

AARON HALFAKER PHD

Email: aaron.halfaker@gmail.com

Website: <http://halfaker.info>

Github: [halfak](#)

EDUCATION

University of Minnesota, Twin Cities Minneapolis, MN <i>Doctor of Philosophy in Computer Science</i>	Fall 2013
The College of St. Scholastica Duluth, MN <i>Bachelor of Arts in Computer Science</i>	Spring 2006

RESEARCH EXPERIENCE

Microsoft Redmond, WA <i>Principal Applied Scientist</i>	Aug 2020 - current
Wikimedia Foundation San Francisco, CA <i>Principal Research Scientist</i>	March 2011 - July 2020
University of Minnesota Minneapolis, MN <i>Senior Scientist</i>	May 2016 - current
GroupLens Research Minneapolis, MN <i>Graduate Research Assistant</i>	Aug. 2007 - Oct. 2013
Google Inc. Mountain View, CA <i>Research Intern</i>	Jan. 2013 - Apr. 2013

REFEREED PUBLICATIONS

- ORES: Lowering Barriers with Participatory Machine Learning in Wikipedia
Halfaker, A. & Geiger, S. *CSCW'20* Minneapolis, MN, USA (2020). [[arxiv](#)]
- Keeping Community in the Loop: Understanding Wikipedia Stakeholder Values for Machine Learning-Based Systems
Smith, C. E., Yu, B., Srivastava, A., **Halfaker, A.**, Terveen, L., & Zhu, H. *CHI'20* Honolulu, HI, USA (2020). [[arxiv](#)]
- With Few Eyes, All Hoaxes are Deep
Asthana, S. & **Halfaker, A.** *CSCW'18* Jersey City, NJ, USA (2018). [[pdf](#)]
- Bot Detection in Wikidata Using Behavioral and Other Informal Cues
Hall, A., Terveen, L., & **Halfaker, A.** *CSCW'18* Jersey City, NJ, USA (2018). [[pdf](#)]
- Value-sensitive algorithm design: Method, case study, and lessons
Zhu, H., Yu, B., **Halfaker, A.**, & Terveen, L. *CSCW'18* Jersey City, NJ, USA (2018). [[pdf](#)]
- Distance and Attraction: Gravity Models for Geographic Content Production
Thebault-Spieker, J., **Halfaker, A.**, Terveen, L. G., & Hecht, B. *CHI'18* Montreal, Canada (2018). [[pdf](#)]
- Evaluating the Impact of the Wikipedia Teahouse on Newcomer Socialization and Retention
Morgan, J. T. & **Halfaker, A.** *OpenSym'18* Paris, France (2018). [[pdf](#)]
- Information Fortification: An Online Citation Behavior
Forte, A., Andalibi, N., Gorichanaz, T., Kim, M. C., Park, T., & **Halfaker, A.** *GROUP'18* Sanibel Island, FL, USA (2018). [[pdf](#)]
- Open Community Health: Workshop Report.
Link, G. J. P., Lombard, K., Damen, N., Rosser, H., Germonprez, M., Goggins, S., Wiggins, A., Ahuja, V. K., Brier, J., Cohoon, J., **Halfaker, A.**, Howison, J., Marti, D., Newman, G., Østerlund, C., Paik, R., Rother, B., & Schecter, A. *Journal of Peer Production* (2018). [[pdf](#)]
- ORES: Facilitating Re-mediation of Wikipedia's Socio-technical Problems.
Halfaker, A., Geiger, R.S., Morgan, J. T., Sarabadani, A., & Wight, A. *Self-published* (2018). [[pdf](#)]
- Simulation Experiments on (the Absence of) Ratings Bias in Reputation Systems.
J Thebault-Spieker, D Kluver, M. A. Klein, **A. Halfaker**, B Hecht & L Terveen *CSCW'17* Jersey City, New Jersey (2017). [[pdf](#)]
- Operationalizing Conflict and Cooperation between Automated Software Agents in Wikipedia: A Replication and Expansion of 'Even Good Bots Fight'.
Geiger, R.S. & **Halfaker, A.** *CSCW'17* Jersey City, New Jersey (2017). [[pdf](#)]

