Sean Goggins, CHAOSS Co-Founder, Augur Maintainer
April 20, 2021

- Defining Open Source Health and Sustainability with over 50 defined CHAOSS Metrics
    - Helping communities, companies, and organizations tell their own stories
    - Recognizing New Frontiers Through Initiatives
    - Constructing tools to express metrics, and identifying beliefs, values, and aesthetics underlying open source software work.
    - Collective, deep experience applying tech to address health and sustainability concerns, including DEI.

- Partners for Advancing Next Generation Open Source Software

Health and Sustainability Metrics Define What we Measure in a Repository

- 2016: No **Common, Shared** Definitions
- 2016: Artisan Activity Metrics

Repo level focus

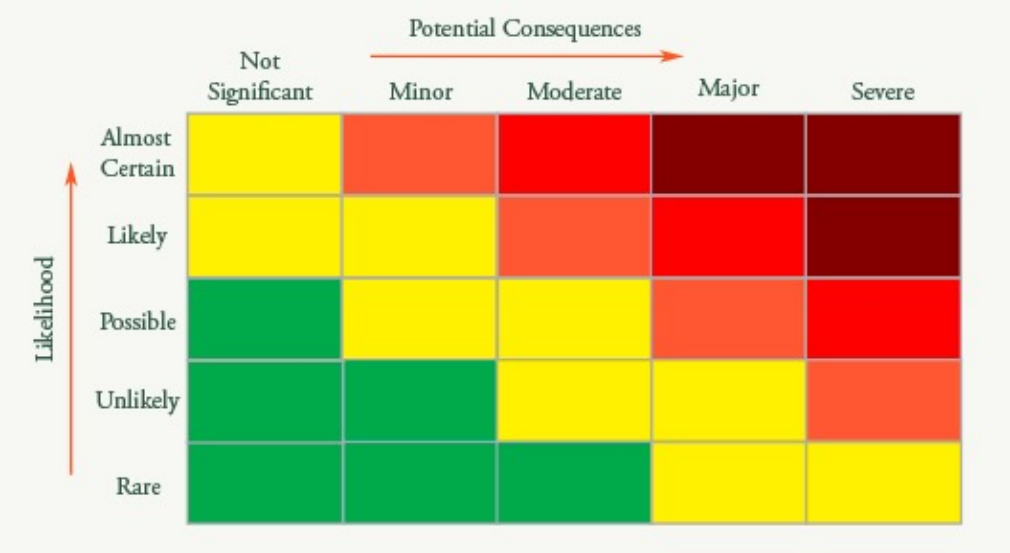
- 2021: 50+ **Defined** Metrics
- 2021: Augur: Examples of Stories about Health and Sustainability

Ecosystem level focus

- 2026: Health and Sustainability Metrics Widely Available
- 2026: Tools Looking Across Projects and Ecosystems.

