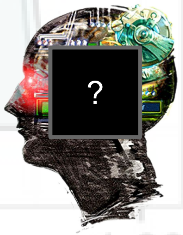
Deploying and maintaining AI in a socio-technical system

Aaron Halfaker
Wikimedia Research



Aaron Halfaker

Principal Research Scientist, Wikimedia Foundation

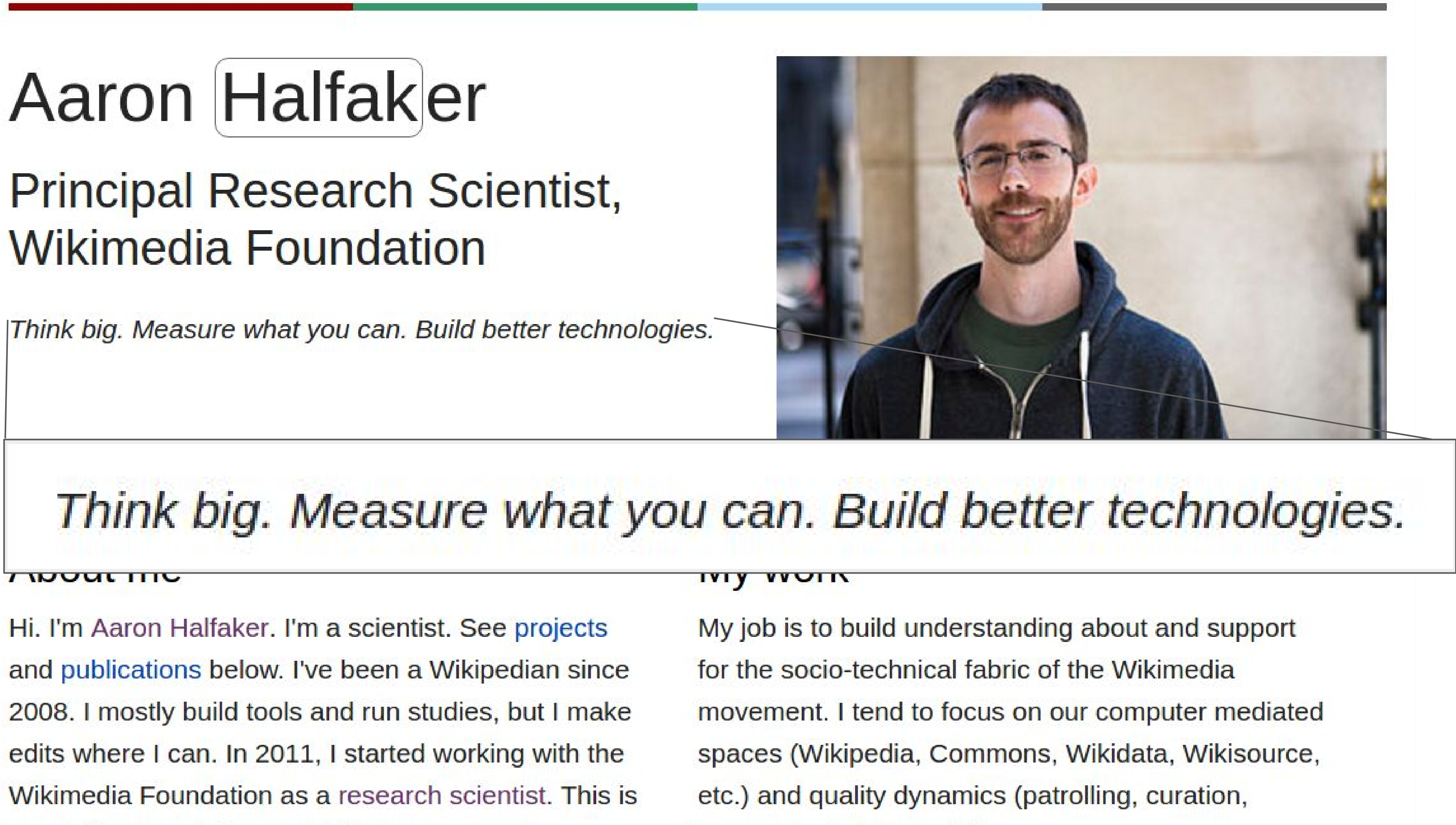
Think big. Measure what you can. Build better technologies.

About me

Hi. I'm Aaron Halfaker. I'm a scientist. See projects and publications below. I've been a Wikipedian since 2008. I mostly build tools and run studies, but I make edits where I can. In 2011, I started working with the Wikimedia Foundation as a research scientist. This is

My work

My job is to build understanding about and support for the socio-technical fabric of the Wikimedia movement. I tend to focus on our computer mediated spaces (Wikipedia, Commons, Wikidata, Wikisource, etc.) and quality dynamics (patrolling, curation,



Aaron Halfaker

Principal Research Scientist, Wikimedia Foundation

Think big. Measure what you can. Build better technologies.

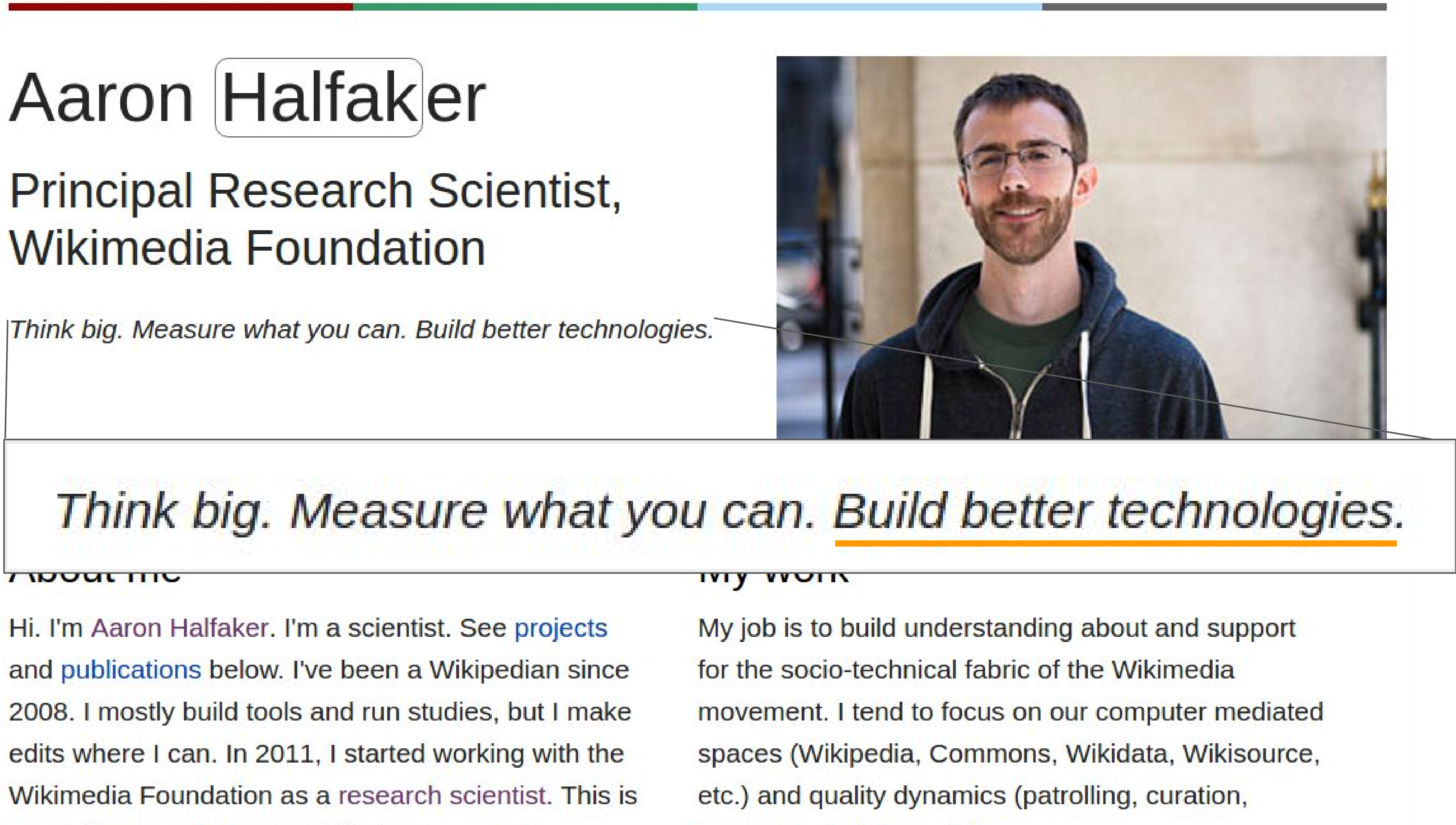
Think big. Measure what you can. Build better technologies.

About me

Hi. I'm Aaron Halfaker. I'm a scientist. See projects and publications below. I've been a Wikipedian since 2008. I mostly build tools and run studies, but I make edits where I can. In 2011, I started working with the Wikimedia Foundation as a research scientist. This is

My work

My job is to build understanding about and support for the socio-technical fabric of the Wikimedia movement. I tend to focus on our computer mediated spaces (Wikipedia, Commons, Wikidata, Wikisource, etc.) and quality dynamics (patrolling, curation,



Aaron Halfaker

Principal Research Scientist, Wikimedia Foundation

Think big. Measure what you can. Build better technologies.

Think big. Measure what you can. Build better technologies.

About me

Hi. I'm Aaron Halfaker. I'm a scientist. See projects and publications below. I've been a Wikipedian since 2008. I mostly build tools and run studies, but I make edits where I can. In 2011, I started working with the Wikimedia Foundation as a research scientist. This is

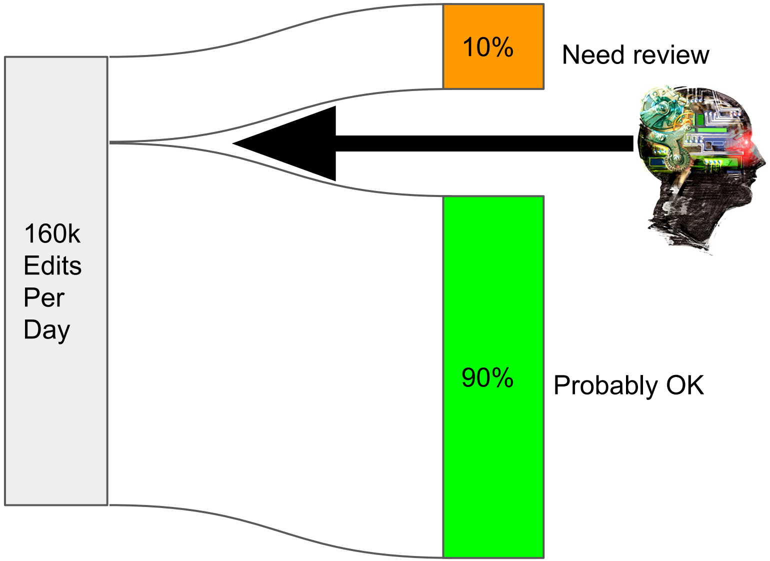
My work

My job is to build understanding about and support for the socio-technical fabric of the Wikimedia movement. I tend to focus on our computer mediated spaces (Wikipedia, Commons, Wikidata, Wikisource, etc.) and quality dynamics (patrolling, curation,

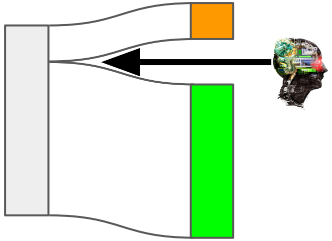
Outline

Outline

1. ORES' vision: Efficiency and innovation

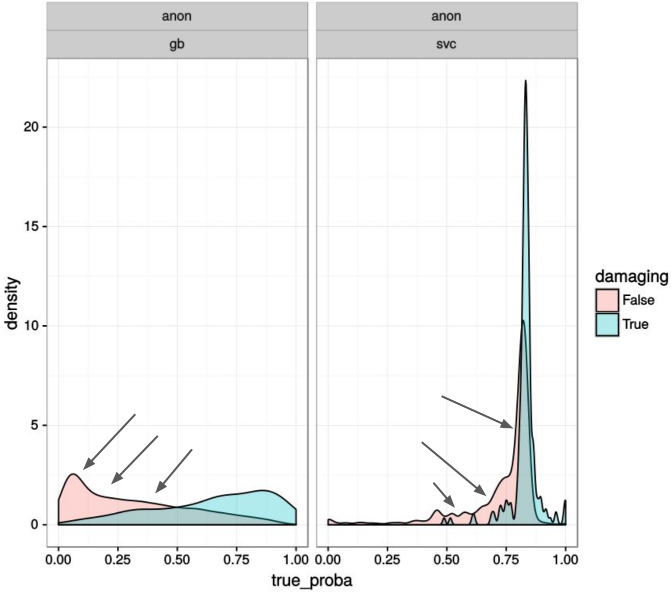


Outline



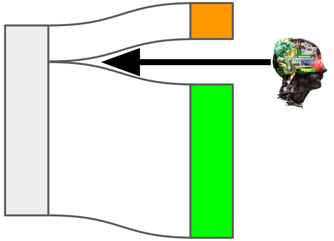
1. ORES' vision: Efficiency and innovation

2. The social dangers of AI realized!

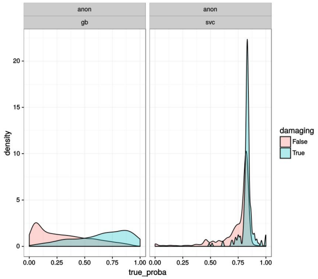


Outline

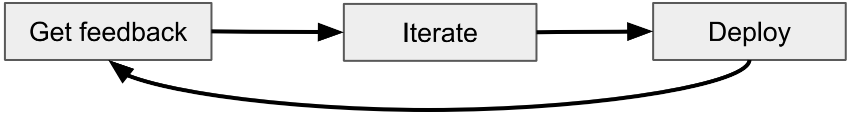
1. ORES’ vision: Efficiency and innovation



2. The social dangers of AI realized!



3. Intuitive evaluation -- AKA, how about we ask the humans?



Get feedback

Iterate

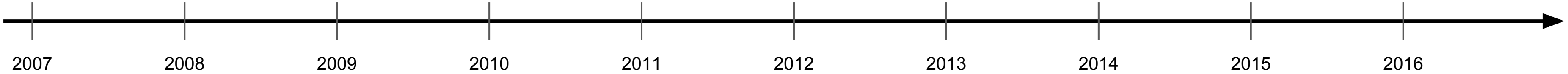
Deploy

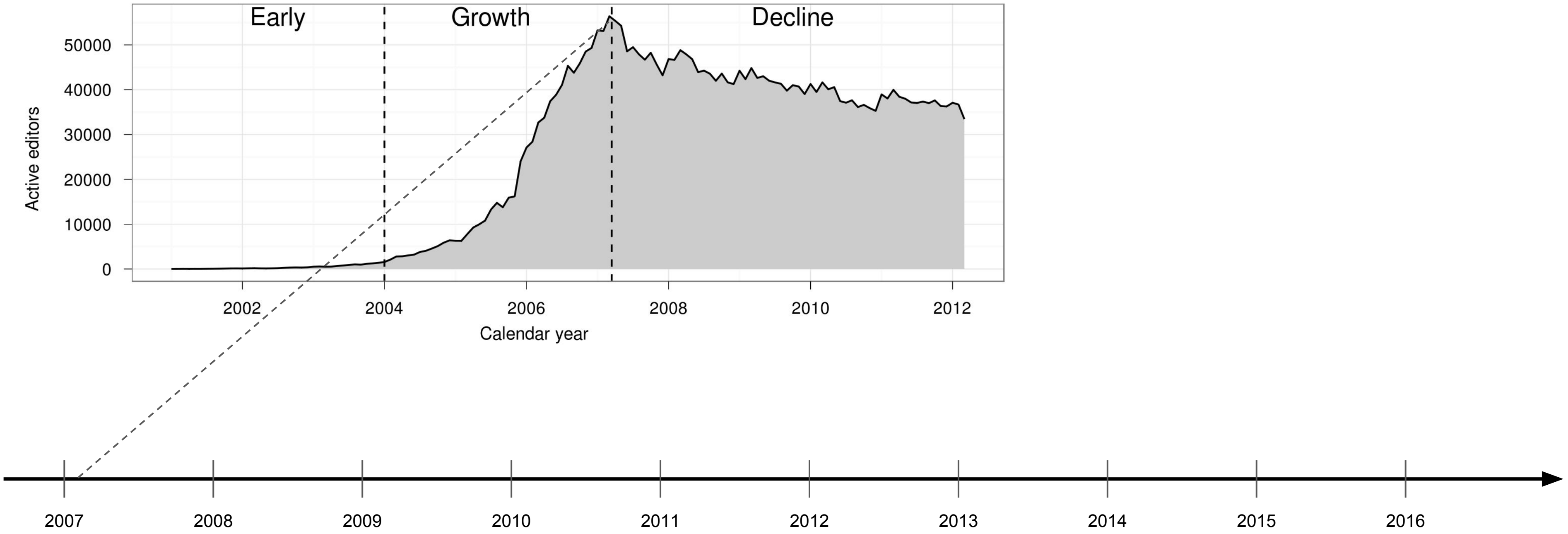
Part 1: ORES’ vision

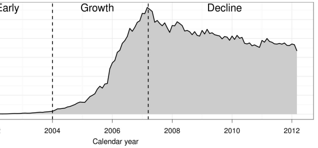
What’s ORES?

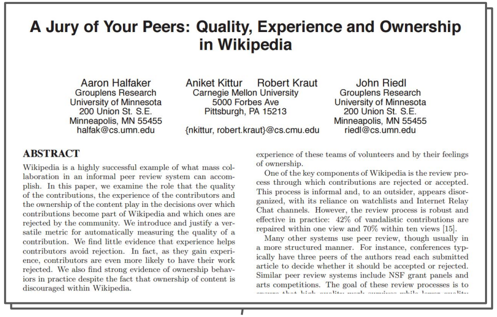
~~What’s ORES?~~

Why is ORES?


2007
2008
2009
2010
2011
2012
2013
2014
2015
2016


Early
Growth
Decline
Active editors
50000
40000
30000
20000
10000
0
2002
2004
2006
2008
2010
2012
Calendar year
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016





A Jury of Your Peers: Quality, Experience and Ownership in Wikipedia

Aaron Halfaker
Grouplens Research
University of Minnesota
200 Union St. S.E.
Minneapolis, MN 55455
halfak@cs.umn.edu

Aniket Kittur Robert Kraut
Carnegie Mellon University
5000 Forbes Ave
Pittsburgh, PA 15213
{nkittur, robert.kraut}@cs.cmu.edu

John Riedl
Grouplens Research
University of Minnesota
200 Union St. S.E.
Minneapolis, MN 55455
riedl@cs.umn.edu

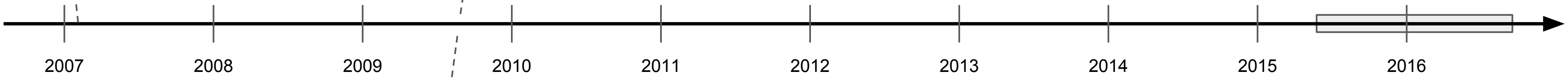
ABSTRACT

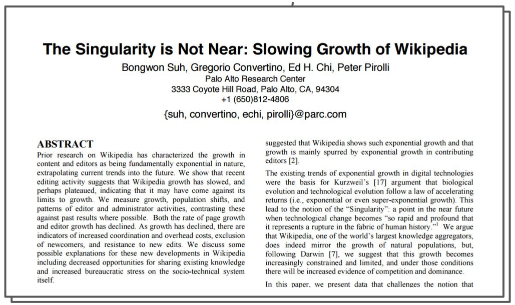
Wikipedia is a highly successful example of what mass collaboration in an informal peer review system can accomplish. In this paper, we examine the role that the quality of the contributions, the experience of the contributors and the ownership of the content play in the decisions over which contributions become part of Wikipedia and which ones are rejected by the community. We introduce and justify a versatile metric for automatically measuring the quality of a contribution. We find little evidence that experience helps contributors avoid rejection. In fact, as they gain experience, contributors are even more likely to have their work rejected. We also find strong evidence of ownership behaviors in practice despite the fact that ownership of content is discouraged within Wikipedia.

experience of these teams of volunteers and by their feelings of ownership.

One of the key components of Wikipedia is the review process through which contributions are rejected or accepted. This process is informal and, to an outsider, appears disorganized, with its reliance on watchlists and Internet Relay Chat channels. However, the review process is robust and effective in practice: 42% of vandalistic contributions are repaired within one view and 70% within ten views [15].

Many other systems use peer review, though usually in a more structured manner. For instance, conferences typically have three peers of the authors read each submitted article to decide whether it should be accepted or rejected. Similar peer review systems include NSF grant panels and arts competitions. The goal of these review processes is to





The Singularity is Not Near: Slowing Growth of Wikipedia

Bongwon Suh, Gregorio Convertino, Ed H. Chi, Peter Pirolli
Palo Alto Research Center
3333 Coyote Hill Road, Palo Alto, CA, 94304
+1 (650)812-4806
{suh, convertino, echi, pirolli}@parc.com

ABSTRACT

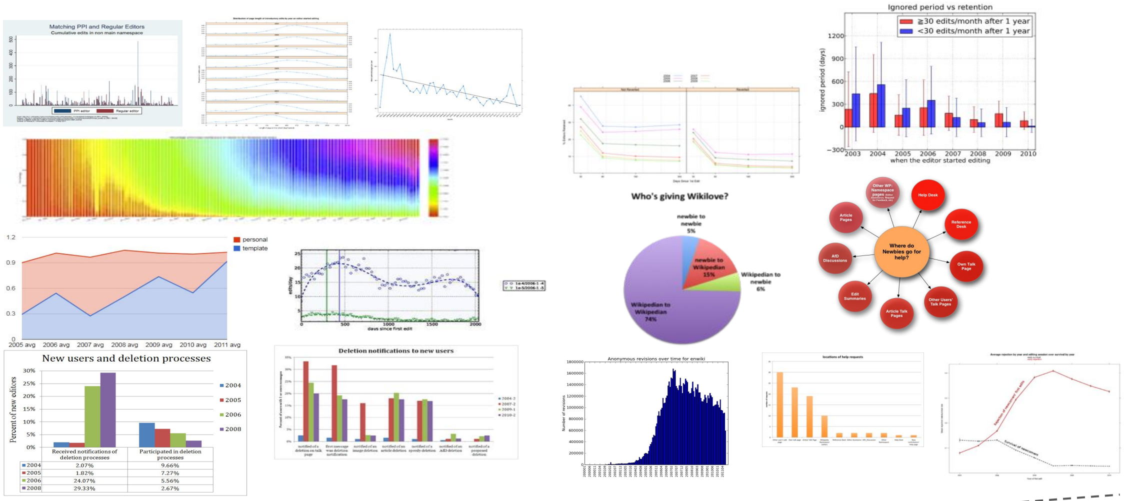
Prior research on Wikipedia has characterized the growth in content and editors as being fundamentally exponential in nature, extrapolating current trends into the future. We show that recent editing activity suggests that Wikipedia growth has slowed, and perhaps plateaued, indicating that it may have come against its limits to growth. We measure growth, population shifts, and patterns of editor and administrator activities, contrasting these against past results where possible. Both the rate of page growth and editor growth has declined. As growth has declined, there are indicators of increased coordination and overhead costs, exclusion of newcomers, and resistance to new edits. We discuss some possible explanations for these new developments in Wikipedia including decreased opportunities for sharing existing knowledge and increased bureaucratic stress on the socio-technical system itself.

suggested that Wikipedia shows such exponential growth and that growth is mainly spurred by exponential growth in contributing editors [2].

The existing trends of exponential growth in digital technologies were the basis for Kurzweil's [17] argument that biological evolution and technological evolution follow a law of accelerating returns (i.e., exponential or even super-exponential growth). This lead to the notion of the "Singularity": a point in the near future when technological change becomes "so rapid and profound that it represents a rupture in the fabric of human history."[1] We argue that Wikipedia, one of the world's largest knowledge aggregators, does indeed mirror the growth of natural populations, but, following Darwin [7], we suggest that this growth becomes increasingly constrained and limited, and under those conditions there will be increased evidence of competition and dominance.

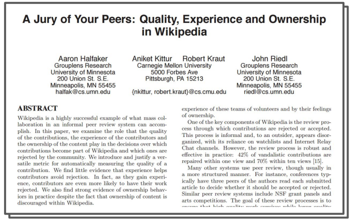
In this paper, we present data that challenges the notion that

WSoR 2011



New users and deletion processes

	Received notifications of deletion processes	Participated in deletion processes
🟦 2004	2.07%	9.66%
🟥 2005	1.82%	7.27%
🟩 2006	24.07%	5.56%
🟪 2008	29.33%	2.67%



A Jury of Your Peers: Quality, Experience and Ownership in Wikipedia

Aaron Halfaker
Grouplens Research
University of Minnesota
200 Union St. S.E.
Minneapolis, MN 55455
halfak@cs.umn.edu

Aniket Kittur Robert Kraut
Carnegie Mellon University
5000 Forbes Ave
Pittsburgh, PA 15213
{nkittur, robert.kraut}@cs.cmu.edu

John Riedl
Grouplens Research
University of Minnesota
200 Union St. S.E.
Minneapolis, MN 55455
riedl@cs.umn.edu

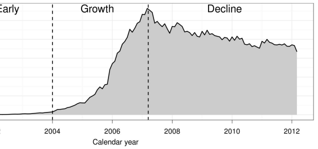
ABSTRACT

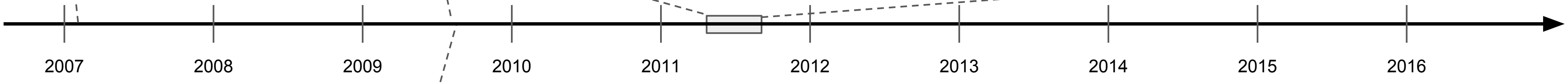
Wikipedia is a highly successful example of what mass collaboration in an informal peer review system can accomplish. In this paper, we examine the role that the quality of the contributions, the experience of the contributors and the ownership of the content play in the decisions over which contributions become part of Wikipedia and which ones are rejected by the community. We introduce and justify a versatile metric for automatically measuring the quality of a contribution. We find little evidence that experience helps contributors avoid rejection. In fact, as they gain experience, contributors are even more likely to have their work rejected. We also find strong evidence of ownership behaviors in practice despite the fact that ownership of content is discouraged within Wikipedia.

