Jacob (Jake) Carlson

CONTACT INFORMATION
jacob_carlson@g.harvard.edu
https://jscarlson.github.io

EDUCATION

Harvard University, Cambridge, Massachusetts
Ph.D. Economics, August 2022 – Present

Yale University, New Haven, Connecticut
B.A. Statistics and Data Science, August 2016 – May 2020
B.A. Ethics, Politics, and Economics, August 2016 – May 2020
Summa cum laude with distinction in both subjects

HONORS AND AWARDS

Presidential Scholar (Harvard University)
Phi Beta Kappa (Yale University)
Forethought Foundation for Global Priorities Research Undergraduate Thesis Prize Winner for undergraduate thesis titled "Culture and Growth: An Empirical Investigation" (2020)

PUBLICATIONS

Dell, Melissa, **Jacob Carlson**, Tom Bryan, Emily Silcock, Abhishek Arora, Zejiang Shen, Luca D'Amico-Wong, Quan Le, Pablo Querubin, Leander Heldring. "American Stories: A Large-Scale Structured Text Dataset of Historical U.S. Newspapers." NeurIPS Datasets and Benchmarks Track (2023).

Bryan, Tom, **Jacob Carlson**, Abhishek Arora, Melissa Dell. "EfficientOCR: An Extensible, Open-Source Package for Efficiently Digitizing World Knowledge." EMNLP System Demonstrations (2023).

Carlson, Jacob, Trevor Incerti, and P. M. Aronow. "Dyadic Clustering in International Relations." Political Analysis (2024).

Shen, Zejiang, Ruochen Zhang, Melissa Dell, Benjamin Lee, **Jacob Carlson**, and Weining Li. "LayoutParser: A Unified Toolkit for Deep Learning Based Document Image Analysis." ICDAR (2021). *Oral presentation.*

WORKING PAPERS

Carlson, Jacob, Tom Bryan, and Melissa Dell. "Efficient OCR for Building a Diverse Digital History." arXiv preprint arXiv:2304.02737 (2023).

ACADEMIC POSITIONS

Harvard University, Cambridge, Massachusetts
Research Assistant May 2023 – Present
- Assisting Isaiah Andrews and Davide Viviano, respectively, with various projects on topics in theoretical and applied econometrics

Harvard University, Cambridge, Massachusetts
Data Science Fellow in Deep Learning September 2020 – May 2022
- Worked with Melissa Dell to develop deep learning methods and tools for social science data curation

Yale University, New Haven, Connecticut
Research Assistant September 2018 – August 2018
- Worked with Yusuke Narita on an applied microeconomics project aimed at better understanding the effectiveness of "evidence-based" policies across the United States

Yale Law School, New Haven, Connecticut
Research Assistant May 2017 – May 2019
- Collaborated with Kate Stith and a small team of law students on a quantitative research project investigating mandatory minimum sentencing trends across DOJ administrations

NON-ACADEMIC POSITIONS

Cornerstone Research, New York, New York
Summer Analyst May 2019 – July 2019
- Performed financial and economic analyses for Fortune 500 clients undergoing complex commercial litigation

TEACHING

Harvard University, Cambridge, Massachusetts
Unleashing Novel Data at Scale (ECON 2355)
- Received an average overall section leader rating of 5.00 ("excellent") and a Harvard University Certificate of Distinction in Teaching
- Selected student comments: "Deeply knowledgeable and good at communicating complex technical concepts"; "Jacob is an incredible TF who was always accessible outside of class, eager to help, and very knowledgeable. One of the best I've had"

PRESENTATIONS

Allied Social Science Associations (ASSA) 2022 Virtual Annual Meeting
- Presented "LayoutParser: A Unified Toolkit for Deep Learning Based Document Image Analysis" at the "Nonstandard Data Sources" paper session on January 8, 2022

SOFTWARE

- *NetworkTargeting*, R package (with Davide Viviano), https://github.com/dviviano/NetworkTargeting
- *EffOCR*, Python package (with Tom Bryan and Melissa Dell), https://pypi.org/project/efficient-ocr
- *dcr*, R package and Stata command (with Trevor Incerti and P.M. Aronow), https://github.com/jscarlson/dcr, https://github.com/jscarlson/stata-dcr
- *LayoutParser*, Python package (with Zejiang Shen, Ruochen Zhang, Melissa Dell, Benjamin Lee, and Weining Li), https://pypi.org/project/layoutparser

LAST UPDATED March 6, 2024