CHAOSS

augur

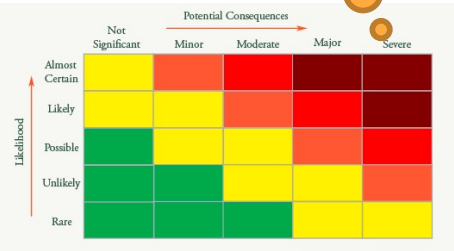


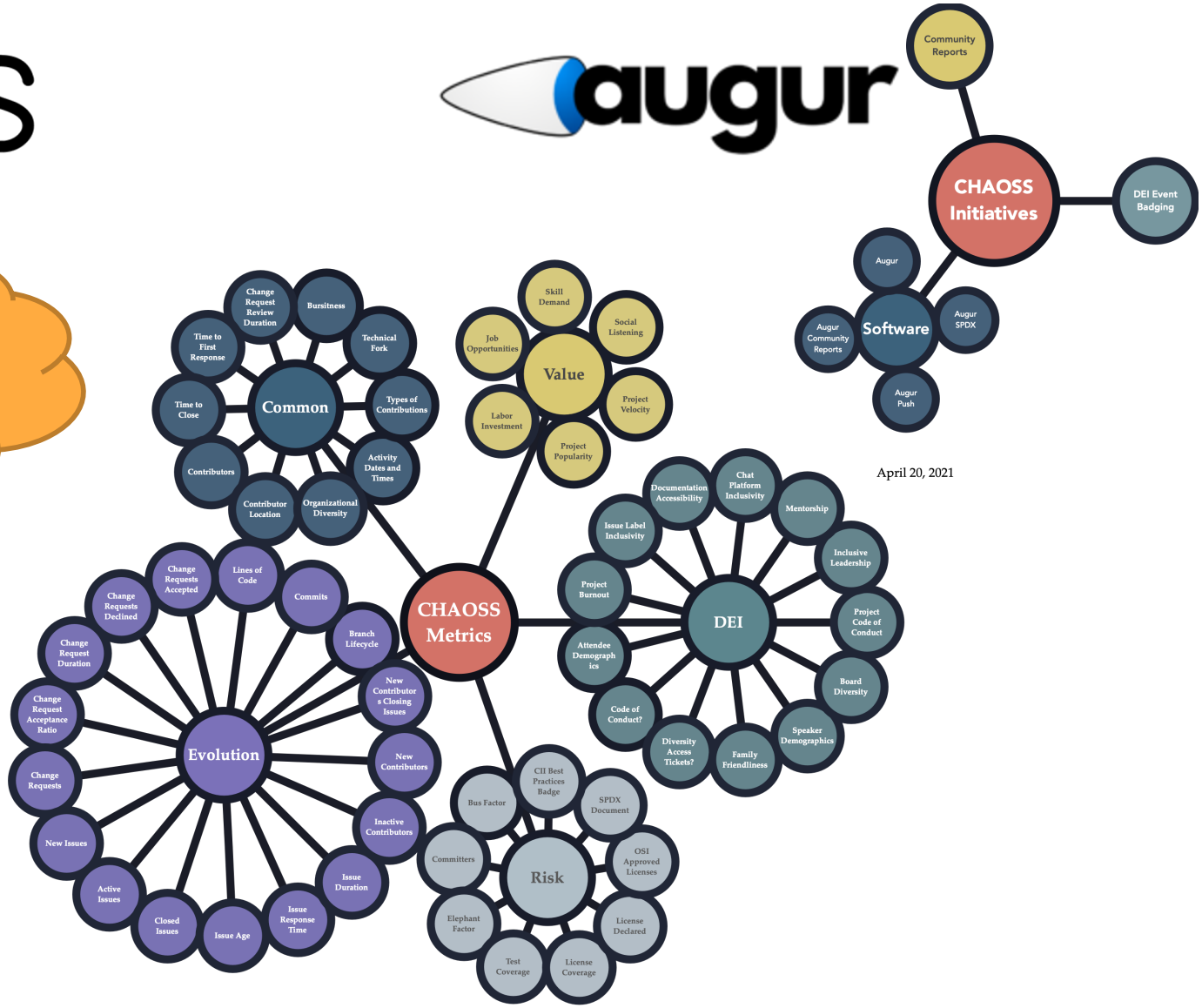
Open source software investors, community members, maintainers, and other stakeholders use metrics, stories, and information about **health and sustainability** to evaluate different types of risks, including risks to the core values that brought them to open source in the first place.

CHAOSS

augur

Each metric lives in a different square on the risk matrix for every project.