- Identifying Semantic Edit Intentions from Revisions in Wikipedia.
Yang, D., **Halfaker, A.**, Kraut, R., & Hovy, E. *EMNLP'17* Copenhagen, Denmark (2017). [[pdf](#)]
- Interpolating quality dynamics in Wikipedia and demonstrating the Keilana effect.
Halfaker, A. *OpenSym'17*, Galway Ireland (2017). [[pdf](#)]
- Building automated vandalism detection tools for Wikidata.
Sarabadani, A., **Halfaker, A.**, & Taraborelli, D. *WWW'17* Perth, Australia (2017). [[pdf](#)]
- Advancing the OCDX: Building Social Computing Infrastructure.
Goggins, S., Germonprez, M., Million, A. J., **Halfaker, A.**, & Hemphill, L. *CSCW'17* Portland, Oregon, USA (2017).
- Not at home on the range: Peer production and the urban/rural divide.
Johnson, I. L., Lin, Y., Li, T. J. J., Hall, A., **Halfaker, A.**, Schöning, J., & Hecht, B. *CHI'16* San Jose, California, USA (2016). [[pdf](#)]
- Who Did What: Editor Role Identification in Wikipedia.
Yang, D., **Halfaker, A.**, Kraut, R. E., & Hovy, E. H. *ICWSM'16* Cologne, Germany (2016). [[pdf](#)]
- Open algorithmic systems: lessons on opening the black box from Wikipedia.
Geiger, R. S., & **Halfaker, A.** *AoIR'16* Berlin, Germany (2016).
- Breaking into new Data-Spaces: Infrastructure for Open Community Science.
Halfaker, A., Morgan, J., Pandian, Y., Thiry, E., Rand, W., Schuster, K., & Laniado, D. (2016). *CSCW'17* San Francisco, California, USA (2016). [[pdf](#)]
- Edit Categories and Editor Role Identification in Wikipedia.
Yang, D., **Halfaker, A.**, Kraut, R. E., & Hovy, E. H. *LREC'16* Portoroz, Slovenia (2016). [[pdf](#)]
- User Session Identification Based on Strong Regularities in Inter-activity Time
Halfaker, A., Keyes, O., Kluver, D., Thebault-Spieker, J., Nguyen, T., Shores, K., ... & Warncke-Wang, M. *WWW*. Florence, Italy (2015). [[pdf](#)]
- Advancing an Industry/Academic Partnership Model for Open Collaboration Research.
Morgan, J. T., **Halfaker, A.**, Taraborelli, D., Hwang, T., & Goggins, S. *CSCW* Vancouver, Canada (2015).
- Accept, decline, postpone: How newcomer productivity is reduced in English Wikipedia by pre-publication review.
Schneider, J., Gelley, B. S., & **Halfaker, A.**, *WikiSym*. Berlin, Germany (2014). [[pdf](#)]
- Snuggle: Designing for Efficient Socialization and Ideological Critique
Halfaker, A., Geiger, R. S. & Terveen, L. *CHI*. Toronto, Canada (2014). [[pdf](#)]
