Solomon Messing

[firstname.lastname] AT gmail DOT com

⌨ https://SolomonMg.github.io
⌓ https://github.com/SolomonMg

Work

Facebook, Core Data Science Team Menlo Park, CA
Principal Research Scientist *2012 - Present*

- Data science lead for external facing research on misinformation, elections, health and well-being. Working with PMs, executives, CEO to coordinate high-visibility program.

- Implementing privacy-preserving technologies: differentially private data releases; systems for analytics/ML/AI on secure data infrastructure. Studying implications of differentially private solutions for precise, global analytics (FLEX/Chorus + Spark) and prediction with deep neural networks (PyTorch/differentially private SGD). Implementing and open-sourcing privacy-oriented open-source URL-santization framework.

- Methodological/data science work estimating heterogenous treatment effects in experiments while limiting false-positives (R); feature modeling using large scale, high dimensional network data (Python/Giraph); adaptive experimentation for online decision-making systems (contextual multi-armed bandits in Python).

- Internal consultation/feedback on experimental design; simulation and data driven power analysis (R); third-party data integration and harmonization.

- Civic engagement work predictive modeling for political campaigns; strong positive impacts on revenue, CTRs, client feedback. Groundwork for Town Hall constituent-communication features. Top tier publications with national media attention.

- Managed & mentored junior staff, taught DataCamp course on statistics and large datasets, instructed joint Facebook-Udacity Exploratory Data Analysis course.

Pew Research Center, Data Labs Washington, DC
Founding Director *2015 - 2018*

- Established and led data science research group focused on machine learning for text (NLP) and images (computer vision), complex experimental designs, R&D, data science consulting for other teams. Worked with senior leadership to craft vision, get stakeholder buy-in, and execute on hiring, infrastructure, research and publication.

- Hired and managed team of 6 data scientists, researchers, engineers. Cross functional work to create data science career track; recruit, evaluate, & "dotted-line" manage data science staff embedded in other teams.

- Supervised development of Python-oriented data science infrastructure with wide adoption: Linux-based AWS cloud servers; custom Django dashboards/labeling/harmonization tools; large scale scraping/API query data collection efforts; PostGreSQL, with S3, BigTable, Hive for larger data sets; ML infrastructure with SKLearn, PyTorch, Keras neural network architectures; Jupyter and RStudio Server web interfaces for analysis; GitHub version control.

- Led publication of collaborative data science-oriented projects with national media attention. See http://www.pewresearch.org/topics/data-labs/.

- Served as public spokesperson for data science research, including interviews with national media and television outlets, presentations to government officials.

- Developed protocols for data science at Pew: machine learning standards, security, data management.

Morning Consult Washington, DC
Advisor *2013 - Present*

- Informal role as consultant for cofounders on overall strategy, innovation in polling methodology.

- Formal advisory role as consultant for executive team and CTO on data science, social media for research and intelligence arms of the company.

Stanford University, Social Science Data and Software (SSDS) Stanford, CA
Consultant *2011 - 2013*

- Guided Stanford researchers through experimental and observational study design, including web experiments, scraping, text analysis, network analysis, statistical modeling, visualization in R.
- Led workshops introducing R, data visualization, scraping & analysis of text data.

Science Applications International   Washington, DC
Media Analytics Team Lead   *2005 - 2011*
- Early-stage word embeddings using Latent Semantic Analysis (LSA/LSI) with substantial NLP post processing and ML for classification, entity resolution, sentiment.
- Qualitative research on misinformation, propaganda in Indonesia, Middle East using primary language internet sources, with readership across the U.S. diplomacy and intelligence communities. More than 50 reports, citations, follow-up research requests from government officials including Undersecretary Hughes, Congressman Markey, and then-Senator Clinton.
- Multi-language quantitative research: sentiment, attribution of responsibility, economic framing; GUI; database; visualizations. Human-coder training, calibration/reliability, report production.

Law-on-line, Inc.   Tucson, AZ
Information Technology Associate/Consultant   *1996 - 2005 (non-continuous)*
- Developed and helped bring to market multiple security software products for applications related to digital signature technology. Java, PGP key infrastructure, web dev.