April 20, 2021

CHAOSS

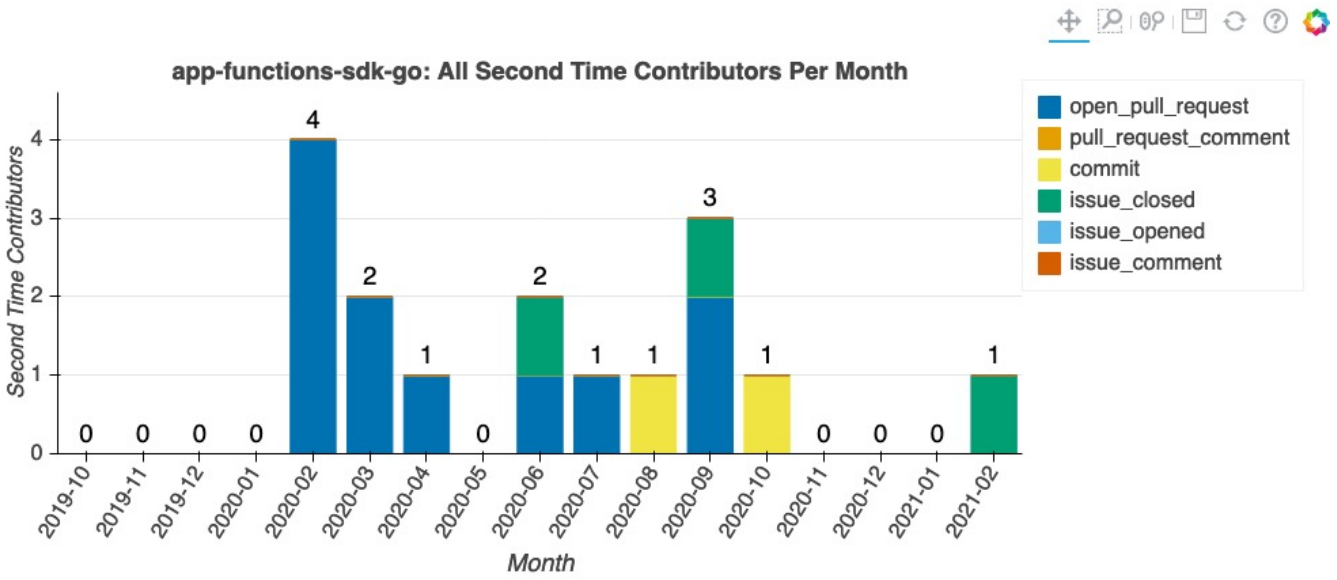
augur

Metrics: Words and Definitions

Beliefs

Values?

