JESSE DODGE

https://jessedodge.github.io/
jessed@allenai.org, dodgejesse@gmail.com

PROFESSIONAL EXPERIENCE

Allen Institute for Artificial Intelligence 2020-09 to Present
Research Scientist, Allen NLP Team
Reproducibility, Efficiency, and Ethics in AI

Allen Institute for Artificial Intelligence 2019-01 to 2019-12
Research Intern, Allen NLP Team
Mentor: Noah A. Smith, Project: Green AI

Google AI 2018-04 to 2018-08
Research Intern, Machine Perception Team
Mentor: Elad Eban, Project: Structured sparsity on neural OCR models [blog]

Facebook AI Research 2015-05 to 2015-10
Research Intern, FAIR
Mentors: Jason Weston, Antoine Bordes. Project: Movie Dialog, WikiMovies Datasets [website]

Johns Hopkins 2011-06 to 2011-08
Research Intern, Frederick Jelinek Memorial Summer Workshop
Mentors: Margaret Mitchell, Project: Visually Descriptive Text [website]

Columbia University 2011-06
Research Intern, Explorations in Statistics Research Summer Workshop [website]

EDUCATION

Carnegie Mellon University

PhD in Language and Information Technology 2020-05
Language Technologies Institute, School of Computer Science
Thesis: *Towards Efficient and Reproducible Natural Language Processing*
Advisor: Noah A. Smith

Master of Language Technologies 2015-05
Language Technologies Institute, School of Computer Science

University of Washington

Bachelor of Science: Computer Science (Honors), Statistics 2013-06
Thesis: *Learning a Context-Dependent Semantic Parser for Temporal Expression Resolution*
Advisor: Luke Zettlemoyer

INVITED TALKS AND PANELS

Learning Workshop 2022-03-04
Invited Talk: Data-first Machine Learning

TVMCon 2021 2021-12-17
Invited Talk: Green AI [video]

NeurIPS Plenary Panel, NeurIPS 2021 [website] 2021-12-10
Invited Panelist: How should a machine learning researcher think about AI ethics?

Workshop on Enormous Language Models, ICLR 2021 [website] 2021-05-07
Invited Talk: Is Brevity the Soul of Wit? What Information to Report About Our Data [video]

Negative Results Workshop, EMNLP 2020 [website] 2020-12-15
Invited Panelist: Leaderboardism in NLP [video]

Allen Institute for Artificial Intelligence 2018-04-18
Invited Talk: Open Loop Hyperparameter Optimization and Determinantal Point Processes [video]

Cambridge University NLP Seminar 2014-07-18
Invited Talk: Context-dependent Semantic Parsing for Time Expressions

SELECTED PRESS

New York Times, "Can a Machine Learn Morality?", Cade Metz 2021-11-19

IEEE Spectrum, "Making Information Tech Greener Can Help Address the Climate Crisis", San Murugesan 2021-10-07

Unite.AI, "Minority Voices 'Filtered' Out of Google Natural Language Processing Models", Martin Anderson 2021-09-24

The Register, AI caramba, those neural networks are power-hungry: Counting the environmental cost of artificial intelligence, Danny Bradbury 2021-09-13

Wired, "The Efforts to Make Text-Based AI Less Racist and Terrible", Khari Johnson 2021-06-17

TechWireAsia, "Is 'Green AI' the same as environmental AI?", Joe Devanesan 2021-01-19

Guest on Practical AI Podcast, Green AI, Chris Benson and Daniel Whitenack 2021-02-04

Wired, "The Dark Side of Big Tech's Funding for AI Research", Tom Simonite 2020-12-10

Forbes, "Deep Learning's Carbon Emissions Problem", Rob Toews 2020-06-17

New York Times, "At Tech's Leading Edge, Worry About a Concentration of Power", Steve Lohr 2019-09-26

PUBLICATIONS

Data Governance in the Age of Large-Scale Data-Driven Language Technology
Yacine Jernite, Huu Nguyen, Stella Biderman, Anna Rogers, Maraim Masoud, Valentin Danchev, Samson Tan, Alexandra Sasha Luccioni, Nishant Subramani, Gérard Dupont, **Jesse Dodge**, Kyle Lo, Zeerak Talat, Dragomir Radev, Somaieh Nikpoor, Aaron Gokaslan, Peter Henderson, Rishi Bommasani, Margaret Mitchell
ACM Conference on Fairness, Accountability, and Transparency (FAccT), 2022

Measuring Machine Learning Software Carbon Intensity in Cloud Instances
Jesse Dodge, Taylor Prewitt, Remi Tachet des Combes, Erika Odmark, Roy Schwartz, Emma Strubell, Alexandra Sasha Luccioni, Noah A. Smith, Nicole DeCario, Will Buchanan
ACM Conference on Fairness, Accountability, and Transparency (FAccT), 2022