- When the Levee Breaks: Without Bots, What Happens to Wikipedia's Quality Control Processes?
Geiger, R. S. & **Halfaker, A.** *WikiSym*. Hong Kong, Hong Kong, (2013). [[pdf](#)]
- Making peripheral participation legitimate: Reader engagement experiments in Wikipedia
Halfaker, A., Keyes, O. & Taraborelli, D. *CSCW*. San Antonio, TX, USA, (2013). [10.1145/2441776.2441872](https://doi.org/10.1145/2441776.2441872) [[summary](#), [pdf](#)]
- Using Edit Sessions to Measure Participation in Wikipedia,
Geiger, R. S. & **Halfaker, A.** *CSCW*. San Antonio, TX, USA, (2013). [10.1145/2441776.2441873](https://doi.org/10.1145/2441776.2441873) [[pdf](#)]
- The Rise and Decline of an Open Collaboration System: How Wikipedia's reaction to sudden popularity is causing its decline,
Halfaker, A., Geiger, R. S., Morgan, J. & John Riedl. *American Behavioral Scientist*. (2013). [10.1177/0002764212469365](https://doi.org/10.1177/0002764212469365) [[summary](#), [pdf](#)]
- What aren't we measuring? Methods for quantifying wiki-work,
Halfaker, A., Keegan, B., Forte, A., Geiger, R. S., Taraborelli, D., Pinchuk, M. & Masli, M. *WikiSym*, Linz, Austria, (2012). [10.1145/2462932.2462983](https://doi.org/10.1145/2462932.2462983)
- Defense Mechanism or Socialization Tactic? Improving Wikipedia's Notifications to Rejected Contributors,
Geiger, R. S., **Halfaker, A.**, Pinchuk, M. & Walling, S. *ICWSM*, Dublin, Ireland, (2012). [[pdf](#)]
- Don't Bite the Newbies: How Reverts Affect the Quantity and Quality of Wikipedia Work,
Halfaker, A., Kittur, A. & Riedl, J., *WikiSym*, Mountain View, CA, (2011). [10.1145/2038558.2038585](https://doi.org/10.1145/2038558.2038585) [[summary](#), [pdf](#)]
- NICE: Social translucence through UI intervention,
Halfaker, A., Song, B., Stuart, D.A.,; Kittur, A. & Riedl, J., *WikiSym*, Mountain View, CA, (2011). [10.1145/2038558.2038575](https://doi.org/10.1145/2038558.2038575) [[summary](#), [pdf](#)]
- A Jury of Your Peers: Quality, Experience and Ownership in Wikipedia,
Halfaker, A., Kittur, N., Kraut, R. & Riedl, J., *WikiSym*, Orlando, FL, USA, (2009). [10.1145/1641309.1641332](https://doi.org/10.1145/1641309.1641332) [[summary](#), [pdf](#)]
- Wikipedians are born, not made: a study of power editors on Wikipedia,
Panciera, K., **Halfaker, A.** & Terveen, L., *GROUP*, Sanibel Island, FL, USA, p.51-60, (2009). [10.1145/1531674.1531682](https://doi.org/10.1145/1531674.1531682) [[pdf](#)]