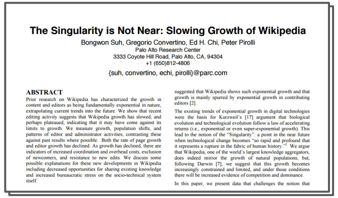
experience of these teams of volunteers and by their feelings of ownership.

One of the key components of Wikipedia is the review process through which contributions are rejected or accepted. This process is informal and, to an outsider, appears disorganized, with its reliance on watchlists and Internet Relay Chat channels. However, the review process is robust and effective in practice: 42% of vandalistic contributions are repaired within one view and 70% within ten views [15].

Many other systems use peer review, though usually in a more structured manner. For instance, conferences typically have three peers of the authors read each submitted article to decide whether it should be accepted or rejected. Similar peer review systems include NSF grant panels and arts competitions. The goal of these review processes is to







The Singularity is Not Near: Slowing Growth of Wikipedia

Bongwon Suh, Gregorio Convertino, Ed H. Chi, Peter Pirolli
Palo Alto Research Center
3333 Coyote Hill Road, Palo Alto, CA, 94304
+1 (650)812-4806
{suh, convertino, echi, pirolli}@parc.com

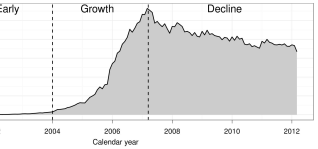
ABSTRACT

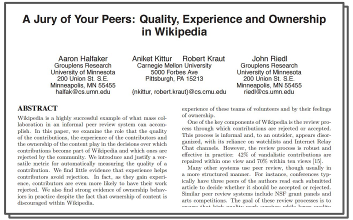
Prior research on Wikipedia has characterized the growth in content and editors as being fundamentally exponential in nature, extrapolating current trends into the future. We show that recent editing activity suggests that Wikipedia growth has slowed, and perhaps plateaued, indicating that it may have come against its limits to growth. We measure growth, population shifts, and patterns of editor and administrator activities, contrasting these against past results where possible. Both the rate of page growth and editor growth has declined. As growth has declined, there are indicators of increased coordination and overhead costs, exclusion of newcomers, and resistance to new edits. We discuss some possible explanations for these new developments in Wikipedia including decreased opportunities for sharing existing knowledge and increased bureaucratic stress on the socio-technical system itself.

suggested that Wikipedia shows such exponential growth and that growth is mainly spurred by exponential growth in contributing editors [2].

The existing trends of exponential growth in digital technologies were the basis for Kurzweil's [17] argument that biological evolution and technological evolution follow a law of accelerating returns (i.e., exponential or even super-exponential growth). This lead to the notion of the "Singularity": a point in the near future when technological change becomes "so rapid and profound that it represents a rupture in the fabric of human history."[1] We argue that Wikipedia, one of the world's largest knowledge aggregators, does indeed mirror the growth of natural populations, but, following Darwin [7], we suggest that this growth becomes increasingly constrained and limited, and under those conditions there will be increased evidence of competition and dominance.

In this paper, we present data that challenges the notion that





A Jury of Your Peers: Quality, Experience and Ownership in Wikipedia

Aaron Halfaker
Grouplens Research
University of Minnesota
200 Union St. S.E.
Minneapolis, MN 55455
halfak@cs.umn.edu

Aniket Kittur Robert Kraut
Carnegie Mellon University
5000 Forbes Ave
Pittsburgh, PA 15213
{nkittur, robert.kraut}@cs.cmu.edu

John Riedl
Grouplens Research
University of Minnesota
200 Union St. S.E.
Minneapolis, MN 55455
riedl@cs.umn.edu

ABSTRACT

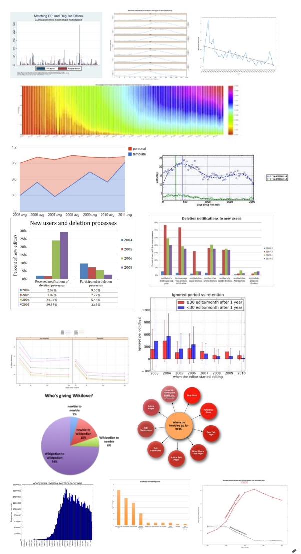
Wikipedia is a highly successful example of what mass collaboration in an informal peer review system can accomplish. In this paper, we examine the role that the quality of the contributions, the experience of the contributors and the ownership of the content play in the decisions over which contributions become part of Wikipedia and which ones are rejected by the community. We introduce and justify a versatile metric for automatically measuring the quality of a contribution. We find little evidence that experience helps contributors avoid rejection. In fact, as they gain experience, contributors are even more likely to have their work rejected. We also find strong evidence of ownership behaviors in practice despite the fact that ownership of content is discouraged within Wikipedia.

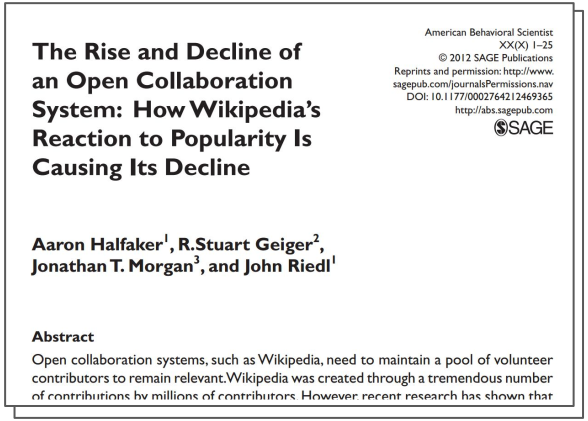
experience of these teams of volunteers and by their feelings of ownership.

One of the key components of Wikipedia is the review process through which contributions are rejected or accepted. This process is informal and, to an outsider, appears disorganized, with its reliance on watchlists and Internet Relay Chat channels. However, the review process is robust and effective in practice: 42% of vandalistic contributions are repaired within one view and 70% within ten views [15].

Many other systems use peer review, though usually in a more structured manner. For instance, conferences typically have three peers of the authors read each submitted article to decide whether it should be accepted or rejected. Similar peer review systems include NSF grant panels and arts competitions. The goal of these review processes is to

WSoR 2011





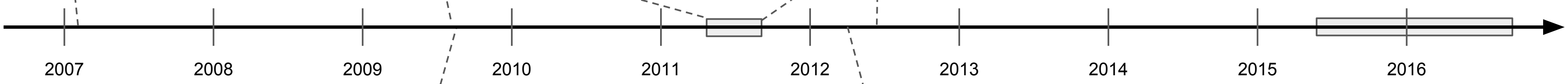
American Behavioral Scientist
XX(X) 1–25
© 2012 SAGE Publications
Reprints and permission: http://www.sagepub.com/journalsPermissions.nav
DOI: 10.1177/0002764212469365
http://abs.sagepub.com
SAGE

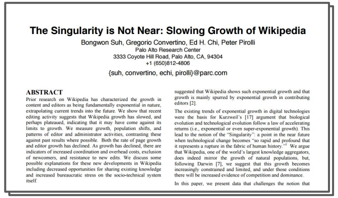
The Rise and Decline of an Open Collaboration System: How Wikipedia's Reaction to Popularity Is Causing Its Decline

Aaron Halfaker[1], R.Stuart Geiger[2], Jonathan T. Morgan[3], and John Riedl[1]

Abstract

Open collaboration systems, such as Wikipedia, need to maintain a pool of volunteer contributors to remain relevant. Wikipedia was created through a tremendous number of contributions by millions of contributors. However, recent research has shown that





The Singularity is Not Near: Slowing Growth of Wikipedia

Bongwon Suh, Gregorio Convertino, Ed H. Chi, Peter Pirolli
Palo Alto Research Center
3333 Coyote Hill Road, Palo Alto, CA, 94304
+1 (650)812-4806
{suh, convertino, echi, pirolli}@parc.com

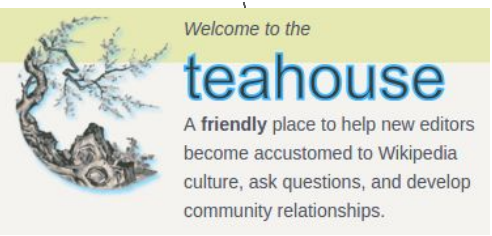
ABSTRACT

Prior research on Wikipedia has characterized the growth in content and editors as being fundamentally exponential in nature, extrapolating current trends into the future. We show that recent editing activity suggests that Wikipedia growth has slowed, and perhaps plateaued, indicating that it may have come against its limits to growth. We measure growth, population shifts, and patterns of editor and administrator activities, contrasting these against past results where possible. Both the rate of page growth and editor growth has declined. As growth has declined, there are indicators of increased coordination and overhead costs, exclusion of newcomers, and resistance to new edits. We discuss some possible explanations for these new developments in Wikipedia including decreased opportunities for sharing existing knowledge and increased bureaucratic stress on the socio-technical system itself.

suggested that Wikipedia shows such exponential growth and that growth is mainly spurred by exponential growth in contributing editors [2].

The existing trends of exponential growth in digital technologies were the basis for Kurzweil's [17] argument that biological evolution and technological evolution follow a law of accelerating returns (i.e., exponential or even super-exponential growth). This lead to the notion of the "Singularity": a point in the near future when technological change becomes "so rapid and profound that it represents a rupture in the fabric of human history."[1] We argue that Wikipedia, one of the world's largest knowledge aggregators, does indeed mirror the growth of natural populations, but, following Darwin [7], we suggest that this growth becomes increasingly constrained and limited, and under those conditions there will be increased evidence of competition and dominance.

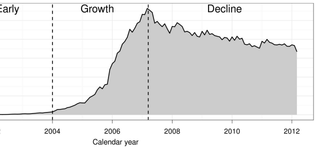
In this paper, we present data that challenges the notion that

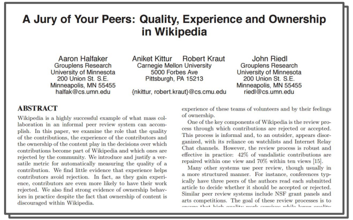


Welcome to the

teahouse

A **friendly** place to help new editors become accustomed to Wikipedia culture, ask questions, and develop community relationships.





A Jury of Your Peers: Quality, Experience and Ownership in Wikipedia

Aaron Halfaker
Grouplens Research
University of Minnesota
200 Union St. S.E.
Minneapolis, MN 55455
halfak@cs.umn.edu

Aniket Kittur Robert Kraut
Carnegie Mellon University
5000 Forbes Ave
Pittsburgh, PA 15213
{nkittur, robert.kraut}@cs.cmu.edu

John Riedl
Grouplens Research
University of Minnesota
200 Union St. S.E.
Minneapolis, MN 55455
riedl@cs.umn.edu

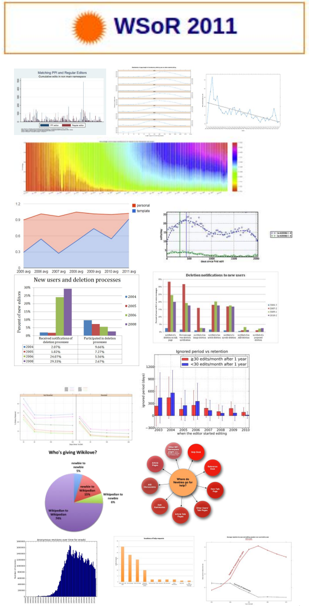
ABSTRACT

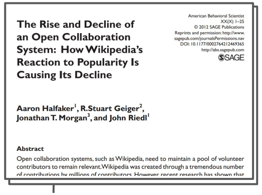
Wikipedia is a highly successful example of what mass collaboration in an informal peer review system can accomplish. In this paper, we examine the role that the quality of the contributions, the experience of the contributors and the ownership of the content play in the decisions over which contributions become part of Wikipedia and which ones are rejected by the community. We introduce and justify a versatile metric for automatically measuring the quality of a contribution. We find little evidence that experience helps contributors avoid rejection. In fact, as they gain experience, contributors are even more likely to have their work rejected. We also find strong evidence of ownership behaviors in practice despite the fact that ownership of content is discouraged within Wikipedia.

experience of these teams of volunteers and by their feelings of ownership.

One of the key components of Wikipedia is the review process through which contributions are rejected or accepted. This process is informal and, to an outsider, appears disorganized, with its reliance on watchlists and Internet Relay Chat channels. However, the review process is robust and effective in practice: 42% of vandalistic contributions are repaired within one view and 70% within ten views [15].

Many other systems use peer review, though usually in a more structured manner. For instance, conferences typically have three peers of the authors read each submitted article to decide whether it should be accepted or rejected. Similar peer review systems include NSF grant panels and arts competitions. The goal of these review processes is to ensure that high quality work survives while lower quality





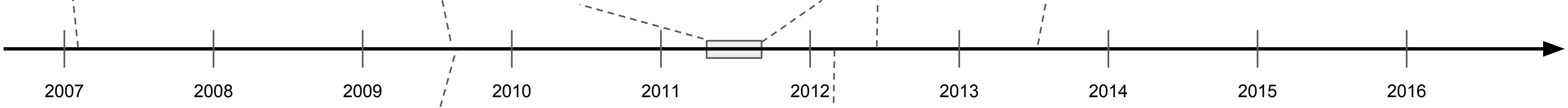
The Rise and Decline of an Open Collaboration System: How Wikipedia's Reaction to Popularity Is Causing Its Decline

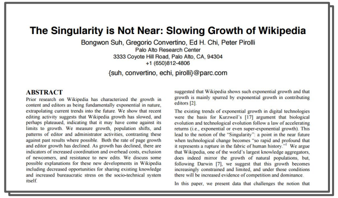
American Behavioral Scientist
XX(X) 1–25
© 2012 SAGE Publications
Reprints and permission: http://www.sagepub.com/journalsPermissions.nav
DOI: 10.1177/0002764212469365
http://abs.sagepub.com
SAGE

Aaron Halfaker[1], R.Stuart Geiger[2],
Jonathan T. Morgan[3], and John Riedl[1]

Abstract

Open collaboration systems, such as Wikipedia, need to maintain a pool of volunteer contributors to remain relevant. Wikipedia was created through a tremendous number of contributions by millions of contributors. However recent research has shown that





The Singularity is Not Near: Slowing Growth of Wikipedia

Bongwon Suh, Gregorio Convertino, Ed H. Chi, Peter Pirolli
Palo Alto Research Center
3333 Coyote Hill Road, Palo Alto, CA, 94304
+1 (650)812-4806
{suh, convertino, echi, pirolli}@parc.com

ABSTRACT

Prior research on Wikipedia has characterized the growth in content and editors as being fundamentally exponential in nature, extrapolating current trends into the future. We show that recent editing activity suggests that Wikipedia growth has slowed, and perhaps plateaued, indicating that it may have come against its limits to growth. We measure growth, population shifts, and patterns of editor and administrator activities, contrasting these against past results where possible. Both the rate of page growth and editor growth has declined. As growth has declined, there are indicators of increased coordination and overhead costs, exclusion of newcomers, and resistance to new edits. We discuss some possible explanations for these new developments in Wikipedia including decreased opportunities for sharing existing knowledge and increased bureaucratic stress on the socio-technical system itself.

suggested that Wikipedia shows such exponential growth and that growth is mainly spurred by exponential growth in contributing editors [2].

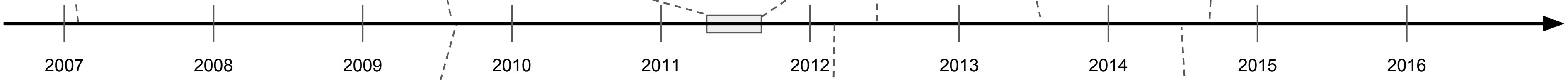
The existing trends of exponential growth in digital technologies were the basis for Kurzweil's [17] argument that biological evolution and technological evolution follow a law of accelerating returns (i.e., exponential or even super-exponential growth). This lead to the notion of the "Singularity": a point in the near future when technological change becomes "so rapid and profound that it represents a rupture in the fabric of human history."[1] We argue that Wikipedia, one of the world's largest knowledge aggregators, does indeed mirror the growth of natural populations, but, following Darwin [7], we suggest that this growth becomes increasingly constrained and limited, and under those conditions there will be increased evidence of competition and dominance.

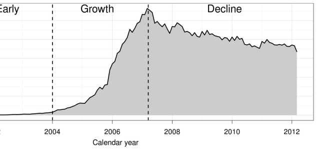
In this paper, we present data that challenges the notion that

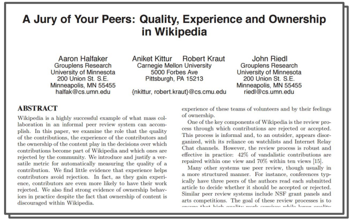
Welcome to the

teahouse

A **friendly** place to help new editors become accustomed to Wikipedia culture, ask questions, and develop community relationships.






A Jury of Your Peers: Quality, Experience and Ownership in Wikipedia

Aaron Halfaker
Grouplens Research
University of Minnesota
200 Union St. S.E.
Minneapolis, MN 55455
halfak@cs.umn.edu

Aniket Kittur Robert Kraut
Carnegie Mellon University
5000 Forbes Ave
Pittsburgh, PA 15213
{nkittur, robert.kraut}@cs.cmu.edu

John Riedl
Grouplens Research
University of Minnesota
200 Union St. S.E.
Minneapolis, MN 55455
riedl@cs.umn.edu

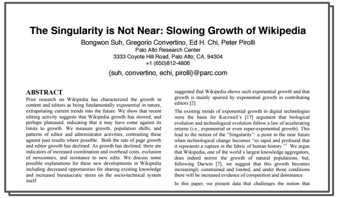
ABSTRACT

Wikipedia is a highly successful example of what mass collaboration in an informal peer review system can accomplish. In this paper, we examine the role that the quality of the contributions, the experience of the contributors and the ownership of the content play in the decisions over which contributions become part of Wikipedia and which ones are rejected by the community. We introduce and justify a versatile metric for automatically measuring the quality of a contribution. We find little evidence that experience helps contributors avoid rejection. In fact, as they gain experience, contributors are even more likely to have their work rejected. We also find strong evidence of ownership behaviors in practice despite the fact that ownership of content is discouraged within Wikipedia.

experience of these teams of volunteers and by their feelings of ownership.

One of the key components of Wikipedia is the review process through which contributions are rejected or accepted. This process is informal and, to an outsider, appears disorganized, with its reliance on watchlists and Internet Relay Chat channels. However, the review process is robust and effective in practice: 42% of vandalistic contributions are repaired within one view and 70% within ten views [15].

Many other systems use peer review, though usually in a more structured manner. For instance, conferences typically have three peers of the authors read each submitted article to decide whether it should be accepted or rejected. Similar peer review systems include NSF grant panels and arts competitions. The goal of these review processes is to


The Singularity is Not Near: Slowing Growth of Wikipedia

Bongwon Suh, Gregorio Convertino, Ed H. Chi, Peter Pirolli
Palo Alto Research Center
3333 Coyote Hill Road, Palo Alto, CA, 94304
+1 (650)812-4806
{suh, convertino, echi, pirolli}@parc.com

ABSTRACT

Prior research on Wikipedia has characterized the growth in content and editors as being fundamentally exponential in nature, extrapolating current trends into the future. We show that recent editing activity suggests that Wikipedia growth has slowed, and perhaps plateaued, indicating that it may have come against its limits to growth. We measure growth, population shifts, and patterns of editor and administrator activities, contrasting these against past results where possible. Both the rate of page growth and editor growth has declined. As growth has declined, there are indicators of increased coordination and overhead costs, exclusion of newcomers, and resistance to new edits. We discuss some possible explanations for these new developments in Wikipedia including decreased opportunities for sharing existing knowledge and increased bureaucratic stress on the socio-technical system itself.

suggested that Wikipedia shows such exponential growth and that growth is mainly spurred by exponential growth in contributing editors [2].

The existing trends of exponential growth in digital technologies were the basis for Kurzweil's [17] argument that biological evolution and technological evolution follow a law of accelerating returns (i.e., exponential or even super-exponential growth). This lead to the notion of the "Singularity": a point in the near future when technological change becomes "so rapid and profound that it represents a rupture in the fabric of human history."[1] We argue that Wikipedia, one of the world's largest knowledge aggregators, does indeed mirror the growth of natural populations, but, following Darwin [7], we suggest that this growth becomes increasingly constrained and limited, and under those conditions there will be increased evidence of competition and dominance.

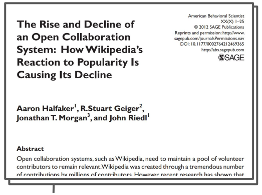
In this paper, we present data that challenges the notion that

WSoR 2011

Welcome to the

teahouse

A **friendly** place to help new editors become accustomed to Wikipedia culture, ask questions, and develop community relationships.

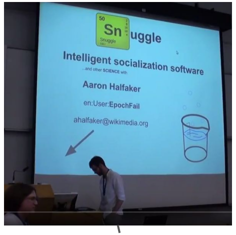

American Behavioral Scientist
XX(X) 1–25
© 2012 SAGE Publications
Reprints and permission: http://www.sagepub.com/journalsPermissions.nav
DOI: 10.1177/0002764212469365
http://abs.sagepub.com
SAGE

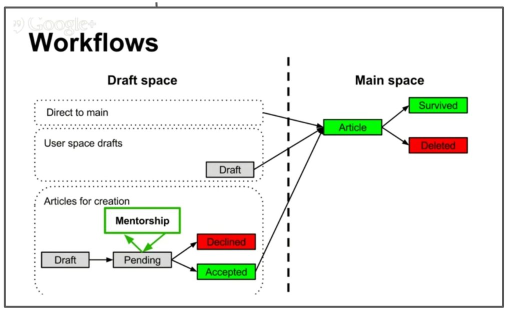
The Rise and Decline of an Open Collaboration System: How Wikipedia's Reaction to Popularity Is Causing Its Decline

Aaron Halfaker[1], R.Stuart Geiger[2], Jonathan T. Morgan[3], and John Riedl[1]

Abstract

Open collaboration systems, such as Wikipedia, need to maintain a pool of volunteer contributors to remain relevant. Wikipedia was created through a tremendous number of contributions by millions of contributors. However recent research has shown that




Google+

Workflows

Draft space

Main space

Direct to main

User space drafts

Draft

Articles for creation

Mentorship

Draft

Pending

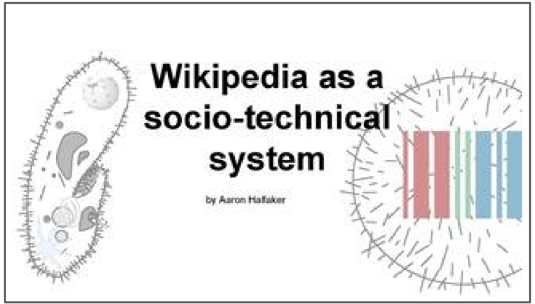
Declined

Accepted

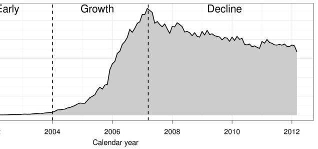
Article

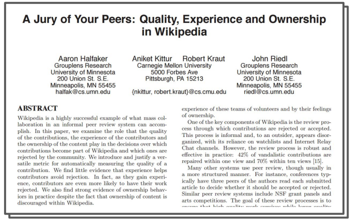
Survived

Deleted


Wikipedia as a socio-technical system

by Aaron Halfaker





A Jury of Your Peers: Quality, Experience and Ownership in Wikipedia

Aaron Halfaker
Grouplens Research
University of Minnesota
200 Union St. S.E.
Minneapolis, MN 55455
halfak@cs.umn.edu

Aniket Kittur Robert Kraut
Carnegie Mellon University
5000 Forbes Ave
Pittsburgh, PA 15213
{nkittur, robert.kraut}@cs.cmu.edu

John Riedl
Grouplens Research
University of Minnesota
200 Union St. S.E.
Minneapolis, MN 55455
riedl@cs.umn.edu

ABSTRACT

Wikipedia is a highly successful example of what mass collaboration in an informal peer review system can accomplish. In this paper, we examine the role that the quality of the contributions, the experience of the contributors and the ownership of the content play in the decisions over which contributions become part of Wikipedia and which ones are rejected by the community. We introduce and justify a versatile metric for automatically measuring the quality of a contribution. We find little evidence that experience helps contributors avoid rejection. In fact, as they gain experience, contributors are even more likely to have their work rejected. We also find strong evidence of ownership behaviors in practice despite the fact that ownership of content is discouraged within Wikipedia.

