

## Josef Valvoda

---

UK Cell Phone: +44 7522 177622 | Email: jv406@cam.ac.uk

- EDUCATION**
- PhD, COMPUTER SCIENCE* 2019 - 2023  
**University of Cambridge**
- Co-advised by Simone Teufel (UoC) and Ryan Cotterell (ETH).
  - Fully funded by Huawei research stipend.
  - Part of the Natural Language Processing and Understanding research group.
  - Part of the ETH AI centre research group - Rycolab.
  - Developing deep learning techniques for robust modelling of legal arguments.
- MPhil, ADVANCED COMPUTER SCIENCE* 2018 - 2019  
**University of Cambridge**
- Graduated with a distinction and a first class in every module.
  - Best paper award at JURIX'18 for legal NLP paper.
- MSc, COMPUTER SCIENCE* 2016 - 2017  
**University of Bristol**
- Received the *Best MSc Computer Science Student* award.
  - Achieved first class grades in all modules and graduated with a distinction.
  - My Thesis was published in the Lecture Notes on Artificial Intelligence journal.
- LLB. LAW* 2013 - 2016  
**University of Exeter**
- Achieved first class for my dissertation: *Attribution Methods in Cyberspace*.
  - Represented England at Jean-Pictet International Law competition.
- PUBLICATION LIST**
- Valvoda J.**, Thompson A., Cotterell R., and Teufel S. (2023) The Ethics of Automating Legal Actors, *Accepted in Transactions of the Association for Computational Linguistics*
- Stoehr N., Hennigen L., **Valvoda J.**, West R., Cotterell R., and Schein A. (2023) An Ordinal Latent Variable Model of Conflict Intensity, *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*
- Valvoda J.**, Cotterell R., Teufel S. (2023) On the Role of Negative Precedent in Legal Outcome Prediction, *Transactions of the Association for Computational Linguistics*
- Pimentel T.\*, **Valvoda J.\***, Stoehr N., Cotterell R. (2022) The Architectural Bottleneck Principle, *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing of the Association for Computational Linguistics*
- Valvoda J.**, Naomi S., Rawski J., Williams A., Cotterell R. (2022) Benchmarking Compositionality with Formal Languages, *Proceedings of the 29th International Conference on Computational Linguistics*
- Valvoda J.**, Yimai, F., Vandyke D. (2022) Prompting for a conversation: How to control a dialog model?, *Proceedings of the Second Workshop on When Creative AI Meets Conversational AI*
- Talat Z., Blix H., **Valvoda J.**, Ganesh M., Cotterell R., Williams A (2022) On the Machine Learning of Ethical Judgments from Natural Language, *In Proceedings of the*

2022 Conference of the North American Chapter of the Association for Computational Linguistics

Batsuren K. et al. (2022) UniMorph 4.0: Universal Morphology, *In Proceedings of the 13th Conference on Language Resources and Evaluation*

**Valvoda J.**, Pimentel T., Stoehr N., Cotterell R., Teufel S. (2021) What About the Precedent: An Information-Theoretic Analysis of Common Law. *In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics*

Pimentel T., **Valvoda J.**, Hall Maudslay R., Zmigrod R., Williams A., Cotterell R. (2020) Information-Theoretic Probing for Linguistic Structure. *In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*

Hall Maudslay R., **Valvoda J.**, Pimentel T. Williams A., Cotterell R. (2020) A Tale of a Probe and a Parser. *In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*

Farag Y., **Valvoda J.**, Yannakoudakis H., Briscoe T. (2020) Analyzing Neural Discourse Coherence Models. *In Proceedings of the First Workshop on Computational Approaches to Discourse; Association for Computational Linguistics*

Vylomova E., White J., Salesky E., Mielke S., Wu S., Ponti E., Maudslay R., Zmigrod R., **Valvoda J.**, Toldova S., Tyers F., Klyachko E., Yegorov I., Krizhanovsky N., Czarnowska P., Nikkarinen I., Krizhanovsky A., Pimentel T., Hennigen L., et al.. (2020) SIGMORPHON 2020 Shared Task 0: Typologically Diverse Morphological Inflection. *In Proceedings of the 17th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*