Education

Stanford University   Stanford, CA
Ph.D. Communication (2013), M.S. Statistics (2012)   *2008 - 2013*

University of California, San Diego   La Jolla, CA
Masters, School of Global Policy and Strategy (Int'l MBA-curriculum)   *2003 - 2005*

University of California, Santa Barbara   Santa Barbara, CA
Bachelor of Arts with Honors in Political Science and Global Studies, Japanese minor

Sophia University (上智大学, 市谷キャンパス)   Tokyo, Japan
Study abroad

Books

The Impression of Influence: Legislator Communication, Representation, and Democratic Accountability
    Grimmer, Westwood & Messing (contributing author). 2014. *Princeton University Press.*
    - Reviewed in: Journal of Politics, Congress & The Presidency, Political Communication, Political Science Quarterly.
    - Subject of roundtable at Western Political Science Association, 2015.
    - Media coverage: Vox (Mischiefs of Faction).

Publications

Projecting Confidence: How the Probabilistic Horse Race Confuses and Demobilizes the Public. Westwood, Messing & Lelkes (joint-first-author).
    - Cited by FiveThirthyEight's Politics Podcast as influential in decision to change forecast presentation.
    - Media coverage: Washington Post, New York Magazine, Political Wire.

"Bots in the Twittersphere: An estimated two-thirds of tweeted links to popular websites are posted by automated accounts – not human beings" Wojcik, Messing, Smith, Rainie (joint-first-author). *Pew Research Center*, 2018.
    - Cited by California State Senator Robert Hertzberg in passage of SB 1001, making it illegal to create bots that misrepresent identity or otherwise deceive people in California.
    - Media coverage: Wired, Vox, USA Today, Observer, NiemanLab, TechCrunch, VentureBeat, FastCompany.

"Estimating Heterogeneous Treatment Effects and the Effects of Heterogeneous Treatments with Ensemble Methods" Grimmer, Messing, and Westwood (contributing author). *Political Analysis*, 2017.
- Replication materials.
- Software implementation (under development): https://github.com/SolomonMg/HetSL

"Partisan Conflict and Congressional Outreach" Messing, van Kessel, Hughes, Judd, Blum (joint-first-author). *Pew Research Center*, 2017.
- Media coverage: Washington Post, Mother Jones.

"Exposure to ideologically diverse news and opinion on Facebook" Bakshy, Messing, & Adamic (joint-first-author). *Science*, 2015.
- Review by David Lazer.
- Supplementary materials.
- Replication materials.
- Media coverage: New York Times, Washington Post, Ars Technica, BBC, CBS News.

"Bias in the Flesh: Skin Complexion and Stereotype Consistency in Political Campaigns" Messing, Plaut & Jabon (first author). *Public Opinion Quarterly*, 2015.
- Replication materials.
- Media coverage: CBS News, Washington Post.

"Quantifying Social Media's Political Space: Estimating Ideology from Publicly Revealed Preferences on Facebook" Bond & Messing (contributing author). *American Political Science Review*, 2015.

"Selective Exposure in the Age of Social Media: Endorsements Trump Partisan Source Affiliation when Selecting News Online" Messing & Westwood (joint-first-author). *Communication Research*, 2013.

"Role of Diffusion Weighted Imaging in the Diagnosis of Pediatric Abdominal Tumors" Gawande, Gonzalez, Messing, Khurana, & Daldrup-Link (contributing author). *Pediatric Radiology*, 2013.

"How Words and Money Cultivate a Personal Vote: The Effect of Legislator Credit Claiming on Constituent Credit Allocation" with Grimmer, Messing & Westwood (joint-first-author). *American Political Science Review*, 2012.

"Intravenous Ferumoxytol Allows Noninvasive MR Imaging of Stem Cell Transplants" with Khurana, Gawande, Lin, Lee, Messing, Castaneda, Derugin, Pisani, Lue, & Heike Daldrup-Link (contributing author). *Radiology*, 2012.

"Do Attitudes about Immigration Predict Willingness to Admit Individual Immigrants? A Cross-National Test of the Person-Positivity Bias" Iyengar, Jackman, Messing, Valentino, Aalberg, Duch, Hahn, Soroka, Harell, & Kobayashi (contributing author). *Public Opinion Quarterly*, 2012.

"Differentiation of the Normal Thymus from Anterior Mediastinal Lymphoma on Pediatric PET-CTs" with Rakhee Gawande, Aman Khurana, Messing, Dong Zhang, Rosalinda Castaneda, Robert Goldsby, Randal Hawkins, Messing & Heike Daldrup-Link (contributing author). *Radiology*, 2011.

"Who is a 'Deserving' Immigrant? An Experimental Study of Norwegian Attitudes" Aalberg, Iyengar, & Messing (contributing author). *Scandinavian Political Studies*, 2011.