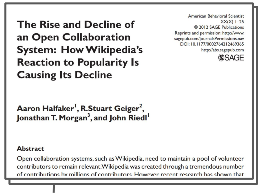
experience of these teams of volunteers and by their feelings of ownership.

One of the key components of Wikipedia is the review process through which contributions are rejected or accepted. This process is informal and, to an outsider, appears disorganized, with its reliance on watchlists and Internet Relay Chat channels. However, the review process is robust and effective in practice: 42% of vandalistic contributions are repaired within one view and 70% within ten views [15].

Many other systems use peer review, though usually in a more structured manner. For instance, conferences typically have three peers of the authors read each submitted article to decide whether it should be accepted or rejected. Similar peer review systems include NSF grant panels and arts competitions. The goal of these review processes is to

WSoR 2011

Matching PPI and Regular Editors
New users and deletion processes
Deletion notifications to new users
Ignored period vs retention
≥30 edits/month after 1 year
<30 edits/month after 1 year
when the editor started editing
Who's giving Wikilove?
personal
template



The Rise and Decline of an Open Collaboration System: How Wikipedia's Reaction to Popularity Is Causing Its Decline

American Behavioral Scientist
XX(X) 1–25
© 2012 SAGE Publications
Reprints and permission: http://www.sagepub.com/journalsPermissions.nav
DOI: 10.1177/0002764212469365
http://abs.sagepub.com
SAGE

Aaron Halfaker[1], R.Stuart Geiger[2], Jonathan T. Morgan[3], and John Riedl[1]

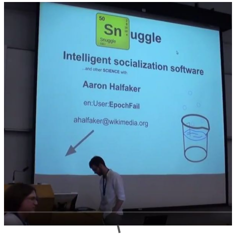
Abstract

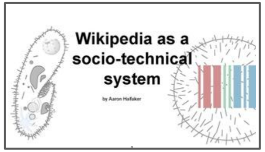
Open collaboration systems, such as Wikipedia, need to maintain a pool of volunteer contributors to remain relevant. Wikipedia was created through a tremendous number of contributions by millions of contributors. However recent research has shown that

ORES

The people's classifier service

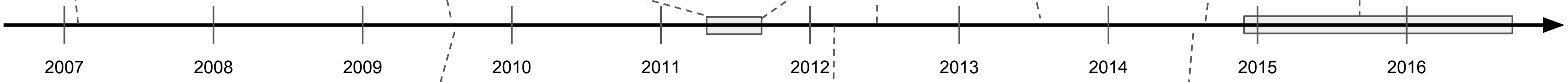
Towards an open model for algorithmic infrastructure

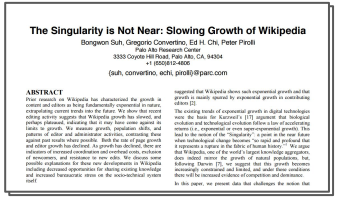




Wikipedia as a socio-technical system

by Aaron Halfaker





The Singularity is Not Near: Slowing Growth of Wikipedia

Bongwon Suh, Gregorio Convertino, Ed H. Chi, Peter Pirolli
Palo Alto Research Center
3333 Coyote Hill Road, Palo Alto, CA, 94304
+1 (650)812-4806
{suh, convertino, echi, pirolli}@parc.com

ABSTRACT

Prior research on Wikipedia has characterized the growth in content and editors as being fundamentally exponential in nature, extrapolating current trends into the future. We show that recent editing activity suggests that Wikipedia growth has slowed, and perhaps plateaued, indicating that it may have come against its limits to growth. We measure growth, population shifts, and patterns of editor and administrator activities, contrasting these against past results where possible. Both the rate of page growth and editor growth has declined. As growth has declined, there are indicators of increased coordination and overhead costs, exclusion of newcomers, and resistance to new edits. We discuss some possible explanations for these new developments in Wikipedia including decreased opportunities for sharing existing knowledge and increased bureaucratic stress on the socio-technical system itself.

suggested that Wikipedia shows such exponential growth and that growth is mainly spurred by exponential growth in contributing editors [2].

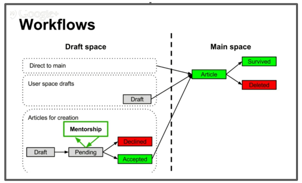
The existing trends of exponential growth in digital technologies were the basis for Kurzweil's [17] argument that biological evolution and technological evolution follow a law of accelerating returns (i.e., exponential or even super-exponential growth). This lead to the notion of the "Singularity": a point in the near future when technological change becomes "so rapid and profound that it represents a rupture in the fabric of human history."[1] We argue that Wikipedia, one of the world's largest knowledge aggregators, does indeed mirror the growth of natural populations, but, following Darwin [7], we suggest that this growth becomes increasingly constrained and limited, and under those conditions there will be increased evidence of competition and dominance.

In this paper, we present data that challenges the notion that

Welcome to the

teahouse

A **friendly** place to help new editors become accustomed to Wikipedia culture, ask questions, and develop community relationships.



Workflows

References

- Suh, B., Convertino, G., Chi, E. H., & Pirolli, P. (2009, October). The singularity is not near: slowing growth of Wikipedia. In *Proceedings of the 5th International Symposium on Wikis and Open Collaboration* (p. 8). ACM.
- Halfaker, A., Kittur, A., Kraut, R., & Riedl, J. (2009, October). A jury of your peers: quality, experience and ownership in Wikipedia. In *Proceedings of the 5th International Symposium on Wikis and Open Collaboration* (p. 15). ACM.
- https://meta.wikimedia.org/wiki/Research:Wikimedia_Summer_of_Research_2011
- https://meta.wikimedia.org/wiki/Research:Teahouse
- Halfaker, A., Geiger, R. S., Morgan, J. T., & Riedl, J. (2012). The rise and decline of an open collaboration system: How Wikipedia's reaction to popularity is causing its decline. *American Behavioral Scientist*, 0002764212469365.
- https://www.youtube.com/watch?v=bwMBLrAHJLM (Snuggle @ Wikimania 2013)
- https://www.mediawiki.org/wiki/File:Wikimedia_Research_%26_Data_Showcase_-_February_2014.webm (Wikipedia article creation analysis)
- https://www.youtube.com/watch?v=AUupsnvV1oA#t=35m26s (Articles for Creation workflow analysis)
- https://www.youtube.com/watch?v=-We4GZbH3Iw#t=34m10s (Wikipedia as a socio-technical system "The Paramecium Talk")
- https://www.youtube.com/watch?v=Hj7o5d-OEis#t=3m25s (ORES -- The people's classifier service)

References

- Suh, B., Convertino, G., Chi, E. H., & Pirolli, P. (2009, October). The singularity is not near: slowing growth of Wikipedia. In *Proceedings of the 5th International Symposium on Wikis and Open Collaboration* (p. 8). ACM.
- Halfaker, A., Kittur, A., Kraut, R., & Riedl, J. (2009, October). A jury of your peers: quality, experience and ownership in Wikipedia. In *Proceedings of the 5th International Symposium on Wikis and Open Collaboration* (p. 15). ACM.
- https://meta.wikimedia.org/wiki/Research:Wikimedia_Summer_of_Research_2011
- https://meta.wikimedia.org/wiki/Research:Teahouse
- Halfaker, A., Geiger, R. S., Morgan, J. T., & Riedl, J. (2012). The rise and decline of an open collaboration system: How Wikipedia's reaction to popularity is causing its decline. *American Behavioral Scientist*, 0002764212469365.
- https://www.youtube.com/watch?v=bwMBLrAHJLM (Snuggle @ Wikimania 2013)
- https://www.mediawiki.org/wiki/File:Wikimedia_Research_%26_Data_Showcase_-_February_2014.webm (Wikipedia article creation analysis)
- https://www.youtube.com/watch?v=AUupsnvV1oA#t=35m26s (Articles for Creation workflow analysis)
- https://www.youtube.com/watch?v=-We4GZbH3Iw#t=34m10s (Wikipedia as a socio-technical system "The Paramecium Talk")
- https://www.youtube.com/watch?v=Hj7o5d-OEis#t=3m25s (ORES -- The people's classifier service)

References

- Suh, B., Convertino, G., Chi, E. H., & Pirolli, P. (2009, October). The singularity is not near: slowing growth of Wikipedia. In *Proceedings of the 5th International Symposium on Wikis and Open Collaboration* (p. 8). ACM.
- Halfaker, A., Kittur, A., Kraut, R., & Riedl, J. (2009, October). A jury of your peers: quality, experience and ownership in Wikipedia. In *Proceedings of the 5th International Symposium on Wikis and Open Collaboration* (p. 15). ACM.
- https://meta.wikimedia.org/wiki/Research:Wikimedia_Summer_of_Research_2011
- https://meta.wikimedia.org/wiki/Research:Teahouse
- Halfaker, A., Geiger, R. S., Morgan, J. T., & Riedl, J. (2012). The rise and decline of an open collaboration system: How Wikipedia's reaction to popularity is causing its decline. *American Behavioral Scientist*, 0002764212469365.
- https://www.youtube.com/watch?v=bwMBLrAHJLM (Snuggle @ Wikimania 2013)
- https://www.mediawiki.org/wiki/File:Wikimedia_Research_%26_Data_Showcase_-_February_2014.webm (Wikipedia article creation analysis)
- https://www.youtube.com/watch?v=AUupsnvV1oA#t=35m26s (Articles for Creation workflow analysis)
- https://www.youtube.com/watch?v=-We4GZbH3Iw#t=34m10s (Wikipedia as a socio-technical system "The Paramecium Talk")
- https://www.youtube.com/watch?v=Hj7o5d-OEis#t=3m25s (ORES -- The people's classifier service)

Why is ORES?

- Wikipedia has socio-technical problems with newbies.

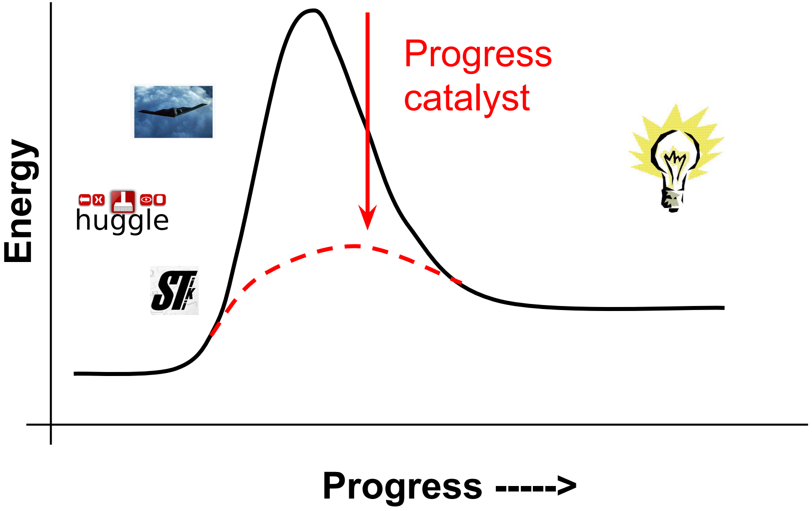
Why is ORES?

- Wikipedia has socio-technical problems with newbies.
- Many of Wikipedia’s problems are due to its scale.

https://commons.wikimedia.org/wiki/File:Flickr_-_Official_U.S._Navy_Imagery_-_Sailor%27s_daughter_operates_a_fire_hose_with_crew_member_assistance..jpg

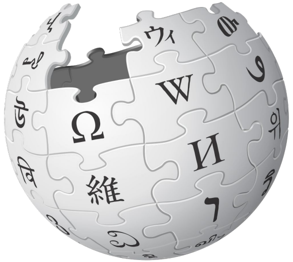
Why is ORES?

- Wikipedia has socio-technical problems with newbies.
- Many of Wikipedia’s problems are due to its scale.
- ORES is an attempt to address both at the same time.



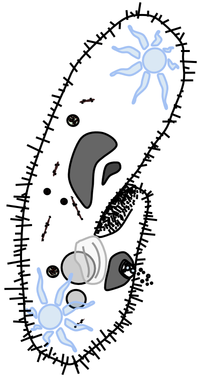
Why is ORES?

- Wikipedia has socio-technical problems with newbies.
- Many of Wikipedia's problems are due to its scale.
- ORES is an attempt to address both at the same time.
- … in a way that accounts for the complex dynamics of Wikipedia.



System with <u>specialized</u> **sub-systems**

~



System with <u>specialized</u> **sub-systems**

Why is ORES?

- Wikipedia has socio-technical problems with newbies.
- Many of Wikipedia’s problems are due to its scale.
- ORES is an attempt to address both at the same time.
- … in a way that accounts for the complex dynamics of Wikipedia.

What is ORES?

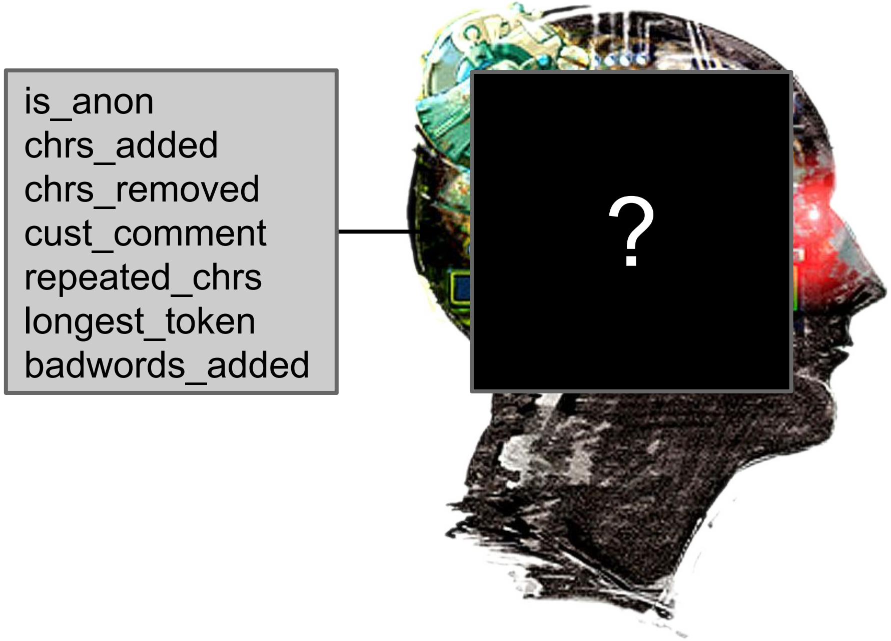
The machine classifier

https://commons.wikimedia.org/wiki/File:Artificial.intelligence.jpg
Public Domain

The machine classifier

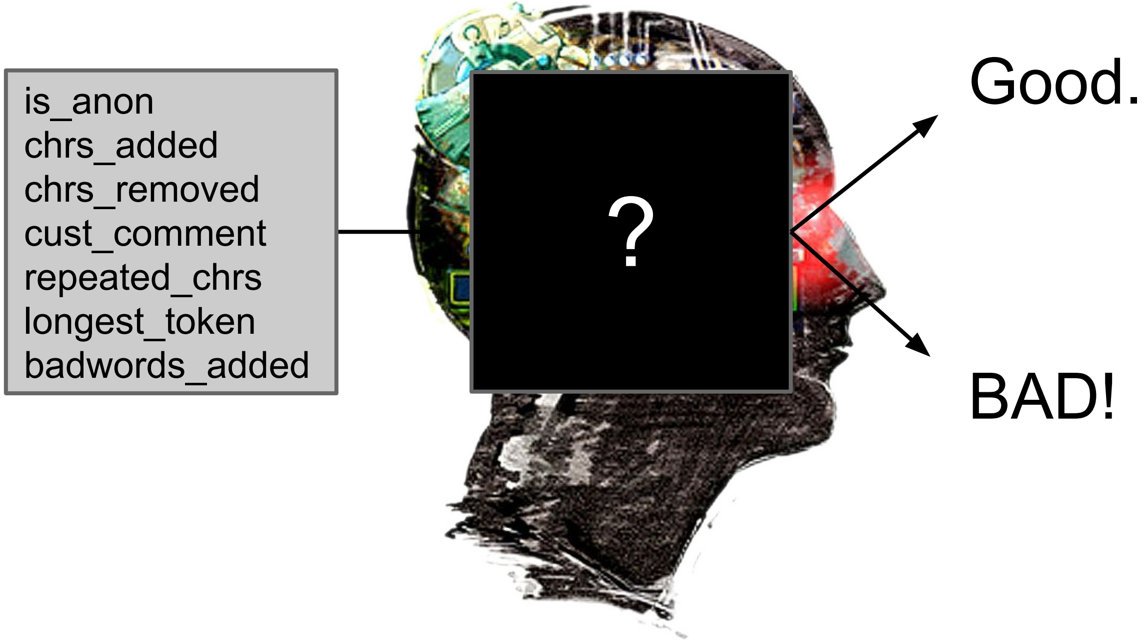
is_anon
chrs_added
chrs_removed
cust_comment
repeated_chrs
longest_token
badwords_added

The machine classifier



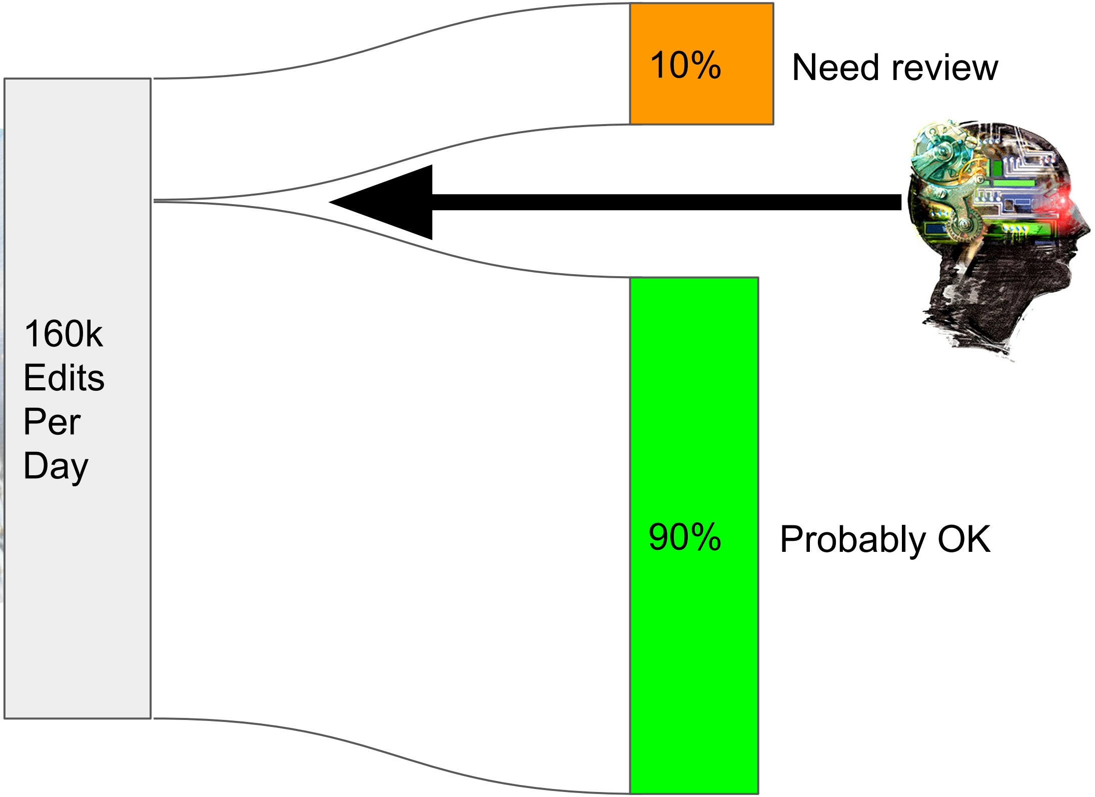
is_anon
chrs_added
chrs_removed
cust_comment
repeated_chrs
longest_token
badwords_added

The machine classifier



is_anon
chrs_added
chrs_removed
cust_comment
repeated_chrs
longest_token
badwords_added

Counter vandalism



160k Edits Per Day

10%

Need review

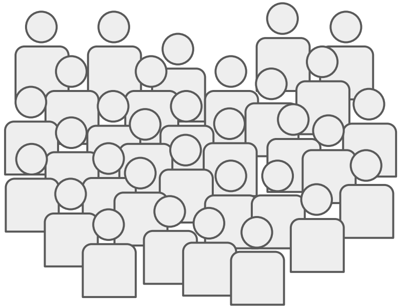
90%

Probably OK

Without ORES: Reviewing 160k edits per day...

267 Hours

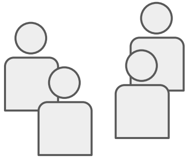
(33 people * 8 hours)



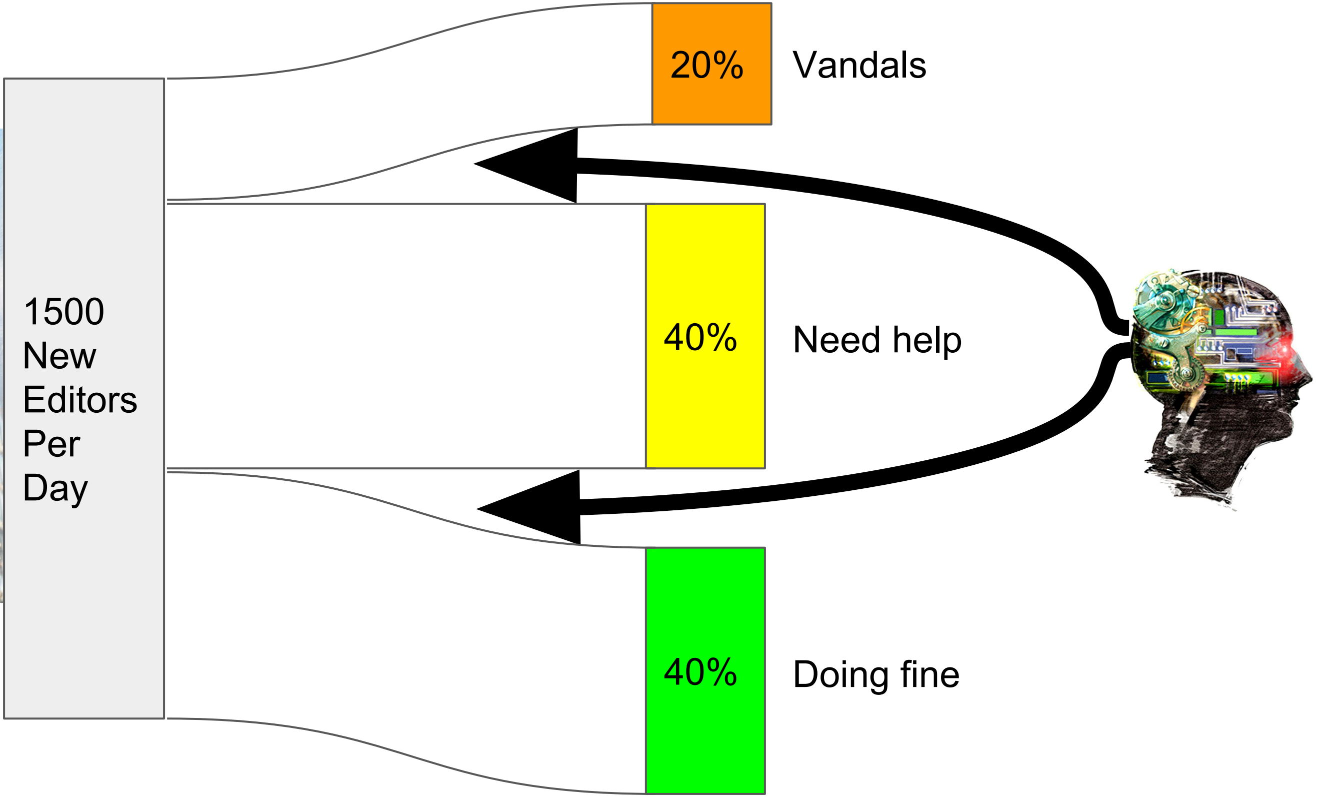
With ORES: Reviewing 16k edits per day…

27 Hours

(4 people * 8 hours)



Newcomer socialization



1500 New Editors Per Day

20% Vandals

40% Need help

40% Doing fine

* New editors = “Newly registered users who have saved at least one edit”

Future work:

Make estimates for the amount of time spent socializing a newcomer…

Some day, ORES will make everything easier...

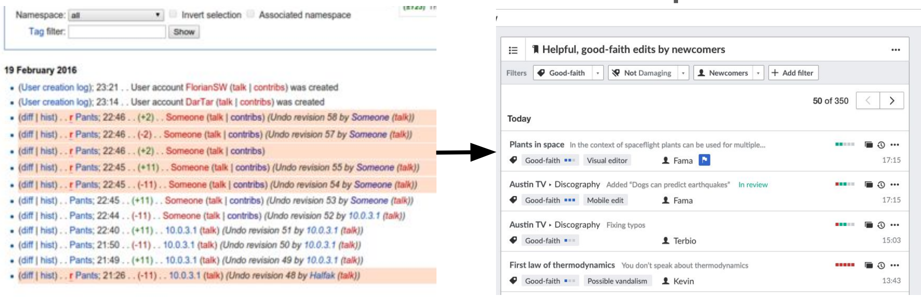
https://commons.wikimedia.org/wiki/File:P%C3%A4ij%C3%A4nne_and_p%C3%A4ij%C3%A4tsalo.jpg (Lake Päijänne at Sysmä, Finland.)

