

# Jack Hessel

LLMs, Evaluation, Data Curation, Multimodality

Email: [jmhessel@gmail.com](mailto:jmhessel@gmail.com)

Website: [www.jmhessel.com](http://www.jmhessel.com)

Publications: [Google Scholar](#)

## Education

### Ph.D., Computer Science

Cornell University

August 2014 - August 2020

*Thesis: Learning from Multimodal Web Data*

### B.A., Computer Science and Math/Statistics

Carleton College

September 2010 - June 2014

*Magna Cum Laude + Computer Science Honors*

## Employment

### Member of Technical Staff, Anthropic

Sep 2025 – Present

### Founding Researcher, Samaya AI

Oct 2023 – Sep 2025

Building, evaluating, and shipping LLM-based knowledge discovery agents

### Postdoc+Research Scientist, Allen Institute for Artificial Intelligence (A1)

July 2020 – Oct 2023

Host: Professor Yejin Choi

Published primarily on vision+language, commonsense reasoning, etc.

### Ph.D. Student, Cornell University

Sep 2014 – June 2020

Advisor: Professor Lillian Lee

## Selected publications ([full list of papers here](#))

As of October 2025, my h-index is 34. Here is a topically+stylistically representative subset of my work:

Orion Weller, Benjamin Van Durme, Dawn Lawrie, Ashwin Paranjape, Yuhao Zhang, **Jack Hessel**. "Promptriever: Instruction-Trained Retrievers Can Be Prompted Like Language Models." ICLR 2025.

Wenting Zhao, Xiang Ren, **Jack Hessel**, Claire Cardie, Yejin Choi, Yuntian Deng. "(InThe)WildChat: 570K ChatGPT Interaction Logs In The Wild." ICLR 2024.

**Jack Hessel**, Ana Marasović, Jena D. Hwang, Lillian Lee, Jeff Da, Rowan Zellers, Robert Mankoff, Yejin Choi. "Do Androids Laugh at Electric Sheep? Humor 'Understanding' Benchmarks from The New Yorker Caption Contest." ACL 2023; *Best Paper Award, one of 3 / 3872 submissions*

Liunian Harold Li, **Jack Hessel**, Youngjae Yu, Xiang Ren, Kai-Wei Chang and Yejin Choi. "Symbolic Chain-of-Thought Distillation: Small Models Can Also 'Think' Step-by-Step." ACL 2023.

Wanrong Zhu\*, **Jack Hessel**\*, Anas Awadalla, Samir Yitzhak Gadre, Jesse Dodge, Alex Fang, Youngjae Yu, Ludwig Schmidt, William Yang Wang, Yejin Choi. "Multimodal C4: An Open, Billion-scale Corpus of Images Interleaved With Text." NeurIPS Datasets+Benchmarks 2023.

Hyunwoo Kim, **Jack Hessel**, Liwei Jiang, Ximing Lu, Youngjae Yu, Pei Zhou, Ronan Le Bras, Malihe Alikhani, Gunhee Kim, Maarten Sap, Yejin Choi. "SODA: Million-scale Dialogue Distillation with Social Commonsense Contextualization." EMNLP 2023. *Outstanding Paper Award, one of 30 / 4909 submissions*

**Jack Hessel**\*, Jena D. Hwang\*, Jae Sung Park, Rowan Zellers, Chandra Bhagavatula, Anna Rohrbach, Kate Saenko, and Yejin Choi. "The Abduction of Sherlock Holmes: A Dataset for Visual Abductive Reasoning." ECCV 2022 (oral).

Peter West, Chandra Bhagavatula, **Jack Hessel**, Jena D. Hwang, Liwei Jiang, Ronan Le Bras, Ximing Lu, Sean Welleck, and Yejin Choi. "Symbolic Knowledge Distillation: from General Language Models to Commonsense Models." NAACL 2022.

Rowan Zellers\*, Ximing Lu\*, **Jack Hessel**\*, Youngjae Yu, Jae Sung Park, Jize Cao, Ali Farhadi, and Yejin Choi. “MERLOT: Multimodal Neural Script Knowledge Models.” NeurIPS 2021 (Oral; top 1% of submissions).

**Jack Hessel**, Ari Holtzman, Maxwell Forbes, Ronan Le Bras, and Yejin Choi. “CLIPScore: A Reference-free Evaluation Metric for Image Captioning.” EMNLP 2021.

**Jack Hessel** and Lillian Lee. “Does my multimodal model learn cross-modal interactions? It’s harder to tell than you might think!” EMNLP 2020.

**Jack Hessel**, Lillian Lee, and David Mimno. “Unsupervised Discovery of Multimodal Links in Multi-image, Multi-sentence Documents.” EMNLP 2019.

**Jack Hessel**, David Mimno, and Lillian Lee. “Quantifying the Visual Concreteness of Words and Topics in Multimodal Datasets.” NAACL 2018.

