year for at least 3 years) • Supervised by Sebastian Riedel and Steffen Petersen • Secured a fully funded studentship (Farr Institute of Health Informatics Research)
MSc in Computational Statistics and Machine Learning, 2013
Themes: supervised, unsupervised & reinforcement machine learning, information retrieval, and machine vision • Secured a funded scholarship (Foundation for Education and European Culture)

National Technical University of Athens, Greece
BSc/MSc in Electrical & Computer Engineering *(diploma)*, 2011
GPA: 8.58/10.0 (28th out of 325 in year of graduation) • I completed this 5-year degree in 4 years • Themes: control theory & signal processing (speech, image, video); computer networks (web application development, protocols, topology, hardware); management (decision support systems, forecasting • Wrote a dissertation on timeseries forecasting and published my findings in 3 follow-up papers in scientific journals • Secured a funded scholarship (Estate of Panagiotis Triadafyllidis)

TECHNICAL SKILLS

Proficient in **Python** and **ML & AI** development, productization and deployment

Languages (Python, Java, Scala, C++, MATLAB, R)• **ML & AI** (Tensorflow, Pytorch)
Software Engineering (Agile, scrum, Kanban, unit testing, CI/CD, git, github)
MLOps & Deployments (Kubernetes, docker, AWS, gcloud)
Databases (SQL, MySQL, PostgreSQL)•**Web & Mobile** (HTML, CSS, JavaScript, Android)

DISTINCTIONS

Scholarships
• PhD Studentship, The Farr Institute of Health Informatics Research, 2013 – 2016
• MSc Scholarship, Foundation for Education and European Culture (IPEP), 2012 – 2013
• BSc/MSc Scholarship, Estate of Panagiotis Triadafyllidis, 2007 – 2011

Grants & Awards
• Prize (for 3rd place out of 50 in stance detection task), Fake News Challenge, 2017
• Grant (for organising public engagement project), UCL Train and Engage, 2017
• Grant (for travelling to present at conference), EMNLP Student Scholarship Programme, 2016
• Award (for exceptional score in nationwide university entrance exams), President of the Hellenic Republic, 2008
• Award (for 1st place among municipality in nationwide university entrance exams), Eurobank EFG Group ("Great Moment for Education" Programme), 2007

OTHER SKILLS & ACTIVITIES

Languages
Greek (Mother Tongue)•English (Proficient/C2)•German (Advanced/C1)•Japanese (Intermediate/JLPT N3)• Spanish (Intermediate)

Volunteering
1. Volunteer for "Dinner of Love" NGO (distributing food and clothes to help homeless people), (1 year) 2020
2. Student volunteer, ACL conference, Melbourne, Australia, (1 week) July 2018
3. Student volunteer, EMNLP conference, Austin, Texas, USA, (1 week) November 2016
4. Environmental restoration, Wales, UK, (2 weeks) 2012
5. Greek translation from Japanese of "My Hiroshima" by Junko Morimoto, 2011

PUBLICATIONS

Conference Proceedings

1. **Spithourakis**, Vulić, Lis, Casanueva, and Budzianowski. EVI: Multilingual Spoken Dialogue Tasks and Dataset for Knowledge-Based Enrolment, Verification, and Identification. *NAACL* 2022
2. Casanueva, Vulić, **Spithourakis**, and Budzianowski. NLU++: A Multi-Label, Slot-Rich, Generalisable Dataset for Natural Language Understanding in Task-Oriented Dialogue. *NAACL* 2022
3. Henderson, Vulić, Casanueva, Budzianowski, Gerz, Coope, **Spithourakis**, Wen, Mrkšić, and Su. PolyResponse: A rank-based approach to task-oriented dialogue with application in restaurant search and booking. *EMNLP* 2019
4. Henderson, Vulić, Gerz, Casanueva, Budzianowski, Coope, **Spithourakis**, Wen, Mrkšić, and Su. Training neural response selection for task-oriented dialogue systems. *ACL* 2019
5. Henderson, Budzianowski, Casanueva, Coope, Gerz, Kumar, Mrkšić, **Spithourakis**, Su, Vulić, and Wen. A repository of conversational datasets. *ACL* 2019
6. **Spithourakis** and Riedel. Numeracy for Language Models: Evaluating their Ability to Predict Numbers. *ACL* 2018
7. Mostafazadeh, Brockett, Dolan, Galley, Gao, **Spithourakis** and Vanderwende. Image-Grounded Conversations: Multimodal Context for Natural Question and Response Generation. *IJCNLP* 2017
8. **Spithourakis**, Augenstein, and Riedel. Numerically Grounded Language Models for Semantic Error Correction. *EMNLP* 2016
9. Li, Galley, Brockett, **Spithourakis**, Gao, and Dolan. A Persona-Based Neural Conversation Model. *ACL*, 2016
10. Theodorou, **Spithourakis**, and Assimakopoulos. Investment portfolio management with trend forecasting methods. Paper presented at *the Operational Research Society Annual Conference (OR54)*, 2012
11. Mountalas, **Spithourakis**, and Assimakopoulos. Empirical application, evaluation and comparison of forecasting methods for the electrical system's System Marginal Price: a case study in Greek electricity market. *23rd National Conference of the Hellenic Operational Research Society*, 2012