Getting there

- Production level web service @ https://ores.wikimedia.org

- Basic counter-vandalism support (ORES review tool) deployed on 6 wikis & active projects with the WMF Collaboration Team and with tool Developers

- 20 wikis and 4 prediction models
  - reverted?
  - good-faith?
  - damaging?
  - wp10 assessment?



Namespace: all □ Invert selection □ Associated namespace

Tag filter: Show

19 February 2016

- (User creation log); 23:21 . . User account FlorianSW (talk | contribs) was created
- (User creation log); 23:14 . . User account DarTar (talk | contribs) was created
- (diff | hist) . . r Pants; 22:46 . . (+2) . . Someone (talk | contribs) *(Undo revision 58 by Someone (talk))*
- (diff | hist) . . r Pants; 22:46 . . (-2) . . Someone (talk | contribs) *(Undo revision 57 by Someone (talk))*
- (diff | hist) . . r Pants; 22:46 . . (+2) . . Someone (talk | contribs)
- (diff | hist) . . r Pants; 22:45 . . (+11) . . Someone (talk | contribs) *(Undo revision 55 by Someone (talk))*
- (diff | hist) . . r Pants; 22:45 . . (-11) . . Someone (talk | contribs) *(Undo revision 54 by Someone (talk))*
- (diff | hist) . . Pants; 22:45 . . (+11) . . Someone (talk | contribs) *(Undo revision 53 by Someone (talk))*
- (diff | hist) . . Pants; 22:44 . . (-11) . . Someone (talk | contribs) *(Undo revision 52 by 10.0.3.1 (talk))*
- (diff | hist) . . Pants; 22:40 . . (+11) . . 10.0.3.1 (talk) *(Undo revision 51 by 10.0.3.1 (talk))*
- (diff | hist) . . Pants; 21:50 . . (-11) . . 10.0.3.1 (talk) *(Undo revision 50 by 10.0.3.1 (talk))*
- (diff | hist) . . Pants; 21:49 . . (+11) . . 10.0.3.1 (talk) *(Undo revision 49 by 10.0.3.1 (talk))*
- (diff | hist) . . r Pants; 21:26 . . (-11) . . 10.0.3.1 (talk) *(Undo revision 48 by Halfak (talk))*

Helpful, good-faith edits by newcomers

Filters: Good-faith | Not Damaging | Newcomers | + Add filter

50 of 350

Today

Plants in space — In the context of spaceflight plants can be used for multiple... — Good-faith — Visual editor — Fama — 17:15

Austin TV ▸ Discography — Added "Dogs can predict earthquakes" — In review — Good-faith — Mobile edit — Fama — 17:15

Austin TV ▸ Discography — Fixing typos — Good-faith — Terbio — 15:03

First law of thermodynamics — You don't speak about thermodynamics — Good-faith — Possible vandalism — Kevin — 13:43

https://commons.wikimedia.org/wiki/File:PEO-monster.svg

Part 2: The social dangers of AI realized

“Subjective algorithms”

"algorithms, often aided by big data, now make decisions in subjective realms where there is **no right decision**, and no anchor with which to judge outcomes."

Tufekci, Z. (2015). Algorithms in our Midst: Information, Power and Choice when Software is Everywhere. CSCW (pp. 1918-1918). ACM.

https://commons.wikimedia.org/wiki/File:Zeynep_Tufekci_2_(5805394290).jpg

“Subjective algorithms”

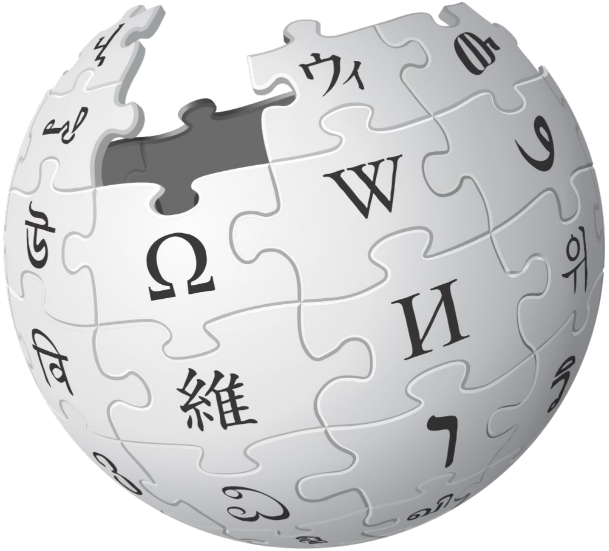
"algorithms, often aided by big data, now make decisions in subjective realms where there is **no right decision**, and no anchor with which to judge outcomes."

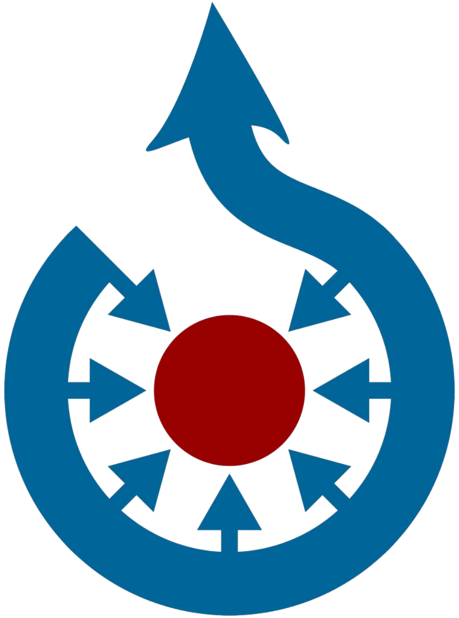
Tufekci, Z. (2015). Algorithms in our Midst: Information, Power and Choice when Software is Everywhere. CSCW (pp. 1918-1918). ACM.

What is good? relevant?

important? desirable? valuable?

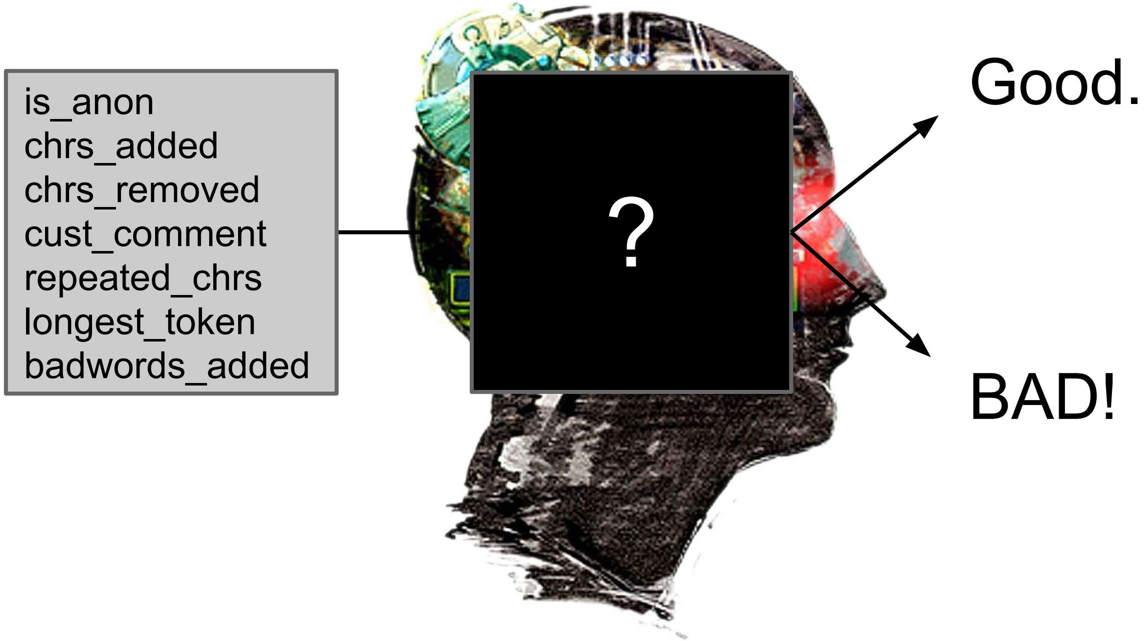
https://commons.wikimedia.org/wiki/File:Zeynep_Tufekci_2_(5805394290).jpg





Who is allowed to participate? Who gets labeled “bad-faith”?

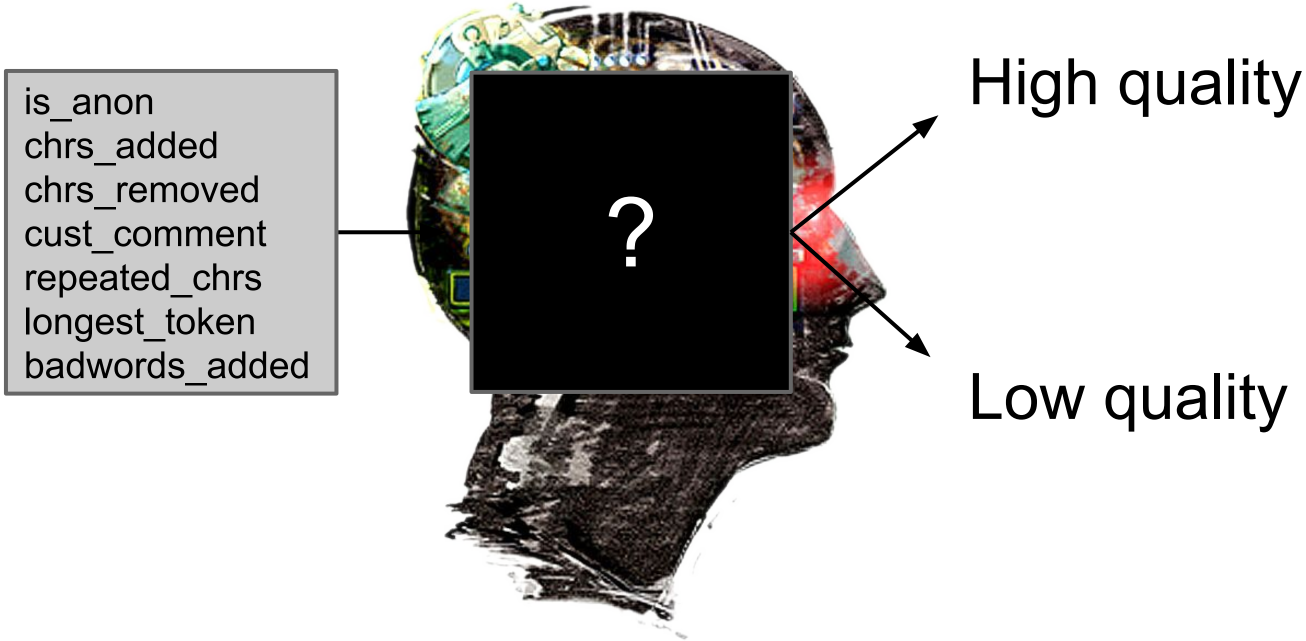
What types of contributions will be labeled “damaging”?



is_anon
chrs_added
chrs_removed
cust_comment
repeated_chrs
longest_token
badwords_added

Good.

BAD!

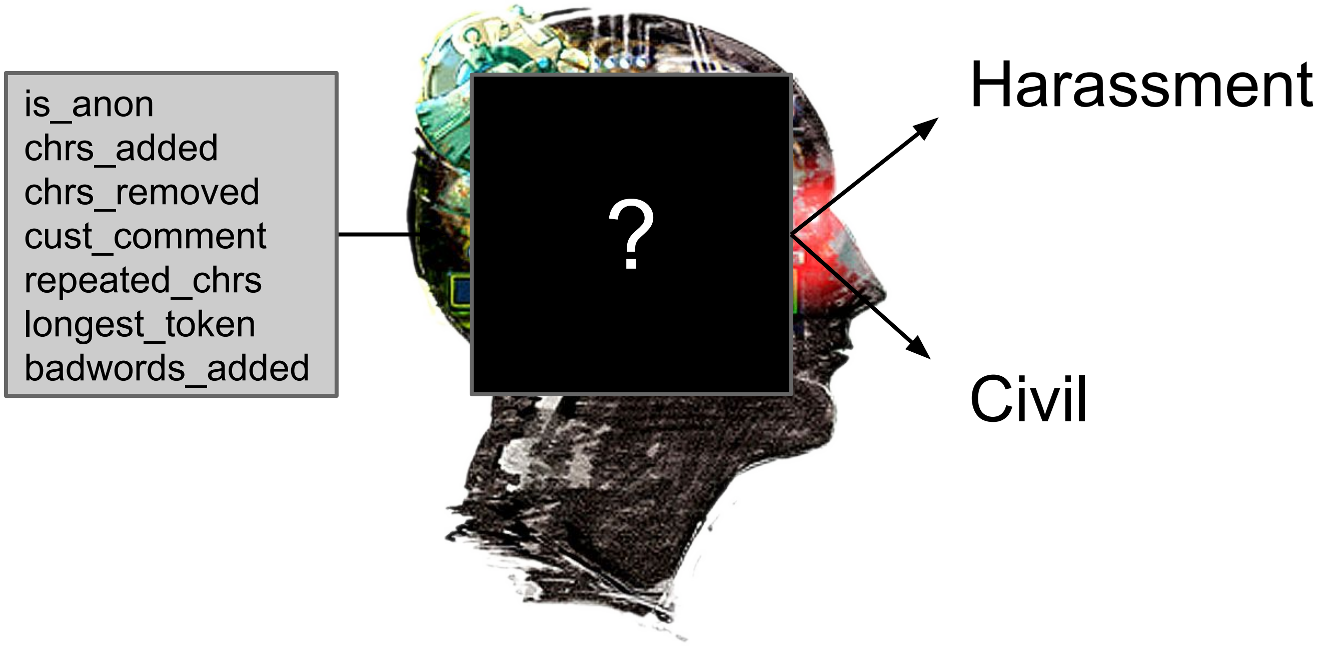


is_anon
chrs_added
chrs_removed
cust_comment
repeated_chrs
longest_token
badwords_added

?

High quality

Low quality



is_anon
chrs_added
chrs_removed
cust_comment
repeated_chrs
longest_token
badwords_added

?

Harassment

Civil

Please exercise *extreme caution* to avoid encoding racism or other biases into an AI scheme. [...] Wnt (talk) 12:58, 20 February 2015 (UTC)

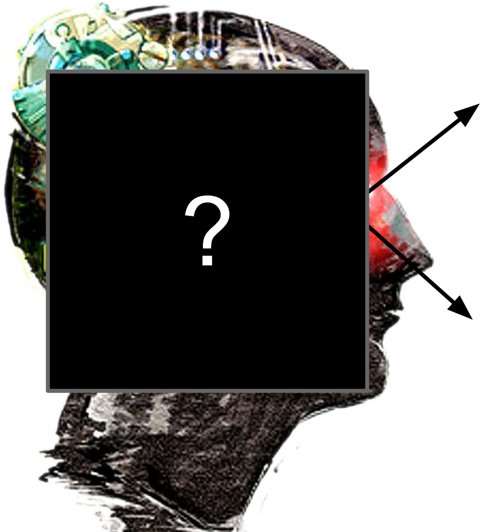
From Wikipedia:Wikipedia_Signpost/2015-02-18/Special_report

Please exercise *extreme caution* to avoid encoding racism or other biases into an AI scheme. [...] Wnt (talk) 12:58, 20 February 2015 (UTC)

From Wikipedia:Wikipedia_Signpost/2015-02-18/Special_report

Please exercise *extreme caution* to avoid encoding racism or other biases into an AI scheme. [...] Wnt (talk) 12:58, 20 February 2015 (UTC)

From Wikipedia:Wikipedia_Signpost/2015-02-18/Special_report

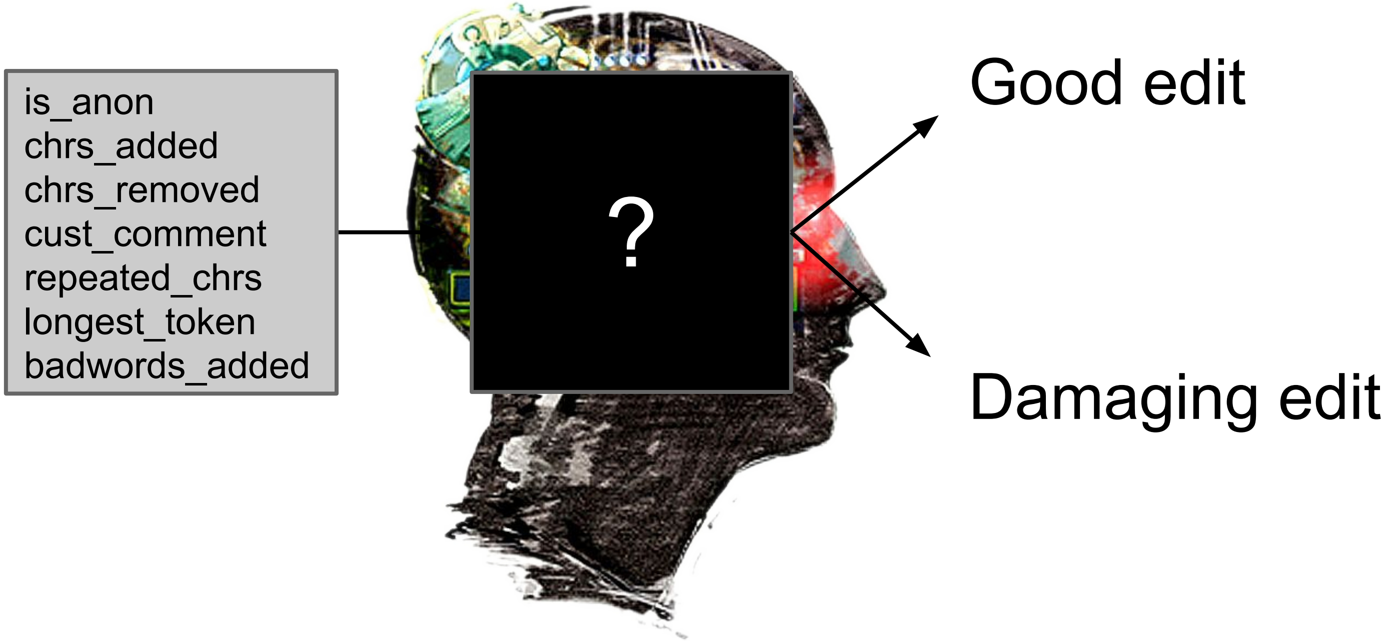


Harassment

and messages from South Africans

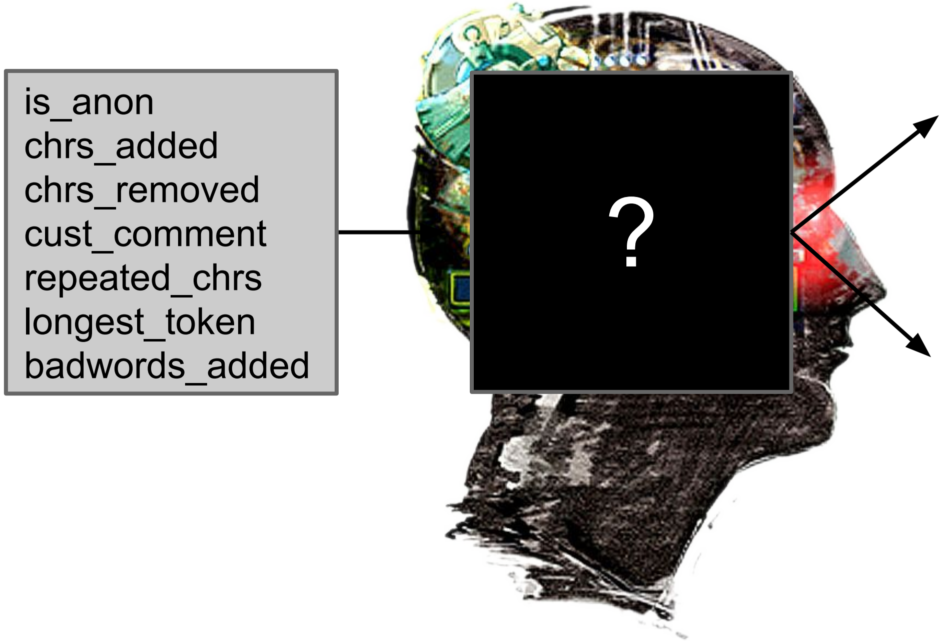
Civil

Two stories



is_anon
chrs_added
chrs_removed
cust_comment
repeated_chrs
longest_token
badwords_added

Two stories



is_anon
chrs_added
chrs_removed
cust_comment
repeated_chrs
longest_token
badwords_added

Good edit

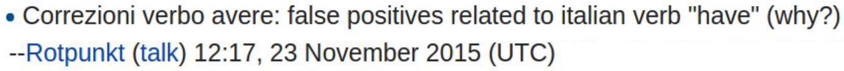
Damaging edit

- The Italian word “ha”
- Anonymous editors

The Italian “ha”

Literally: Not a laughing matter

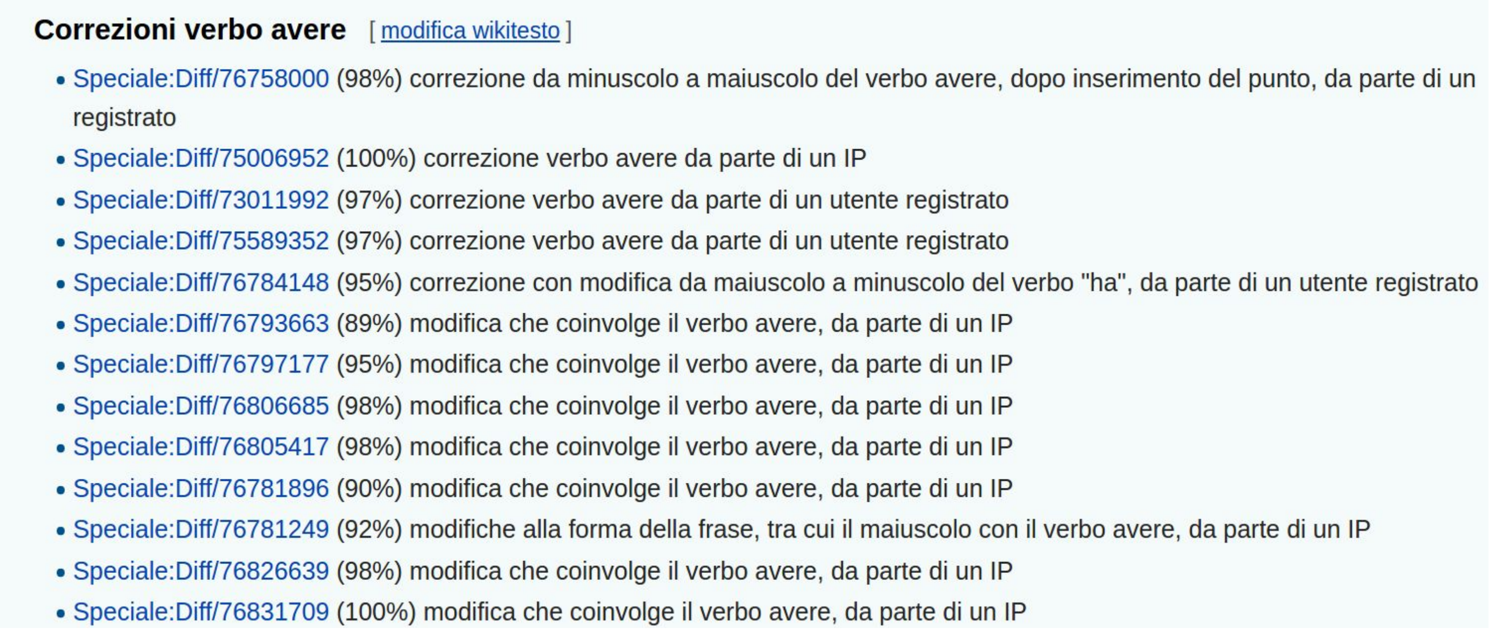
:m:Talk:ORES#Checklist for itwiki setup.



- Correzioni verbo avere: false positives related to italian verb "have" (why?)