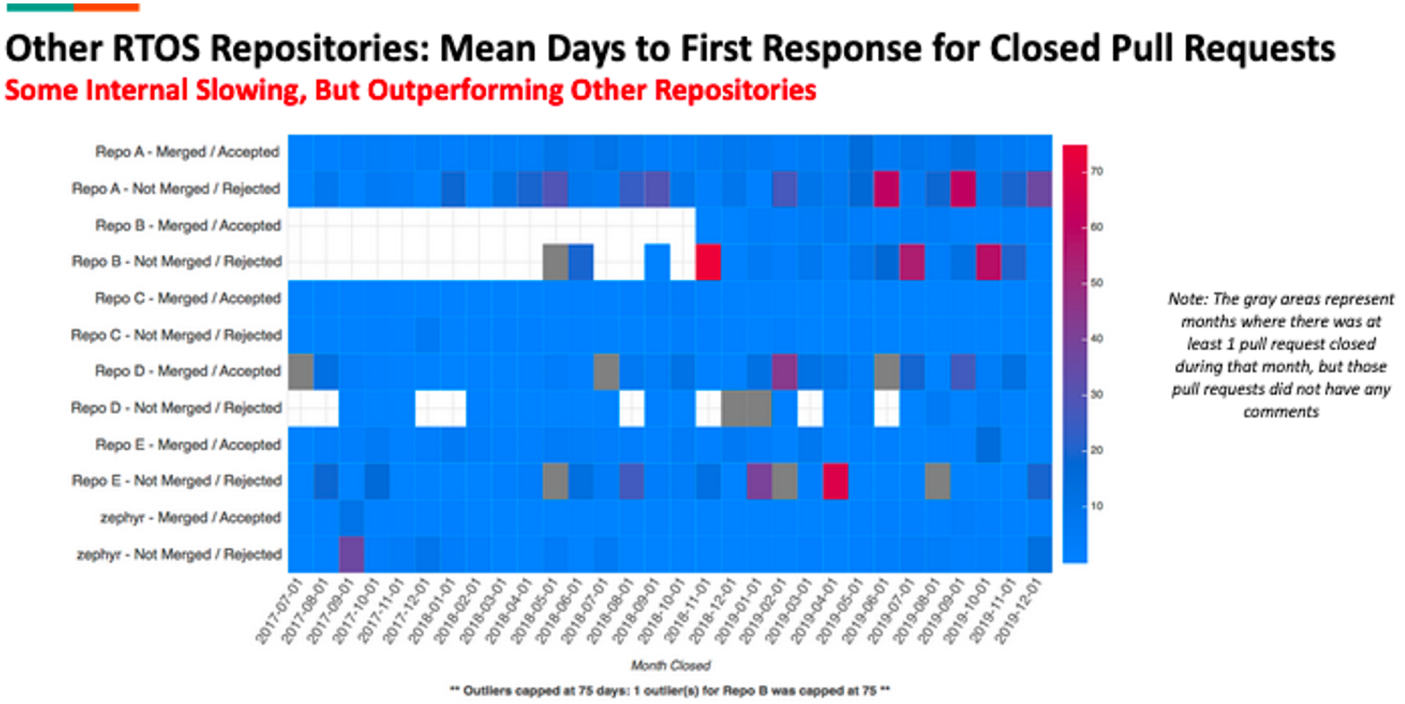
Composite Metrics Today



This graph shows the second contribution of all first time contributors in the specified time period.

Does a project engage new contributors?

Composite Metrics Today



How quickly does my project respond to outside code contributions compared to its competition?

Composite Metrics Today

License Declared

Click on a license for a description of the license

Click on a license count to download a list of associated files

Short Name	Count
Apache-2.0	86999
Artistic-2.0	2
BSD-2-Clause	903
BSD-3-Clause	82347
BSD-3-Clause-Attribution	118
BSD-3-Clause-Clear	752
BSD-4-Clause	12
BSD-4-Clause-UC	76
BSL-1.0	20