## Awards + Invited Talks

### *Awards*

ACL Best Theme Paper Award	2024
ACL Best Paper Award	2023
EMNLP Outstanding Paper Award	2023

### *Invited Talks*

<i>Vision in the Wild @ CVPR</i> : “Visual Details Don’t Matter (Until They Do).”	2024
<i>Carnegie Mellon University</i> : “Hurdles for vision+language models.”	2023
<i>Cornell University</i> : “Hurdles for vision+language models.”	2023
<i>NICE CVPR Workshop</i> : “Recognition isn’t enough for image captioning.”	2023
<i>Caltech guest lecture with Yejin Choi</i> : “Common Sense: The Dark Matter of Language Intelligence”	2023
<i>The AI Talks (NUS/NTU Singapore)</i> : “The Case for Reasoning Beyond Recognition”	2023
<i>Procter &amp; Gamble</i> : “The Case for Reasoning Beyond Recognition”	2022
<i>Seoul National University</i> : “The Case for Reasoning Beyond Recognition”	2022
<i>Adobe Research</i> : “The Case for Reasoning Beyond Recognition”	2022
<i>University of Washington</i> : Two Lectures (Linguistics, Honors Col.) “New Frontiers in Multimodal Grounding”	2022
<i>University of Pittsburgh</i> : “(at least) Two Conceptions of Visual-Textual Grounding”	2020
<i>Allen Institute for AI</i> : “The Promise and Perils of Learning Grounding from Visual-Textual Web Data.”	2020
<i>UNC Chapel Hill</i> : “The Promise and Perils of Learning Grounding from Visual-Textual Web Data.”	2020
<i>Rutgers University</i> : “Multimodal Grounding from User-generated Web Content.”	2019
<i>SRI International</i> : “Multimodal Grounding from User-generated Web Content.”	2019
<i>Cornell University</i> : PhD Colloquium, “Unsupervised Learning From Multimodal Documents.”	2019
<i>University of Pittsburgh</i> : “Grounding Images from a Digital Library in their Textual Contexts.”	2018
<i>Cornell University</i> : Two Guest Lectures for CS4300, “Practical Unsupervised Learning”	2015
<i>Carleton College</i> : “The Role of Altruism on Kickstarter”	2014

## Other Employment and Academic Service

### *Teaching+Internships*

**Research Intern, Google Research**

Summer 2019, Summer 2018

Hosts: Bo Pang and Zhenhai Zhu.

Worked with the natural language understanding team on video language joint learning: this work was published at *CoNLL 2019* and *EMNLP 2020*.

**Invited Visiting Faculty of Computer Science, Carleton College**

Spring 2019

Lead instructor of i) Natural Language Processing and ii) Discrete Math; 30+ students in each class

**Research Intern, Facebook, Inc.**

Summer 2017

Host: Amit Bahl.

Worked with the Core Data Science team on personalized language modeling, and cross-modal retrieval.

**Research Intern, Twitter, Inc.**

Summer 2016

Host: Clément Farabet.

Worked with the Cortex Team as their first intern on large-scale/multimodal node embeddings in graphs, language modeling, and engagement prediction.

**Teaching Assistant, Cornell University**

Various

Language and Information, 2016; Machine Learning for Data Science, 2015; Intro to Computer Graphics, 2014

**Research Intern, Washington University, St. Louis**

Summer 2013

Host: Kilian Weinberger.

Contributed to a GPU support vector machine package that accompanies Tyree et al.'s "Parallel Support Vector Machines in Practice." 2014.

### *Program Committees/Reviewing/(Senior)Area Chairing*

Since 2015, I have regularly reviewed, (senior) area chaired, ++ for major NLP/CV/ML conferences+journals like ACL, EMNLP, NeurIPS, etc.

## Development Experience

### *Open Source Contributions*

I regularly open source datasets and code to accompany my academic publications. In addition, I occasionally do fun, standalone open source work like:

Developed a TreeLSTM in TensorFlow2: this neural network dynamically changes its topology on a per-example basis ([https://github.com/jmhessel/recursive\\_nn\\_tf2](https://github.com/jmhessel/recursive_nn_tf2))

Developed fmpytorch (150+ stars on GitHub): a cythonized implementation of second order factorization machines in pytorch, forward/backward passes implemented in C (<https://github.com/jmhessel/fmpytorch>)

Developed fightingwords for comparing word usage rate differences between corpora; used in several refereed publications (<https://github.com/jmhessel/FightingWords>)

Pull requests merged to Keras, and Gensim, and tensorflow

### *Technical Skills*

*Machine Learning Skills:* Various machine learning/statistical toolkits/languages (e.g. sklearn, Tensorflow, PyTorch, R, etc.). Experience working with large, multi-faceted datasets, both in academic and industry contexts.

*Development Skills:* Object-oriented programming (Python mostly, but also Java, C++); experience shipping production-grade machine learning systems; parallel programming experience on CPUs + GPUs + TPUs; experience with various languages, development environments, version control systems, operating systems.

## References

Available upon request