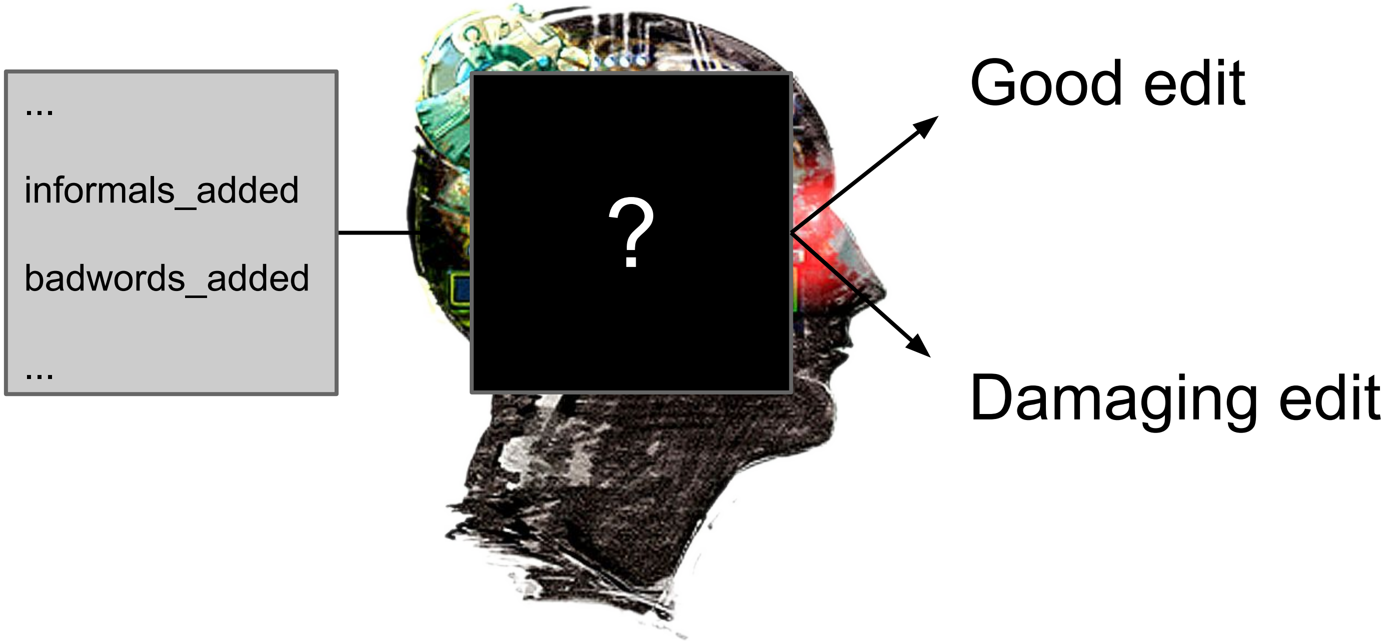
--Rotpunkt (talk) 12:17, 23 November 2015 (UTC)

:it:Progetto:Patrolling/ORES



Correzioni verbo avere [modifica wikitesto]

- Speciale:Diff/76758000 (98%) correzione da minuscolo a maiuscolo del verbo avere, dopo inserimento del punto, da parte di un registrato
- Speciale:Diff/75006952 (100%) correzione verbo avere da parte di un IP
- Speciale:Diff/73011992 (97%) correzione verbo avere da parte di un utente registrato
- Speciale:Diff/75589352 (97%) correzione verbo avere da parte di un utente registrato
- Speciale:Diff/76784148 (95%) correzione con modifica da maiuscolo a minuscolo del verbo "ha", da parte di un utente registrato
- Speciale:Diff/76793663 (89%) modifica che coinvolge il verbo avere, da parte di un IP
- Speciale:Diff/76797177 (95%) modifica che coinvolge il verbo avere, da parte di un IP
- Speciale:Diff/76806685 (98%) modifica che coinvolge il verbo avere, da parte di un IP
- Speciale:Diff/76805417 (98%) modifica che coinvolge il verbo avere, da parte di un IP
- Speciale:Diff/76781896 (90%) modifica che coinvolge il verbo avere, da parte di un IP
- Speciale:Diff/76781249 (92%) modifiche alla forma della frase, tra cui il maiuscolo con il verbo avere, da parte di un IP
- Speciale:Diff/76826639 (98%) modifica che coinvolge il verbo avere, da parte di un IP
- Speciale:Diff/76831709 (100%) modifica che coinvolge il verbo avere, da parte di un IP



...

informals_added

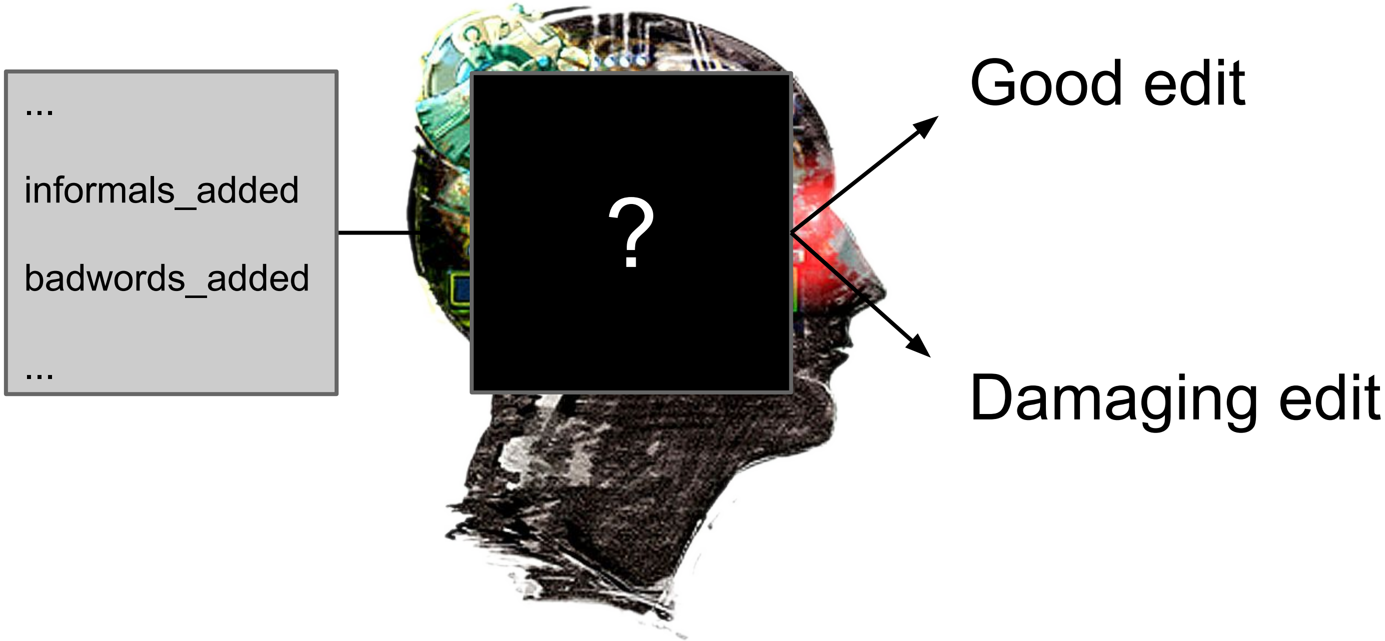
badwords_added

...

?

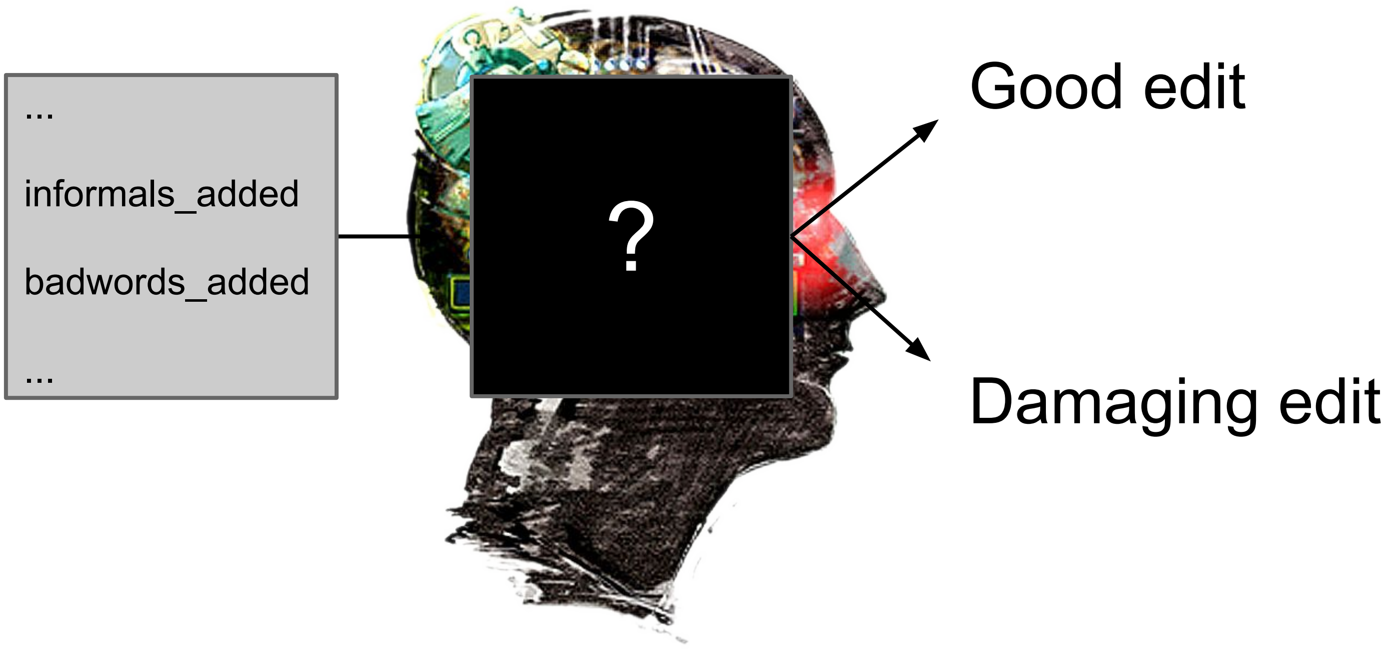
Good edit

Damaging edit



Badwords: Curse words, racial slurs and other offensive terminology

Informals: Casual speak that would be welcome on a talk page, but not within an article.



...

informals_added

badwords_added

...

?

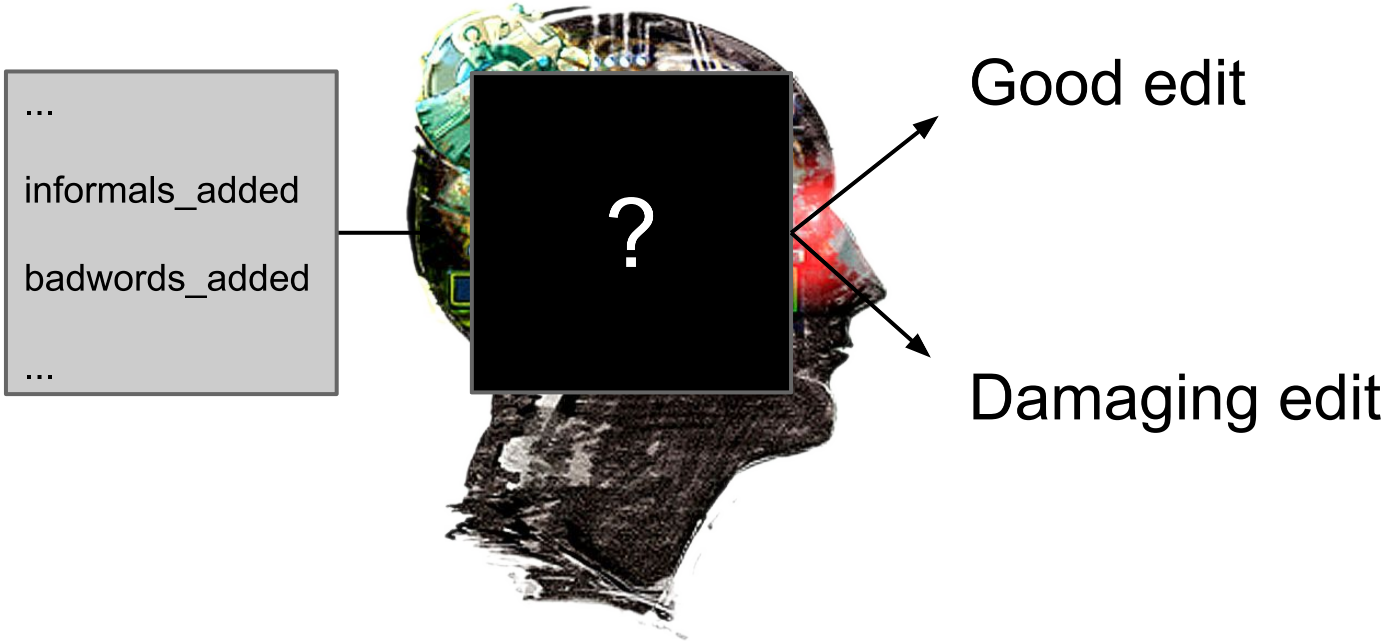
Good edit

Damaging edit

Badwords: Curse words, racial slurs and other offensive terminology

Informals: Casual speak that would be welcome on a talk page, but not within an article.

For example "hello" or "hahaha".

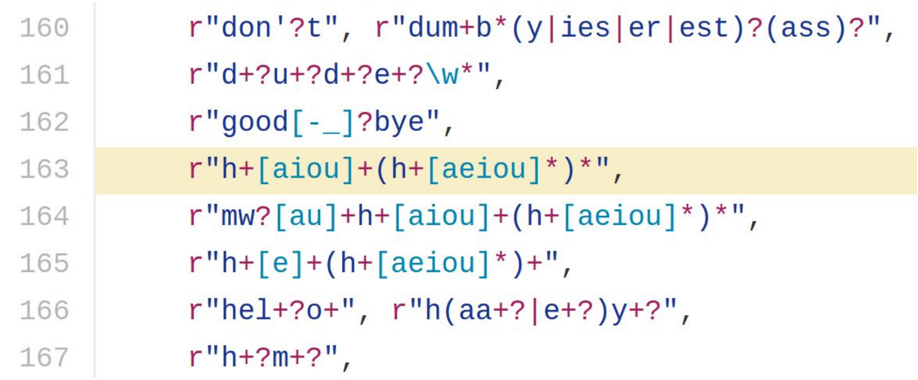


Badwords: Curse words, racial slurs and other offensive terminology

Informals: Casual speak that would be welcome on a talk page, but not within an article.

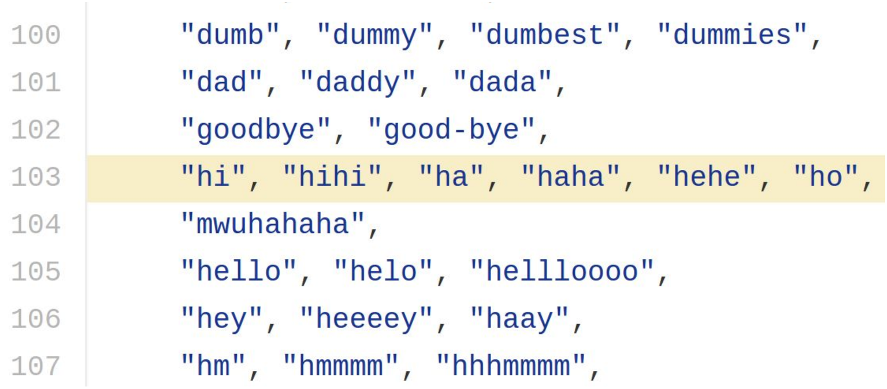
For example "hello" or "hahaha".

revscoring/languages/english.py#L163



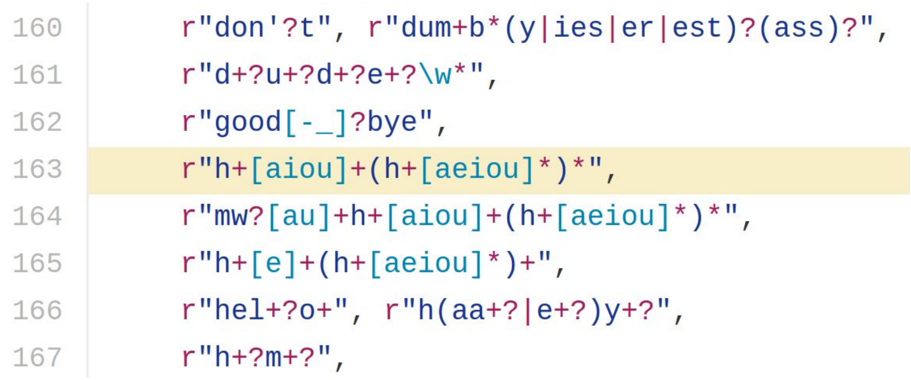
```
160         r"don'?t", r"dum+b*(y|ies|er|est)?(ass)?",
161         r"d+?u+?d+?e+?\w*",
162         r"good[-_]?bye",
163         r"h+[aiou]+(h+[aeiou]*)*",
164         r"mw?[au]+h+[aiou]+(h+[aeiou]*)*",
165         r"h+[e]+(h+[aeiou]*)+",
166         r"hel+?o+", r"h(aa+?|e+?)y+?",
167         r"h+?m+?",
```

revscoring/languages/tests/test_english.py#L103



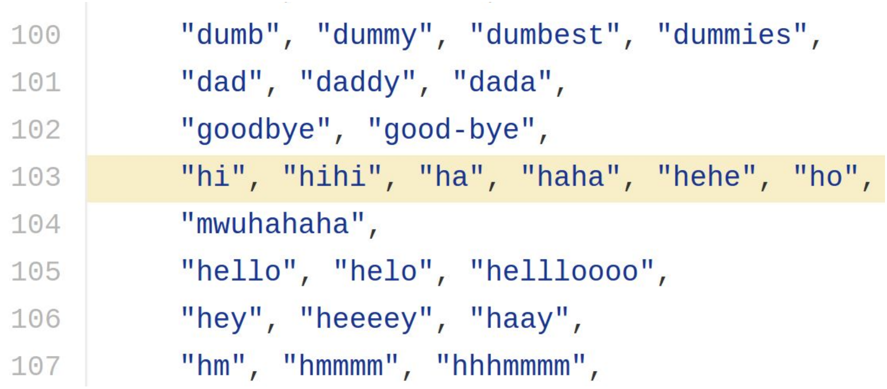
```
100         "dumb", "dummy", "dumbest", "dummies",
101         "dad", "daddy", "dada",
102         "goodbye", "good-bye",
103         "hi", "hihi", "ha", "haha", "hehe", "ho",
104         "mwuhahaha",
105         "hello", "helo", "hellloooo",
106         "hey", "heeeey", "haay",
107         "hm", "hmmmm", "hhhmmmm",
```

revscoring/languages/english.py#L163



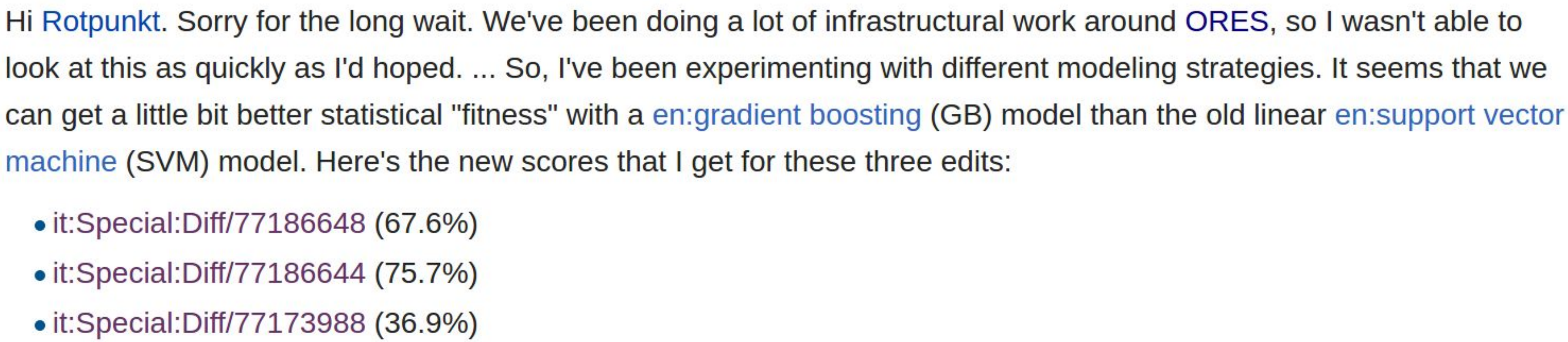
```
160        r"don'?t", r"dum+b*(y|ies|er|est)?(ass)?",
161        r"d+?u+?d+?e+?\w*",
162        r"good[-_]?bye",
163        r"h+[aiou]+(h+[aeiou]*)*",
164        r"mw?[au]+h+[aiou]+(h+[aeiou]*)*",
165        r"h+[e]+(h+[aeiou]*)+",
166        r"hel+?o+", r"h(aa+?|e+?)y+?",
167        r"h+?m+?",
```

revscoring/languages/tests/test_english.py#L103



```
100        "dumb", "dummy", "dumbest", "dummies",
101        "dad", "daddy", "dada",
102        "goodbye", "good-bye",
103        "hi", "hihi", "ha", "haha", "hehe", "ho",
104        "mwuhahaha",
105        "hello", "helo", "hellloooo",
106        "hey", "heeeey", "haay",
107        "hm", "hmmmm", "hhhmmmm",
```

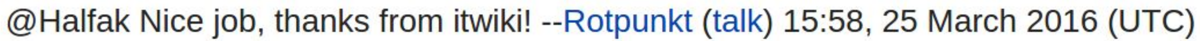
"Ha" is laughing in English, but "Ha" is **not** laughing in Italian!



Hi Rotpunkt. Sorry for the long wait. We've been doing a lot of infrastructural work around ORES, so I wasn't able to look at this as quickly as I'd hoped. ... So, I've been experimenting with different modeling strategies. It seems that we can get a little bit better statistical "fitness" with a en:gradient boosting (GB) model than the old linear en:support vector machine (SVM) model. Here's the new scores that I get for these three edits:

- it:Special:Diff/77186648 (67.6%)
- it:Special:Diff/77186644 (75.7%)
- it:Special:Diff/77173988 (36.9%)

… <snip> ...

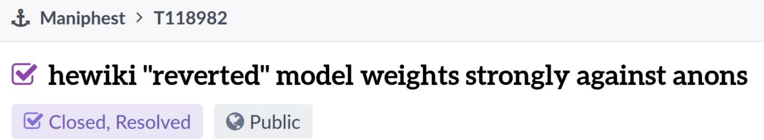


@Halfak Nice job, thanks from itwiki! --Rotpunkt (talk) 15:58, 25 March 2016 (UTC)

https://commons.wikimedia.org/wiki/File:Enduring_Freedom_120707-A-DL472-175.jpg Public Domain (US Gov)

Anonymous editors

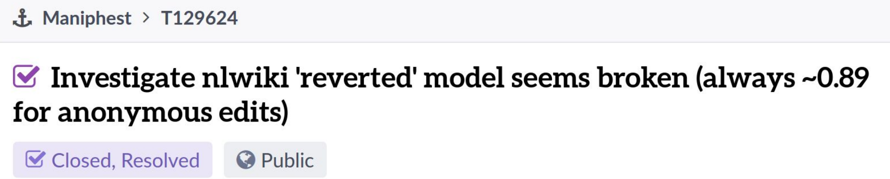
Anonymous editors



Manipheșt > T118982

☑ hewiki "reverted" model weights strongly against anons

☑ Closed, Resolved | Public



Maniphest > T129624

☑ Investigate nlwiki 'reverted' model seems broken (always ~0.89 for anonymous edits)

☑ Closed, Resolved | Public

Otherwise, anons seemed to dominate false-positive reports from every wiki

… maybe anons are really bad.

… maybe anons are really bad.

- Generally, anon edits are **twice** as likely to be vandalism

… maybe anons are really bad.

- Generally, anon edits are **twice** as likely to be vandalism
- **90% of anonymous edits are good**

https://ores.wmflabs.org/v2/scores/enwiki/damaging/642345235?feature.revision.user.**is_anon=false**

https://ores.wmflabs.org/v2/scores/enwiki/damaging/642345235?feature.revision.user.**is_anon=false**

```
{"prediction": false,
 "probability": {"false": 0.656,
                 "true": 0.344}}
```

https://ores.wmflabs.org/v2/scores/enwiki/damaging/642345235?feature.revision.user.**is_anon=false**

```
{"prediction": false,
 "probability": {"false": 0.656,
                 "true": 0.344}}
```

https://ores.wmflabs.org/v2/scores/enwiki/damaging/642345235?feature.revision.user.**is_anon=true**

```
{"prediction": false,
 "probability": {"false": 0.541,
                 "true": 0.459}}
```

https://ores.wmflabs.org/v2/scores/enwiki/damaging/642345235?feature.revision.user.**is_anon=false**

```
{"prediction": false,
 "probability": {"false": 0.656,
                 "true": 0.344}}
```

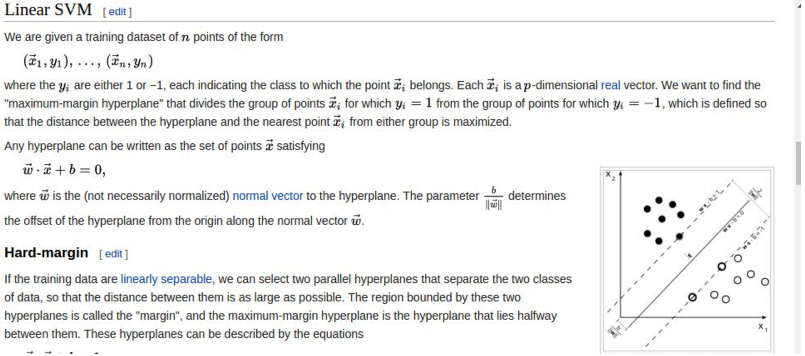
https://ores.wmflabs.org/v2/scores/enwiki/damaging/642345235?feature.revision.user.**is_anon=true**

```
{"prediction": false,
 "probability": {"false": 0.541,
                 "true": 0.459}}
```

Just by being “anon”, we score this edit 11.5% more likely to be damaging to the article.

Modeling strategies

https://en.wikipedia.org/wiki/Support_vector_machine#Linear_SVM



Linear SVM [edit]

We are given a training dataset of n points of the form

$$(\vec{x}_1, y_1), \ldots, (\vec{x}_n, y_n)$$

where the y_i are either 1 or −1, each indicating the class to which the point $\vec{x}_i$ belongs. Each $\vec{x}_i$ is a p-dimensional real vector. We want to find the "maximum-margin hyperplane" that divides the group of points $\vec{x}_i$ for which $y_i = 1$ from the group of points for which $y_i = -1$, which is defined so that the distance between the hyperplane and the nearest point $\vec{x}_i$ from either group is maximized.