OTHER PUBLICATIONS

- WikiProjects Machine Readable Dataset
Asthana, S. & Halfaker, A. *figshare*. (2018). [10.6084/m9.figshare.5503819](https://doi.org/10.6084/m9.figshare.5503819)
- Known Undisclosed Paid Editors (English Wikipedia)
T., Heilman, J., Henry, B., & Halfaker, A. *figshare*. (2018). [10.6084/m9.figshare.6176927.v1](https://doi.org/10.6084/m9.figshare.6176927.v1)
- Citations with contexts in Wikipedia.
Halfaker, A., Kim, M. C., Forte, A. & Taraborelli, D. *figshare*. (2017). [10.6084/m9.figshare.5588842.v1](https://doi.org/10.6084/m9.figshare.5588842.v1)
- [Wikidata item quality labels](#).
Yapinus, G. & Halfaker, A. *figshare*, (2017). [10.6084/m9.figshare.5035796.v1](https://doi.org/10.6084/m9.figshare.5035796.v1)
- [Scholarly article citations in Wikipedia](#).
Halfaker, A. & Taraborelli, D. *figshare*, (2017). [10.6084/m9.figshare.1299540.v9](https://doi.org/10.6084/m9.figshare.1299540.v9)
- [Wikipedia's very active editor numbers have stabilized](#).
Elder, J., & Erhart, E. (edited by Halfaker, A.) *Wikimedia Blog*, (2017).
- [Deleted Wikipedia articles \(spam/vandalism/attack\)](#).
Halfaker, A. & Rogers, J. *figshare*, (2017). [10.6084/m9.figshare.4245035.v1](https://doi.org/10.6084/m9.figshare.4245035.v1)
- [New dataset shows fifteen years of Wikipedia's quality trends](#).
Halfaker, A., Sarabadani, A., & Taraborelli, D. *Wikimedia Blog*, (2016).
- [Monthly Wikipedia article quality predictions](#).
Halfaker, A. & Sarabadani, A. *figshare*, (2016). [10.6084/m9.figshare.3859800.v3](https://doi.org/10.6084/m9.figshare.3859800.v3)
- [Introducing the unique devices dataset: a new way to estimate reach on Wikimedia projects](#).
Ruiz, N., A., Viswanatha, M., & Halfaker, A. *Wikimedia Blog*, (2016).
- [Artificial intelligence service "ORES" gives Wikipedians X-ray specs to see through bad edits](#).
Halfaker, A. & Taraborelli, D. *Wikimedia Blog*, (2016).
- [Monthly Wikipedia article quality predictions](#).
Halfaker, A., Keyes, O., Kluver, D., Thebault-Spieker, J., Nguyen, T., Shores, K., Uduwage, A. & Warncke-Wang, M. *figshare*, (2015). [10.6084/m9.figshare.1291033.v1](https://doi.org/10.6084/m9.figshare.1291033.v1)
- [Wikipedia's very active editor numbers have stabilized](#).
Erhart, E. & Halfaker, A. *Wikimedia Blog*, (2015).
- [Robots and Cyborgs: Wikipedia's Immune System](#).
Halfaker, A., Riedl, J., *IEEE Computing*, 45(3), 79-82, (2012).
- [Kids these days: The quality of new Wikipedia editors over time](#).
Halfaker, A. *Wikimedia Blog*, (2012).
- Wikimedia Foundation technical documents:
 - [AFC process and productivity](#). Halfaker, A., et al. (2015)
 - [Asking anons to register](#). Halfaker, A., et al. (2014)
 - [Task recommendations research](#). Halfaker, A., et al. (2014)
 - [Onboarding rolloout](#). Halfaker, A., et al. (2014)
 - [Wikipedia article creation](#). Halfaker, A. (2014)
 - [Impact and volume of anonymous editor contribution](#). Halfaker, A. (2013)
 - [Onboarding \(v.6\) experiment](#). Halfaker, A. (2013)
 - [Module storage performance](#). Halfaker, A. (2013)
 - [Newcomer survival models](#). Halfaker, A. (2013)
 - [VisualEditor's effect on newly registered editors](#). Halfaker, A. & Taraborelli, D. (2013, 2015)
 - [Notifications: Experiment #1](#). Halfaker, A. (2013)
 - [Article Feedback Tool Quality Analysis](#). Halfaker, A., Keyes, O., Taraborelli, D. (2012).
 - [Article Feedback Tool Prominence Analysis](#). Halfaker, A. (2012).
 - [Article Feedback Tool Newcomer Conversion Analysis](#). Halfaker, A. (2012).
- [Wikimedia Summer 2011 of Research](#) technical documents:
 - [Newbie reverts and article length](#). Halfaker, A. (2011).
 - [New page patroller work load](#). Halfaker, A. (2011).
 - [Multiprocessing software for Wikidumps](#). Halfaker, A. (2011).
 - [Rejection in the first edit session](#). Halfaker, A. (2011).
 - [Vandal fighter work load](#). Halfaker, A. (2011).

SOFTWARE ENGINEERING EXPERIENCE

Wikimedia Foundation San Francisco, CA <i>Tech lead for the Scoring Platform team (ORES)</i>	May 2015 - Current
Thomson-Reuters Eagan, MN <i>Software Engineer Intern</i>	May 2006 - Aug. 2006 May 2007 - Aug. 2007
The College of St. Scholastica Duluth, MN <i>Web Application Developer</i>	Oct. 2004 - May 2006
3M Maplewood, MN <i>Software Engineer Intern</i>	May 2005 - Aug. 2005

OPEN SOURCE SOFTWARE

Note that this section only contains software I actively maintain.