Honors and Awards

2010 $60,000 Google Research Award (with Sean Westwood, supervised by Dan McFarland & Jeremy Bailenson) "Harnessing the social graph: Social cues and reliability in content selection and evaluation"

2010 $1,000 A-REP Research Grant (with Sean Westwood), "The impact of social news ratings on media consumption heuristics and agenda setting"

2008 Department of Communication Fellowship

2007 Annual Achievement Award for Excellence in Science and Technology - Technical Collaboration, Science Applications International

2007 Selected for Future Leaders program, Science Applications International

2004–2005 Dean's Fellow, University of California, San Diego IR/PS

2001 Phi Beta Kappa, University of California, Santa Barbara

2000 Golden Key, University of California, Santa Barbara

Invited Talks and Workshops

Catalist, Election forecasting, Washington, DC, April 10, 2018.

Social Media and Political Participation Global Conference, Shang Hai, March 11, 2018.

Congressional Management Foundation, Facebook and the U.S. Congress, Washington, DC, January 17, 2018.

USC, Media Exposure and Opinion Formation workshop, Los Angeles, November 9, 2017.

Social Media and Political Participation Global Conference, New York, November 3, 2017.

World Bank, Big Data in Government, Washington, May 16, 2017.

Data and Society, New York, March 31, 2017.

Social Media and Political Participation Global Conference, Abu Dhabi, February 9, 2017.

Analyst Institute Social Networks Roundtable, March 9, 2016.

GAO CG Forum on 21st Century Data and Analytics, January 28, 2016.

Duke Network Analysis Center (DNAC), October 20, 2015.

Hewlett Foundation, News Funders Conference, June 19, 2015.

Spatial Computational Social Science at Stanford, Oct 20, 2014.

UCLA Institute for Politics in Math, April 22, 2014.

Stanford Political Science American Politics Workshop, April 22, 2013.

Stanford Computational Social Science, Analyzing large data sets, September 20, 2012.

Stanford Political Science Methods Workshop, May 2, 2012.

Software

HetEffects with Justin Grimmer. R package utilizing ensemble machine learning methods to estimate high-dimensional causal effects that vary by treatment-group combination. Development version at https://github.com/SolomonMg/HetSL. See Estimating Heterogeneous Treatment Effects and the Effects of Heterogeneous Treatments with Ensemble Methods.

ImageMetrics R package to facilitate the analysis of image data, designed to facilitate analysis described in "Bias in the flesh" (see above). See https://github.com/SolomonMg/ImageMetrics.

`NetCluster, triads, NetData` with Mike Nowak, Sean Westwood & Dan McFarland. R packages to determine node-level triad type membership for the 16 types of triads that occur in a directed network, facilitate network clustering and evaluation of cluster configurations, and provide example data. Examples in "Social network analysis labs in R."

Computing

R Statistical Programming Language Data visualization, conventional statistical analysis/econometrics, Bayesian statistics, text-as-data, image, GIS, and network data, including authoring multiple software packages.

Python Data analysis (Pandas, NumPy, SciPy), Machine Learning (SciKit Learn, Gensim), parsing data from the web, computer vision, text-as-data, deep-learning (text data), data management.

SQL Experience utilizing Hive (HQL) with petabyte scale data sets, PostGreSQL, database design with MySQL, SQLite.

Java Undergraduate & natural language processing courses taken; some instrustry coding experience.

Web Dev Working knowledge of JavaScript, PHP, JSON, RegEx, HTTP, Linux shell scripts, C-sharp/.NET.

Journal Service

Reviewer, *Science*

Reviewer, *American Political Science Review*

Reviewer, *American Journal of Political Science*

Reviewer, *Political Analysis*

Reviewer, *Public Opinion Quarterly*

Reviewer, *Political Behavior*

Assistant Editor, Reviewer, *Political Communication*

Founding Editor, Reviewer, *Journal of International Policy Solutions*

Guest Editor, Reviewer, *Journal of Public and International Affairs*

Languages

Indonesian/Malay Bahasa 2 years professional experience as Indonesian analyst/linguist; 2 years intensive Indonesian graduate study; Certified 3/2 Written/Spoken on the Interagency Language Roundtable (ILR) scale used by the U.S. Foreign Service.

Japanese 3 years undergraduate language study, 1 year study abroad in Tokyo, Japan.

Arabic Digital courses taken; introductory-level translation experience.