Any hyperplane can be written as the set of points $\vec{x}$ satisfying

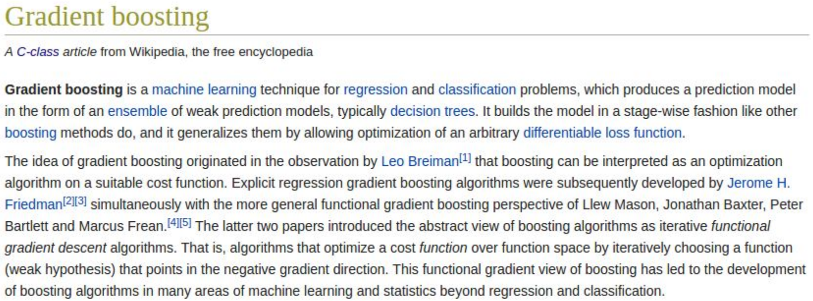
$$\vec{w} \cdot \vec{x} + b = 0,$$

where $\vec{w}$ is the (not necessarily normalized) normal vector to the hyperplane. The parameter $\frac{b}{\|\vec{w}\|}$ determines the offset of the hyperplane from the origin along the normal vector $\vec{w}$.

Hard-margin [edit]

If the training data are linearly separable, we can select two parallel hyperplanes that separate the two classes of data, so that the distance between them is as large as possible. The region bounded by these two hyperplanes is called the "margin", and the maximum-margin hyperplane is the hyperplane that lies halfway between them. These hyperplanes can be described by the equations

https://en.wikipedia.org/wiki/Gradient_boosting



Gradient boosting

A C-class article from Wikipedia, the free encyclopedia

Gradient boosting is a machine learning technique for regression and classification problems, which produces a prediction model in the form of an ensemble of weak prediction models, typically decision trees. It builds the model in a stage-wise fashion like other boosting methods do, and it generalizes them by allowing optimization of an arbitrary differentiable loss function.

The idea of gradient boosting originated in the observation by Leo Breiman[1] that boosting can be interpreted as an optimization algorithm on a suitable cost function. Explicit regression gradient boosting algorithms were subsequently developed by Jerome H. Friedman[2][3] simultaneously with the more general functional gradient boosting perspective of Llew Mason, Jonathan Baxter, Peter Bartlett and Marcus Frean.[4][5] The latter two papers introduced the abstract view of boosting algorithms as iterative *functional gradient descent* algorithms. That is, algorithms that optimize a cost *function* over function space by iteratively choosing a function (weak hypothesis) that points in the negative gradient direction. This functional gradient view of boosting has led to the development of boosting algorithms in many areas of machine learning and statistics beyond regression and classification.

User classes

Anon editor

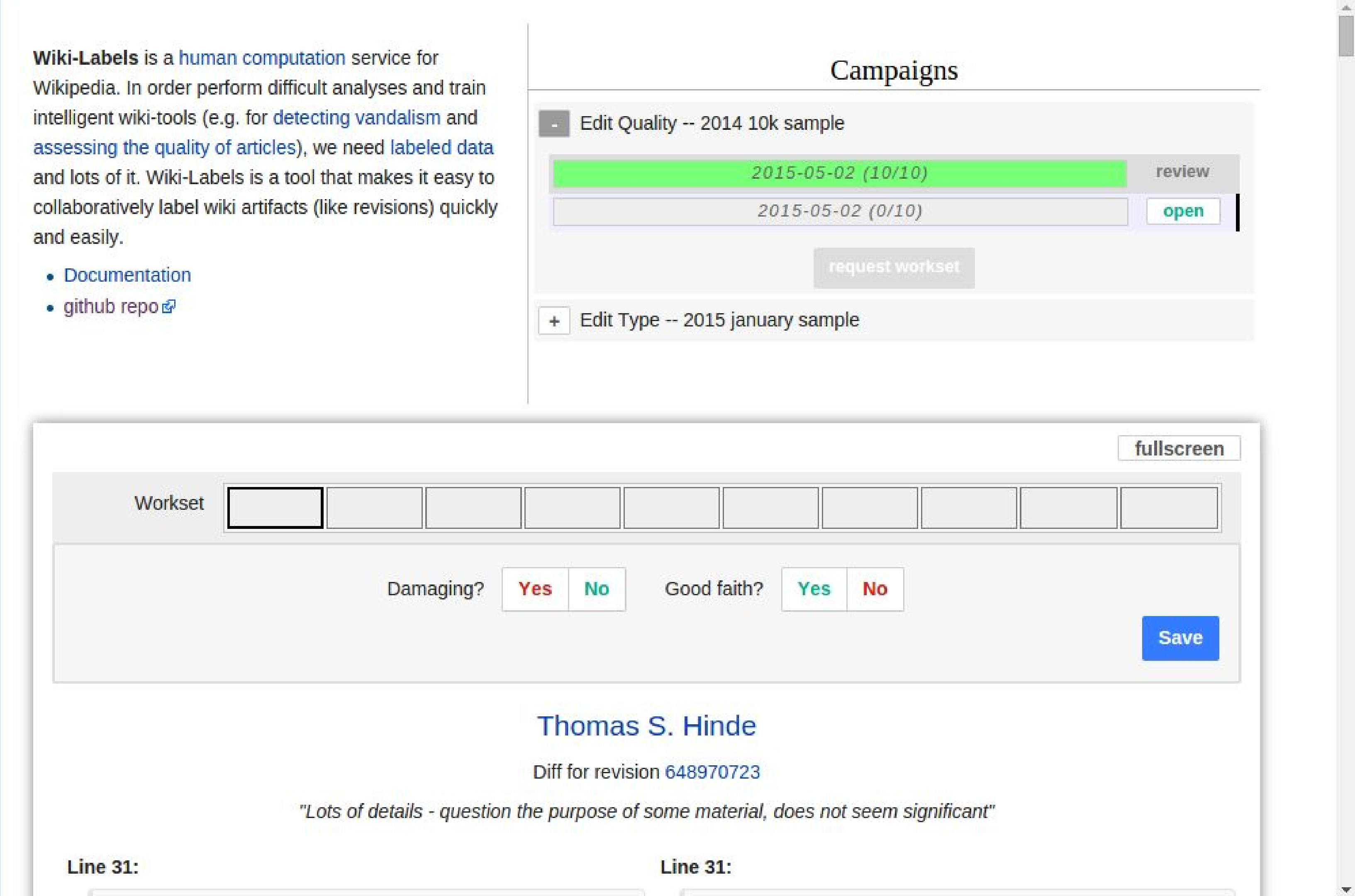
```
{"feature.revision.user.is_anon":  true,
 "feature...seconds_since_registration":  0,
 "feature.revision.user.has_advanced_rights":  false,
 "feature.revision.user.is_admin":  false,
 "feature.revision.user.is_bot":  false,
 "feature.revision.user.is_curator":  false}
```

New editor (2h since registration)

```
{"feature.revision.user.is_anon":  false,
 "feature...seconds_since_registration":  18000,
 "feature.revision.user.has_advanced_rights":  false,
 "feature.revision.user.is_admin":  false,
 "feature.revision.user.is_bot":  false,
 "feature.revision.user.is_curator":  false}
```

User:EpochFail (8 years since registration)

```
{"feature.revision.user.is_anon":  false,
 "feature...seconds_since_registration":  257995021,
 "feature.revision.user.has_advanced_rights":  false,
 "feature.revision.user.is_admin":  false,
 "feature.revision.user.is_bot":  false,
 "feature.revision.user.is_curator":  false}
```



Wiki-Labels is a human computation service for Wikipedia. In order perform difficult analyses and train intelligent wiki-tools (e.g. for detecting vandalism and assessing the quality of articles), we need labeled data and lots of it. Wiki-Labels is a tool that makes it easy to collaboratively label wiki artifacts (like revisions) quickly and easily.

- Documentation
- github repo

Campaigns

- Edit Quality -- 2014 10k sample

2015-05-02 (10/10)	review
2015-05-02 (0/10)	open

+ Edit Type -- 2015 january sample

Damaging? Yes No   Good faith? Yes No

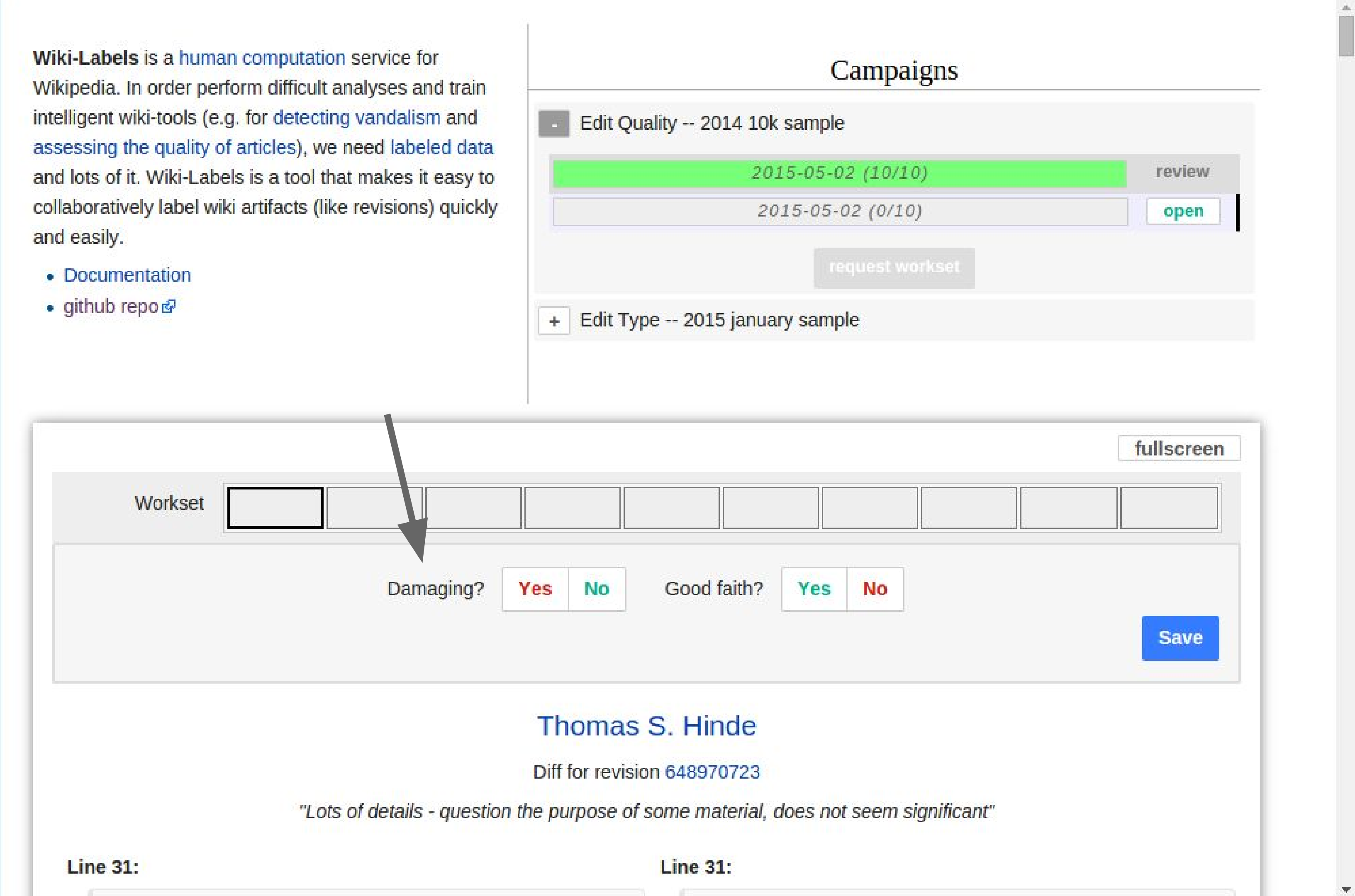
Thomas S. Hinde

Diff for revision 648970723

"Lots of details - question the purpose of some material, does not seem significant"

Line 31:

Line 31:



Wiki-Labels is a human computation service for Wikipedia. In order perform difficult analyses and train intelligent wiki-tools (e.g. for detecting vandalism and assessing the quality of articles), we need labeled data and lots of it. Wiki-Labels is a tool that makes it easy to collaboratively label wiki artifacts (like revisions) quickly and easily.

- Documentation
- github repo

Campaigns

- Edit Quality -- 2014 10k sample

2015-05-02 (10/10)	review
2015-05-02 (0/10)	open

+ Edit Type -- 2015 january sample

Damaging? Yes No Good faith? Yes No

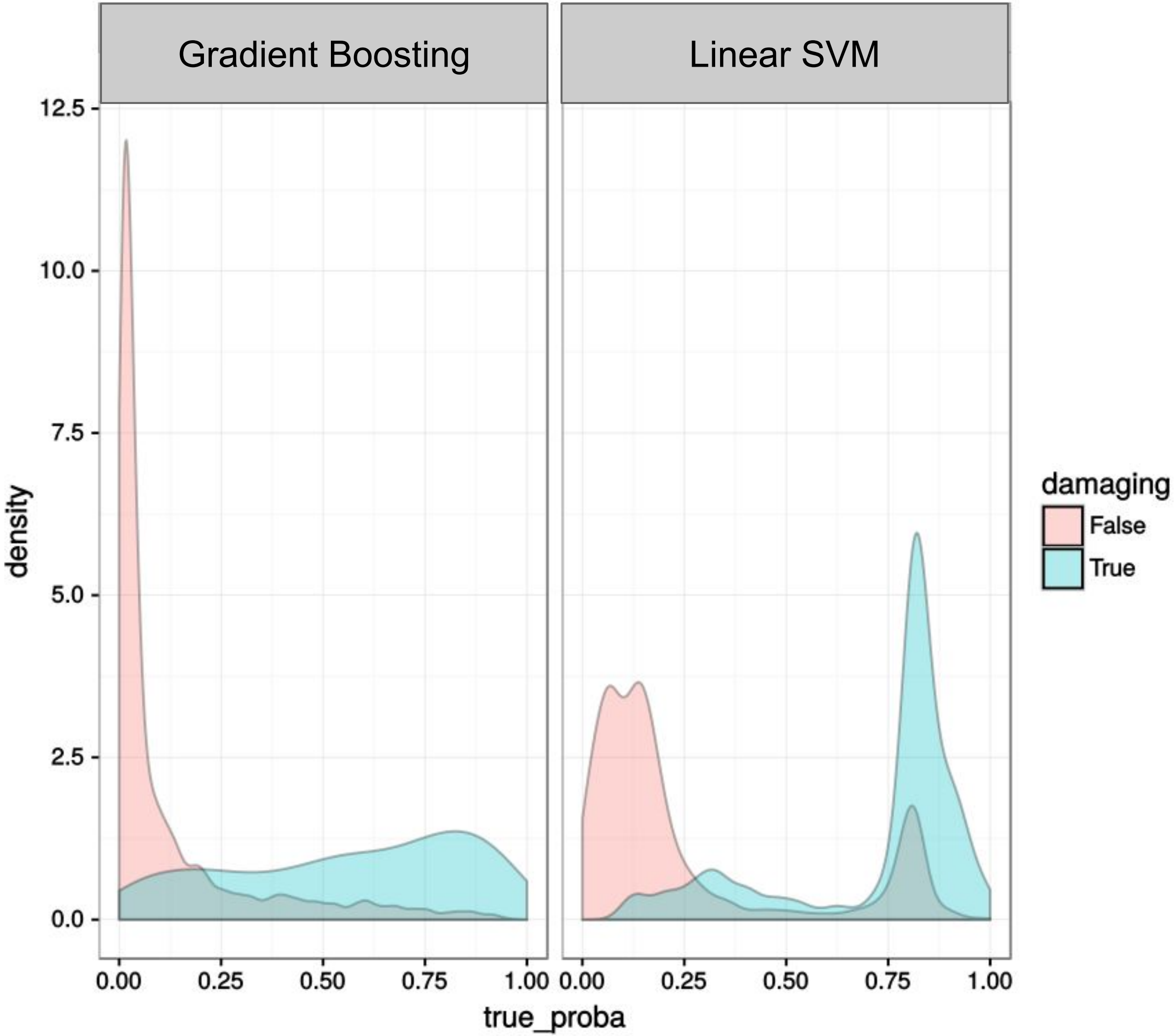
Thomas S. Hinde

Diff for revision 648970723

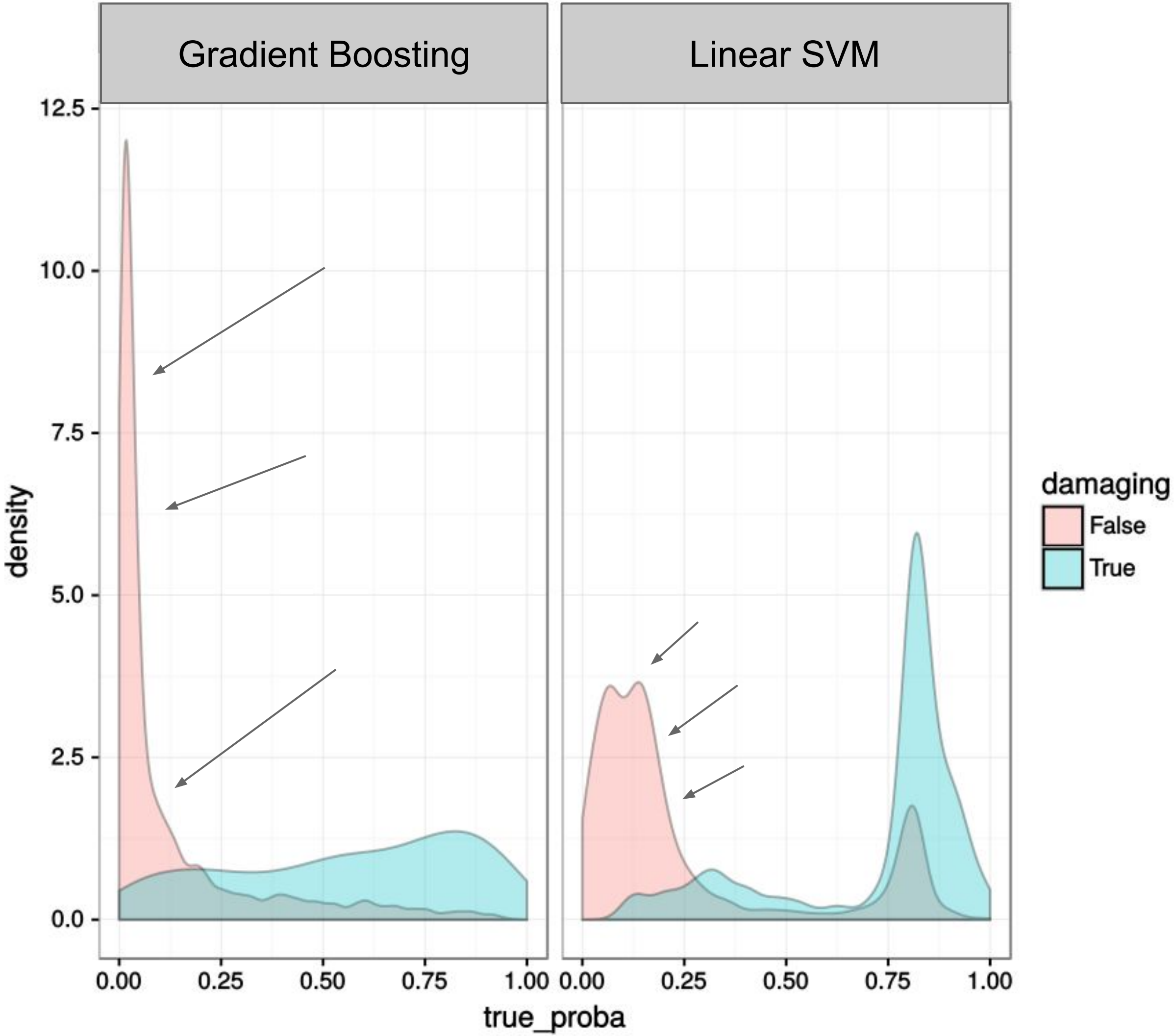
"Lots of details - question the purpose of some material, does not seem significant"

