Georgios Spithourakis

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

SUMMARY

Machine Learning Scientist with hands-on software engineering experience in big tech, universities, and start-ups.

I have channelled my fascination with computers and languages (both programming and human) into specialist academic training and a professional career in computer science, machine learning and artificial intelligence, and natural language processing. I have thoroughly worked with neural networks, deep learning, and conversational AI to build and deploy spoken dialogue systems for business applications, and I attain pro-active thought-leadership in the field by organising workshops in top-tier conferences and by publishing scientific papers and patents. My most valuable skills are persistence and creativity, which allow me to take on challenging, high-reward problems and come up with innovative solutions to stay ahead of the competition.

EMPLOYMENT

PolyAI, London, UK

Senior Machine Learning Scientist, December 2021 – now

Prototyped and deployed novel ML models for inferring customer metadata (satisfaction level and demographic information). This enabled the company to generate its first at-scale reports about the demographic segmentation of customers and provide insights to clients ⬩ Co-authored 2 papers and 1 patent

Machine Learning Scientist, December 2018 – December 2021

Solved the difficult problem of knowledge-based identification and verification of customers over spoken dialogue in noisy and ambiguous settings. This enabled the company to expand their product offering (state-of-the-art identification feature) and market share (1 year after I delivered the solution, the company secured 5 new business clients using this technology) ⬩ Trained, evaluated and deployed ML models for NLU and ASR for dialogue systems ⬩ Co-authored 3 papers and 1 patent

Machine Learning Engineer, August 2018 – December 2018

Trained, evaluated and deployed ML models for NLU for dialogue systems ⬩ Developed core software and ML infrastructure (data collection, text processing, intent classification) in early-stage start-up (fewer than 10 people).

Department of Computer Science, University College London, UK

Postgraduate Teaching Assistant, *October 2013 – December 2018*

Facilitated practical computer lab sessions for multiple software engineering BSc/MSc modules ⬩ Provided guidance and feedback to students on their projects and assessed their coursework ⬩ Among my peer cohort of 10 assistants, I was promoted to senior in 2015 and selected to lead a student team in an exchange programme with a Japanese university in 2016.

Amazon, Cambridge, UK

Research Scientist Intern, July 2016 – February 2017

Question answering with neural networks for dialogue systems (related to Alexa product)

Microsoft Research, Redmond, Washington, US

Research Intern, June 2015 – August 2015

Conversation modelling with neural networks (related to Cortana product)

Forecasting and Strategy Unit, National Technical University of Athens, Greece

Research Associate, August 2011 – August 2012

Deployed and evaluated forecasting simulation platform for teaching forecasting at MSc level ⬩ Conducted and published research on supply chain forecasting ⬩ Drafted research proposals

Research Software Engineer, September 2009 – August 2011

Developed web-based tools for time-series forecasting

Georgios Spithourakis

EDUCATION

University College London, UK

PhD in Natural Language Processing, 2018

My work introduced the concept of "numeracy" (understanding of numbers) for "language model" neural networks and catalysed an increased interest in the topic (just after my original publication, scientific 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 to scientific journals ⬩ Secured a funded scholarship (Estate of Panagiotis Triadafyllidis)

TECHNICAL SKILLS

Proficient in Python

Previous experience with:

ML (Tensorflow, Pytorch) ⬩ **MLOps/Deployments** (Kubernetes/kubectl, docker)
Cloud Storage/Compute (AWS, S3, gsutil, gcloud) ⬩ **Software engineering** (Agile, scrum, Kanban, unit testing, CI)
Web & Mobile Development (HTML, CSS, JavaScript, ASP.NET, AJAX, Android) ⬩ **Version control** (git, github)
Programming languages (Python, Java, Scala, VB.NET, C++, MATLAB, R) ⬩ **Databases** (SQL, MySQL, PostgreSQL)

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. Volunteer in environmental programme, Wales, UK, (2 weeks) 2012
5. Pro-bono Japanese-to-Greek translation of "My Hiroshima" by Junko Morimoto, 2011