Staged Training for Transformer Language Models [pdf]
Sheng Shen, Pete Walsh, Kurt Keutzer, **Jesse Dodge**, Matthew E. Peters, Iz Beltagy
under review, 2022

Efficient Hierarchical Domain Adaptation for Pretrained Language Models [pdf]
Alexandra Chronopoulou, Matthew E. Peters, **Jesse Dodge**
North American Chapter of the Association for Computational Linguistics (NAACL), 2022

Documenting Large Webtext Corpora: A Case Study on the Colossal Clean Crawled Corpus [pdf]
Jesse Dodge, Maarten Sap, Ana Marasović, William Agnew, Gabriel Ilharco, Dirk Groeneveld, Margaret Mitchell, Matt Gardner
Empirical Methods on Natural Language Processing (EMNLP), 2021

Competency Problems: On Finding and Removing Artifacts in Language Data [pdf]
Matt Gardner*, William Merrill*, **Jesse Dodge**, Matthew E. Peters, Alexis Ross, Sameer Singh, Noah A. Smith
Empirical Methods on Natural Language Processing (EMNLP), 2021
* denotes equal contribution

Expected Validation Performance and Estimation of a Random Variable's Maximum [pdf]
Jesse Dodge, Suchin Gururangan, Dallas Card, Roy Schwartz, Noah A. Smith
Findings of Empirical Methods on Natural Language Processing (EMNLP Findings), 2021

Probing Language Models for Commonsense Knowledge using Template Variations
Jesse Dodge, Karishma Mandyam, Akari Asai, Hannaneh Hajishirzi, Noah A. Smith
2020

Towards Efficient and Reproducible Natural Language Processing [pdf]
Jesse Dodge
PhD Thesis, Carnegie Mellon University, 2020

Fine-Tuning Pretrained Language Models: Weight Initializations, Data Orders, and Early Stopping [pdf]
Jesse Dodge, Gabriel Ilharco, Roy Schwartz, Ali Farhadi, Hannaneh Hajishirzi, Noah A. Smith
arXiv, 2020

The Right Tool for the Job: Matching Model and Instance Complexities [pdf]
Roy Schwartz, Gabriel Stanovsky, Swabha Swayamdipta, **Jesse Dodge**, Noah A. Smith

Association for Computational Linguistics (ACL), 2020

Green AI [pdf]
Roy Schwartz*, **Jesse Dodge***, Noah A. Smith, Oren Etzioni
Communications of the ACM (CACM), 2020
* denotes equal contribution

Show Your Work: Improved Reporting of Experimental Results [pdf]
Jesse Dodge, Suchin Gururangan, Dallas Card, Roy Schwartz, Noah A. Smith
Empirical Methods on Natural Language Processing (EMNLP), 2019

RNN Architecture Learning with Sparse Regularization [pdf]
Jesse Dodge, Roy Schwartz, Hao Peng, Noah A. Smith
Empirical Methods on Natural Language Processing (EMNLP), 2019

Open Loop Hyperparameter Optimization and Determinantal Point Processes [pdf]
Jesse Dodge, Kevin Jamieson, Noah A. Smith
AutoML Workshop at International Conference on Machine Learning (AutoML at ICML), 2017

Key-Value Memory Networks for Directly Reading Documents [pdf]
Alexander Miller, Adam Fisch, **Jesse Dodge**, Amir-Hossein Karimi, Antoine Bordes, Jason Weston
Empirical Methods on Natural Language Processing (EMNLP)*, 2016*

Evaluating Prerequisite Qualities for Learning End-to-end Dialog Systems [pdf] [poster]
Jesse Dodge*, Andreea Gane*, Xiang Zhang*, Antoine Bordes, Sumit Chopra, Alexander Miller, Arthur Szlam, Jason Weston
International Conference on Learning Representations (ICLR)*,* 2016
* denotes equal contribution

Retrofitting Word Vectors to Semantic Lexicons [pdf] [code]
Manaal Faruqui, **Jesse Dodge**, Sujay Kumar Jauhar, Chris Dyer,Eduard Hovy, and Noah A. Smith.
North American Chapter of the Association for Computational Linguistics (NAACL), 2015
Won Best Student Paper Award

Large scale retrieval and generation of image descriptions [pdf]
Vicente Ordonez, Xufeng Han, Polina Kuznetsova, Girish Kulkarni, Margaret Mitchell, Kota Yamaguchi, Karl Stratos, Amit Goyal, **Jesse Dodge**, Alyssa Mensch, Hal Daumé III, Alexander C Berg, Yejin Choi, Tamara L Berg
International Journal of Computer Vision, 2015