Line 31: **Line 31:**

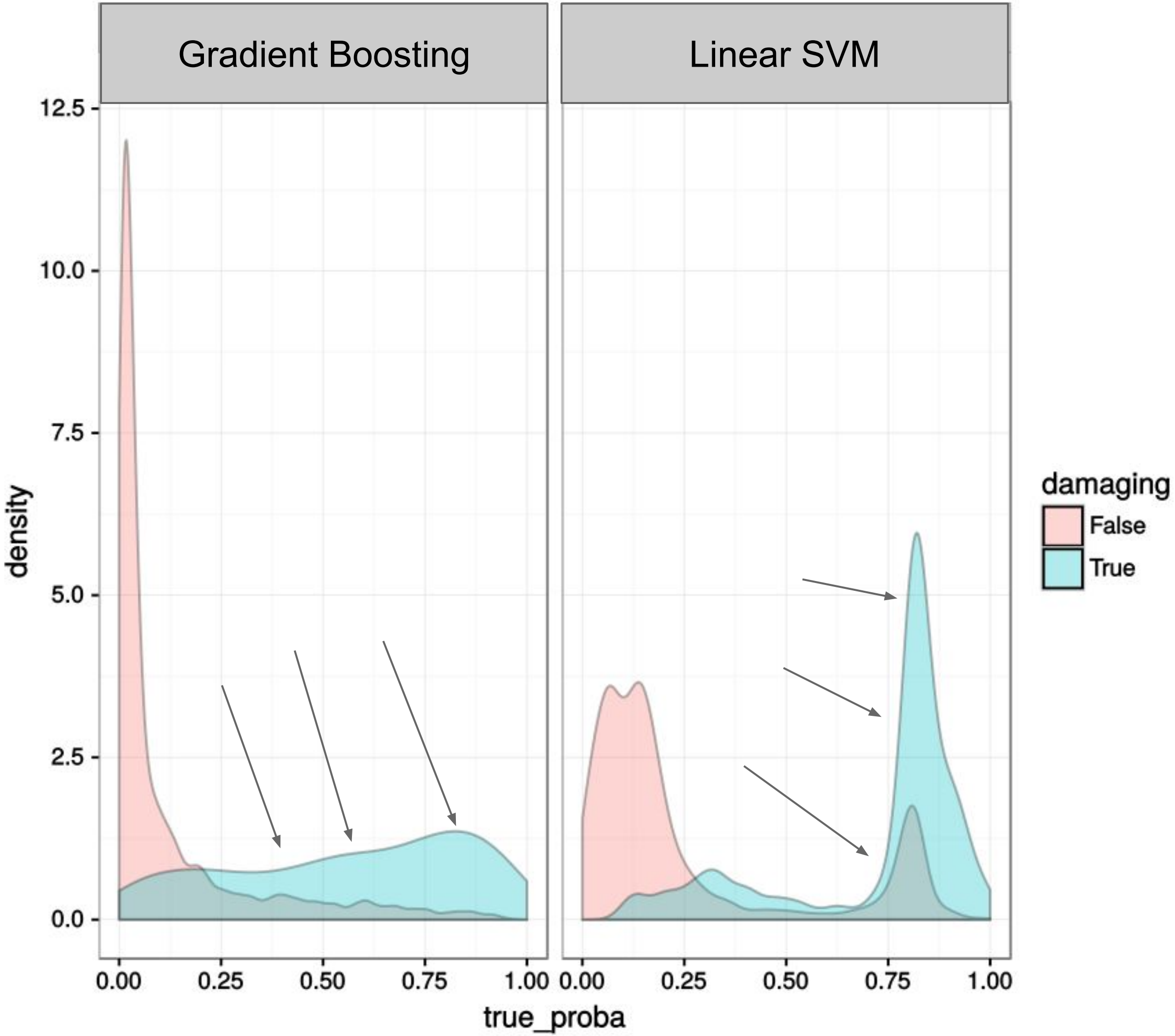
Natural user class



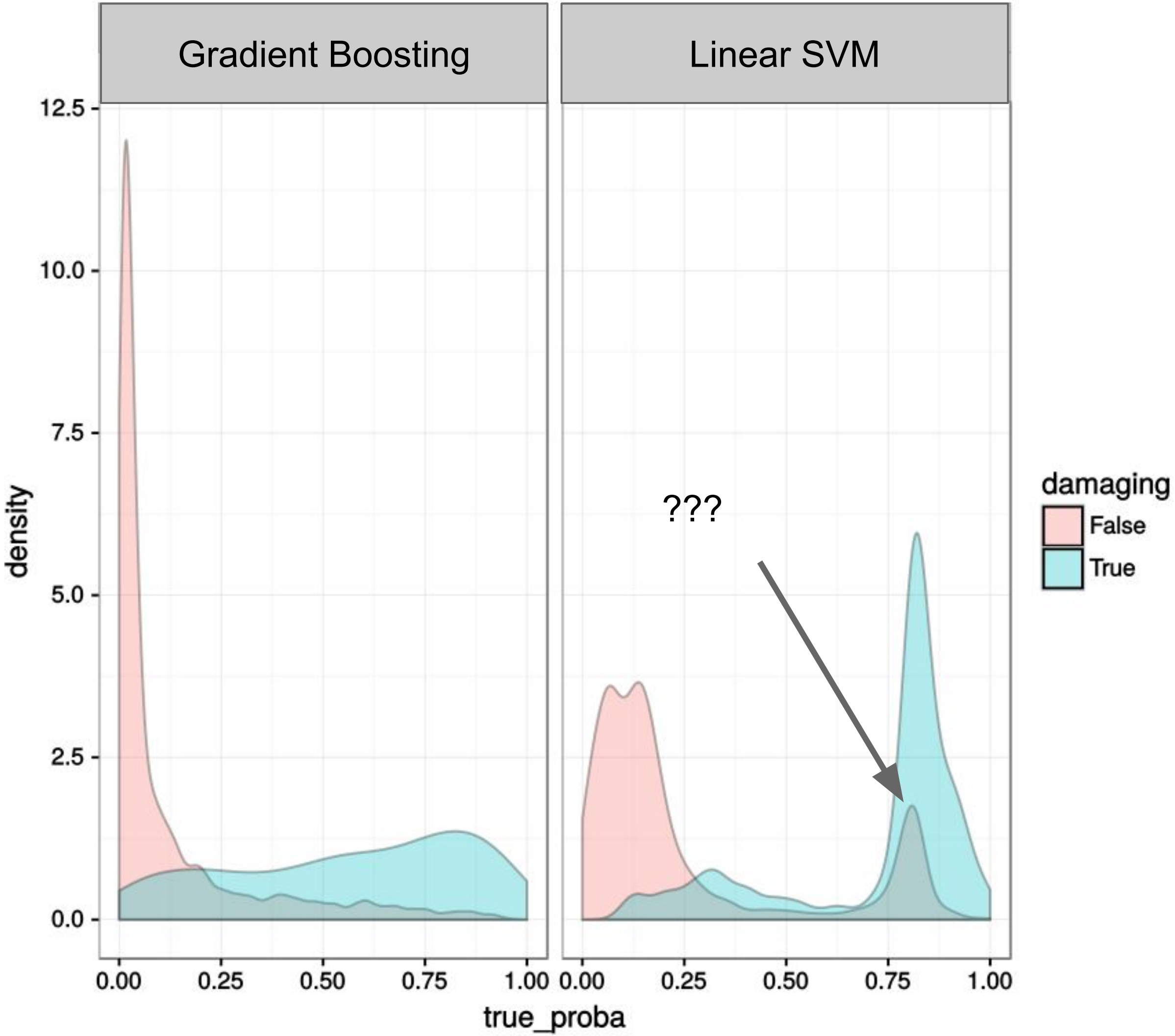
Natural user class



Natural user class



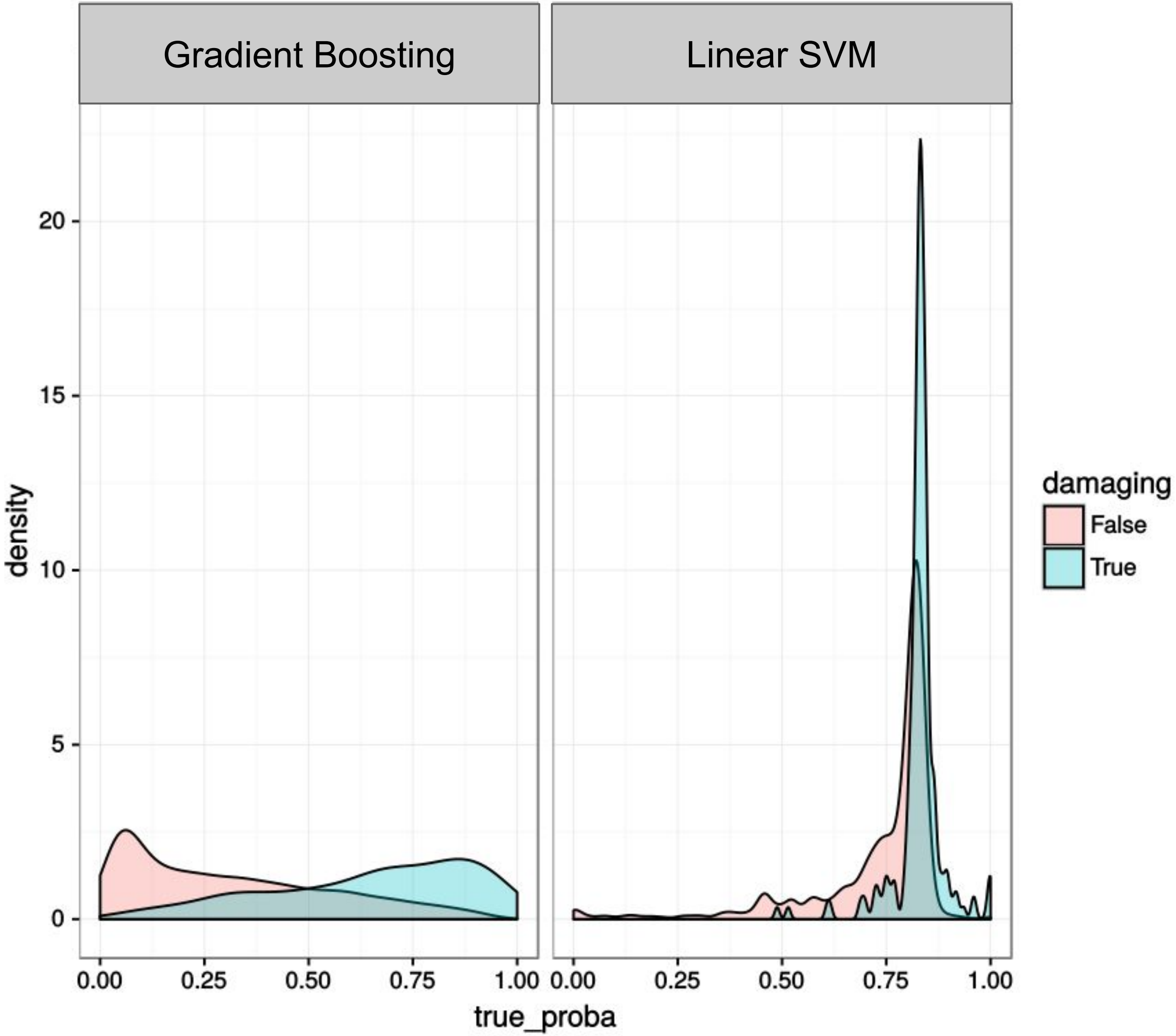
Natural user class



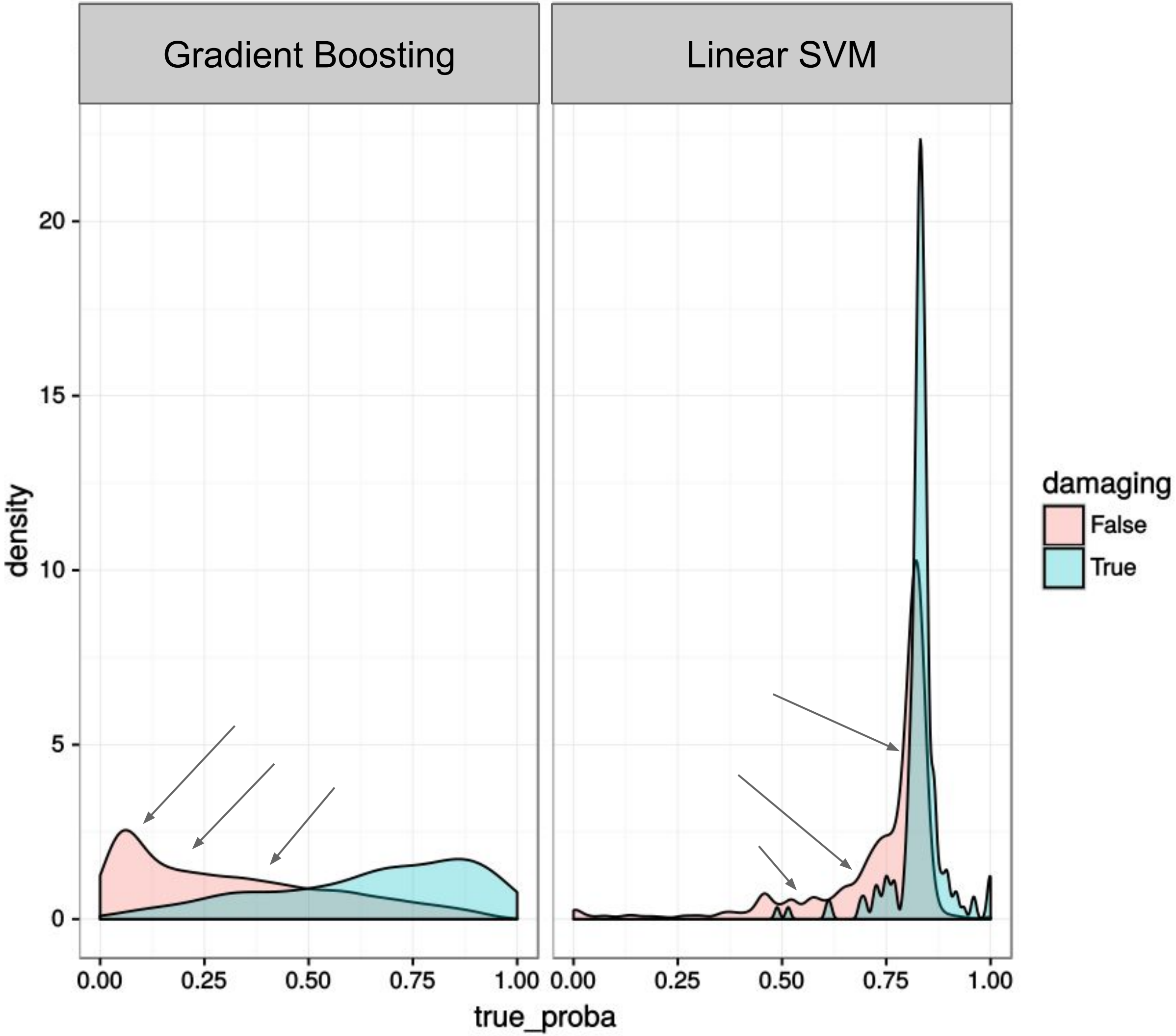
???

OK. What if every edit was anon?

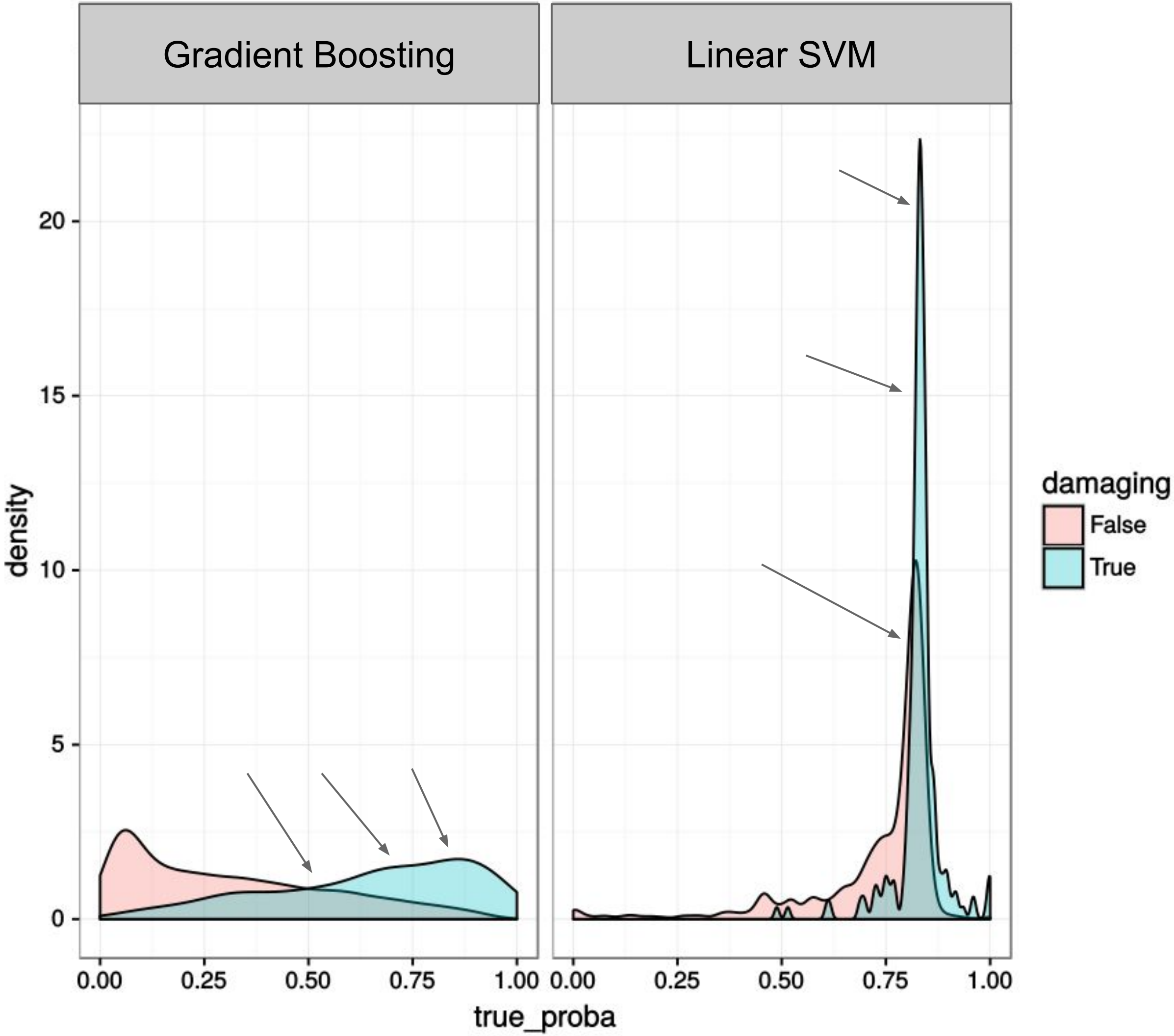
Anon user class



Anon user class

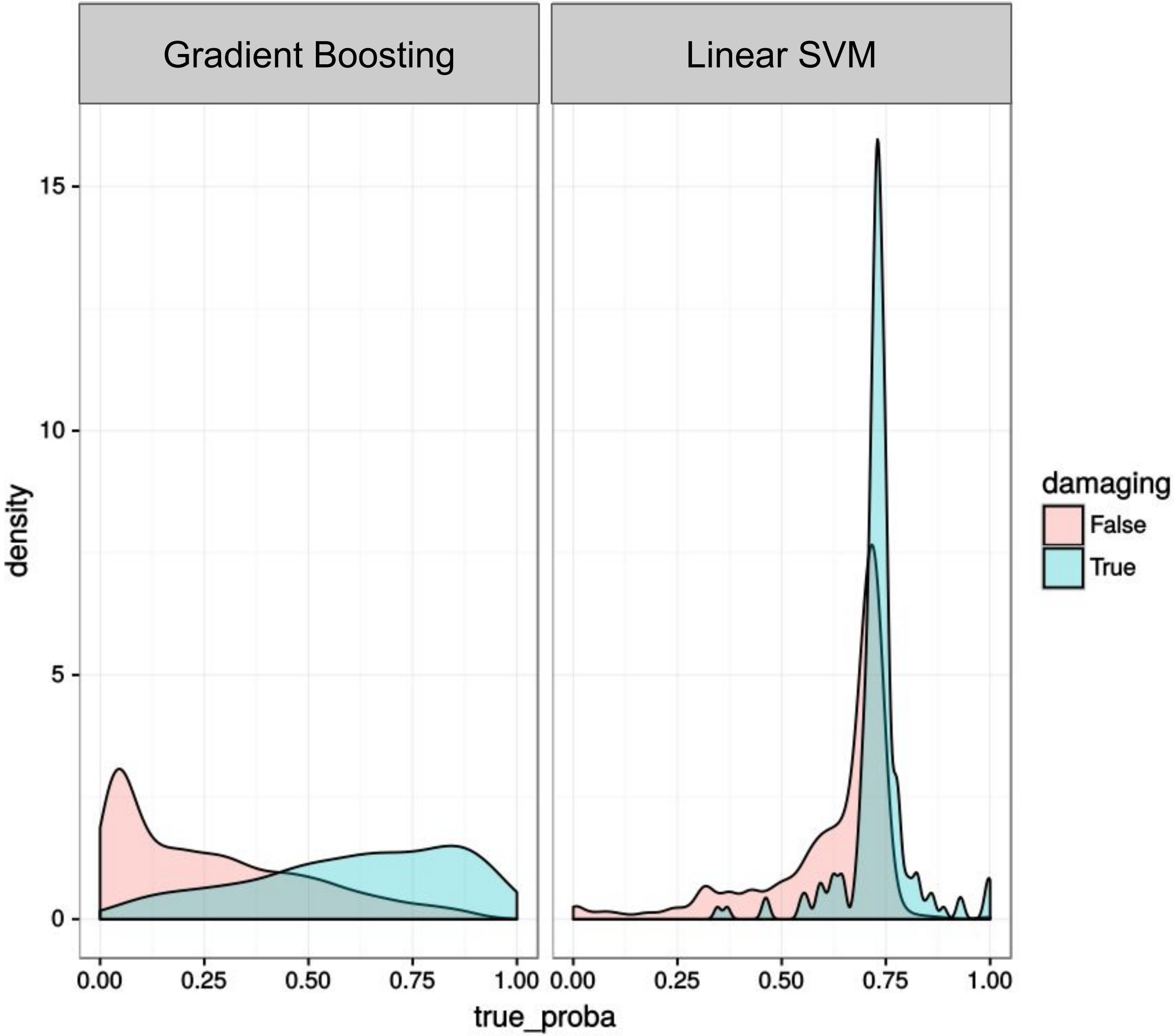


Anon user class

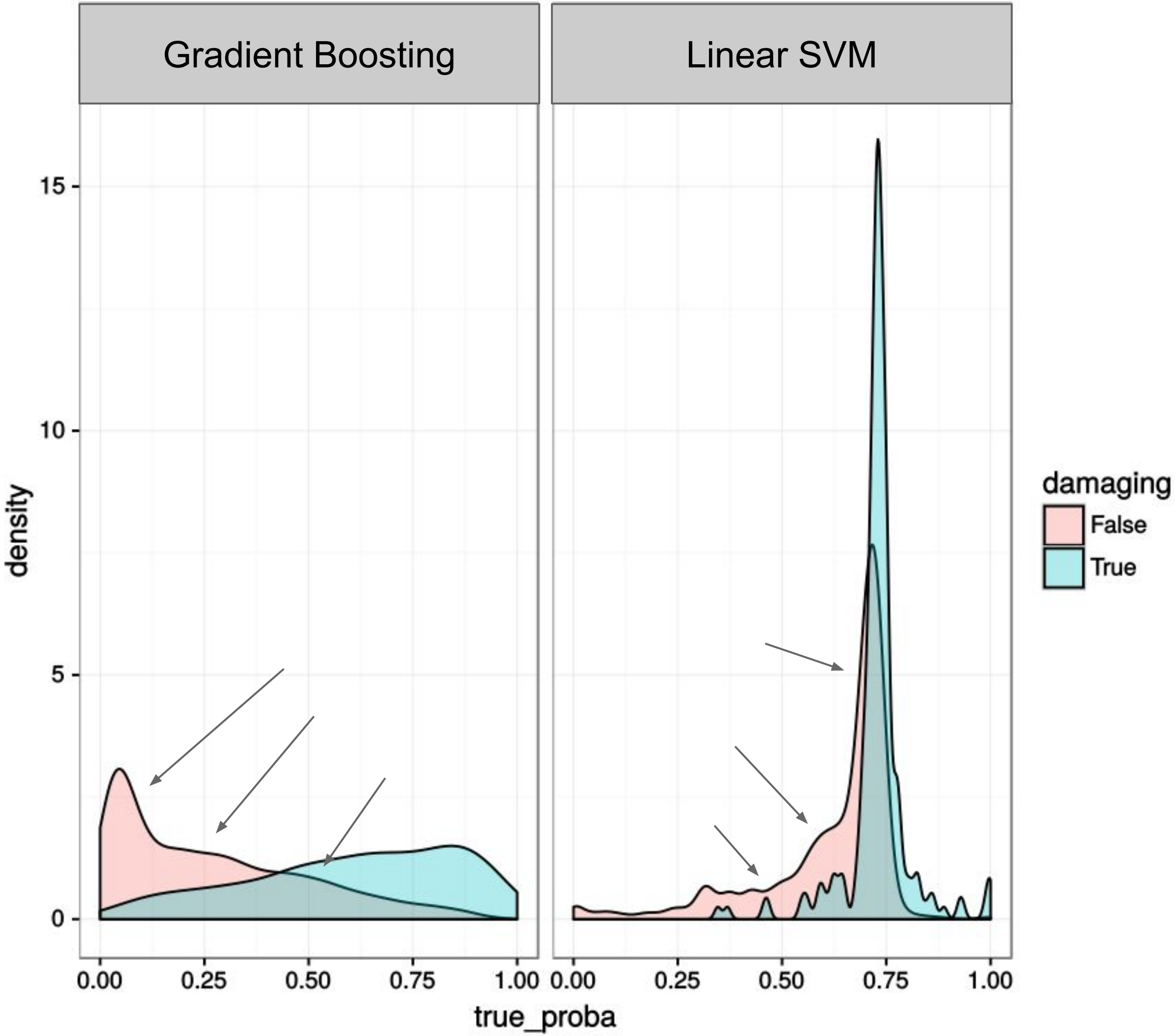


What if every edit were from a newcomer?