Workshop Proceedings

1. Nouri, Rastogi, **Spithourakis**, Liu, Chen, Li, Albalak, Hiromi Wakaki, and Papangelis (organisers). *Proceedings of the 6th Workshop on NLP for Conversational AI*, ACL 2024
2. Rastogi, **Spithourakis**, Chen, Liu, Li, Nouri, Albalak, and Papangelis (organisers). *Proceedings of the 5th Workshop on NLP for Conversational AI*, ACL 2023
3. Liu, Papangelis, Ultes, Rastogi, Chen, **Spithourakis**, Nouri, and Shi (organisers). *Proceedings of the 4th Workshop on NLP for Conversational AI,* ACL 2022
4. **Spithourakis**, Petersen and Riedel. Clinical Text Prediction with Numerically Grounded Conditional Language Models. *LOUHI workshop*, *EMNLP* 2016
5. **Spithourakis,** Petersen, and Riedel. Harnessing the predictive power of clinical narrative to resolve inconsistencies and omissions in EHRs. *2nd Workshop on Machine Learning for Clinical Data Analysis, Healthcare and Genomics, NIPS* 2014

Journals

1. **Spithourakis,** Petropoulos, Nikolopoulos and Assimakopoulos. Amplifying the learning effect via a forecasting and foresight support system. *International Journal of Forecasting*, 2015.
2. *Spithourakis*, Petropoulos, Nikolopoulos and Assimakopoulos. A systemic view of the ADIDA framework. *IMA Journal of Management Mathematics*, 2014.
3. *Petropoulos, Nikolopoulos, **Spithourakis** and Assimakopoulos. Empirical heuristics for improving intermittent demand forecasting. *Industrial Management & Data Systems,* 2013
4. ***Spithourakis,** Petropoulos, Babai, Nikolopoulos and Assimakopoulos. Improving the performance of popular supply chain forecasting techniques: an empirical investigation. *Supply Chain Forum: an International Journal*, 2012

Miscellaneous

1. **Patent**. *System and method for identification and verification*, US11861521B2, 2022
2. **Patent**. *Response retrieval system and method*, US10664527B1, 2020
3. **PhD Thesis.** Numeracy of Language Models: Joint Modelling of Words and Numbers. Examined by Ivan Titov and Spiros Denaxas. PhD in Computer Science, UCL, 2018
4. Preprint. Riedel, Augenstein, **Spithourakis** and Riedel. A simple but tough-to-beat baseline for the Fake News Challenge stance detection task. arXiv, 2017
5. **MSc Dissertation**. Column generation in loopy models. Supervised by Sebastian Riedel, *MSc in Computational Statistics and Machine Learning, UCL*, 2013
6. **BSc Dissertation**. Application of aggregate-disaggregate forecasting methodology to non-intermittent demand data. Supervised by Vasilis Assimakopoulos, *BSc/MSc in Electrical & Computer Engineering, NTUA*, 2011

TALKS & EVENTS

Organiser
1. **Programme Chair**, *6th NLP4ConvAI workshop*. ACL, Bangkok, 2024
2. **Programme Chair**, *5th* NLP4ConvAI workshop. ACL, Toronto, 2023
3. **Publicity Chair**, *4th NLP4ConvAI workshop*. ACL, Dublin, 2022
4. **Public engagement**, *Human Code Computer Tongue* (event series) " (event series for teaching about neural networks through poetry generation, in collaboration with Zena Edwards and Apples & Snakes, funded by UCL Train and Engage). UCL, London, 2017

Invited Speaker
1. Tutorial on Conversational AI. ATILA, Antwerp, October 2019
2. Human Code Computer Tongue: A Human Neural Network. *Welcome/EPSRC Centre for Interventional and Surgical Sciences (WEISS)*, London, UK, March 2018
3. Human Code Computer Tongue: A Human Neural Network. *UCL Culture*, London, UK, January 2018
4. Numerically Grounded Language Models. Cambridge NLIP seminar series, *University of Cambridge*, UK, 2017
5. Text Mining and Sentiment Analysis. Operations Analytics course, *University of Cardiff*, UK, March 2015

TEACHING & MENTORING EXPERIENCE

Teaching Assistant (modules)
1. Natural Language Processing (MSc), UCL, autumn terms 2013 – 2018
2. Object-Oriented Programming (BSc), UCL, spring terms 2014 – 2015
3. Mobile App Design (MSc), UCL, autumn terms 2013 – 2015
4. Introductory Programming (MSc), UCL, autumn terms 2013 – 2014, 2015 (promoted to senior)
5. E-Health Analytics using Hadoop (one-week project), NII, Tokyo, Japan, 2016
6. Forecasting Techniques (MSc), NTUA, spring terms 2010 – 2012

Mentoring & student supervision
1. Alexander Hoyle. Detecting statements that can be fact checked. Co-supervised with Jeff Mitchel. MSc in Machine Learning, UCL, 2018
2. Finneas Jacob Robson Catling. Towards automated clinical coding. Co-supervised with Sebastian Riedel, MSc in Business Analytics (specialisation in Computer Science), UCL, 2017
3. Jiayu Sherri Lee. Question-to-statement conversion using pivot translations. Co-supervised with Sebastian Riedel, MSc in Computer Science, UCL, 2017
4. Benjamin Riedel. Simple yet powerful: A system for news article stance detection. Co-supervised with Sebastian Riedel and Isabelle Augenstein, MSc in Machine Learning, UCL, 2017
5. Dhruv Ghulati. Distant Supervision and Cost Sensitive Classification for Weakly Supervised Claim Detection. Co-supervised with Sebastian Riedel and Isabelle Augenstein, MSc in Computer Science, UCL, 2016
6. Emanuel Mountalas. Empirical application, evaluation and comparison of forecasting methods for the electrical system's System Marginal Price: a case study in Greek electricity market. BSc/MSc in Electrical Engineering, NTUA, 2012
7. Triantafyllos Nikolaos Theodorou. Investment portfolio management with trend forecasting methods. BSc/MSc in Electrical Engineering, NTUA, 2012
8. Andreas Desalermos. Application and evaluation of optimisation techniques for exponential smoothing parameters. Co-supervised with Fotios Petropoulos, BSc/MSc in Electrical Engineering, NTUA, 2011