**Valvoda J.**, Ray O. (2018) From Case Law to Ratio Decidendi. *In New Frontiers in Artificial Intelligence. JSAI-isAI 2017. Lecture Notes in Computer Science, vol 10838. Springer, Cham*

**Valvoda J.**, Ray O., Satoh K. (2018) Using agreement statements to identify majority opinion in UKHL case law. *In Legal Knowledge and Information Systems - JURIX 2018: The Thirty-first Annual Conference, Groningen, The Netherlands, 12-14 December 2018*

\* shared first authorship

## WORK EXPERIENCE

APPLIED SCIENTIST INTERNSHIP  
Amazon Seattle, USA

August - November 2023

- Working with the PXT Central Science team.
- Project focus on long document processing with LLMs.
- Paper submitted for review at NAACL'24

AI/ML - MACHINE LEARNING INTERNSHIP  
Apple, Cambridge

June - September 2021

- Worked on the Siri experience team.
- Researched prompting LLMs for dialogue modelling.
- *Prompting For a Conversation* paper presented at COLING'22

*VISITING RESEARCHER* March - June 2022  
**ETH Zurich, Switzerland**

- Visiting Ryan Cotterell's lab.
- Developing new methods for probing language models.
- *The Architectural Bottleneck Principle* paper presented at EMNLP'22

*NLP RESEARCHER* July - September 2020  
**TechWolf, Ghent**

- Build a hierarchical transformer language model for job/vacancy matching.
- Lead a weekly NLP reading group at the company.

*NLP SUPERVISOR* October 2019 (ongoing)  
**University of Cambridge, Cambridge**

- Supervise students in classes on Natural Language Processing, Machine Learning and Real World Data.
- Teach techniques ranging from Naive Bayes to Doc2Vec.

*NLP RESEARCHER* July - September 2019  
**Lexsnap, London**

- Developed a legal document search engine using BERT.
- Lead a team of interns for product development.

*RESEARCH PLACEMENT* 3-6 Months Spring 2018 & 2020 & 2023  
**National Institute of Informatics, Tokyo**

- Best paper award at JURIX'18 conference.
- Used ML and NLP techniques to automate the legal research task of majority opinion identification to a near human-level accuracy.

**INVITED  
TALKS**

*AllenNLP - AI2* November 2023  
**Invited Talk** organised by Yanai Elazar

*FTP AI Center* April 2023  
**Invited Talk** organised by Truong Son Hy

*National Institute of Advanced Industrial Science and Technology* April 2023  
**Invited Talk** organised by Edison Marrese-Taylor

*Tokyo Institute of Technology* March 2023  
**Invited Talk** organised by Professor Hiroaki Yamada

*Seminars on Formal Languages and Neural Networks* January 2023  
**FLaNN Seminars** organised by Will Merrill

*EMNLP - Natural Legal Language Processing Workshop* December 2022  
**NLLP Workshop** organised by Professor Nikolaos Aletras

*ETH Zurich* July 2021  
**Seminar in Economics + Data Science** organised by Professor Elliott Ash

*University of Sheffield* June 2021  
**NLLP Talk Series** organised by Professor Nikolaos Aletras

**ACADEMIC  
SERVICE**

*CONFERENCE ORGANISATION*

**COLIEE Shared Task** Program Committee 2021/2022/2023

**NLLP Workshop** Program Committee 2021/2022/2023

**JURIX Workshop** Program Committee 2022/2023

**SIGTYP** Session Chair 2021

*REVIEWING*

I review for **ARR (ACL, NAACL), EMNLP, NEURIPS, JURIX, JURISIN, NLLP** and **COLIEE**

**PROFICIENCY**

*PROGRAMMING*

I code in **Python** and have worked on the above publications with the **Pytorch** and **Transformers** libraries.

*LANGUAGES*

Czech (first language), English (fluent), German (beginner), French (beginner)