Georgios Spithourakis

PUBLICATIONS

Conference Proceedings

1. **G. Spithourakis**, I. Vulić, M. Lis, I. Casanueva, and P. Budzianowski. EVI: Multilingual Spoken Dialogue Tasks and Dataset for Knowledge-Based Enrolment, Verification, and Identification. *NAACL* 2022.
2. I. Casanueva, I. Vulić, **G. Spithourakis**, and P. Budzianowski. NLU++: A Multi-Label, Slot-Rich, Generalisable Dataset for Natural Language Understanding in Task-Oriented Dialogue. *NAACL* 2022.
3. M. Henderson, I. Vulić, I. Casanueva, P. Budzianowski, D. Gerz, S. Coope, G. Spithourakis, T. Wen, N. Mrkšić, and P. Su. PolyResponse: A rank-based approach to task-oriented dialogue with application in restaurant search and booking. *EMNLP* 2019
4. M. Henderson, I. Vulić, D. Gerz, I. Casanueva, P. Budzianowski, S. Coope, **G. Spithourakis**, T. Wen, N. Mrkšić, and P. Su. Training neural response selection for task-oriented dialogue systems. *ACL* 2019.
5. M. Henderson, P. Budzianowski, I. Casanueva, S. Coope, D. Gerz, G. Kumar, N. Mrkšić, **G. Spithourakis**, P. Su, I. Vulić, and T. Wen. A repository of conversational datasets. *ACL* 2019.
6. **G. Spithourakis** and S. Riedel. Numeracy for Language Models: Evaluating their Ability to Predict Numbers. *ACL* 2018.
7. N. Mostafazadeh, C. Brockett, B. Dolan, M. Galley, J. Gao, **G. Spithourakis** and L. Vanderwende. Image-Grounded Conversations: Multimodal Context for Natural Question and Response Generation. *IJCNLP* 2017.
8. **G. Spithourakis**, I. Augenstein and S. Riedel. Numerically Grounded Language Models for Semantic Error Correction. *EMNLP* 2016.
9. J. Li, M. Galley, C. Brockett, **G. Spithourakis**, J. Gao, B. Dolan. A Persona-Based Neural Conversation Model. *ACL*, 2016.
10. N. Theodorou, **G. Spithourakis** and V. Assimakopoulos. Investment portfolio management with trend forecasting methods. Paper presented at *the Operational Research Society Annual Conference (OR54)*, 2012.
11. E. Mountalas, **G. Spithourakis** and V. Assimakopoulos. Empirical application, evaluation and comparison of forecasting methods for the electrical system's System Marginal Price: a case study in Greek electricity market. In *Proceedings of the 23rd National Conference of the Hellenic Operational Research Society*, pages 9-13, 2012.

Workshop Proceedings

1. B. Liu, A. Papangelis, S. Ultes, A. Rastogi, Y. Chen, **G. Spithourakis**, E. Nouri, and W. Shi (organisers). *Proceedings of the 4th Workshop on NLP for Conversational AI,* ACL 2016.
2. **G. Spithourakis**, S. Petersen and S. Riedel. Clinical Text Prediction with Numerically Grounded Conditional Language Models. *LOUHI workshop*, *EMNLP* 2016.
3. **G. Spithourakis,** S. Petersen, and S. 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. **G. Spithourakis,** F. Petropoulos, K. Nikolopoulos and V. Assimakopoulos. Amplifying the learning effect via a forecasting and foresight support system. *International Journal of Forecasting*, 31(1):20-32, 2015.
2. **G. Spithourakis**, F. Petropoulos, K. Nikolopoulos and V. Assimakopoulos. A systemic view of the ADIDA framework. *IMA Journal of Management Mathematics*, 25(2): 125-137, 2014.
3. F. Petropoulos, K. Nikolopoulos**, G. Spithourakis** and V. Assimakopoulos. Empirical heuristics for improving intermittent demand forecasting. *Industrial Management & Data Systems*, 113(5):683-696, 2013.
4. **G. Spithourakis,** F. Petropoulos, M.Z. Babai, K. Nikolopoulos and V. Assimakopoulos. Improving the performance of popular supply chain forecasting techniques: an empirical investigation. *Supply Chain Forum: an International Journal*, 12(4):16-25, 2012.

Miscellaneous

1. **Patent**. System for identifying and verifying callers, (under examination) 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. B. Riedel, I. Augenstein, **G. Spithourakis** and S. Riedel. A simple but tough-to-beat baseline for the Fake News Challenge stance detection task. arXiv, 2017.
5. **Dissertation**. **G. Spithourakis.** Column generation in loopy models. Supervised by Sebastian Riedel, *MSc in Computational Statistics and Machine Learning, UCL*, 2013.
6. **Dissertation**. **G. Spithourakis.** 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. **Workshop co-organiser & Programme Chair**, *5th NLP4ConvAI workshop*. ACL, Toronto, 2023 (scheduled)
2. **Workshop co-organiser & Diversity Chair**, *4th NLP4ConvAI workshop*. ACL, Dublin, 2022
3. **Public engagement co-organiser**, *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 (1 performance and 2 all-day workshops).

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.