Joe Davison – curriculum vitae

Phone	+1 (702) 608 4062
Email	joeddav@gmail.com
GitHub	GITHUB.COM/JOEDDAV
Semantic Scholar	JOE-DAVISON

Education

Degrees

2018-2020 **Harvard University**
M.S. Data Science

2015-2018 **Brigham Young University**
B.S. Computer Science

Highlighted Coursework

Harvard	Machine Learning for Natural Language	taught by	Alexander Rush
	Topics in Machine Learning: Deep Bayesian Models		Finale Doshi-Velez
	Machine Learning: Advances in Uncertainty Quantification, Structured Prediction, and Large-Scale Models		Jasper Snoek et al.
BYU	Probabilistic Machine Learning		David Wingate
	Deep Learning: Theory & Practice		David Wingate
	Tools for Machine Learning		Tony Martinez
	Intro to Artificial Intelligence		Jacob Crandall

Work & Internship Experience

Jul. 2021 – Dec. 2022 **Enveda Biosciences**
Senior Machine Learning Scientist / Manager

Developed machine learning system for crucial component of drug discovery platform, providing industry-leading internal tools for molecular structure elucidation.

Feb. 2020 – Jul. 2021 **Hugging Face**
Research Engineer

Implemented state-of-the-art zero-shot classification tools, enabling text classification via the open-source TRANSFORMERS library and Hugging Face API endpoint in 100 languages without supervised training.

Jun. 2019 – Sep. 2019 **IBM Research**
Research Intern

Joe Davison, Kristen Severson, and Soumya Ghosh. **CROSS-POPULATION VARIATIONAL AUTOENCODERS**. In *4th workshop on Bayesian Deep Learning (NeurIPS)*, 2019b. URL `http://bayesiandeeplearning.org/2019/papers/96.pdf`

May 2018 – Aug. 2018 **Pluralsight**
Data Science Intern

Improved search and recommendation systems by developing universal model for embedding 5 educational content formats into common vector space. Created dashboard for comparison of models including DOC2VEC, FASTTEXT, LDA, and TF-IDF.

Sep. 2017 – Apr. 2018 **Zeff**
Machine Learning Engineer *(Part-time)*
Developed image embedding model trained with distributed MXNet framework on over 40 million images.

May 2017 – Aug. 2017 **Microsoft**
Software Engineer Intern
Expanded CNTK deep learning framework by creating R bindings to existing Python interface.

Aug. 2016– May 2017 **Qualtrics**
Software Engineer *(Part-time)*
Developed Redis microservice for logging and maintenance of internal export utility.

May 2016– Aug. 2016 **Instructure**
Software Engineer Intern
Converted Pages module in Canvas app from outdated Objective C to more modern Swift, updating with modern styling.

Aug. 2015– Apr. 2016 **Rollins Center for Entrepreneurship & Technology — BYU**
Web Developer / IBMC Student Co-Lead *(Part-time)*
Designed new website and led student team for 2016 *International Business Model Competition* held at Microsoft compus in Seattle.

Jan. 2015 Aug. 2015 **Dept. of Microbiology & Molecular Biology — BYU**
Web Designer *(Part-time)*
Designed interactive website to increase student interest in department's programs.

Publications

First Author

1. **Joe Davison**, Joshua Feldman, and Alexander Rush. **Commonsense knowledge mining from pretrained models**. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 1173–1178, Hong Kong, China, November 2019a. Association for Computational Linguistics. doi: 10.18653/v1/D19-1109. URL `https://aclanthology.org/D19-1109`

   EMNLP Oral Presentation: youtube.com/watch?v=NBYL7s8cVfw

2. **Joe Davison**, Kristen Severson, and Soumya Ghosh. **Cross-population variational autoencoders**. In *4th workshop on Bayesian Deep Learning (NeurIPS)*, 2019b. URL `http://bayesiandeeplearning.org/2019/papers/96.pdf`

   Workshop Poster: dropbox.com/.../CPVAE Poster.pdf

Contributing Author

3. Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, **Joe Davison**, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander Rush. **Transformers: State-of-the-art natural language processing**. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45, Online, October 2020. Association for Computational Linguistics. doi: 10.18653 / v1 / 2020.emnlp - demos.6. URL `https://aclanthology.org/2020.emnlp-demos.6`

4. Quentin Lhoest, Albert Villanova del Moral, Yacine Jernite, Abhishek Thakur, Patrick von Platen, Suraj Patil, Julien Chaumond, Mariama Drame, Julien Plu, Lewis Tunstall, **Joe Davison**, Mario Šaško, Gunjan Chhablani, Bhavitvya Malik, Simon Brandeis, Teven Le Scao, Victor Sanh, Canwen Xu, Nicolas Patry, Angelina McMillan-Major, Philipp Schmid, Sylvain Gugger, Clément Delangue, Théo Matussière, Lysandre Debut, Stas Bekman, Pierric Cistac, Thibault Goehringer, Victor Mustar, François Lagunas, Alexander Rush, and Thomas Wolf. **Datasets: A community library for natural language processing**. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 175–184, Online and Punta Cana, Dominican Republic, November 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.emnlp-demo.21. URL `https://aclanthology.org/2021.emnlp-demo.21`

5. Gennady Voronov, Rose Lightheart, **Joe Davison**, Christoph A Krettler, David Healey, and Thomas Butler. **Multi-scale sinusoidal embeddings enable learning on high resolution mass spectrometry data**. *arXiv preprint arXiv:2207.02980*, 2022

6. Mark Hamilton, Sudarshan Raghunathan, Akshaya Annavajhala, Danil Kirsanov, Eduardo Leon, Eli Barzilay, Ilya Matiach, **Joe Davison**, Maureen Busch, Miruna Oprescu, Ratan Sur, Roope Astala, Tong Wen, and ChangYoung Park. **Flexible and scalable deep learning with mmlspark**. In Claire Hardgrove, Louis Dorard, and Keiran Thompson, editors, *Proceedings of The 4th International Conference on Predictive Applications and APIs*, volume 82 of *Proceedings of Machine Learning Research*, pages 11–22. PMLR, 24–25 Oct 2018. URL `https://proceedings.mlr.press/v82/hamilton18a.html`

Open Source Projects

2017 **joeddav/devol**
★ 950
Proof-of-concept repository enabling genetic search of ConvNet architectures represented by fixed-length vectors.

2017 **microsoft/CNTK-R**
★ 32
Proof-of-concept repository enabling genetic search of ConvNet architectures represented by fixed-length vectors.

Note: Developed as 2017 Microsoft internship project

2016 **joeddav/get_smarties**
★ 20
Standalone tool for dummy variable generation in Python with full sklearn fit/transform compatibility.

Personal

Books	The Way of Kings	by Brandon Sanderson
	Shōgun	by James Clavell
	Born a Crime	by Trevor Noah
	Educated	by Tara Westover
Theater	Hadestown (2019)	by Anaïs Mitchell
	Camelot (1960)	by Alan Jay Lerner
	Bright Star (2014)	by Steve Martin and Edie Brickell
Languages	English	native
	Russian	conversational
Hobbies	piano, guitar, ukulele, racquetball, skiing, photography (nature & wildlife), hiking, reading	