CMU: Arc-Factored, Discriminative Semantic Dependency Parsing [pdf]
Sam Thomson, Brendan O'Connor, Jeffrey Flanigan, David Bamman, **Jesse Dodge**, Swabha Swayamdipta, Nathan Schneider, Chris Dyer, and Noah A. Smith

International (COLING) Workshop on Semantic Evaluations (SemEval)*, 2014.*

Context-dependent Semantic Parsing for Time Expressions [pdf] [demo] [code] [tool]
Kenton Lee, Yoav Artzi, **Jesse Dodge**, Luke Zettlemoyer
Association for Computational Linguistic (ACL)*, 2014.*

Detecting Visual Text [pdf]
Jesse Dodge, Amit Goyal, Xufeng Han, Alyssa Mensch, Margaret Mitchell, Karl Stratos, Kota Yamaguchi, Yejin Choi, Hal Daumé III, Alexander C. Berg, Tamara L. Berg
North American Chapter of the Association for Computational Linguistics (NAACL)*, 2012.*

Midge: Generating Image Descriptions From Computer Vision Detections [pdf]
Margaret Mitchell, **Jesse Dodge**, Amit Goyal, Kota Yamaguchi, Karl Sratos, Xufeng Han, Alysssa Mensch, Alexander C. Berg,Tamara L. Berg, Hal Daumé III
European Chapter of the Association for computational Linguistics (EACL)*, 2012.*

Understanding and Predicting Importance in Images [pdf]
Alexander C. Berg, Tamara L Berg, Hal Daumé III, **Jesse Dodge**, Amit Goyal, Xufeng Han, Alyssa Mensch, Margaret Mitchell, Aneesh Sood, Karl Stratos, Kota Yamaguchi
Computer Vision and Pattern Recognition (CVPR)*, 2012.*

BLOG POSTS

NAACL 2022: NAACL 2022 Reproducibility Track
Allen Institute for AI Blog 2021: Empowering cloud providers and AI practitioners to make greener decisions
Microsoft Green Tech Blog 2021: Charting the path towards sustainable AI with Azure Machine Learning resource metrics
Reproducibility Challenge 2021: The Reproducibility Challenge as an Educational Tool
EMNLP 2020: Guest Post: Reproducibility at EMNLP 2020
Google AI Blog, 2019: MorphNet: Towards Faster and Smaller Neural Networks

SERVICE

Reproducibility: Created reproducibility checklist for EMNLP 2020 [blog]. Checklist was used at EMNLP 2020, NAACL 2021, ACL 2021, EMNLP 2021
Reproducibility Chair at NAACL 2022
Organizer of the Reproducibility Challenge 2020, 2021

Workshop Organization:
Machine Learning Retrospectives, ICML 2020 [website]
ML-Retrospectives, Surveys & Meta-Analyses, NeurIPS 2020 [website]
Setting up ML Evaluation Standards to Accelerate Progress, ICLR 2022 [website]

Tutorial Organization:
Reproducibility, ACL 2022

Program Committees:
2016: NAACL, NAACL Student Research Workshop, EMNLP
2017: ACL, ACL Demo, EMNLP, ACL RoboNLP Workshop
2018: NAACL, NAACL Student Research Workshop, ACL, CoNLL, EMNLP Demo
2019: ICML, ACL Demo, NeuralGen workshop, CoNLL, EMNLP-IJCNLP Demo, JAIR, NeurIPS (top 50% of reviewers)
2020: AAAI, ICML, UAI, ACL, NeurIPS, EMNLP (outstanding reviewer), TACL, Patterns (Cell Press), NeurIPS Pre-registration workshop
2021: ICLR, ICML, NeurIPS, Patterns (Cell Press)

Area Chair:
2021: Green NLP track at EACL, Resources and Evaluation track at ACL-IJCNLP, Green NLP track at NAACL, ARR Action Editor 2021

Senior Area Chair:
2021: Efficient Methods for NLP at EMNLP

TEACHING

Teaching Assistant:

CSE 599D1: Advanced Natural Language Processing, UW CSE.	2016-03 to 2016-06
CSE 517: Natural Language Processing, UW CSE.	2016-01 to 2016-03
CSE 142: Computer Programming 1, UW CSE.	2010-06 to 2010-12

Creation and Management of Class Project:

CSE 517: Natural Language Processing, UW CSE	2019-01 to 2019-03

PROGRAMMING LANGUAGES

Experienced: Python. Exposed: Java, R, Matlab