Sol Messing
solomon DOT [lastname] AT gmail DOT com
https://SolomonMg.github.io
https://github.com/SolomonMg

Summary

Engineering manager building teams that deliver strategic guidance and high-impact data products. Industry experience: recommender systems/ML, complex experimentation/causal inference, ads experimentation, modeling, HCI, algorithm audits, differential privacy. Built data science research groups at Twitter, Acronym, Pew Research Center. Publications in top social science & medical journals. Technical experience working with distributed computing platforms (Hive/Hadoop/GCP/K8/etc.), Keras/TensorFlow/Pytorch, Django, Tidyverse, etc.

Work History

Twitter Twin Cities, MN
Discovery DS Lead (Sr EM) *2021 - Present*

- DS lead for Home Ranking, Experience, Relevance Platform, Search, & Applied Sciences (40+ person org).
- Built Applied Sciences team, working on some of the hardest problems at Twitter: (1) quantify the value people generate when they create content or connect with a contact; (2) how to tradeoff engagement and problematic discourse; (3) estimate long-term value with short-term experiments. Scaled team from 3 to 10 highly experienced PhD researchers & engineers.
- Supervising company-wide strategic shift in goaling emphasizing content production, ecosystem value.
- Led award-winning hackweek project: What Moves Metrics. "Most impactful to the bottom line."
- Supervised exploratory analysis of geolocation system and creation of prototype CNN that flexibly models 2-dimensional geographies using various signals. Ads application: 75 bps sustained causal revenue impact, $300m projected impact (NPV); notifications application: over 140k additional daily active users.
- Supervised consumer lifetime value model allowing optimization across programs throughout the company, including customer acquisition.
- Driving effort to increase mutual following relationships among new users in critical first week using notifications and online model predicting mutual follow behavior; strong impact on connections, content production, retention.

ACRONYM/Lockwood Washington, DC
Chief Scientist *2020 - 2021*

- Shaped org-wide experiment-informed 2020 election strategy with executive team. Demonstrated effectiveness of "boosted" (promoted) news content over conventional advertising for persuasion, unlocking vast content sourcing & cost-savings; Identified and modeled low-information audiences who were more responsive to ads and news; crafted battleground state projections that drove decision to spend in Georgia.
- Led build-out of digital field experiment analysis system to identify most effective real-world messaging strategy; supervised design & execution of hundreds of in-field messaging-track RCTs. Incremental improvements took turnaround time from 5-10 days to 2 hours via automation, APIs, scripting (R + Python).
- Directed development of surrogacy metrics using machine learning system to map behavioral data to persuasive effects observed in RCTs.
- Supported technical effort to build dashboard & daily budget reallocation systems using surrogacy metrics to allow clients to reallocate & optimize budget & audience for ads and other messaging tracks. Wide adoption in ecosystem.
- Grew measurement org from 3 to 10 staff-level data scientists, software engineers, other staff.

Facebook, Core Data Science Team Menlo Park, CA
Research Scientist (IC6) *2012 - 2020*

- Partnered with PMs & executives to lead data science for high-visibility external facing research on misinformation, elections, health and well-being.

- Led R&D, application, launch of largest ever data sharing effort (reflecting >1 exabyte) to facilitate study of misinformation—global view- and interaction data on external content on Facebook since 2017. Groundwork w/ engineering teams on analysis platform. Implemented & worked with SWEs to refactor & open-source privacy-oriented URL-privacy framework.
- Evaluation of privacy solutions: reidentification risk metrics; implications for precise, global analytics; privacy-preserving analytics/ML (FLEX/Chorus + Spark) on secure data infrastructure; deep neural networks (PyTorch/differentially private SGD); external vendor solutions.
- Complex experimentation frameworks: estimating heterogenous treatment effects in experiments while limiting false-positives (R); adaptive experimentation for online decision-making systems (contextual multi-armed bandits in Python).
- Internal consultation/feedback on experimental design; simulation & data driven power analysis (R).
- Feature modeling for political campaigns using large scale, high dimensional network data (Python/Giraph); 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 & large datasets, instructed joint Facebook-Udacity Exploratory Data Analysis course.

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

- Established & led data science research group focused on machine learning for text (NLP) & 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, & execute on hiring, infrastructure, research & publication.
- Hired & managed team of data scientists, researchers, engineers, growing org from 2 to 20, including dotted-line reporting relationship. Cross functional work to create data science career track; recruit, evaluate, & manage data science staff embedded in other teams.
- Supervised development of R & 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 & 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 & 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, data science, innovation in polling methodology as company grew from 2 people to digital media unicorn (> billion dollar valuation).
- Formal advisory role as consultant for executive team & CTO on data science, social media for research & intelligence arms of the company.

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

- Guided Stanford researchers through experimental & 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 & 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 & intelligence communities. More than 50 reports, citations, follow-up research requests from government officials including Undersecretary Hughes, Congressman Markey, & 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
Software Engineer/Consultant *1996 - 2005 (non-continuous)*

- Developed & 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 & Strategy (Int'l MBA-curriculum) *2003 - 2005*

University of California, Santa Barbara Santa Barbara, CA
BA (Honors) in Political Science, Japanese minor, 2 yrs CS/eng coursework

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

Software

`URL-Sanitization` with Runchao Jiang & Da Li. Algorithms to make URL data privacy-safe. Removes query parameters unrelated to content navigation and removes query parameters often related to user PII by using string matching and modeling. Part of privacy-first data sharing launch of Facebook Privacy-Protected Full URLs Data Set. Live version at `https://github.com/facebookresearch/URL-Sanitization`.

`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.

Books

The Impression of Influence: Legislator Communication, Representation, & Democratic Accountability Grimmer, Westwood & Messing. 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 & Demobilizes the Public. Westwood, Messing & Lelkes. *Journal of Politics, forthcoming.*
- 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. *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.

"Very liberal or conservative legislators most likely to share news on Facebook" Messing. *Pew Research Center*, 2018.

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

"Sharing the News in a Polarized Congress" Messing, van Kessel, Hughes. *Pew Research Center*, 2017.
- Media coverage: Politico, Washington Post, Axios, Quartz.

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

"Exposure to ideologically diverse news and opinion on Facebook" Bakshy, Messing, & Adamic. *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. *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. *American Political Science Review*, 2015.

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

"Role of Diffusion Weighted Imaging in the Diagnosis of Pediatric Abdominal Tumors" Gawande, Gonzalez, Messing, Khurana, & Daldrup-Link. *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. *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. *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. *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. *Radiology*, 2011.

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

Invited Talks and Workshops

NULab for Texts, Maps, and Networks, Northeastern University, What do the data tell us about election 2020?, Dec 11, 2020

CODE@MIT, Fireside Panel, Technology in Political Campaigns and Activism, Nov 19, 2020.

Analyst Institute, Meeting the Moment, Washington DC, Oct 14, 2020.

Analyst Institute, Approaches to Message Testing, Washington DC, Jun 16, 2020.

Georgetown Massive Data Institute, Effects of Election Forecasting, Washington, DC, Nov 18, 2019.

UVA Data Science Institute, Differential Privacy & Corporate Data, Charlottesville, VA, Sep 6, 2019.

University of Amsterdam International Conference on Computational Social Science, Amsterdam, NL, Jul 18, 2019.

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

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

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

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

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

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

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

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

Analyst Institute Social Networks Roundtable, Mar 9, 2016.

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

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

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

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

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

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

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

Stanford Political Science Methods Workshop, May 2, 2012.

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

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.