Revision Scoring / ORES

A framework and set of services for building horizontally scalable machine classifications systems for MediaWiki

- [ORES](#) Scalable machine classification service
- [revscoring](#) A framework for building machine classifier containers
- [editquality](#) A collection of models assessing damage and intent of edits in Wikipedia
- [articlequality](#) A collection of models assessing content quality in Wikipedia
- [draftquality](#) A collection of models assessing the quality of article drafts in Wikipedia
- [drafttopic](#) A collection of models assessing general topic space of new articles in Wikipedia

MediaWiki Utilities

A collection of libraries for processing MediaWiki content, APIs, and XML dumps efficiently

- [mwxml](#) Efficient processing of Wikipedia and Wikia's gigantic XML dumps
- [mwxml](#) A framework for accessing Wikipedia and Wikia's web APIs
- [mwoauth](#) An OAuth handshaker that smooths over MediaWiki's peculiarities
- [mwreverts](#) Efficient detection of revert actions in Wikipedia
- [mwsessions](#) Efficient extraction of contiguous activities in Wikipedia
- [mwpersistence](#) Efficient tracking of "word persistence"/authorship in Wikipedia
- [mwdiffs](#) Efficient application of diff algorithms in Wikipedia
- [mwrefs](#) Extraction of references and surrounding information.
- [mwcli](#) Utilities for creating unix-y command-line tools
- [mwtypes](#) A set of basic types for MediaWiki entites

Etc.

Other notable utilities and software systems

- [deltas](#) A framework for and a collection of advanced "diff" algorithms
- [deep_merge](#) A framework for accessing Wikipedia and Wikia's web APIs
- [yamlconf](#) Flexible configuration files based on YAML
- [para](#) Distributed processing framework built on python's Multiprocessing
- [kasami](#) Framework and models for Probabalistic Context-Free Grammars
- [snuggle](#) A server, model, and UI for tracking good-faith newcomers in Wikipedia and visualizing their activities

INVITED TALKS/VISITS/PRESENTATIONS

- International Symposium on Open Collaboration
 - Keynote: Cooking data with care -- Aug 2018 [\[video\]](#)
- Stanford Human-Computer Interaction Seminar
 - Productive Efficiency, Ideology, and AI in Wikipedia -- Nov 2017 [\[video\]](#)
- Boise State Computing Colloquium
 - Engineering at the Intersection of Productive Efficiency, Ideology, and Ethical AI in Wikipedia -- Oct 2017 [\[video\]](#)
- Berkely Institute for Data Science
 - Engineering at the Intersection of Productive Efficiency, Ideology, and Ethical AI in Wikipedia -- Jan 2017 [\[video\]](#)
- Northwestern University's Technology & Social Behavior speaker series
 - Engineering open production efficiency at scale -- May 2016
- Berkman Center
 - Engineering open production efficiency at scale -- Feb. 2016 [\[video\]](#)
 - AI and Ethics in Wikipedia -- Nov. 2018
 - Deploying and maintaining AI in a socio-technical system -- Jun. 2019
- University of Missouri
 - Engineering open production efficiency at scale -- Oct. 2015
- Oxford Internet Institute
 - The Practice of Managing Open Production at Scale -- Aug. 2014
- The French research institute for digital sciences
 - The Practice of Managing Open Production at Scale -- Aug. 2014