CC0-1.0	11
CC-BY-3.0	2
CC-BY-SA-3.0	8
EPL-1.0	272
GFDL-1.2	2
GPL-1.0	2
GPL-2.0	1090
GPL-2.0+	3241
GPL-3.0	24
GPL-3.0+	68
Intel	4
ISC	566
JSON	2
LGPL-2.0	22
LGPL-2.0+	848
LGPL-2.1	8766
LGPL-2.1+	302
LGPL-3.0	2
MIT	9608
MPL-1.1	2
NCSA	2152
No Assertion	23264
OFL-1.1	10
OpenSSL	4
Zlib	3128

What license declarations are included in a project's files?

Percent OSI-Approved Licenses by File

99.49%

OSI Approved: **200378** Files
View Approved Licenses

Not OSI Approved: **1021**
View Non-Approved licenses

What is a project's license coverage?

CHAOSS

augur

ALL IN

CHAOSS

augur

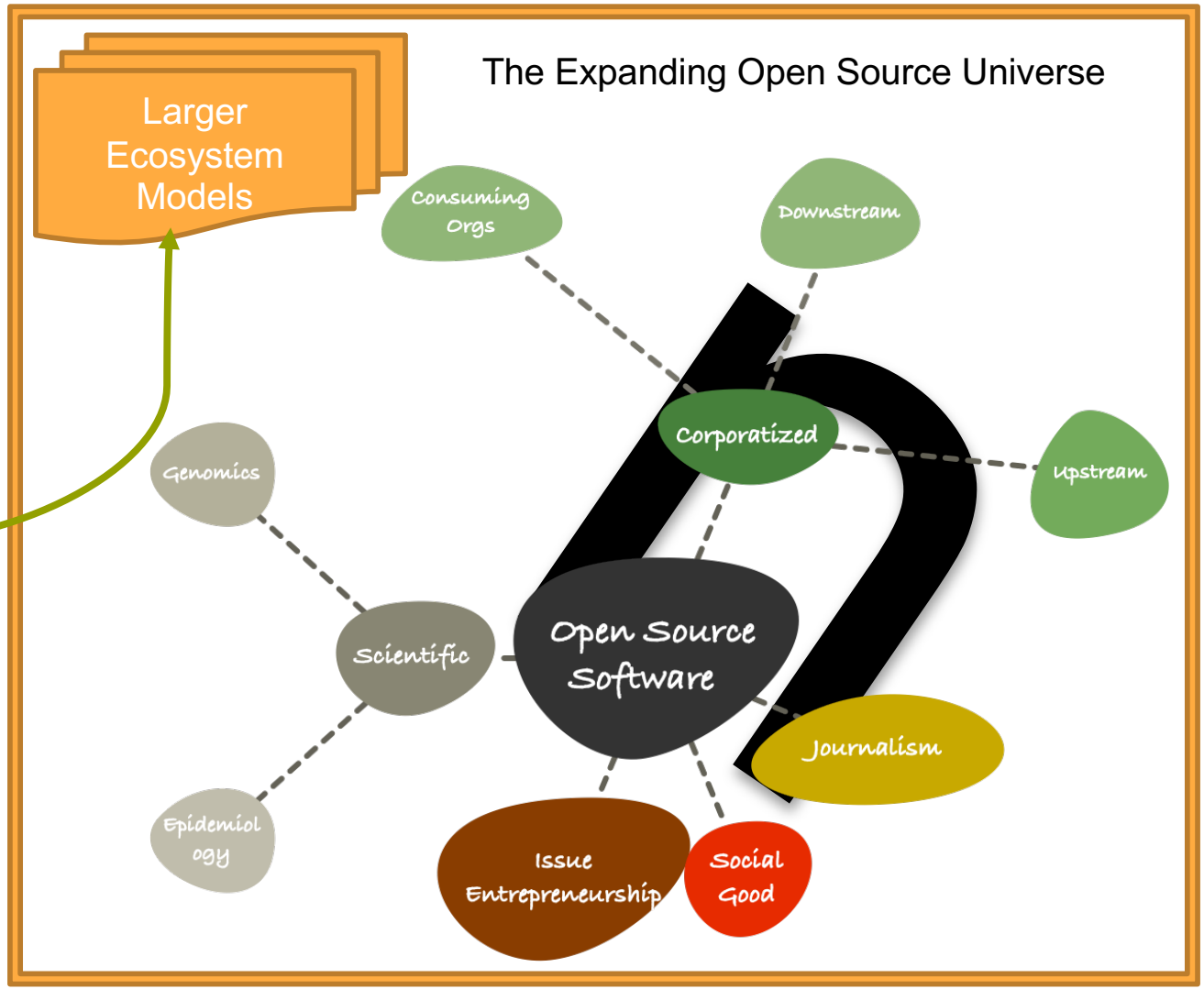
Computational Tools & Models for Supporting DEI