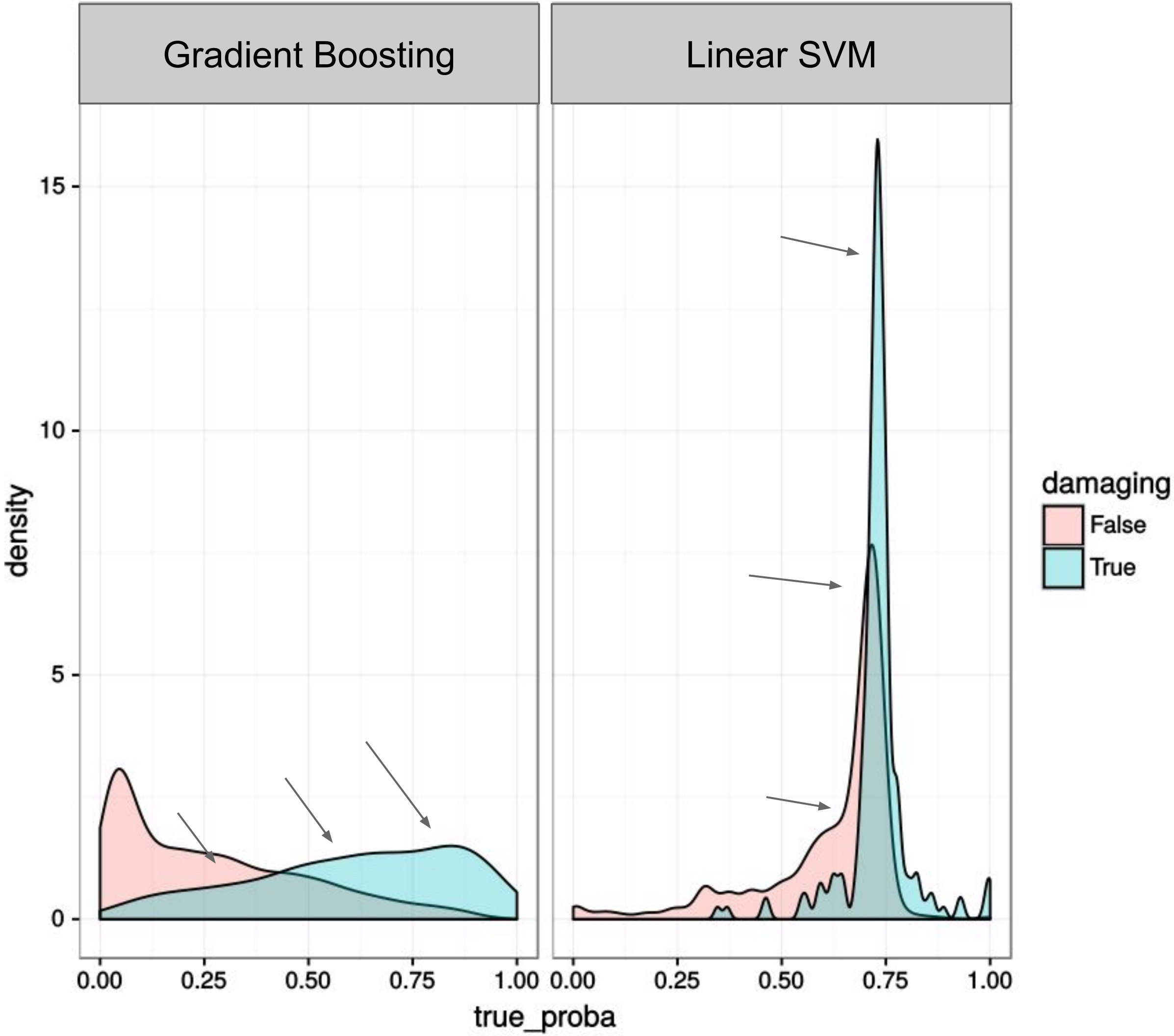
Newbie user class



Newbie user class

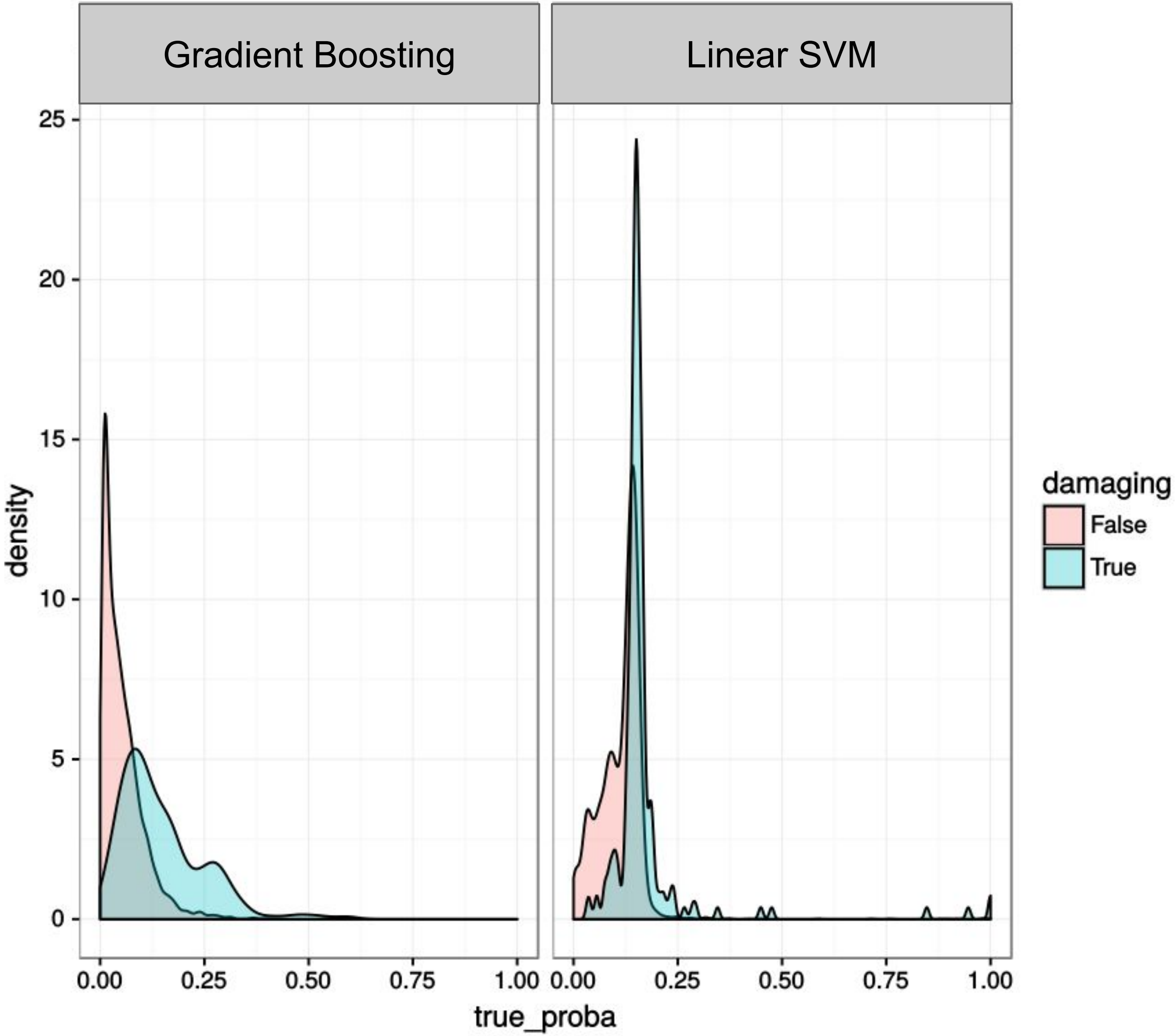


Newbie user class

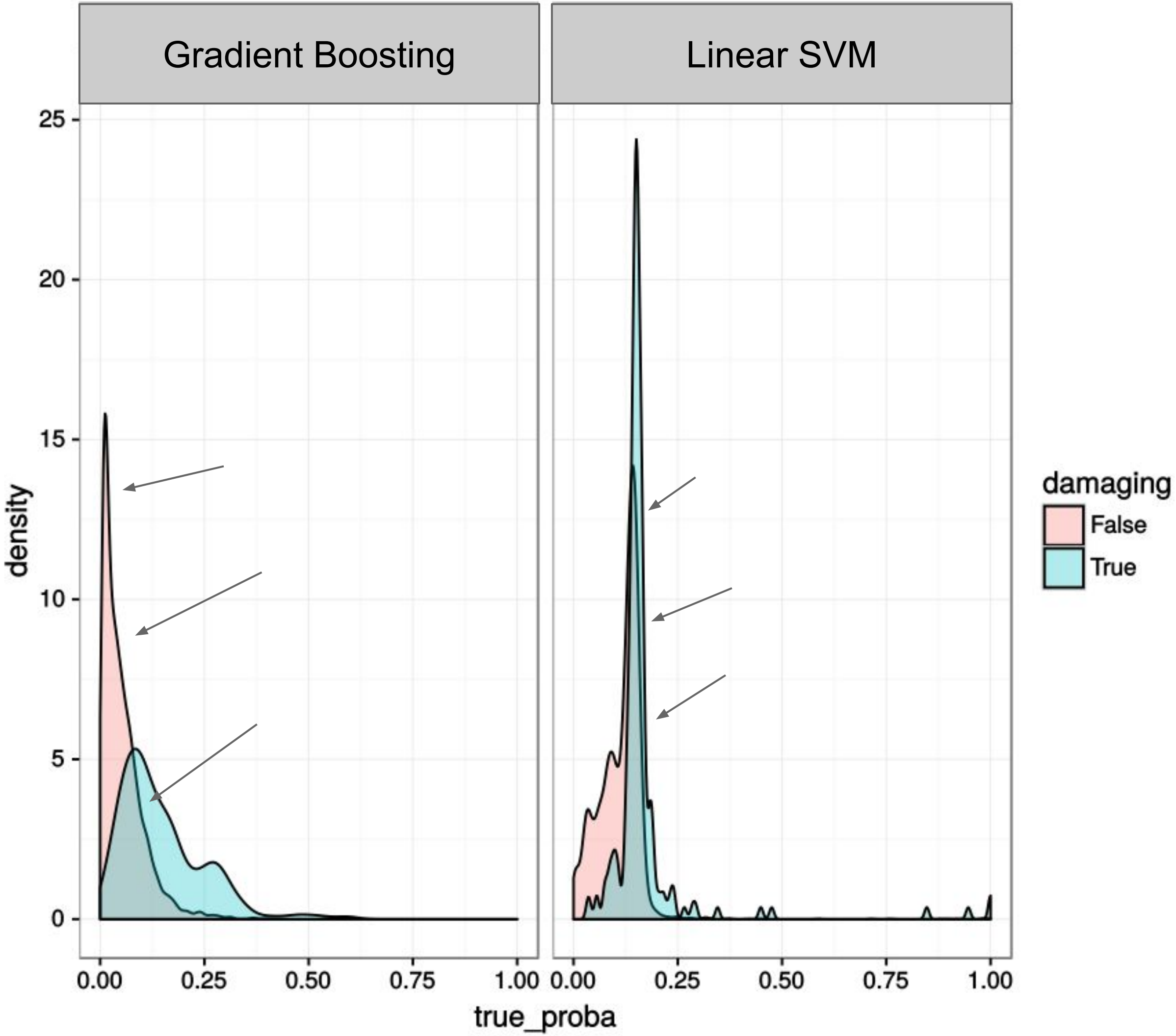


What if I (EpochFail) saved all the edits?

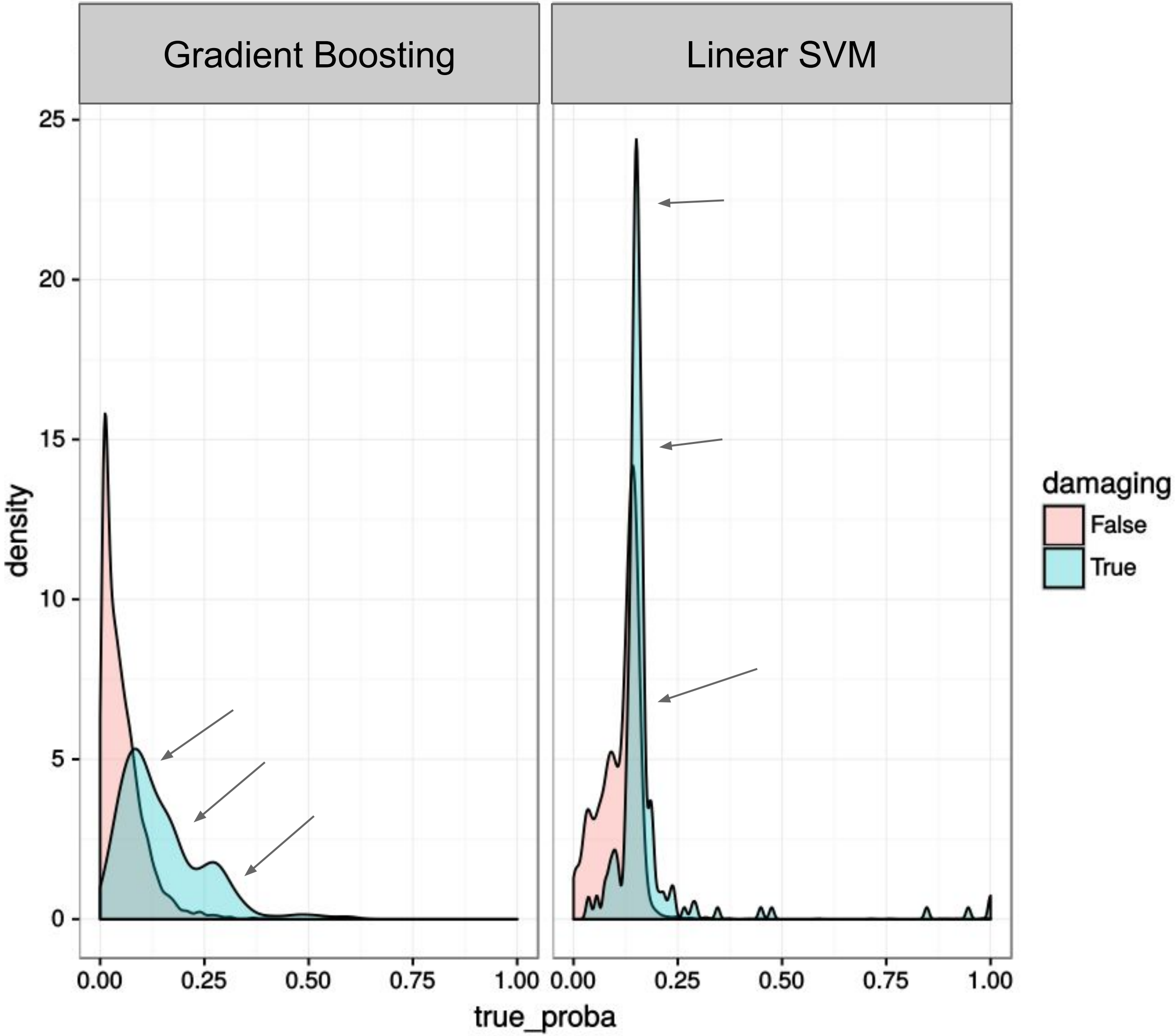
EpochFail user class



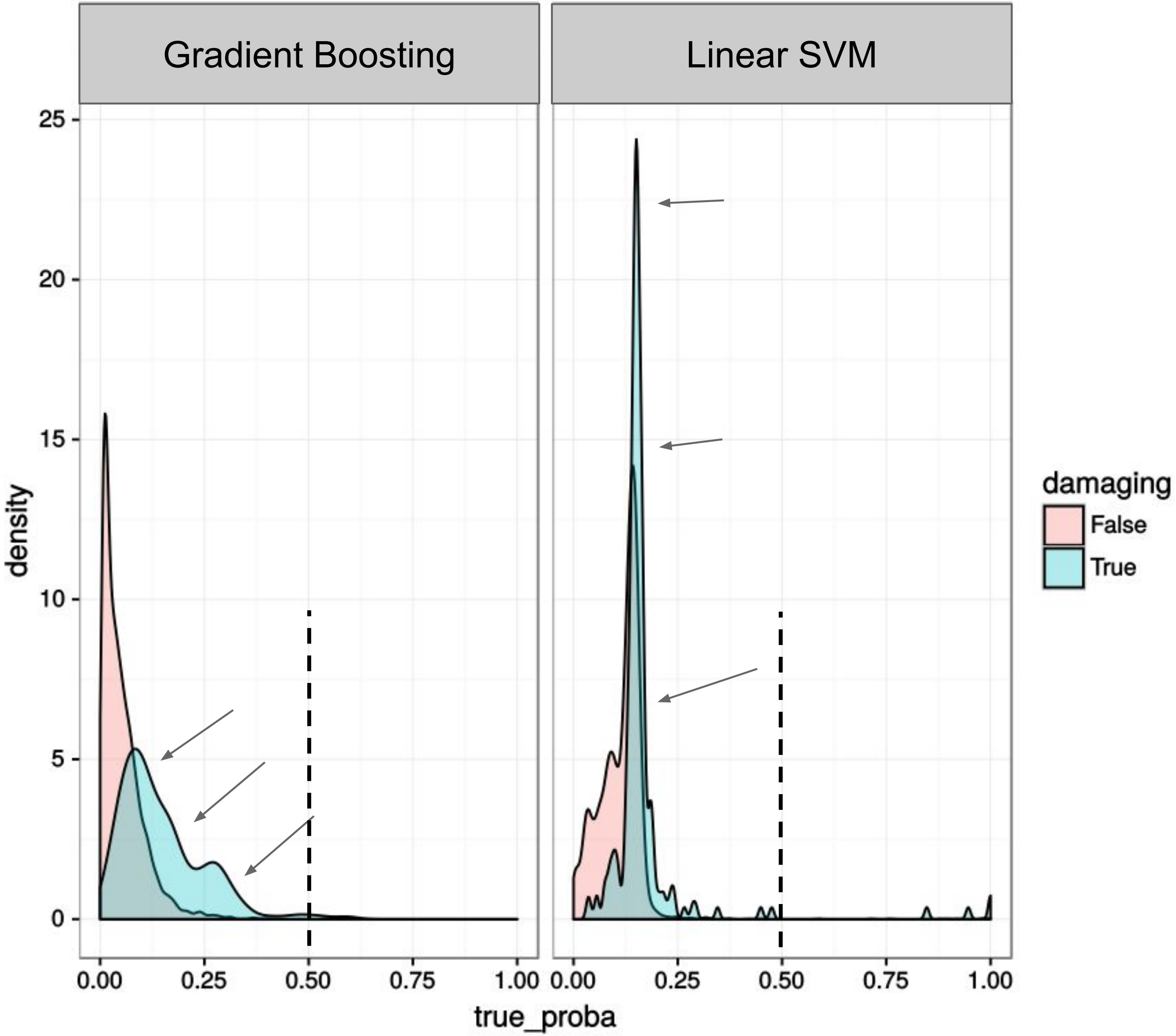
EpochFail user class



EpochFail user class



EpochFail user class



Mwahahahaha!

Dec. 2015:

- Gradient boosting deployed

Dec. 2015:

- Gradient boosting deployed

Still needs work:

- Bias against anons/newcomers lessened, but not gone
- New sources of signal
  - HashingVectorization
  - Probabilistic Context-free Grammars

Part 3: A call to action

How about we ask the humans?

Evaluation of AI

Evaluation of AI

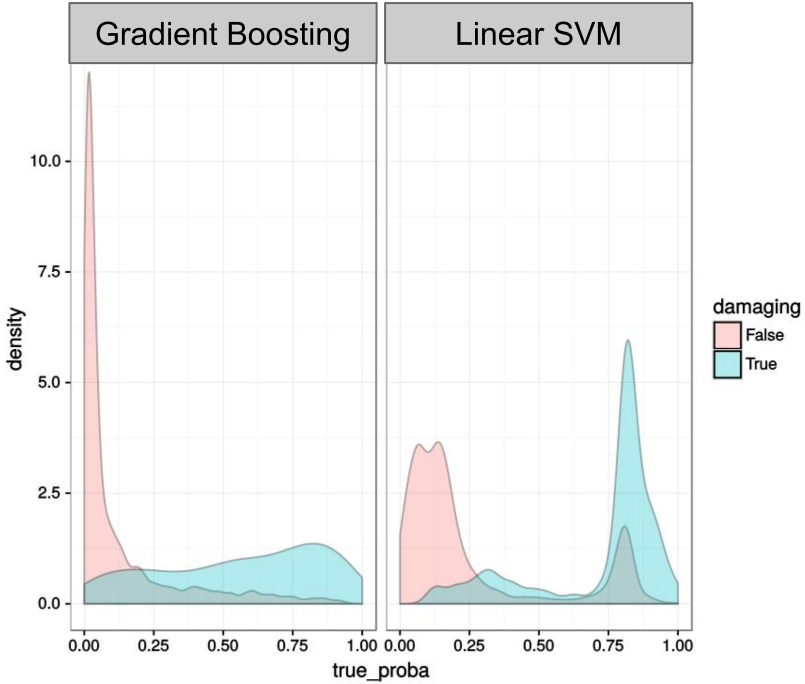
- Historically, focused on quantitative metrics:
  - Accuracy, Precision, Recall
  - F-score
  - ROC-AUC
  - PR-AUC
  - Etc.

Evaluation of AI

- Historically, focused on quantitative metrics:
  - Accuracy, Precision, Recall
  - F-score
  - ROC-AUC
  - PR-AUC
  - Etc.

“How well does my classifier work in general?”

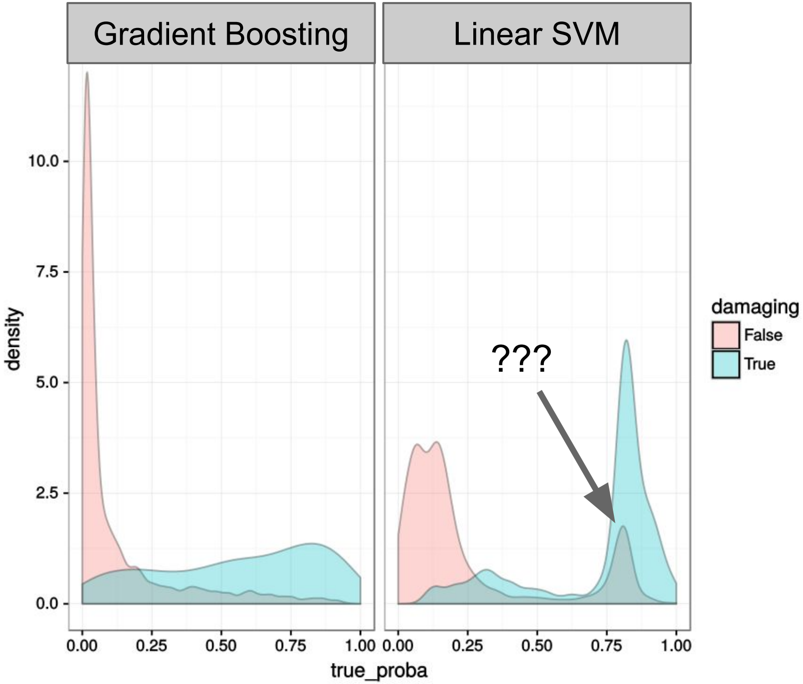
Evaluation of AI



Both work pretty good in general...

“How well does my classifier work in general?”

Evaluation of AI



Both work pretty good in general...

“Does my classifier behave strangely sometimes?”

Step 1:

“

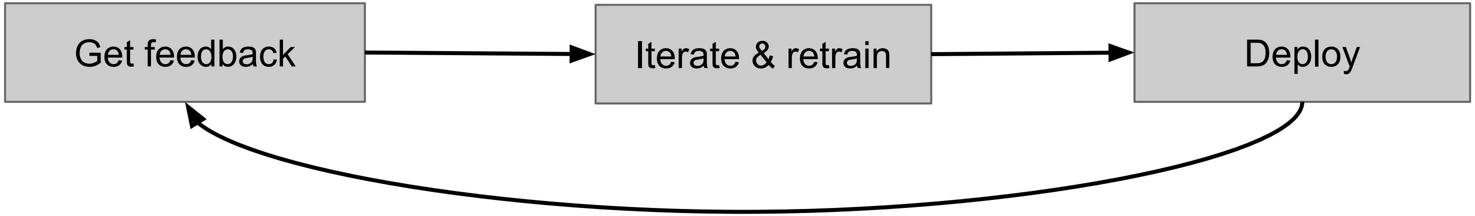
ORES is an experimental technology. We encourage you to take advantage of it but also to be skeptical of the predictions made. It's a tool to support you – it can't replace you. **Please reach out to us with your questions and concerns.**

”

Step 1:

> “
>
> ORES is an experimental technology. We encourage you to take advantage of it but also to be skeptical of the predictions made. It's a tool to support you – it can't replace you. **Please reach out to us with your questions and concerns.**
>
> ”

Step 2:

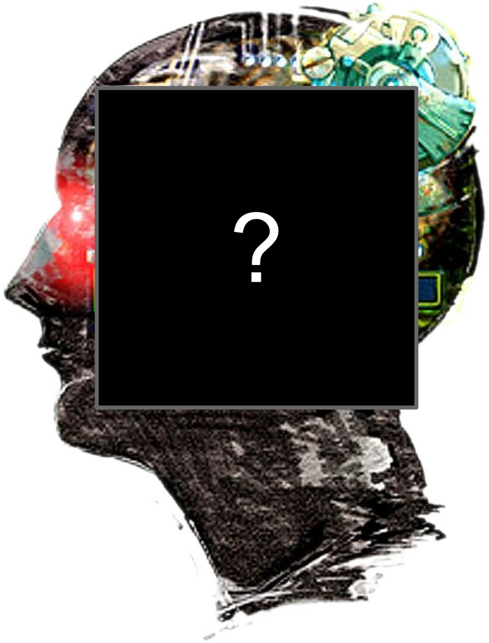


Insight:

“Despite the black-box nature of AI, humans are really good at noticing patterns and trends.”

https://commons.wikimedia.org/wiki/File:Bill_Gates_mugshot.png

https://commons.wikimedia.org/wiki/File:Bill_Gates_mugshot.png

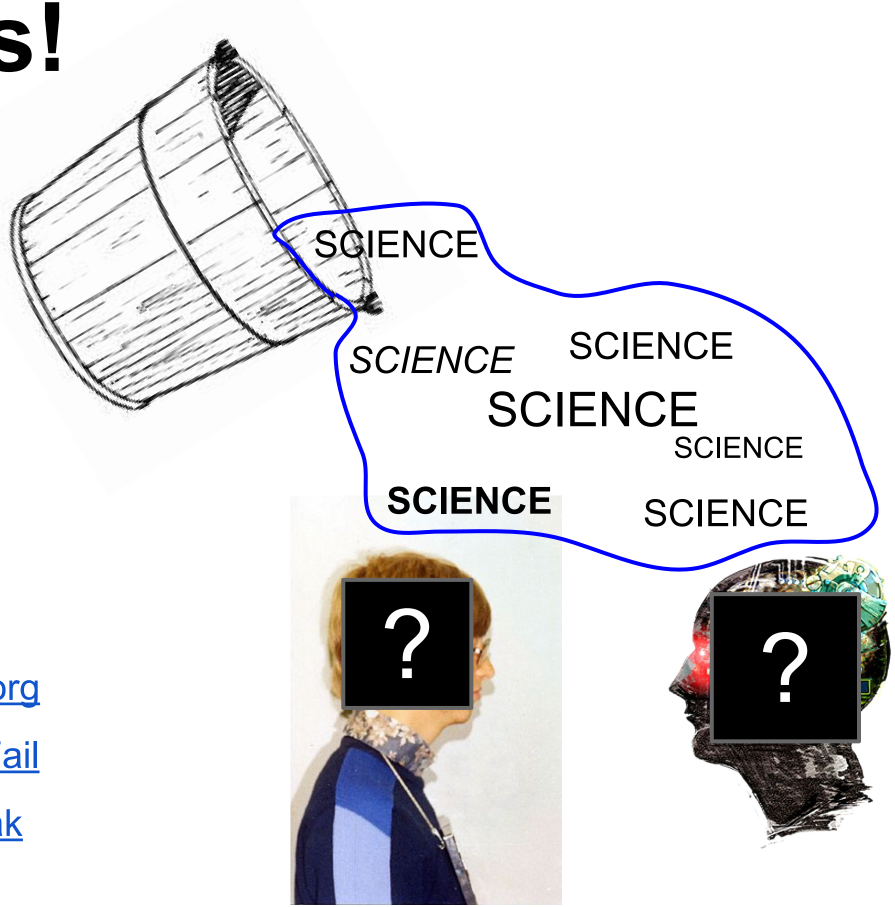


Human evaluation of AI’s bias should be part of standard practice for AI designers working in social contexts.

Thanks!

Props to my collaborators

- User:Ladsgroup
- User:Aetilley
- Sabya
- User:Rotpunkt
- User:Eran



SCIENCE

SCIENCE

SCIENCE

SCIENCE

SCIENCE

SCIENCE

SCIENCE

Aaron Halfaker

ahalfaker@wikimedia.org

enwp.org/User:EpochFail

https://twitter.com/halfak