- Macalester College
 - What is Wikipedia? Why study it? -- Nov. 2013
 - Wikipedia as a socio-technical system -- Nov. 2015
 - Engineering Open Production Efficiency at Scale -- Feb. 2018
- The Wikimedia Foundation
 - Using AI to Keep Wikipedia Open -- Jul. 2018 [\[video\]](#)
 - Case studies in the Appropriation of ORES -- May. 2018 [\[video\]](#)
 - Using machine classification to clean up the new page backlog -- Feb. 2018 [\[video\]](#)
 - The Keilana Effect: Visualizing the closing coverage gaps with ORES -- Jul. 2017 [\[abstract\]](#)
 - English Wikipedia Quality Dynamics and the Case of WikiProject Women Scientists -- Dec. 2016 [\[video\]](#)
 - Deploying and maintaining AI in a socio-technical system. Lessons learned -- Aug. 2016 [\[video\]](#)
 - Anon productivity and productive efficiency in English Wikipedia -- Jan. 2016 [\[video\]](#)
 - The impact of the Wikipedia Teahouse on new editor retention -- Oct. 2015 [\[video\]](#)
 - Wikipedia's Health: A Socio-Technical Overview -- Jul. 2015 [\[abstract\]](#)
 - The Revision Scoring Service: Exposing Quality to Wiki Tools -- Jul. 2015 [\[abstract\]](#)
 - VisualEditor's effect on newly registered users -- Jul. 2015 [\[video\]](#)
 - The people's classifier: Towards an open model for algorithmic infrastructure -- May. 2015 [\[video\]](#)
 - User Session Identification Based on Strong Regularities in Inter-activity Time -- Mar. 2015 [\[video\]](#)
 - Wikipedia as a socio-technical system -- Oct. 2014 [\[video\]](#)
 - WikiCredit: Calculating and presenting Value-added to Wikipedia -- Aug. 2014 [\[video\]](#)
 - Trends in Wikipedia Article Creation -- Aug. 2014 [\[video\]](#)
 - Halfak's wiki research libraries (v0.0.1) -- Jul. 2014 [\[video\]](#)
 - Collaboration patterns in Articles for Creation -- May 2014 [\[video\]](#)
 - Wikipedia: maintaining production efficiency -- Mar. 2014 [\[video\]](#)
 - Wikipedia Article Creation Research -- Feb. 2014 [\[video\]](#)
 - On the nature of Anonymous Editors -- Dec. 2013 [\[slides, results\]](#)
 - Snuggle: Software support for wiki-mentors -- Aug. 2013 [\[video\]](#)
 - Newcomer socialization with Snuggle -- Jan. 2013
 - The state of newcomers in Wikipedia -- Oct. 2012
- MAXIMA - Interdisciplinary Research Experience for Undergraduates
 - The Rise and Decline of Wikipedia -- Jun. 2012
- The Colege of St. Scholastica
 - Engineering Open Production Efficiency at Scale -- Mar. 2016
 - Why go to graduate school? -- Mar. 2007
 - AJAX: why it matters and how you use it. -- Dec. 2006
- The University of Minnesota
 - The Rise and Decline of an Open Collaboration System -- Sept. 11, 2014
 - Wikipedia as a Socio-Technical System -- Sept. 23, 2016
 - Engineering Open Production Efficiency at Scale -- May 2018
- The Internation Wiki Science Conference
 - Keynote: Engineering at the Intersection of Productive Efficiency, Ideology, and Ethical AI in Wikipedia -- Mar. 2018
 - Keynote: The Great Promise of Wiki Science -- Mar. 2019

ACTIVITIES AND LEADERSHIP

Wikimedia Hackathon *(global event)*

- Organize mentors and the mentoring program
- Organize project matching sessions
- Develop common infrastructure to share information between mentors/newcomers
- Train new coordinators

Newcomer programs
 coordinator: May 2015 - May
 2020

Midtown Cycling *Minneapolis, MN*

- Organize routes and lead group rides
- Plan and execute in-person events
- Train newcomers and address deviant behavior

Board Member: May 2019 -
 current

L² Wiki Research Hackathon *(global event)*

- Outreach to academic researchers
- Facilitating community/researcher collaboration

Event organizer: [Nov. 2013](#)
 Event organizer: [Aug. 2014](#)

Wikimedia Research Committee Wikimedia Foundation

- Vet subject recruitment proposals
- Advise researchers on methods, datasets, etc.
- Write policies to govern research activities

Subject Recruitment Chair:
 2011 - current

Computer Science Graduate Student Association University of Minnesota

- Student support and organization

President: 2011-2012
 Secretary: 2009-2010

- Interface between department and students

Adaptive Sports Club University of Minnesota

- Organize games, repair wheelchairs and publicize events
- Recruitment of new players

Recruiter: 2010-2013
President: 2008-2010

Young Scientist Round Table Volunteer University of Minnesota

- Encourage elementary school students to pursue math and science

Oct 2006