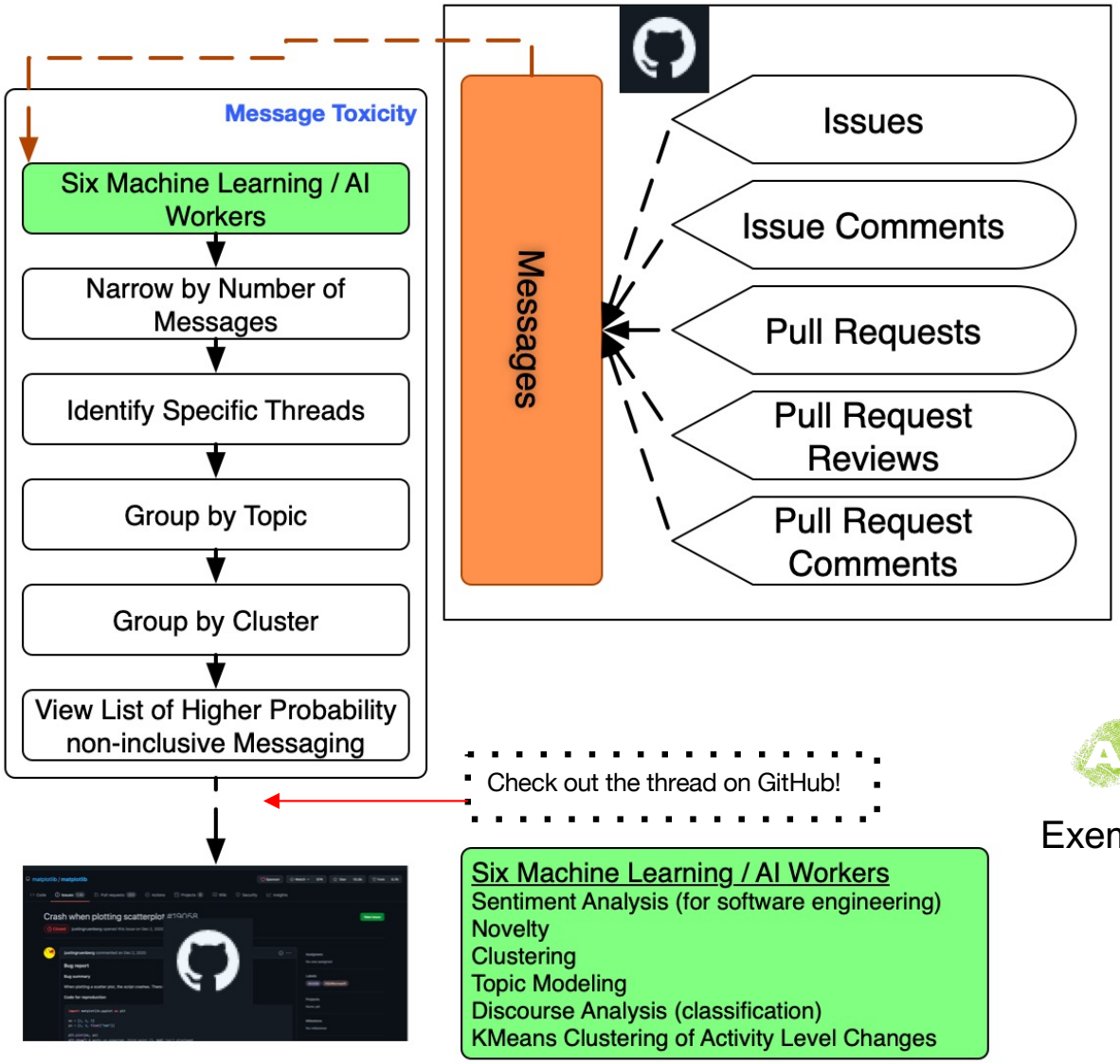
Advancing DEI in Open Source

ALL IN

ALL IN Project Models

Desperately Seeking Scalable Insight





augur

CHAOSS

A computational linguistic pipeline working toward identification of communities who are more, or less ready for advancing DEI.

Check out the thread on GitHub!

Six Machine Learning / AI Workers
Sentiment Analysis (for software engineering)
Novelty
Clustering
Topic Modeling
Discourse Analysis (classification)
KMeans Clustering of Activity Level Changes

Exemplar Models

Scalable Across Open Source

CHAOSS

augur

Metrics:
Words and
Definitions

Moving beneath
the surface

- Defining Open Source Health and Sustainability with over 50 defined CHAOSS Metrics
  - Helping communities, companies, and organizations tell their own stories
  - Recognizing New Frontiers Through Initiatives
  - Constructing tools to express metrics, and identifying beliefs, values, and aesthetics underlying open source software work.
  - Collective, deep experience applying tech to address health and sustainability concerns, including DEI.
- Partners for Advancing Next Generation Open Source Software

https://chaoss.community/community-reports/
https://chaoss.community/diversity-and-inclusion-badging/
https://www.augurlabs.io/

Thank You

https://chaoss.community/metrics

Epilogue

First, thank you. During questions, I enjoyed an important conversation with everyone on the ethics of data use, machine learning, and its histrionic twin, “artificial intelligence”. The ethics of algorithms and data are something I personally think about a great deal. GDPR, for example does important work codifying a value I share: that each person should be the owner of their own data. Frankly, I’m a computer scientist more because of accidents of fate than intention. Sometimes I wonder if my colleagues would really know what I meant if I said that most of my research is “engaged field work.” Matt and I first met as mentors in a summer institute for sociotechnical researchers, many of whom are also interested and concerned about how algorithms carry with them the value system of their creators, and too often reify a status quo that measures the wrong things.

Specifically, as I drive down this path in open source software, my hope is to use “AI” or “ML” or their equivalent buzzwords at any point in time to identify exemplar communities, like the one Demetris at GitHub is working to create. Communities like these in open source software can help us to create computational models for highlighting communication that advances diversity, equity, and inclusion. Ambitiously, perhaps even our acknowledgement that when we are working with data about people, we have the power to advance, or impede our shared humanity.

Epilogue

Today's questions caused Matt and I to wonder whether or not the AI/ML components of Augur, or any equivalent technology, should be "fully open sourced"? Should there be an open source license that requires commitment to ethical use of data, and some form of training on the human effects of these technologies, the embedding of our value systems in algorithms, and the ethical use of data? To my knowledge, ethical use is not incorporated in such a specific way in any existing open source software licenses. If I am wrong about that, we will change our license after training our own contributors on the importance of doing so. Finally, thank you again. I thought you might appreciate, or perhaps be reassured by my present reading list.